Bayesian Method for Recovering Surface and Illuminant Properties from Photosensor Responses

David H. Brainard
Department of Psychology, University of California, Santa Barbara
William T. Freeman
Mitsubishi Electric Research Laboratories. Inc., Cambridge, MA

Abstract

The goal of computational color constancy is to recover the physical properties of illuminants and surfaces from photosensor responses. We formulate computational color constancy as a statistical estimation problem. We assume that the likelihood that any particular illuminant or surface will occur in a scene is governed by a prior probability distribution. In particular, we assume that illuminant spectral power distributions are drawn according to multivariate. Normal distribution over the weights of a finite dimensional linear model, and similarly for surface reflectance functions. Given a set of photosensor responses, Hayes rule may be applied to derive the posterior distribution for illuminants and surfaces. We discuss how to use the posterior to estimate the illuminant. We use simulation to compare the performance a Bayesian algorithm to that of two previously reported color constancy algonthms. For our simulation conditions, the Bayesian algorithm results in the smallest expected estimation error.

Introduction

The surface reflectance function of an object specifies, as a function of wavelength, what fraction of incident light is reflected from the object. The surface reflectance function provides a spectral signature which may help identify the object. The light reflected from an object to an imaging device, however, confounds the object's surface reflectance function with the spectral power distribution of the illuminant: changes in the spectrum of the reflected light may be due to changes in surface reflectance or to changes in illumination. To estimate object surface reflectance, a visual system must factor the image into illuminant and surface components.

A visual system that can perform such a factorization, so that the color descriptors assigned to surfaces are independent of the illumination, is referred to as *color constant*. Over a wide range of viewing conditions, the human visual system exhibits at least partial color constancy.[1-3] A complete theory of human color vision must explain how color constancy, or even partial color constancy, is possible. The goal of computational color constancy is to provide this explanation.

Why Color Constancy is Difficult

To understand why color constancy requires a computational explanation, we must examine how images are formed. Consider the special case of a collection of N_1 matte surfaces, diffusely illuminated by a single illuminant. For this viewing geometry, we may characterize each surface by its reflectance function. We specify the reflectance function of the jth surface with the column vector s_j. The entries of s_j specify the fraction of incident light reflected in N_λ evenly spaced wavelength bands throughout the visible spectrum. Similarly, we specify the illuminant spectral power distribution with the column vector e. The entries of e specify the radiant power in each of the wavelength bands. The spectral power distribution of the light reaching the imaging device is given by the vector $c_j = e.* s_j$, where we use the notation .* to denote entry-by-entry multiplication.

A visual system does not typically have direct access to the N_1 spectra c_j. Rather, the imaging device samples each c_j with N_r classes of linear photosensor. Each photosensor is characterized by a spectral sensitivity function, which specifies how strongly it responds to light as a function of wavelength. We specify the spectral sensitivities with an N_r by N_λ dimensional matrix R. The klth element of R specifies the sensitivity of the kth sensor class to light in the lth wavelength band. We let the N_r dimensional column vector r_j represent the responses from all N_r sensor classes to the spectrum c_j. In the absence of sensor noise, we have the *rendering equation* for a single surface:

$$r_j = R\ c_j = R\ (e.*s_j). \qquad (1)$$

This equation specifies the relation between the *data* available to the visual system at an image location (r_j) and the *scene parameters* it must estimate (e and s_j). In this paper we consider methods for estimating the scene parameters from the data. This type of estimation is often referred to as an *inverse problem*.

The inverse problem specified by Eq. 1 is difficult for two reasons: it is underdetermined and it is non-linear. An inverse problem is underdetermined if there are more scene parameters than there are degrees of freedom in the data. The spectral functions for surfaces and illuminants are typically represented using wavelength bands with widths between 1 nm and 10 nm throughout the visible spectrum

(roughly 380 nm to 780 nm). Even if we use sparse 10 nm wavelength sampling in the limited wavelength range 400 nm to 700 nm, $\mathbf{e}$ and the $\mathbf{s}_j$ are of dimension 31. On the other hand, a typical imaging device (e.g. an RGB camera or the human eye), contains only a few classes of photosensor. Indeed, the $\mathbf{r}_j$ are often of dimension 3. If we have data from N_l image locations and assume one illuminant, then we have $(N_l N_r)$ measurements available to estimate $(N_\lambda(N_l + 1))$ scene parameters.

An inverse problem is non-linear if the scene parameters and the data are non-linearly related. In Eq. 1 some scene parameters (i.e. the entries of $\mathbf{e}$) are multiplied by others (i.e. the entries of the $\mathbf{s}_j$). Although standard methods exist for solving underdetermined linear inverse problems,[4] no general solution is available for the non-linear case.

Previous Work

Because the inverse problem posed by color constancy is highly underdetermined, previous work investigated ways to reduce the number of parameters required to describe illuminants and surfaces. This will be possible if it can be established a priori that not all possible illuminant and surface functions will be encountered in the environment where the visual system will operate.

A commonly used approach has been to describe spectral functions using finite-dimensional linear models. Let $\mathbf{B}_e$ be an N_λ by N_e dimensional matrix. We say that the illuminant spectra are constrained to lie within the linear model $\mathbf{B}_e$ if we can write $\mathbf{e} = \mathbf{B}_e \mathbf{w}_e$, where $\mathbf{w}_e$ is an N_e dimensional column vector. We call the columns of $\mathbf{B}_e$ the basis functions of the linear model, since the matrix product $\mathbf{B}_e \mathbf{w}_e$ expresses a weighted sum of these columns. The dimension of the linear model is N_e. The N_e entries of $\mathbf{w}_e$ are the linear model weights for the illuminant $\mathbf{e}$. Similarly, we say that the surface spectra are constrained to lie within the linear model $\mathbf{B}_S$ if we can write $\mathbf{s}_j = \mathbf{B}_S \mathbf{w}_{Sj}$, where $\mathbf{B}_S$ is an N_λ by N_S dimensional matrix and $\mathbf{w}_{Sj}$ is an N_S dimensional column vector.

If we assume that a spectral function is constrained to lie within an N dimensional linear model, then we can parameterize the function by specifying the model weights. (The basis functions are assumed to be fixed and known.) When N is small, linear models provide very compact descriptions of spectra. Many naturally occurring spectra are well-described by small-dimensional linear models. The number of dimensions required to describe a large sample of measured daylights is approximately four.[5] The number required for large samples of measured surfaces lies somewhere between six and eight.[6,7] Moreover, linear models with dimension as low as three capture a large percentage of the variance of the measured spectra.

Most modern color constancy algorithms incorporate the assumption that surfaces and illuminants may be parameterized by the weights of small-dimensional linear models. By incorporating linear model constraints, it has been possible to show several sets of conditions under constancy may be achieved. To illustrate the idea, we discuss two classic color constancy algorithms in some detail.

Maloney and Wandell[8] showed that when there are N_r classes of photoreceptors, the rendering equation may be inverted if the illuminants are constrained to lie within an N_r dimensional linear model and that the surfaces are con-strained to lie within N_r-1 dimensional linear model. We refer to their algorithm as the subspace algorithm. When the linear model constraints hold, the sensor responses $\mathbf{r}_j$ lie within an N_r-1 dimensional subspace of the N_r dimensional sensor space. The particular subspace depends on the illuminant, so that identifying the subspace that contains the $\mathbf{r}_j$ leads to an estimate of the illuminant.[8,9]

An elegant feature of Maloney and Wandell's algorithm is that the estimate of the illuminant does not depend on the particular surfaces in the scene. For human vision, however, there are only three classes of cone photosensors, so that Maloney and Wandell's analytical result applies directly only if the surfaces lie within a two-dimensional linear model. Since this is not the case for natural surfaces, the practical value of the subspace algorithm depends on how it behaves under violations of the assumptions on which it is based. We used simulation to investigate this question (see below) and conclude that the original subspace algorithm is not sufficiently robust to serve as a model for human color constancy.

Buchsbaum[10] developed an algorithm that is less strict about the dimension of the linear model constraints it requires. Both the illuminants and surfaces need only be constrained to lie within N_r dimensional linear models. Buchsbaum's algorithm assumes that the spatial mean of the surface reflectances $\mathbf{s}_j$ is constant across scenes and equal to some known reflectance function $\mathbf{s}_0$. This assumption is often referred to as the gray world assumption, and we therefore refer to Buchsbaum's algorithm the gray world algorithm. When the gray world assumption holds, the spatial mean of the sensor responses $\mathbf{r}_j$ can be shown to carry sufficient information to allow estimation of the illuminant.

The gray world assumption is somewhat unsatisfactory, as it implies that color constancy must break down for images where mean reflectance function differs from $\mathbf{s}_0$.[11] This breakdown does not seem to be characteristic of human color vision, although there is only limited psychophysical data available on this point. As a practical matter, we can ask how much error in the illuminant estimate is introduced by statistical variation in the mean of N_l surfaces drawn from a population mean with reflectance function $\mathbf{s}_0$. We used simulation to investigate this question and report the results below.

The subspace and gray world algorithms both work by extracting a summary statistic from the sensor responses and then using this statistic to estimate the illuminant. For the subspace algorithm, the statistic is the identity of the subspace that contains the responses. For the gray world algorithm, the statistic is the mean sensor responses. The algorithms demonstrate that these statistics carry information about the illuminant. On the other hand, there is no reason to use only one statistic. If both statistics carry information, then we might expect improved performance from an algorithm that combines information from both.

To capitalize on this intuition, however, we must know how to combine the information in an effective manner.

Bayesian Framework

In this paper, we apply Bayesian estimation methods to computational color constancy. The Bayesian approach gives us a prescription for how to use all of the information

about the illuminant contained in the sensor responses $\mathbf{r}_j$, including the information used by the subspace and gray world algorithms. See Berger[12] for a general discussion of Bayesian methods. See Trussell and Vrhel[13,14] and D'Zmura and Iverson[15] for related statistical approaches to color constancy.
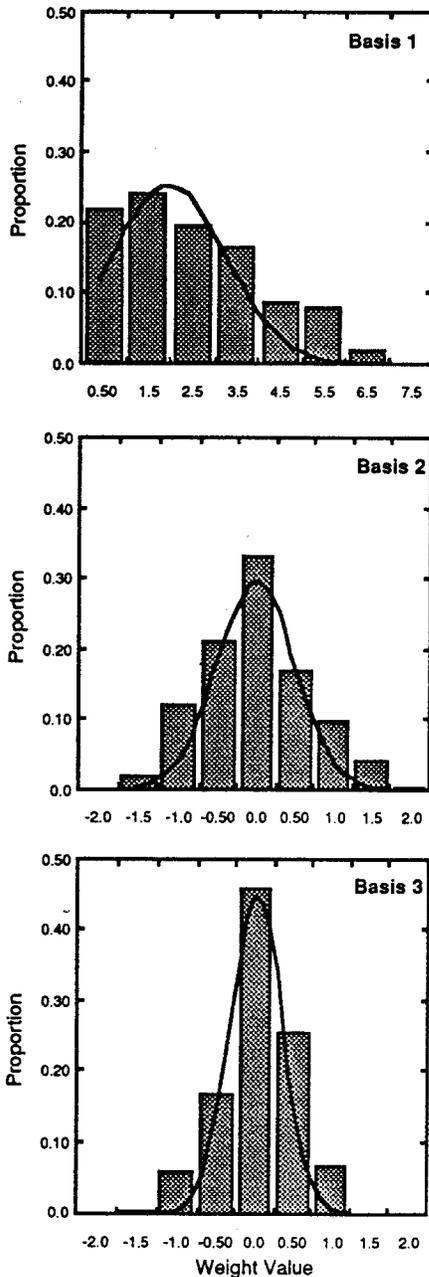
Figure 1. Histograms of surface linear model weights. Each panel shows weights for one basis function. Solid lines show Normal probability density with the same mean and variance as the weights.

In the next section, we introduce the Bayesian approach to parameter estimation. We discuss how this approach may be applied to color constancy and review a simple illustrative example. We then derive an analytic

result that allows us to use the Bayesian approach to estimate the illuminant from the sensor responses. Finally, we use simulation to compare the performance of the Bayesian algorithm to that of the subspace and gray world algorithms.

Bayesian Estimation

The Bayesian approach to estimation is very simple. Prior information about the parameters is expressed as a probability distribution. If we are trying to estimate parameters described by the vector $\mathbf{x}$, then the prior information is given by the probability density $p(\mathbf{x})$. The relation between the parameters $\mathbf{x}$ and the data $\mathbf{y}$ is also expressed as a probability density $p(\mathbf{y}|\mathbf{x})$, often referred to as the likelihood function. (We adopt the convention that the notation p( ) is used to denote different probability density functions. The particular function in any context is indicated by the argument.)

Given the prior $p(\mathbf{x})$ and the likelihood $p(\mathbf{y}|\mathbf{x})$, the probability of any set of the parameter values, given the data, is computed using Bayes rule

$$p(\mathbf{x}|\mathbf{y}) = C\, p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x}). \qquad (2)$$

In this paper, we will use C to indicate any expression constant over the variables of interest. Here, C is a normalizing constant that depends on the data $\mathbf{y}$ but not on the parameters $\mathbf{x}$. The distribution $p(\mathbf{x}|\mathbf{y})$ is referred to as the posterior. The posterior gives the probability that the parameter $\mathbf{x}$ generated the data $\mathbf{y}$.

To go from the posterior to a single estimate $\hat{\mathbf{x}}$ for the parameters $\mathbf{x}$ we need to specify a loss function $L(\tilde{\mathbf{x}}\,;\mathbf{x})$. This function specifies the penalty for choosing $\tilde{\mathbf{x}}$ when the actual parameters are $\mathbf{x}$. Given the posterior and a loss function, we may compute the loss expected in choosing $\tilde{\mathbf{x}}$, called the Bayes risk, which we write

$$\mathbf{R}(\tilde{\mathbf{x}}|\mathbf{y}) = \int_{\mathbf{x}} L(\tilde{\mathbf{x}},\mathbf{x})p(\mathbf{x}|\mathbf{y})d\mathbf{x}. \qquad (3)$$

We choose $\hat{\mathbf{x}}$ to minimize the Bayes risk.

Bayesian estimation provides a principled way to choose an optimal estimate. It may be applied to any problem where the prior information can be cast in the form of a probability distribution over the scene parameters and where a suitable loss function can be specified.

Probabilistic Constraints

To apply the Bayesian framework to computational color constancy, we must express our knowledge about illuminants and surfaces as a probability distribution. Figure 1 illustrates how we do this. We have taken a large set of measured surface reflectances and found the three dimensional linear model that provides the best approximation to the data. The figure shows histograms of the model weights required to fit each reflectance. The weights for each basis function cluster around a central value. We will assume that the model weights are distributed according to a Normal distribution with the same mean and variance as the measured weights. In the figure, we have shown as solid lines the Normal distributions corresponding to each basis function. It is quite possible that some other form could better model the weight distributions.

Bilinear Problems

Although the color constancy rendering equation (Eq. 1) is non-linear, it does exhibit considerable structure. In particular, the relation between the photosensor responses r_j, the illuminant parameters e, and the surface parameters s_j is *bilinear*. This means that the relation between r_j and s_j is linear when e is held fixed and that the relation between r_j and e is linear when s_j is held fixed. To emphasize this, we note that we can write

$$r_j = L(e)s_j = L(s_j)e \qquad (4)$$

where $L(e) = R \, \mathrm{diag}(e)$ is an N_r by N_λ dimensional matrix that depends only on e and $L(s_j) = R \, \mathrm{diag}(s_j)$ is an N_r by N_λ dimensional matrix that depends only on s_j.

When we reparameterize in terms of linear model weights, we do not perturb the bilineanty of the rendering equation. By letting $M(w_e) = L(B_e w_{sj})B_S$ and $M(w_{sj}) = L(B_S w_{sj})B_e$ we have

$$r_j = M(w_e) \, w_{sj} = M(w_{sj})w_e. \qquad (5)$$

It is convenient to extend Eq. 5 to express the relation between the data and scene parameters at all locations simultaneously. Let r be the $(N_l N_r)$ dimensional column vector obtained by stacking the vectors r_j. Let w_S be the $(N_l N_s)$ dimensional column vector obtained by stacking the vectors w_{sj}. Then we can write the overall rendering equation as

$$r = N(w_e) \, w_S = N(w_S)w_e. \qquad (6)$$

Here the matrix $N(w_e)$ is the $(N_l N_r)$ by $(N_l N_S)$ block diagonal matrix with $M(w_e)$ repeated along its diagonal, while the matrix $N(w_S)$ is the $(N_l N_r)$ by N_e matrix obtained by stacking the matrices $M(w_{sj})$.

As with the rendering equations for individual image locations, Eq. 6 expresses a bilinear relation. Any results that hold for generally for bilinear rendering equations may be applied to the specific problem of color constancy.

The Product Example

The simplest case of a bilinear inverse problem occurs when the data and the scene parameters are all scalars. For example, we can consider the rendering equation $r = es + n$, where the random variable n represents additive Gaussian noise. The estimation problem is to reason from r back to estimates of e and s. This example shares with the color constancy problem its bilinearity and the fact that it is underdetermined.
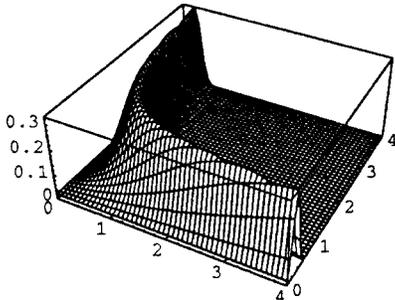
Figure 2. The posterior distribution for the product example.

Suppose we know that e and s were each drawn from a uniform distribution on the range [0,4]. Figure 2 shows the bivariate posterior distribution on e and s when the observed data $r = 1$ and when the added noise is Normally distributed with mean 0 and variance 0.18. Over the plotted domain, the posterior is given by

$$p(e, s|r = 1) = C \exp \left\{ \frac{[1 - es]^2}{2[0.42]^2} \right\}. \qquad (7)$$

The posterior makes perfect sense. Points (e,s) whose product is close to 1 have a large posterior probability. Points whose product is quite different from 1 have a low posterior probability. The set of points with highest probability form a hyperbolic ridge along the locus of points (e,s) such that es = 1.

From a deterministic point of view, it is hopeless to try to estimate e and s given r. There are many pairs of points (e,s) that are consistent with the data, and the uniform prior does not distinguish between them. Within the Bayesian framework, however, some estimates are better than others. Which estimate is best depends on the loss function. The left column of Figure 3 shows four different loss functions. The first is the quadratic loss function:

$$L_q(\tilde{e}, \tilde{s}; e, s) = (\tilde{e} - e)^2 + (\tilde{s} - s)^2. \qquad (8)$$

This function produces a penalty that grows at an accelerating rate with the size of the estimation error. The second is a delta loss function:

$$L_d(\tilde{e}, \tilde{s}; e, s) = -\delta\big((\tilde{e}, \tilde{s}) - (e, s)\big). \qquad (9)$$

This loss function produces a constant penalty except for when the estimate is exactly correct. Both of these loss functions are commonly used. It is not clear, however, that either actually describes the loss for many real estimation problems. Often, coming fairly close to the true value is satisfactory and the loss saturates once the estimation error is sufficiently large. The third panel in the left column of Figure 3 plots this sort of loss function. We call this the local mass loss function since it integrates probability density over a local area. It is given by

$$L_{lm}(\tilde{e}, \tilde{s}; e, s) = -\exp \frac{-1}{2\mu^2} \Big[(\tilde{e} - e)^2 + (\tilde{s} - s)^2 \Big]. \qquad (10)$$

The local mass loss function rewards estimates close to the true value and provides for a saturating penalty for squared errors much greater than μ^2. Yuille and Bulthoff[16] have independently noted the usefulness of this loss function for related problems.

The final panel of the top row shows an example of an asymmetric loss function

$$L_a(\tilde{e}, \tilde{s}; e, s) = -\delta(\tilde{e} - e). \qquad (11)$$

This type loss function is appropriate in situations where we are more interested in some of the scene parameters than in others. For example, we may be much more interested in correctly estimating surface properties than in estimating the illuminant or vice versa. In the figure, we show an asymmetric version of the delta loss function. This loss function is an extreme example in that no weight at all

is placed on estimation error for one parameter. It is straightforward to construct asymmetric versions of the quadratic or local mass loss functions where estimation error for different parameters is not weighted equally.

The right column of Figure 3 shows the Bayes risk corresponding to each loss function as a function of $\hat{e}$ and $\hat{s}$. The risks were computed using the posterior shown in Figure 2. From top to bottom, the points of minimum risk are (1.30,1.30), (-,-), (0.99,0.99), and (0.35,-), where the symbol - denotes an indeterminate minimum.

For two of the loss functions, there is a point of minimum risk. Even for highly underdetermined problems, some solutions are better than others. Also note that the minimum risk estimate depends on the loss function. For example, by comparing the second and fourth rows of the figure, we see that modifying the loss function to ignore the estimation error for $\hat{s}$ has a large effect on the how the risk depends on $\hat{e}$. The choice of loss function has a substantive effect on the estimate obtained from the posterior.

The computational cost of finding the minimum risk estimate depends on the loss function. In general, it is easier to minimize the Bayes risk when the risk can be evaluated using calculations that are local in the parameter space. For the delta loss function, finding the risk associated with any estimate $(\tilde{e},\tilde{s})$ only requires evaluating the posterior at the point $(\tilde{e},\tilde{s})$. For the local mass loss function, the calculation involves integration of a function that is localized in the parameter space. For some cases, the Bayes risk integral for both the asymmetric and the local mass loss functions may be approximated, yielding a local analytic calculation. See Freeman[17,18] for examples using an asymmetric loss function. For the quadratic loss function, finding the risk requires a global calculation in the parameter space. But for this loss function, the point of minimum risk is equal to the mean of the posterior distribution; it is not necessary to search numerically for the point of minimum risk.

We can make a second general point with the scalar product example: prior information helps. Suppose that rather than choosing uniform priors we choose $p(e) = N(u_e,\sigma^2_e)$ and $p(s) = N(u_s, \sigma^2_s)$, with $p(e,s) = p(e)p(s)$. For large variances σ^2_e and σ^2_s, the distribution of e and s over the region [0,4] will be almost uniform. But as the variance of even one of the variables shrinks, the prior information will start to influence the posterior. The effect of changing the prior variance may be quantified calculating the Bayes risk as a function of the prior variances. Figure 4 shows plots the results of such a calculation. Smaller prior variances result in lower Bayes risks.

Figure 3. Left column: Four loss functions for the product example. From top to bottom, the quadratic, delta, local mass, and asymmetric loss functions. Each of the loss functions is shift-invariant and is plotted as a function of e and s with $\tilde{e}$ and $\tilde{s}$ held constant. The constant 1 was added to the last three loss functions for visual clarity. For the local mass function, $\mu = 0.2$. Right column: The Bayes risks corresponding to each loss function for the posterior of Figure 2. For visual clarity, theses functions are plotted with risk increasing towards the bottom of the page.

Figure 4. Effect of prior variance. The plot shows how the Bayes risk varies with the prior variance for the product example. The x-axis shows the prior vanance σ^2_e. The y-axis shows the risk corresponding to the minimum risk estimate. The risks were computed with the local mass function, $r = 1$, $u_e = u_s = 2$, and $\sigma^2_s >> 2$.

Bayesian Color Constancy

We now develop a Bayesian approach to inverting the color constancy rendering equation. Our solution applies to any bilinear inverse problem with multivariate Normal priors where the two (e.g. illuminant and surface) sets of parameters are drawn independently.

Priors

We assume that the illuminant prior is determined by a multivariate Normal distribution on the weights of a linear model, so that $\mathbf{e} = \mathbf{B}_e \mathbf{w}_e$ and

$$p(\mathbf{w}_e) = N(\mathbf{u}_e, \mathbf{K}_e). \qquad (12)$$

Similarly, the prior for each surface is determined by a multivariate Normal distribution on the weights of a linear model, so that $\mathbf{s}_j = \mathbf{B}_S \mathbf{w}_S$ and

$$p(\mathbf{w}_{sj}) = N(\mathbf{u}_{sj}, \mathbf{K}_{sj}). \qquad (13)$$

We assume that the surfaces at each location are independent and identically distributed, so that we may write

$$p(\mathbf{w}_S) = N(\mathbf{u}_S, \mathbf{K}_S) \qquad (14)$$

where $\mathbf{u}_S$ is obtained by stacking N_l copies of $\mathbf{u}_{sj}$ and $\mathbf{K}_S$ is the $(N_l N_S)$ by $(N_l N_S)$ block diagonal matrix with $\mathbf{K}_{sj}$ along the diagonal. Assuming that the illuminant and the surfaces are statistically independent, we have

$$p(\mathbf{w}_e, \mathbf{w}_S) = p(\mathbf{w}_e) \, p(\mathbf{w}_S). \qquad (15)$$

Data Formation

From the rendering equation (Eq. 6) we know the relation between $\mathbf{r}$, $\mathbf{w}_e$ and $\mathbf{w}_S$. We assume that the vector of sensor responses is perturbed by additive Gaussian noise with mean vector $\mathbf{u}_n$ and covariance matrix $\mathbf{K}_n$. We will study the case where the noise for each sensor is independent and identically distributed with zero mean, so that $\mathbf{u}_n = 0$ and $\mathbf{K}_n$ is a multiple of the identity matrix. We can write

$$
\begin{aligned}
p(\mathbf{r} \mid \mathbf{w}_e, \mathbf{w}_S) &= N(\mathbf{A}(\mathbf{w}_e)\,\mathbf{w}_S + \mathbf{u}_n, \mathbf{K}_n) \\
&= N(\mathbf{A}(\mathbf{w}_S)\,\mathbf{w}_e + \mathbf{u}_n, \mathbf{K}_n). \qquad (16)
\end{aligned}
$$

The Posterior

From Bayes' Rule, we can compute the posterior distribution for the illuminants and surfaces, given the data.

$$p(\mathbf{w}_e, \mathbf{w}_S \mid \mathbf{r}) = C\, p(\mathbf{r} \mid \mathbf{w}_e, \mathbf{w}_S)\, p(\mathbf{w}_e) p(\mathbf{w}_S). \qquad (17)$$

The posterior captures our knowledge about the scene parameter given the sensor responses and the priors.

Choosing a Point Estimate

In general, minimizing the Bayes Risk requires minimizing the expected value of a loss function for different choices of estimates. Because the dimensionality of $\mathbf{w}_S$ grows with the number of image locations, it can be quite high. Thus computing the expectation of the loss function for any given parameter estimate and searching for the parameter estimates that minimize this expected loss is computationally very intensive. Here we report on an analytic method for computing the marginal posterior distribution $p(\mathbf{w}_e \mid \mathbf{r})$. Because the dimensionality of $\mathbf{w}_e$ is reduced from that of the full problem, numerical methods may be used to find point estimates from the marginal. For the quadratic loss function, the minimum risk estimate for $\mathbf{w}_e$ is given by the mean of the marginal distribution $p(\mathbf{w}_e \mid \mathbf{r})$. This estimate for $\mathbf{w}_e$ is the same as would be obtained by finding the minimum quadratic risk estimate $(\mathbf{w}_e, \mathbf{w}_S)$ from the full posterior $p(\mathbf{w}_e, \mathbf{w}_S \mid \mathbf{r})$. As the scalar product example illustrates, these properties do not generalize to other loss functions.

Marginal Distributions

To compute the marginal distribution on illuminants, we integrate the joint posterior distribution (Eq. 17) over the variable $\mathbf{s}$.

$$
\begin{aligned}
p(\mathbf{w}_e \mid \mathbf{r}) &= \int_s p(\mathbf{w}_e, \mathbf{w}_s \mid \mathbf{r}) ds \\
&= C\, p(\mathbf{w}_e) \int_s p(\mathbf{r} \mid \mathbf{w}_e, \mathbf{w}_s) p(\mathbf{w}_s) ds \\
&= C\, \exp\left(-\frac{1}{2} \left| \mathbf{K}_{we}^{-1/2}(\mathbf{w}_e, \mathbf{u}_e) \right|^2 \right) \times \qquad (18) \\
&\quad \int_s \exp\left(-\frac{1}{2} \left| \mathbf{K}_n^{-1/2}(\mathbf{r} - \mathbf{A}(\mathbf{w}_e)\mathbf{s} - \mathbf{u}_n) \right|^2 \right) \times \\
&\quad \exp\left(-\frac{1}{2} \left| \mathbf{K}_{ws}^{-1/2}(\mathbf{s} - \mathbf{u}_s) \right|^2 \right) ds
\end{aligned}
$$

where the constant C is independent of $\mathbf{w}_e$ and $\mathbf{w}_s$. Let $\mathbf{s}_{ls}$ denote the least squares solution to the set of linear equations $\mathbf{r} = \mathbf{A}(\mathbf{w}_e)\mathbf{s}$. Then we may write

$$
\begin{aligned}
\left| \mathbf{K}_n^{-1/2}(\mathbf{r} - \mathbf{A}(\mathbf{w}_e)\mathbf{s} - \mathbf{u}_n) \right| &= \\
= \left| \mathbf{K}_n^{-1/2}(\mathbf{r} - \mathbf{A}(\mathbf{w}_e)\mathbf{s}_{ls}) \right|^2 &+ \qquad (19) \\
+ \left| \mathbf{K}_n^{-1/2}(\mathbf{A}(\mathbf{w}_e)\mathbf{s} - \mathbf{s}_{ls}) \right|^2
\end{aligned}
$$

where we used the facts that $(\mathbf{r} - \mathbf{A}(\mathbf{w}_e)\mathbf{s}_{ls}) \perp (\mathbf{A}(\mathbf{w}_e)(\mathbf{s} - \mathbf{s}_{ls}))$ for a least squares solution $\mathbf{s}_{ls}$, that $\mathbf{K}_n$ is a multiple of the identity matrix, and that $\mathbf{u}_n = 0$. We use Eq. 19 to rewrite Eq. 18.

$$
\begin{aligned}
p(\mathbf{w}_e \mid \mathbf{r}) &= C\, \exp\left(-\frac{1}{2} \left| \mathbf{K}_e^{-1/2}(\mathbf{w}_e - \mathbf{u}_e) \right|^2 \right) \times \\
&\quad \exp\left(-\frac{1}{2} \left| \mathbf{K}_n^{-1/2}(\mathbf{r} - \mathbf{A}(\mathbf{w}_e)\mathbf{s}_{ls}) \right|^2 \right) \times \\
&\quad \int_s \exp\left(-\frac{1}{2} \left| \mathbf{K}_n^{-1/2}(\mathbf{A}(\mathbf{w}_e)(\mathbf{s} - \mathbf{s}_{ls})) \right|^2 \right) \times \quad (20) \\
&\quad \exp\left(-\frac{1}{2} \left| \mathbf{K}_s^{-1/2}(\mathbf{s} - \mathbf{u}_s) \right|^2 \right) ds.
\end{aligned}
$$

We can solve the above integral analytically if we employ a matrix version of "completing the square". Let $\mathbf{K}_1 = [\mathbf{A}(\mathbf{w}_e)^T \mathbf{K}_n^{-1} \mathbf{A}(\mathbf{w}_e)]^{-1}$, $\mathbf{K}_2 = \mathbf{K}_S$, $\mathbf{u}_1 = \mathbf{s}_{ls}$, and $\mathbf{u}_2 = \mathbf{u}_S$. We can rewrite the integral in Eq. 20 and simplify

$$\int_s \exp\left(-\frac{1}{2}\left|\mathbf{K}_n^{-1/2}(\mathbf{A}(\mathbf{w}_e)(\mathbf{s}-\mathbf{s}_{ls})\right|^2\right)\times$$

$$\exp\left(-\frac{1}{2}\left|\mathbf{K}_{ws}^{-1/2}(\mathbf{s}-\mathbf{u}_{ws})\right|^2\right)d\mathbf{s}$$

$$=\int_s \exp\left(-\frac{1}{2}\left|\mathbf{K}_1^{-1/2}(\mathbf{s}-\mathbf{u}_1)\right|^2\right)\times$$

$$\exp\left(-\frac{1}{2}\left|\mathbf{K}_2^{-1/2}(\mathbf{s}-\mathbf{u}_s)\right|^2\right)d\mathbf{s} \qquad (21)$$

$$=\exp\left(\frac{1}{2}C_K\right)\int_s \exp\left(-\frac{1}{2}\left|\mathbf{K}^{-1/2}(\mathbf{s}-\mathbf{u})\right|^2\right)d\mathbf{s}$$

where $\mathbf{K}^1 = \mathbf{K}_1^{-1}+\mathbf{K}_2^{-1}$, $\mathbf{K}_1 = \left[\mathbf{A}(\mathbf{w}_e)^T \mathbf{K}_n^{-1}\mathbf{A}(\mathbf{w}_e)\right]^{-1}$, and

$$C_K = \mathbf{u}_1^T\mathbf{K}_1^{-1}\mathbf{K}\mathbf{K}_2^{-1}\mathbf{u}_2 - \mathbf{u}_1^T\mathbf{K}_1^{-1}\mathbf{K}\mathbf{K}_2^{-1}\mathbf{u}_1 +$$
$$\mathbf{u}_2^T\mathbf{K}_2^{-1}\mathbf{K}\mathbf{K}_1^{-1}\mathbf{u}_1 - \mathbf{u}_2^T\mathbf{K}_2^{-1}\mathbf{K}\mathbf{K}_1^{-1}\mathbf{u}_2 \qquad (22)$$

The equivalence of the integrals in Eq. 21 may be verified by direct algebraic expansion. The last integral in Eq. 21 may be evaluated analytically, and we use this to derive for the marginal posterior on $\mathbf{w}_e$

$$p(\mathbf{w}_e|\mathbf{r}) = C\ \exp\left(-\frac{1}{2}\left|\mathbf{K}_{we}^{-1/2}(\mathbf{w}_e,\mathbf{u}_e)\right|^2\right)\times$$

$$\exp\left(-\frac{1}{2}\left|\mathbf{K}_n^{-1/2}(\mathbf{r}-\mathbf{A}(\mathbf{w}_e)\mathbf{s}_{ls})\right|^2\right)\times$$

$$\exp\left(-\frac{1}{2}C_K\right)\times$$

$$\frac{1}{\sqrt{\left|\mathbf{K}^{-1}\right|}} \qquad (23)$$

where we have absorbed some additional constants into C. Note that we have achieved significant computational savings by this analytic integration. We have summed the posterior density over the surface parameters at all locations, leaving a lower dimensional marginal density over the illuminant parameters.

Simulations

To evaluate the potential of the Bayesian approach, we performed simulations. We computed the sensor data corresponding to simulated scenes and used a Bayesian algorithm to estimate the illuminant from the data. We compared the performance of this algorithm with that of the subspace and gray world algorithms.

Methods

To perform the simulations, we created three dimensional linear models for surfaces and illuminants. We choose a linear model for surfaces by performing principal components analysis on the data of Kelly *et al.*,[19,20] who measured the reflectance functions of 462 Munsell papers. We then found the best fitting model weights for each individual surface in the data set. The distribution of these weights is shown in Figure 1 above. We used the weight means as the entries the prior mean $\mathbf{u}_{sj}$. We formed $\mathbf{K}_{sj}$ as the diagonal matrix whose diagonal entries were the three weight variances. We also investigated prior distributions where the standard deviation of the third basis component was set to 0.1 or 0.5 of its full value.

Large sets of measured illuminants are not readily available. Summaries of such data sets, in the form of linear models, do exist. We took our three dimensional linear model for illuminants to be the CIE linear model for daylights.[21] To generate a prior distribution on the weights of this linear model, we generated a set of randomly drawn CIE daylights with correlated color temperatures drawn according to a Normal distribution with mean 6500° K and standard deviation 4000° K. (Any draws outside the range 3000° to 25000° K were rejected.) We then perturbed the intensities of these illuminants by scale factors drawn uniformly between 1 and 10. Using our data set, we computed the linear model weights on each function and formed $\mathbf{u}_e$ and $\mathbf{K}_e$ as the mean and covariance of these weights.

We simulated scenes consisting of 8 randomly drawn surfaces under a single randomly drawn illuminant. In drawing surfaces and illuminants, we enforced physical realizability constraints. Any surface whose reflectance was negative or exceeded unity at any wavelength was rejected, while any illuminant with negative power at any wavelength was rejected. We used the rendering equation and the Smith-Pokorny estimates of the human cone sensitivities[22] to compute the cone responses. We perturbed the cone responses by additive Gaussian noise with mean zero and standard deviations of either 0.01 or 0.1 times the cone responses of the mean surface under the mean illuminant.

We compared three estimation methods. For the Bayesian algorithm we used Eq. 23 to evaluate the marginal posterior on the illuminant. We then used Monte Carlo methods to estimate the mean of this posterior distribution.[23] This procedure is equivalent to minimizing the Bayes risk with respect to a quadratic loss function. In computing the posterior mean, we added a physical realizability constraint on the posterior. We set the posterior for any illuminant with negative power at any wavelength to zero.

For comparison, we also estimated the illuminant using the subspace and gray world algorithms. Our implementation of the subspace method algorithm follows Wandell.[9] (We have not investigated a recent suggestion by Marimont and Wandell[24] that might lead to improved performance from this algorithm.) Our implementation of the gray world algorithm follows Buchsbaum.[10]

The algorithm of Maloney and Wandell is only designed to recover the relative spectrum of the illuminant. In some of our comparisons, therefore, we scale each estimated illuminant to provide the best fit to the true simulated illuminant. This scaling allows us to evaluate how well the algorithms estimate relative spectra.

As a summary measure of performance, we use the fractional squared estimation error

$$\varepsilon = \frac{|\mathbf{e}-\hat{\mathbf{e}}|^2}{|\mathbf{e}|^2}. \qquad (24)$$

Bayesian

Gray World

Subspace
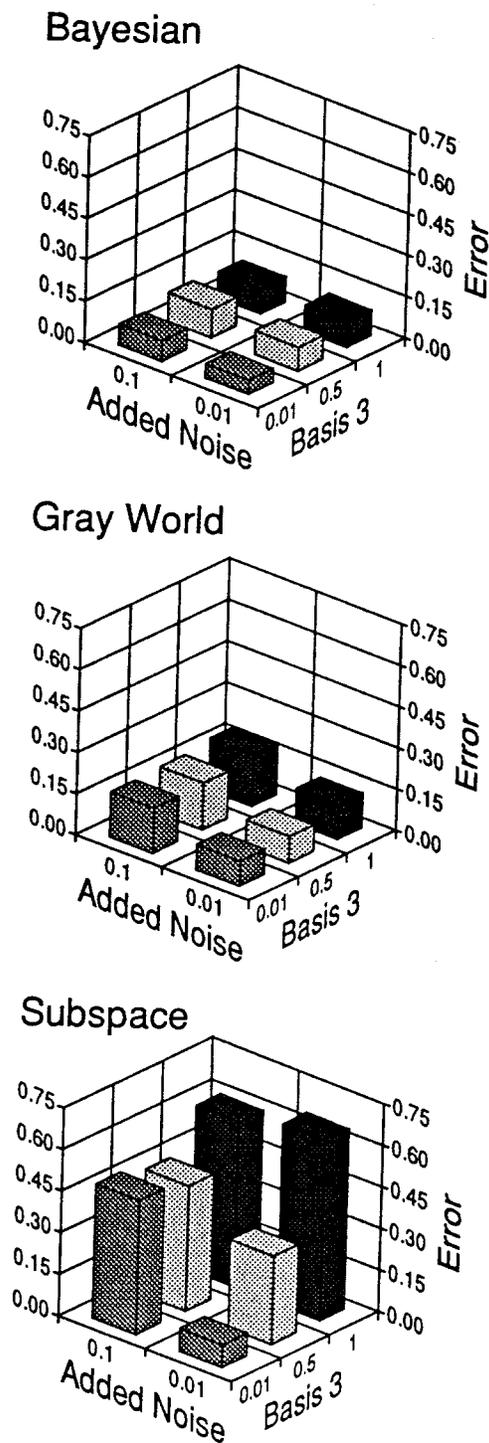
Figure 5. Fixed illuminant simulation. The solid line in each panel shows the spectral power distribution of the simulated illuminant. The dashed lines show the upper and lower envelopes of 10 illuminant estimates. The mean error for the 10 estimates is given in the upper right of each panel. Simulation details: full surface linear model, sensor noise 0.01, results scaled to provide best fit to simulated illuminant.

Results

Figure 5 illustrates the performance of the three algorithms for a fixed illuminant. Each panel of the figure illustrates performance for one of the three algorithms. To generate each panel of the figure, we fixed the simulated illuminant and repeated the simulation for 10 different sets of 8 randomly drawn surfaces. The solid line in each panel shows the simulated illuminant. The dashed lines show the upper and lower envelopes of 10 estimates generated by each algorithm. To generate this figure, each estimated illuminant was scaled in intensity to provide the best fit to the true illuminant.

The subspace algorithm does very poorly for these conditions. This is because the assumptions on which it is based are violated for our simulations. The gray world algorithm does better, capturing the approximate shape of the illuminant. The Bayesian algorithm does the best of all. The fractional error for the Bayesian algorithm is roughly half that of the gray world algorithm.

To get a sense of the conditions under which the algorithms performed well or poorly, we conducted further simulations. In each of these, we simulated performance for 100 draws of a simulated illuminant and a collection of simulated surfaces. For each set of 100 draws, we summarized the performance of each algorithm by the mean fractional error. Again, we scaled each estimated illuminant to the simulated illuminant before computing the error. We performed this set of simulations for a number of different choices of surface prior and level of added sensor noise. The surface priors varied according to the standard deviation of the weight of the third basis function as described above. Figure 6 summarizes the results. Each panel of the figure shows the mean estimation error of one of the algorithms for six combinations of third basis weight standard deviation and levels of added noise. To help understand the error measure, we note that for our illuminant prior, simply estimating the relative illuminant spectrum as the prior mean, without examining the sensor data, leads to a mean fractional error of roughly 0.20.

Again we see that the Bayesian algorithm performs best for all conditions. When the standard deviations of the third basis component and the added noise are small, the subspace algorithm also performs well, as expected. The rapid loss in performance as either of these standard deviations is increased is indicative of the robustness problems encountered with this algorithm. The gray world algorithm is robust, but never performs as well as the Bayesian algorithm.

Figure 7 illustrates the effect of scaling on the estimates. Both the Bayesian and the gray world algorithms are designed to recover not only the relative spectrum of the illuminant but also the absolute spectrum. In Figure 7 we plot, in the same manner as in Figure 5, the unscaled illuminant estimates for these two algorithms. Performance degrades relative to that shown in Figure 5, but for the Bayesian algorithm the mean error is at the respectable level of 0.10

Discussion

The Bayesian Framework

It should not be too surprising that the Bayesian algorithm outperforms the subspace and gray world algorithms. The Bayesian algorithm is constructed to minimize the expected error. As we noted above, there are two reasons we can expect it to do better. First, it takes advantage of all of

the information in the sensor data; it is not restricted to use only the information available in a particular summary statistic. Second, our assumption that illuminants and surfaces follow a Normal distribution is an added assumption not incorporated into the other two algorithms. Since this assumption held true for our simulation conditions, its incorporation was advantageous..
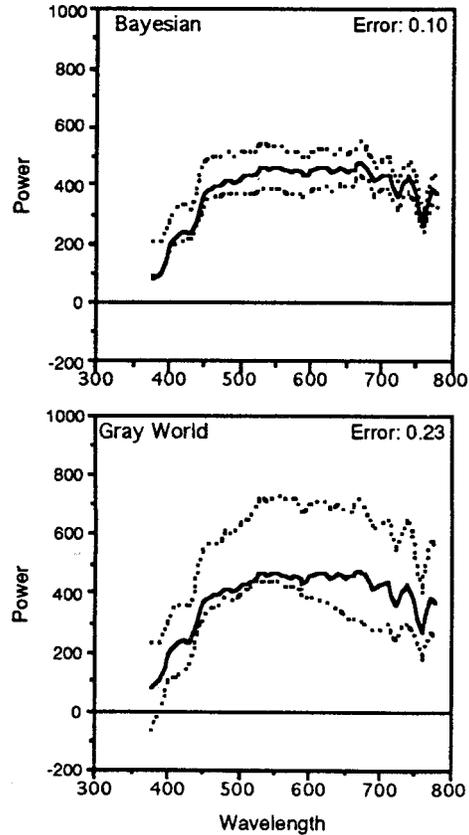
Figure 7. Fixed illuminant simulations. The solid line in each panel shows the spectral power distribution of the simulated illuminant. The dashed lines show the upper and lower envelopes of 10 illuminant estimates. The mean error for the 10 estimates is given in the upper right of each panel. Simulation details: full surface linear model, sensor noise 0.01, no scaling of results.

The distribution of weights shown in Figure 1 strongly suggests that it is reasonable to make some type of assumption about the distribution of linear model weights. One reason to expect the weights to cluster is that surfaces and illuminants are subject to physical realizability constraints. Illuminants may not have negative power. Surface reflectances must lie between zero and unity. Forsyth[25] argued that combining linear model and physical realizability constraints leads to improved color constancy algorithms. D'Zmura and Iverson[15] have recently reached similar conclusions by using a Bayesian approach to combine the two types of constraints. Our use of a Normal prior distribution captures the type of weight clustering that as is imposed by physical realizability constraints, a point also made by Trussell and Vrhel.[13]

The Normal distribution also captures a second feature of the weight distributions shown in Figure 1, a feature that is not captured by a physical realizability constraint: the variance on the weights of the three basis functions is not uniform. There is no reason in principle that physical realizability constraints cannot be added to our description of the weight priors. The difficulty is computational. The simplification provided by Eq. 23 does not apply for non-Normal distributions.

Figure 6. Multiple illuminant simulations. The mean error for each algorithm is plotted for six simulation conditions. The conditions vary in the size of the standard deviation of the third basis component and in the bvel of added noise

Other Approaches to Computational Color Constancy

Rather than seeking to invert the rendering model described by Eqs. 1-5, other authors have considered more complex viewing models. These authors conclude that additional information about the illuminant may be gained by considering interreflection between surfaces,[26] specular highlights,[27,28] and variation in the illumination.[29,30] Our approach complements this work in that we examine how better performance may be achieved for a simple viewing geometry. We expect that insight gained from our work could be extended to the more complex viewing geometries.

Future Directions

There are three keys to successfully using a Bayesian approach. The first is to formulate priors appropriate to the problem at hand. We view our use of the Normal prior as only a first step towards doing this. In the absence of additional data about the distribution of natural surfaces and illuminants, one can investigate how robust Bayesian algorithms are with respect to misspecification of the prior distribution. Second, as illustrated by the product example, the choice of loss function can be critical. We do not believe that either the quadratic nor the delta loss functions are the best choice for perception problems. The local mass loss function may be better. A related question of particular interest is how strongly estimates of illuminant and surface spectra vary with asymmetries in the loss function. This is intriguing in part because the question of whether human performance exhibits such variation could be investigated psychophysically. Finally, the use of Bayesian methods depends on finding computationally efficient methods for minimizing the Bayes risk.

Acknowledgments

References

1. Evans, R.M. *The Perception of Color*, (Wiley: NY, 1974).
2. McCann, J. J., McKee, S. P. and Taylor, T. H., "Quantitative studies in retinex theory: A comparison between theoretical predictions and observer responses to the 'Color Mondrian' experiments," *Vis. Res.* **16**, 445-458 (1976).
3. Brainard, D. H., Wandell, B. A. and Chichilnisky, E. J., "Color constancy: from physics to appearance," *Current Directions in Psychological Science* **2**, 165-170 (1993).
4. Menke, W. Geophysical data analysis: discrete inverse theory, (Academic Press: San Diego, 1989).
5. Judd, D. B., MacAdam, D. L. and Wyszecki, G. W., "Spectral distribution of typical daylight as a function of correlated color temperature," *J. Opt. Soc. Am.* **54**, 1031 (1964).
6. Maloney, L. T., "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *J. Opt. Soc. Am. A* **3**, 1673-1683 (1986).
7. Jaaskelainen, T., Parkkinen, J. and Toyooka, S., "A vector-subspace model for color representation," *J. Opt. Soc. Am. A* **7**, 725-730 (1990).
8. Maloney, L. T. and Wandell, B. A., "Color constancy: a method for recovering surface spectral reflectances," *J. Opt. Soc. Am. A* **3**, 29-33 (1986).
9. Wandell, B. A., "The synthesis and analysis of color images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-9**, 2- 13 (1987).
10. Buchsbaum, G., "A spatial processor model for object colour perception," *Journal of the Franklin Institute* **310**, 1 -26 (1980).
11. Brainard, D. H. and Wandell, B. A., "Analysis of the retinex theory of color vision," *J. Opt. Soc. Am. A* **3**, 1651-1661 (1986).
12. Berger, T. O. *Statistical Decision Theory and Bayesian Analysis*, (Springer-Verlag: New York, 1985).
13. Trussell, H. J. and Vrhel, M. J., "Estimation of illumination for color correction," *Proceedings of ICASSP* 2513-2516 (1991).
14. Vrhel, M. J. and Trussell, H. J., "Color correction using principal components," *Col. Res. Appl.*, **17**, 329-337 (1992).
15. D'Zmura, M., Iverson, G. and Singer, B., "Probabilistic color constancy." In *Geometric Representations of Perceptual Phenomena: Papers in Honor of Tarow Indow's 70th Birthday*; R. D. Luce et al., eds., (Lawrence Erlbaum Associates: Hillsdale, NJ, in press).
16. Yuille, A. L. and Bulthoff, H. H. "Bayesian decision theory and psychophysics," Max-Planck-Institut fur biologische Kybernetik, Tubingen, Germany, *Technical Report* **MPI-BK** Bu2 (1993).
17. Freeman, W. T., "The generic viewpoint assumption in a framework for visual perception," *Nature* (in press).
18. Freeman, W. T., "Exploiting the generic view assumption to estimate scene parameters," *4th Intl. Conf. Computer Vision, IEEE* 347 - 356 (1993).
19. Kelly, K. L., Gibson, K. S. and Nickerson, D., "Tristimulus specification of the Munsell Book of Color from spectrophotometric measurements," *J. Opt. Soc. Am.* **33**, 355-376 (1943).
20. Nickerson, D. "Spectrophotometric data for a collection of Munsell samples," *U.S. Dept. of Agriculture Report*, 1957.
21. CIE "*Colorimetry*," 2nd ed., Bureau Central de la CIE, 1986.
22. DeMarco, P., Pokorny, J. and Smith, V. C., "Full-spectrum cone sensitivity functions for X-chromosome-linked anomalous trichromats," *J. Opt. Soc. Am. A* **9**, 1465-1476 (1992).
23. Press, W. H., Flannery, B. P., Vetterling, W. T. and Teukolsky, S. A. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed., (Cambridge University Press: Cambridge, 1992).
24. Marimont, D. H. and Wandell, B. A., "Linear models of surface and illuminant spectra," *J. Opt. Soc. Am. A* **9**, 1905-1913 (1992).
25. Forsyth, D. A., "A novel approach to colour constancy," *International Conference Comp. Vis*. 9-18 (1988).
26. Drew, M. S. and Funt, B. V., "Variational Approach To Interreflection In Color Images," *J. Opt. Soc. Am. A* **9**, 1255-1265 (1992).
27. Lee, H., "Method for computing the scene-illuminant chromaticity from specular highlights," *J. Opt. Soc. Am. A* **3**, 1694-1699 (1986).
28. D'Zmura, M. and Lennie, P., "Mechanisms of color constancy," *J. Opt. Soc. Am. A* **3**, 1662-1672 (1986).
29. D'Zmura, M. and Iverson, G., "Color constancy. I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces," *J. Opt. Soc. Am. A* **10**, 2148-2165 (1993).
30. D'Zmura, M. and Iverson, G., "Color constancy. II. Results for two-stage linear recovery of spectral descriptions for lights and surfaces," *J. Opt. Soc. Am. A* **10**, 2166-2180 (1993).