Bayesian color constancy

David H. Brainard

Department of Psychology, University of California, Santa Barbara, California 93106

William T. Freeman

MERL, a Mitsubishi Electric Research Laboratory, Cambridge, Massachusetts 02139

Received August 30, 1996; revised manuscript received December 23, 1996; accepted January 7, 1997

The problem of color constancy may be solved if we can recover the physical properties of illuminants and surfaces from photosensor responses. We consider this problem within the framework of Bayesian decision theory. First, we model the relation among illuminants, surfaces, and photosensor responses. Second, we construct prior distributions that describe the probability that particular illuminants and surfaces exist in the world. Given a set of photosensor responses, we can then use Bayes's rule to compute the posterior distribution for the illuminants and the surfaces in the scene. There are two widely used methods for obtaining a single best estimate from a posterior distribution. These are maximum *a posteriori* (MAP) and minimum mean-squared-error (MMSE) estimation. We argue that neither is appropriate for perception problems. We describe a new estimator, which we call the maximum local mass (MLM) estimate, that integrates local probability density. The new method uses an optimality criterion that is appropriate for perception tasks: It finds the most probable approximately correct answer. For the case of low observation noise, we provide an efficient approximation. We develop the MLM estimator for the color-constancy problem in which flat matte surfaces are uniformly illuminated. In simulations we show that the MLM method performs better than the MAP estimator and better than a number of standard color-constancy algorithms. We note conditions under which even the optimal estimator produces poor estimates: when the spectral properties of the surfaces in the scene are biased. [S0740-3232(97)01607-4]

1. INTRODUCTION

Vision tells about object properties. This is difficult because there is no simple mapping between an object's intrinsic properties and the light that reaches the observer. Extrinsic factors, such as where the object is located, how it is oriented, and how it is illuminated, also influence the image. To provide reliable information about objects, a visual system must separate the contributions of intrinsic and extrinsic factors. Often, however, there is no unique factorization: Many combinations of intrinsic and extrinsic factors can result in the same set of sensor measurements. In spite of this ambiguity, our own perceptual systems perform remarkably well.

Bayesian decision theory[1–3] provides a framework for designing solutions to perceptual problems. Bayesian methods combine the information contained in the image with information given *a priori* about likely physical configurations of the world. This prior information can resolve ambiguities in the image data.

In this paper we provide a Bayesian treatment of the perceptual problem of determining object color.[4] In the case of object color, the light reflected from an object depends not only on the object's intrinsic surface reflectance but also on the spatial and spectral power distribution of the illumination and the orientation of the object relative to sources of illumination and the viewer. The ability of a visual system to maintain object color appearance across variations in factors extrinsic to the object is called color constancy. Human vision exhibits approximate color constancy.[10–15]

A. Why Color Constancy is Difficult

1. Problem Statement

Consider the special case of a collection of N matte surfaces diffusely illuminated by a single illuminant, as illustrated in Fig. 1. For this viewing geometry, we may characterize each surface by its spectral reflectance function. We represent the reflectance function of the jth surface by a column vector $\mathbf{s}_j$. The entries of $\mathbf{s}_j$ specify the fraction of incident light reflected in N_l evenly spaced wavelength bands throughout the visible spectrum. Similarly, we specify the illuminant spectral power distribution with the column vector $\mathbf{e}$. The entries of $\mathbf{e}$ 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 $\mathbf{c}_j = \mathbf{e} .* \mathbf{s}_j$, where we use the notation $.*$ to denote entry-by-entry multiplication.

The imaging device samples each $\mathbf{c}_j$ with N_r classes of linear photosensors. Each photosensor is characterized by a spectral sensitivity function. We specify the spectral sensitivities with an $(N_r \times N_l)$-dimensional matrix $\mathbf{R}$. The klth element of $\mathbf{R}$ specifies the sensitivity of the kth sensor class to light in the lth wavelength band. We let the N_r-dimensional column vector $\mathbf{r}_j$ represent the responses from all N_r sensor classes to the spectrum $\mathbf{c}_j$. In the absence of sensor noise, we have the rendering equation for a single surface:

$$\mathbf{r}_j = \mathbf{R}\mathbf{c}_j = \mathbf{R}(\mathbf{e} .* \mathbf{s}_j). \tag{1}$$

This equation specifies the relation between the data available to the visual system at an image location (pho-

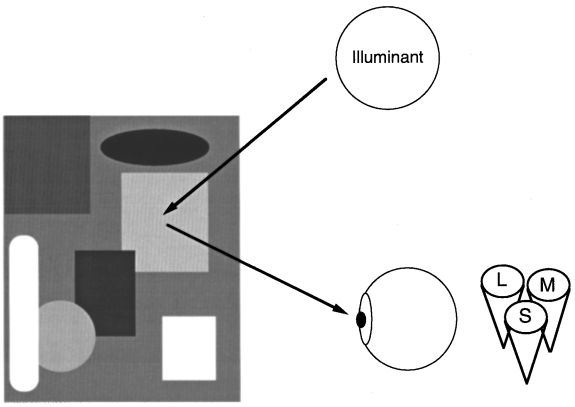


Fig. 1. Image formation for a simple geometry. The image is formed when light from an illuminant reflects off a collection of surfaces to the imaging device. We assume that the illuminant is diffuse and spatially uniform, so that it may be characterized by a single spectral power distribution. We assume that each surface is flat and matte, so that it may be characterized by a single spectral reflectance function. The spectral power distribution of the light reaching the observer from each surface is given as the wavelength-by-wavelength product of the illuminant spectral power distribution and the surface reflectance function. At each location this light is sampled by a small number of types of photosensors. For the case of human vision (as shown), these are the long-, medium-, and short-wavelength (LMS) sensitive cones. For a typical digital camera, these are the red, green, and blue (RGB) sensors.

tosensor responses $\mathbf{r}_j$) and the scene parameters that it must estimate (illuminant spectrum $\mathbf{e}$ and surface reflectance spectra $\mathbf{s}_j$).

We cast color constancy as the problem of estimating $\mathbf{e}$ and the $\mathbf{s}_j$ from the ensemble of sensor responses $\mathbf{r}_j$. Once the estimates are obtained, color-constant descriptors for surfaces can be constructed by using any fixed function of the surface reflectance estimates or the reflectance estimates themselves.

Forsyth[16] and Brainard *et al.*[17] (see also McCann *et al.*,[11] Land and McCann,[18] and Land[19,20]) have noted that constancy may be achieved even when the estimates of the physical parameters are not correct: All that is required is that the estimates not vary with the illuminant. Nonetheless, the ability to estimate the physical parameters with precision is sufficient to achieve constancy. We return to this point in Section 4.

2. *Why It Is Difficult*

The inverse problem specified by Eq. (1) is difficult for two reasons: It is underdetermined and it is nonlinear. 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 by 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 $N_l = 31$. On the other hand, a typical imaging device (e.g., a RGB camera or the human eye), contains only a few classes of photosensors. Indeed, the $\mathbf{r}_j$ are often of dimension 3. If we have data from N image locations (say, 10) and assume one illuminant, then we have NN_r measurements (e.g., $10 \times 3 = 30$) available to estimate $N_l(N + 1)$ scene parameters [e.g., $31 \times (10 + 1) = 341$].

Equation (1) represents a nonlinear inverse problem because some scene parameters multiply others. Although standard methods exist for solving linear inverse problems,[21] no general solution is available for the nonlinear case.

A general solution to color constancy must handle the geometrical richness of natural image formation. Human observers, however, show some constancy for the simple geometry of Fig. 1,[11,14,22–29] and this situation has received considerable attention in the computational literature.[7–9,16,19,20,30–34] Insights gained from studying the simple geometry may help us to extend the analyses of color constancy to natural images. In addition, the underdetermined and nonlinear character of the simple color-constancy problem may make it a good model for other computational vision problems. The aperture problem in motion perception,[35] the bas-relief illusion of shape perception,[36] the interpretation of line drawings,[37,38] the separation of shading and reflectance,[39] and the selection of perceptual grouping[40] are examples of other underdetermined vision problems for which a similar analysis may be useful.

B. Previous Work

To address the underdeterminancy of color constancy, previous investigators have described spectral functions by using low-dimensional linear models. Brainard[41] reviews the use of linear models in computational color vision. Let $\mathbf{B}_e$ be an $(N_l \times N_e)$-dimensional matrix. 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. The columns of $\mathbf{B}_e$ are 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, 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 $(N_l \times N_s)$-dimensional matrix and $\mathbf{w}_{sj}$ is an N_s-dimensional column vector.

If we assume that a population of spectra lie within an N_m-dimensional linear model, then we can parameterize the spectra by specifying the model weights. (The basis functions are assumed to be fixed and known.) When N_m is small, linear models provide compact descriptions of spectra. Naturally occurring spectra are well described by low-dimensional linear models. A four-dimensional model characterizes a large sample of measured daylights.[42] The dimension required for large samples of measured surfaces lies somewhere between 6 and 8.[43,44] Moreover, linear models with dimension as low as 3 capture a large percentage of the variance of the measured surface spectra. We lose no generality by reparameterizing the color-constancy problem in terms of linear model weights. If we choose $\mathbf{B}_e$ and $\mathbf{B}_s$ as the identity matrices, the linear model parameterization reduces to the spectral parameterization.

Buchsbaum[30] developed an algorithm that requires that the illuminants and the surfaces 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 $\mathbf{s}_0$. In such a gray world, the illuminant can be estimated from the spatial mean of the sensor responses.

Algorithms based on the gray world assumption break down for images for which the mean reflectance function differs from $\mathbf{s}_0$.[45] That is, these algorithms cannot distinguish changes of illumination from changes in the collection of surfaces in a scene. Both changes affect the spatial mean of the sensor responses and are interpreted as a change in illumination. The human visual system, on the other hand, seems capable of distinguishing true illuminant changes from changes in the collection of surfaces in a scene.[15,22,46,47] There is also evidence that color appearance is affected by factors other than the mean of the photosensor responses.[48–52]

Maloney and Wandell[31] and Wandell[53] showed that the rendering equation may be inverted exactly if the illuminants are constrained to lie within an N_r-dimensional linear model and that the surfaces are constrained to lie within an $(N_r - 1)$-dimensional linear model. 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.

An important feature of Maloney and Wandell's subspace algorithm is that the estimate of the illuminant does not depend on the particular collection surfaces in the scene. For human vision, however, there are only three classes of cone photosensors, so that the 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 simulations to investigate this question[5] and concluded that the subspace algorithm is not robust (see Subsection 3.B).

Forsyth[16] used linear models in conjunction with the constraints that real illuminants cannot have negative power at any wavelength and that real (nonfluorescent) surface reflectance functions must take on values between 0 and 1 at all wavelengths. He showed that examining the convex hull of the sensor responses $\mathbf{r}_j$ provided information about the illuminant. Others have also argued that physical realizability constraints provide useful information for estimating the illuminant.[9,34]

The gray world, subspace, and physical realizability algorithms all work by extracting a summary statistic from the sensor responses and then using this statistic to estimate the illuminant. For the gray world algorithm, the statistic is the mean sensor response. For the subspace algorithm, the statistic is the identity of the subspace that contains the responses. For the physical realizability algorithm, the statistic is the convex hull of the sensor responses. The algorithms demonstrate that each of these statistics carries information about the illuminant. On the other hand, there is no reason to use only one statistic. If different statistics carry different information, then we might expect improved performance from an algorithm that uses information from all of them. To capitalize on this intuition, we must know how to combine the information in an effective manner.

2. BAYESIAN FRAMEWORK

The Bayesian framework provides 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 gray world, subspace, and physical realizability algorithms. Three probability distributions play key roles. These are the prior, the posterior, and the likelihood. The prior probability describes what is known about the parameters before observing the data, while the posterior probability describes what is known after observing the data. If we are trying to estimate parameters described by the vector $\mathbf{x}$, then the prior information is the probability density $p(\mathbf{x})$.[54] The likelihood $p(\mathbf{y}|\mathbf{x})$ expresses the relation between the data $\mathbf{y}$ and the parameters $\mathbf{x}$. The likelihood may be thought of as the rendering equation expressed as a probability distribution.

Given the prior $p(\mathbf{x})$ and the likelihood $p(\mathbf{y}|\mathbf{x})$, the posterior probability $p(\mathbf{x}|\mathbf{y})$ is computed by using Bayes's rule:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = Cp(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \tag{2}$$

In this paper we will use the symbol 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 $\mathbf{x}$.

To go from the posterior to a single estimate $\tilde{\mathbf{x}}$ 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 Bayesian expected loss[1]:

$$\bar{L}(\tilde{\mathbf{x}}|\mathbf{y}) = \int_{\mathbf{x}} L(\tilde{\mathbf{x}},\mathbf{x})p(\mathbf{x}|\mathbf{y})\mathrm{d}\mathbf{x}. \tag{3}$$

We choose $\tilde{\mathbf{x}}$ to minimize the expected loss. Often the loss function is shift invariant, so that the loss depends only on the difference $\tilde{\mathbf{x}} - \mathbf{x}$. In this case the expected loss is simply the posterior convolved by the loss function with its argument negated.

Bayesian estimation provides a principled way to choose an optimal estimate that uses all of the information contained in the data (including the information carried by the mean of the data, the subspace containing the data, and the convex hull of the data). As a practical matter, a number of difficulties can arise. First, it may be difficult to specify a prior distribution that adequately captures what is known about the structure of the parameters. In the case of color constancy, the prior must specify how likely it is that particular illuminant spectral power distributions and surface reflectances will occur. Second, it may be difficult to specify a loss function that captures how costly errors of various type are. Finally, it may be computationally difficult to minimize the expected

loss. For the case of color constancy, however, we have developed a Bayesian solution.

A. Bilinear Structure

The color-constancy rendering equation [Eq. (1)] is bilinear.[27,55] This means that the relation between $\mathbf{r}_j$ and $\mathbf{s}_j$ is linear when $\mathbf{e}$ is held fixed and that the relation between $\mathbf{r}_j$ and $\mathbf{e}$ is linear when $\mathbf{s}_j$ is held fixed. Any results that hold generally for bilinear rendering equations[56–58] apply to color constancy.

To demonstrate the bilinearity, we rewrite the rendering equation [Eq. (1)] as

$$\mathbf{r}_j = \mathbf{L}(\mathbf{e})\mathbf{s}_j = \mathbf{L}(\mathbf{s}_j)\mathbf{e}, \tag{4}$$

where $\mathbf{L}(\mathbf{e}) = \mathbf{R}\,\mathrm{diag}(\mathbf{e})$ is an $(N_r \times N_l)$-dimensional matrix that depends only on $\mathbf{e}$ and $\mathbf{L}(\mathbf{s}_j) = \mathbf{R}\,\mathrm{diag}(\mathbf{s}_j)$ is an $(N_r \times N_l)$-dimensional matrix that depends only on $\mathbf{s}_j$. Expressing the rendering equation in terms of linear model weights does not alter its bilinearity. By letting $\mathbf{M}(\mathbf{w}_e) = \mathbf{L}(\mathbf{B}_e\mathbf{w}_e)\mathbf{B}_s$ and $\mathbf{M}(\mathbf{w}_{sj}) = \mathbf{L}(\mathbf{B}_s\mathbf{w}_{sj})\mathbf{B}_e$, we have

$$\mathbf{r}_j = \mathbf{M}(\mathbf{w}_e)\mathbf{w}_{sj} = \mathbf{M}(\mathbf{w}_{sj})\mathbf{w}_e. \tag{5}$$

Equation (5) may be extended to express the relation between the data and scene parameters at all N locations simultaneously. Let $\mathbf{r}$ be the NN_r-dimensional column vector obtained by stacking the vectors $\mathbf{r}_j$. Let $\mathbf{w}_s$ be the NN_s-dimensional column vector obtained by stacking the vectors $\mathbf{w}_{sj}$. Then we can write the overall rendering equation as

$$\mathbf{r} = \mathbf{N}(\mathbf{w}_e)\mathbf{w}_s = \mathbf{N}(\mathbf{w}_s)\mathbf{w}_e. \tag{6}$$

Here the matrix $\mathbf{N}(\mathbf{w}_e)$ is the $NN_r \times NN_s$ block diagonal matrix with $\mathbf{M}(\mathbf{w}_e)$ repeated along its diagonal, while the matrix $\mathbf{N}(\mathbf{w}_s)$ is the $NN_r \times N_e$ matrix obtained by stacking the matrices $\mathbf{M}(\mathbf{w}_{sj})$.

B. Product Example

The simplest case of a bilinear inverse problem occurs when the data and the scene parameters are all scalars. Suppose that we observe a number y and are told that it is the product of two other numbers, $y = ab$. The problem is to estimate the two numbers a and b.

This problem is clearly underdetermined. Let us say that the observation is $y = 1$ and that we know that $0 < a, b < 4$. From the constraint $ab = 1$, we can say only that the solution must lie on the ridge of the surface shown in Fig. 2(a).

A Bayesian analysis yields more. We start by finding the posterior probability $p(\mathbf{x}|\mathbf{y})$ of the parameter vector $\mathbf{x} = (a \;\; b)^T$ given the observation vector $\mathbf{y} = y$. If we write the rendering equation as $\mathbf{y} = f(\mathbf{x}) = ab$ and allow for normally distributed observation noise with mean 0 and variance σ^2, then we can write the likelihood as

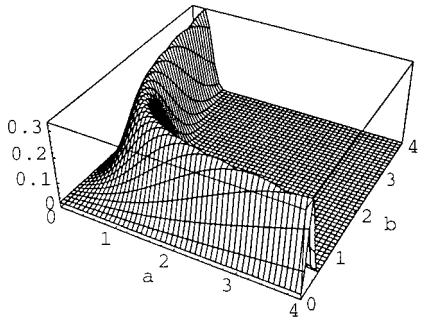
$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{|\mathbf{y} - f(\mathbf{x})|^2}{2\sigma^2}\right]. \tag{7}$$

If we assume a uniform prior $p(\mathbf{x}) = 1/16$ over the range $[0, 4] \times [0, 4]$ and 0 elsewhere, Bayes's rule [Eq. (2)] gives us the posterior for the case $y = 1$ and $\sigma^2 = 0.18$:

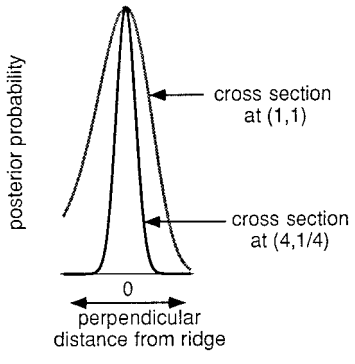
$$p(\mathbf{x}|\mathbf{y}) = \begin{cases} C \exp\left[\dfrac{|1 - ab|^2}{2(0.18)}\right], & 0 < a, b < 4 \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

where C is a normalizing constant. This posterior distribution is shown in Fig. 2(a). Points (a, b) 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 forms a ridge along the hyperbola $ab = 1$. Note from Fig. 2(b) that, while the ridge has equal height everywhere, it is wider near (1, 1) than at other points.

The posterior probability distribution provides a complete description of what we know, given the data and the prior. In a typical estimation problem, however, the goal is to choose one best estimate of the scene parameters. For the product example, we seek an estimate $\tilde{\mathbf{x}} = (\tilde{a} \;\; \tilde{b})^T$. In practice, two decision rules are almost universally used. These are to choose the scene parameter value that maximizes the posterior distribution (called the maximum *a posteriori*, or MAP, estimate) or to choose the mean of the posterior distribution (called the minimum mean-squared-error, or MMSE, estimate).



(a)



(b)

Fig. 2. Bayesian analysis of the product example: (a) posterior probability for the observed data $ab = 1$ for Gaussian observation noise of variance $\sigma^2 = 0.18$ and uniform prior probabilities over the plotted region, (b) cross section through the posterior at two different locations. Note the different thicknesses of the ridge; some local regions have more probability mass than others, even though the entire ridge has a constant maximum height.

The MAP estimator is closely related to maximum-likelihood methods and has been widely used in computational vision.[3,7,9,59–65]

The MAP rule corresponds to minimizing the expected loss with respect to the minus delta loss function:

$$L(\tilde{\mathbf{x}}, \mathbf{x}) = -\delta(\tilde{\mathbf{x}} - \mathbf{x}). \tag{9}$$

Since convolving the posterior with the minus delta loss function leaves the relative shape of the posterior unchanged, the estimate that maximizes the posterior also minimizes the corresponding expected loss. For the product example, the minus delta loss function and its corresponding expected loss are shown in Figs. 3(a) and 3(d), respectively. (In this and related figures, decreasing loss is plotted upward to show the extrema more clearly.) Every point along the hyperbolic ridge in the figure has equal expected loss, so the MAP rule does not provide a unique estimate. Such ridges can occur in underdetermined estimation problems when many alternatives account for the observations equally well. MAP estimation ignores variations in the shape of the ridge, which can be a significant source of information.

The MMSE rule is to choose the posterior mean as the best estimate for the scene parameters. This estimator is also in wide use; it is, for example, the basis of Kalman filtering.[66] It is easily shown[67] that the MMSE rule corresponds to minimizing the expected loss with respect to the squared-error loss function:

$$L(\tilde{\mathbf{x}}, \mathbf{x}) = |\tilde{\mathbf{x}} - \mathbf{x}|^2. \tag{10}$$

For the product example, the squared-error loss function and its corresponding expected loss are shown in Figs. 3(b) and 3(e), respectively. Note that the estimate that minimizes the expected loss is $\tilde{a} = \tilde{b} = 1.3$, a strange result given the observation $ab = 1$. The MMSE rule is sensitive to the structure of the posterior, but it leads to an estimate that is very unlikely to have generated the data. Furthermore, computing the MMSE estimate may require a computationally intensive integration over the entire scene parameter space.

When the posterior probability mass is well localized in the scene parameter space, both the MAP and MMSE rules provide intuitively appealing estimates that tend to agree with each other. When the posterior is less simple, as in our product example, the MAP and MMSE rules can disagree and both can be unsatisfactory unless the actual loss function for the estimation problem matches that implied by the estimator. We believe that complicated posterior probability distributions, where these effects matter, occur in real computational vision problems. Therefore we believe that it is worth considering what is an appropriate loss function for perception.

The minus delta function loss function implies that small estimation errors are as bad as large ones. The squared-error loss function provides for a loss that accelerates with the size of the estimation error. But in perception an estimate that is approximately correct will often do, and once the estimation error is sufficiently large, the loss may saturate. For example, if we are trying to catch a ball thrown to us, small errors in the perceived size or velocity of the ball will not cause us to drop the ball. Once the errors are large enough so that we fail to catch the ball, it does not really matter how large they are. We define the local mass loss function as

$$L(\tilde{\mathbf{x}}, \mathbf{x}) = -\exp[-|\mathbf{K}_L^{-1/2}(\tilde{\mathbf{x}} - \mathbf{x})|^2], \tag{11}$$

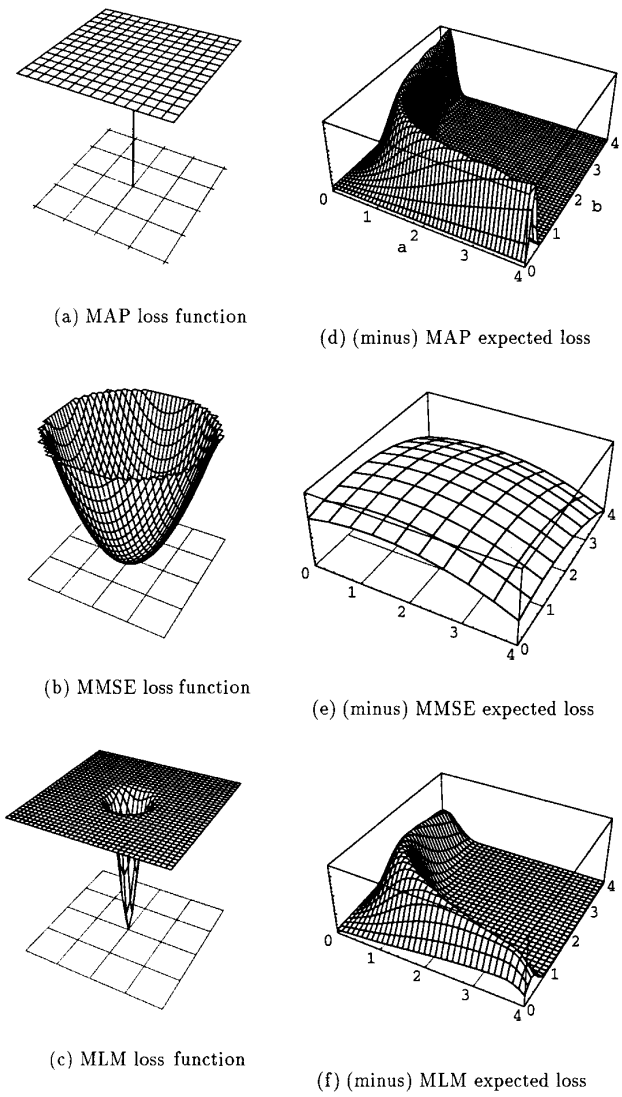


Fig. 3. Product example loss functions and corresponding expected losses. (a)–(c): Three loss functions. The plots show the penalty for guessing parameter values offset from the actual value, taken to be the plot center. Each loss function is shift invariant. (a) Minus delta function loss, implicit in MAP estimation. The penalty is constant except for the correct estimate. (b) Squared-error loss (a parabola), implicit in MMSE estimation. Very inaccurate estimates can carry inordinate influence. (c) Local mass loss function. Nearly correct estimates are rewarded, while all others carry nearly equal penalty. (d)–(f): Corresponding expected loss for the product example. Note that the plots are inverted, so that increasing height on the plots represents decreasing loss. This convention allows better visualization of the location of minimum expected loss. (d) The expected loss for the MAP estimator is minus the posterior probability. There is no unique point of minimum expected loss. (e) Expected loss for the MMSE estimator. The MMSE estimate (1.3, 1.3) does not lie along the ridge of solutions to $ab = 1$. (f) Expected loss for the MLM estimator. The local mass loss favors the point (1.0, 1.0), where the ridge of high probability is widest. There is the most probability mass in the local neighborhood of this estimate.

where we adopt the notation that $|\mathbf{K}^{-1/2}\mathbf{x}|^2 = \mathbf{x}^T\mathbf{K}^{-1}\mathbf{x}$. For matrices $\mathbf{K}_L$ of sufficiently small eigenvalues, this loss function rewards approximately correct estimates and penalizes all grossly incorrect estimates equally. We believe that the local mass loss function describes the actual loss for perception problems better than either the minus delta loss function or the squared-error loss function. We refer to the estimator obtained by minimizing the expected loss for the local mass loss function as the maximum local mass (MLM) estimator.

For the product example, the local mass loss function and its corresponding expected loss are shown in Figs. 3(c) and 3(f), respectively. The expected loss is obtained from the posterior by integrating the probability mass in the neighborhood of that estimate. Note that the MLM estimate is $\tilde{a} = \tilde{b} = 1.0$, which accounts well for the observation $ab = 1$. This is the estimate that is most probable to be approximately correct. The ridge of the posterior is widest near (1.0, 1.0) (see Fig. 2). More parameter values near (1.0, 1.0) could have caused the observed datum than those near any other estimate.

The loss function is important and can change the optimal estimate. The MAP estimator is sensitive only to the height of the posterior and ignores important information in its structure. For example, when applied to underdetermined inverse problems, MAP estimation leads to a unique solution only when prior information favors one explanation of the observed data over the others. In contrast, the MLM rule can in some cases provide a unique estimate even when the prior information does not distinguish among parameter values that can explain the data well. This is because the local mass loss function exploits information from both the height and the shape of the posterior.

The MMSE estimator, while optimal in the MSE sense, can give estimates that are nonsensical from a perceptual standpoint. In contrast, the MLM rule leads to estimates that are consistent with the observed data because the local mass function is set to be local in the parameter space. In addition, we show in Appendixes A and B that the MLM estimate can be computed efficiently.

3. BAYESIAN COLOR CONSTANCY

To apply the MLM estimator to color constancy requires finding a computable expression for the expected loss. Once this is obtained, numerical minimization methods may be used to estimate illuminant and surface spectra from photosensor responses. In Appendixes A and B, we derive an analytic approximation to the expected loss for the local mass loss function. Appendix A derives a general result that may be used with any rendering function and prior. The expression that we derive there depends on the derivatives of the rendering function. In Appendix B we calculate these derivatives for the special case in which the rendering function is bilinear.

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 characterized the MLM estimator and compared its performance with that of other constancy algorithms.

A. General Simulation Methods

1. Choice of Priors

We express our knowledge about illuminants and surfaces as a probability distribution. To do so, we build on the low-dimensional linear model approach. We took a set of Munsell paper surface reflectances[68,69] and used principal-components analysis to find the three-dimensional linear model that provided the best approximation to the spectra.[70] Figure 4 shows histograms of the model weights required to fit each reflectance. The weights for each basis are not distributed uniformly over all possible values; they cluster around a central value. This suggests that it is reasonable to make additional assumptions about surface reflectances beyond the fact that they are well described by low-dimensional linear models. The solid curves in the figure show normal probability-density functions with the same mean and variance as those of the weights. The normal densities fit the histograms well, except that the histogram for the first basis function is severely truncated relative to the corresponding normal density function. This truncation may reflect the fact that surface reflectance functions are constrained to lie between 0 and 1.

The analysis shown in Fig. 4 suggests that we can obtain a good description of naturally occurring surfaces by using truncated normal distributions over the weights of low-dimensional linear models. We assume that the prior for each surface is determined by a truncated multivariate normal distribution on the weights of a linear model, so that $\mathbf{s}_j = \mathbf{B}_s\mathbf{w}_{sj}$ and

$$p(\mathbf{w}_{sj}) = \begin{cases} N(\mathbf{u}_{sj}, \mathbf{K}_{sj}), & \mathbf{B}_s\mathbf{w}_{sj} \text{ realizable} \\ 0, & \text{otherwise} \end{cases}. \quad (12)$$

For surfaces, realizable means that each entry of $\mathbf{B}_s\mathbf{w}_{sj}$ lies between 0 and 1. Similarly, we assume that the illuminant prior is determined by a truncated multivariate normal distribution on the weights of a linear model, so that $\mathbf{e} = \mathbf{B}_e\mathbf{w}_e$ and

$$p(\mathbf{w}_e) = \begin{cases} N(\mathbf{u}_e, \mathbf{K}_e), & \mathbf{B}_e\mathbf{w}_e \text{ realizable} \\ 0, & \text{otherwise} \end{cases}. \quad (13)$$

For illuminants, realizable means that each entry of $\mathbf{e} = \mathbf{B}_e\mathbf{w}_e$ is greater than 0. We assume that the surface reflectances at each location are independent and identically distributed, so that we may write

$$p(\mathbf{w}_s) = \prod_{j=1}^{N_s} p(\mathbf{w}_{sj}). \quad (14)$$

Finally, by assuming that the illuminant and the surfaces are statistically independent, we have

$$p(\mathbf{w}_e, \mathbf{w}_s) = p(\mathbf{w}_e)p(\mathbf{w}_s). \quad (15)$$

Our use of truncated normal priors over linear model weights captures three types of information about surfaces and illuminants. First, our priors generalize the notion of linear model constraints that have been used in previous work. Similar generalizations have also been

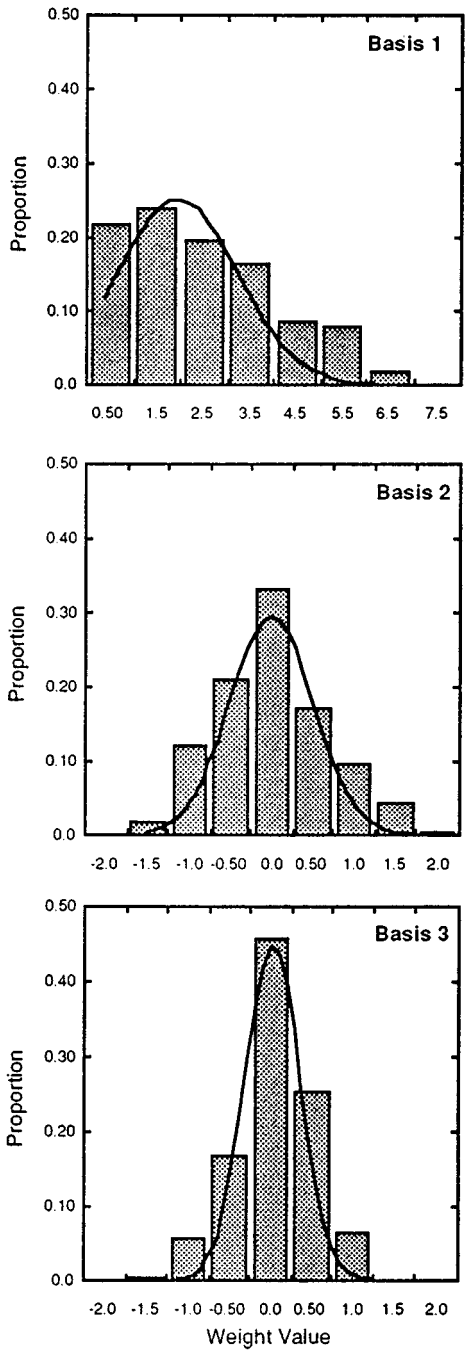


Fig. 4. Distribution of surface weights. The histograms show the distribution of linear model weights derived from the measurements of Kelly *et al.*[68] and Nickerson.[69] Each histogram corresponds to one basis vector. The solid curves show the fit of a truncated trivariate normal distribution to the weights.

used by Trussell and Vrhel[7] and by D'Zmura *et al.*[9] Second, our priors incorporate physical realizability constraints on surfaces and illuminants. Physical realizability constraints have a long history in color science[71–78] and have been used in previous analyses of color constancy.[9,16,34,79] Finally, our priors capture the fact that within the gamut of physically realizable spectra, linear model weights do not seem to be distributed uniformly. The mean and the covariance of the truncated normal distributions describe this nonuniformity. We note, however, that any other priors could be used in the Bayesian algorithm that follows.

To perform the simulations, we created three-dimensional linear models for surfaces and illuminants. Our surface linear model was the one described above, computed from the data of Kelly *et al.*[68] and Nickerson.[69] Using this linear model, we found the best-fitting model weights for each individual surface in the data set. We used the sample mean and covariance of these as the $\mathbf{u}_{sj}$ and the $\mathbf{K}_{sj}$ for our surface prior. We incorporated physical realizability as indicated by Eq. (12).

Large sets of measured illuminants are not readily available. Summaries of such data sets, in the form of linear models, do exist. We used the CIE linear model for daylights as our three-dimensional linear model for illuminants.[80] To generate a prior distribution on the weights of this linear model, we produced a set of CIE daylights with correlated color temperatures drawn according to a truncated univariate normal distribution with mean 6500 K and standard deviation 4000 K. (Any draws outside the range 3000–25,000 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 the covariance of these weights. We incorporated physical realizability as indicated by Eq. (13).

2. Simulated Scenes

We simulated scenes consisting of eight 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 assumed three classes of sensors and used Smith–Pokorny estimates of the human cone sensitivities[81] to compute the sensor responses. In the simulations presented here, we did not perturb the simulated sensor responses with additive noise.

3. Maximum Local Mass Method

To implement the MLM method, we used the expression developed in Appendixes A and B to define the expected loss and then used numerical search routines from the MATLAB Optimization Toolbox[82] to locate the point of minimum expected loss. Our approximation requires that we specify $\mathbf{K}_n$ and $\mathbf{K}_L$ and evaluate for large values of the limit parameter τ (see the appendixes). We set $\mathbf{K}_n$ to represent additive noise with a standard deviation equal to 1% of the mean sensor responses to the modal prior surface rendered under the modal prior illuminant. We set $\mathbf{K}_L$ to be 1/1000 of the joint prior covariance for the illuminant and surface scene parameters. For numerical calculations we set $\tau = 10{,}000$. As described in the appendixes, our search was constrained to parameter values along points of maximum likelihood. That is, we evaluated the expected loss only for parameter values that explained the data exactly. This simplification greatly reduces the dimensionality of the search while not omitting any possible solutions to the low-noise case.

4. Comparison Methods

We also estimated the illuminant by using a number of comparison methods. These include a Bayesian MAP algorithm (MAP), two algorithms based on physical realizability constraints (Realizability 1 and 2), a gray world algorithm (Gray World), and a subspace algorithm (Subspace). We describe each of these algorithms below.

MAP. We found Bayesian MAP estimates by using a numerical search over the joint illuminant–surface posterior. As with our implementation of the MLM method, we searched only over parameter values that were consistent with the data. We used the same priors and the same $\mathbf{K}_n$ as for our MLM method.

Realizability 1 and 2. Our main interest in simulating physical realizability algorithms was not in testing previous algorithms[9,16,34] *per se* but rather in establishing the level of performance obtainable when only physical realizability constraints are used. We therefore implemented two simple physical realizability algorithms. These algorithms found a physically realizable illuminant such that both (1) the illuminant was constrained to lie within a three-dimensional linear model for illuminants and (2) the corresponding surface reflectance functions were constrained to lie within a three-dimensional linear model for surfaces and were also physically realizable. We used the same linear models as those for the MLM method. Constraints (1) and (2) alone are not sufficient to determine a unique illuminant estimate. For example, we can take any solution for a physically realizable illuminant and corresponding surfaces and create a new solution by multiplying the illuminant spectral power distribution by a factor greater than unity and dividing the surface reflectances by the same factor. We used two different heuristics to choose a unique estimate. For our Realizability 1 algorithm, we sought the physically realizable solution with minimum power. For our Realizability 2 algorithm, we sought the physically realizable solution with minimum power at the longest wavelength. In both cases we used numerical search to find the estimate.

Gray World. Our implementation of the Gray World algorithm follows Buchsbaum.[30] For trichromatic sensing this algorithm requires specification of three-dimensional linear models for illuminants and surfaces. We used the same linear models as those for the MLM method. Because our actual simulations drew surfaces from normal distributions truncated by the constraint of physical realizability, we used the sample mean of a large set of actual draws as the population mean for the Gray World algorithm. (Using the mean of the untruncated normal distribution produces a noticeable bias in the estimates.)

Subspace. Our implementation of the subspace method algorithm follows Wandell.[53] For trichromatic sensing this algorithm requires specification of a three-dimensional linear model for illuminants and a two-dimensional linear model for surfaces. We used the same linear model for illuminants as that for our Bayesian algorithms. For the surface linear model, we used the first two principal components of the three-dimensional linear model for surfaces that we constructed for the MLM method. The algorithm of Maloney and Wandell is designed to recover only the relative spectrum of the illuminant. To scale the absolute intensity of the estimates returned by this algorithm, we performed the following steps. First, we computed the mean response of the image and computed the ratio of this response to the mean response expected given our illuminant and surface priors. We then scaled the recovered illuminant spectral power distribution so that its scalar ratio (found by univariate regression) to the mean illuminant was in this same ratio.

Implementation of numerical search. Our implementations of the MLM, MAP, Realizability 1, and Realizability 2 algorithms rely on numerical parameter search to minimize an objective function (e.g., the expected loss for the MLM method). Numerical parameter search is something of an art. Our interest here is in evaluating the performance possible with Bayesian methods rather than in developing algorithms for implementing them. In our implementation we used the general purpose search routines provided in the MATLAB Optimization Toolbox.[82] These routines are useful but by no means perfect. In addition, they are not tailored to our problems. To reduce the probability that our search would terminate at a local rather than a global minimum, we chose a number of random starting points for each search. We always checked that the objective function value for our obtained estimates was lower than its value for the actual simulated illuminant. In the few cases in which this condition was not met, we trimmed the result from our analyses and plots.

B. Simulation Results

1. Qualitative Analysis

Figure 5 shows how well each of six algorithms estimates the mode of the illuminant prior when the surfaces are drawn at random from the actual prior distribution on surfaces. Each plot illustrates performance for one of the algorithms. To generate each plot, we fixed the simulated illuminant and repeated the simulation for 15 different sets of eight randomly drawn surfaces. The solid curve in each plot shows the simulated illuminant. The dashed curves show the individual estimates of this illuminant produced by each algorithm.

We can draw a number of conclusions from the figure. First, the MLM algorithm provides the best estimates. The Gray World algorithm also provides reasonable estimates, although their scatter around the true illuminant is greater than that for the MLM algorithm. Presumably, this difference in clustering arises because the Bayesian MLM algorithm is able to take advantage of information in the sensor responses that is not carried by their spatial mean.

The MAP algorithm returns estimates whose relative spectral power distribution is approximately correct. The overall intensity of the estimates returned by this algorithm is too high, however. For the color-constancy problem, this bias seems to be a consistent property of the MAP algorithm.

The Realizability 1 algorithm also returns estimates that are approximately correct. In contrast to the MAP algorithm, however, the overall intensity of these estimates is consistently too low. Note that the perfor-

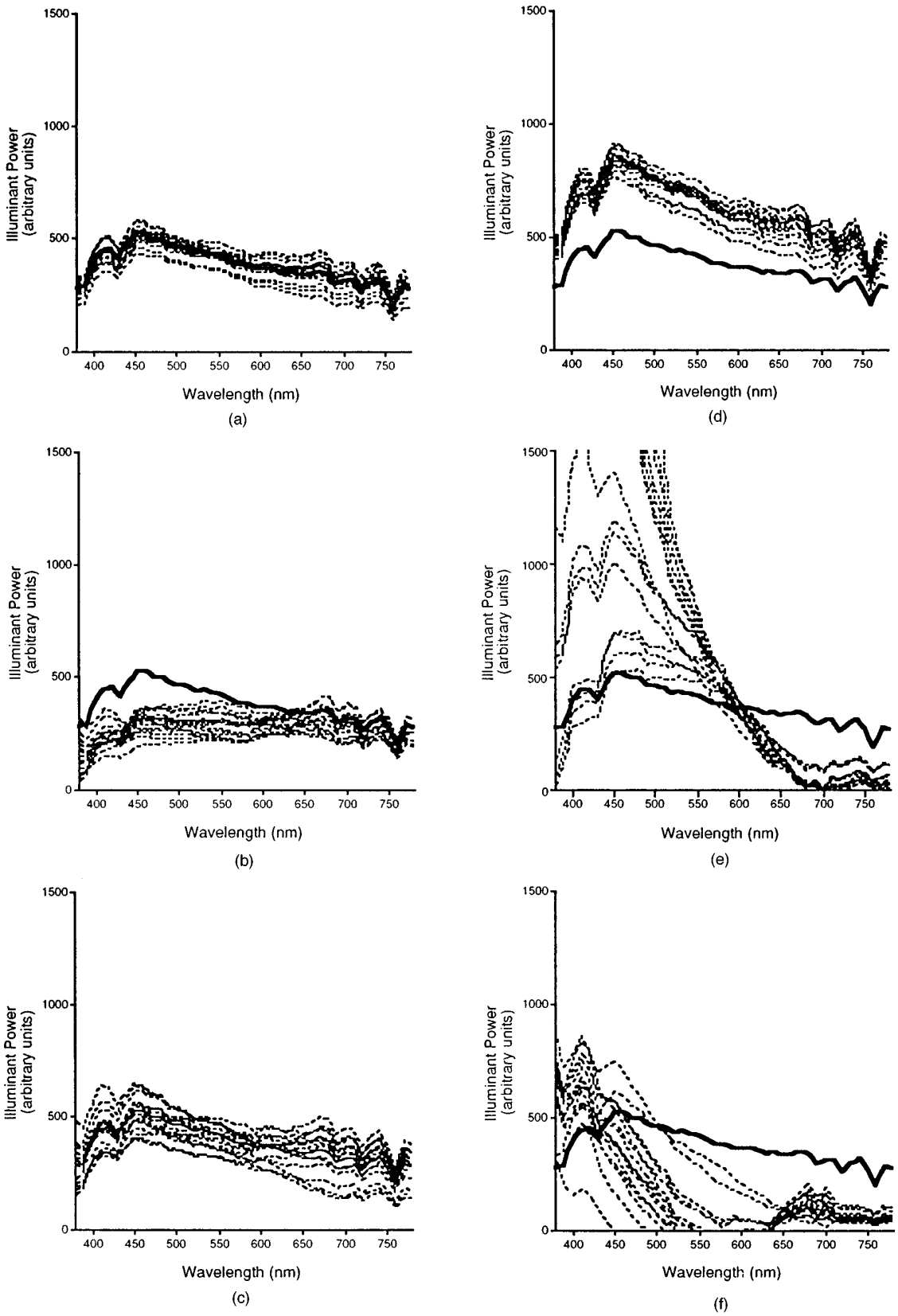


Fig. 5. Basic algorithm performance. Each plot illustrates performance for one of the algorithms. To generate each panel, we fixed the simulated illuminant and repeated the simulation for 15 different sets of eight randomly drawn surfaces. The solid curve in each plot shows the simulated illuminant. The dashed curves show the individual estimates of this illuminant produced by each algorithm. The simulated illuminant was a CIE 6500-K daylight, which is the mode of the illuminant prior. The surfaces were drawn at random from the surface prior. The six algorithms compared in this figure are (a) MLM, (b) Realizability 1, (c) Gray World, (d) MAP, (e) Realizability 2, and (f) Subspace.

mances of the Realizability 1 and Realizability 2 algorithms differ vastly. This tells us that the physical realizability constraint alone is not strong enough to provide an accurate estimate of the illuminant. Recall that the only difference between these two algorithms was the heuristic used to choose a unique realizable solution from the set of all such solutions. Apparently, the good performance of previous algorithms incorporating physical realizability constraints[9,16,34] is not driven by the realizability constraint *per se*. This point is bolstered further when one notes that MLM and MAP may also be considered realizability algorithms, since the priors that we used incorporate the constraint of physical realizability. In Subsection 4.c we consider why the minimum-power heuristic incorporated in our Realizability 1 algorithm may be a reasonable one.

It is not surprising that the Subspace algorithm does poorly for these simulations, where its assumptions are violated. For a visual system with three classes of sensors, this algorithm is guaranteed to work only when the surfaces are described exactly by a two-dimensional linear model. Apparently, the Subspace algorithm is not robust against violations of its linear model constraints.

Table 1. Simulated Illuminants[a]

Illuminant Index	CCT (K)	x	y	Y (Arbitrary Units)
1	4000	0.38	0.38	5.32×10^6
2		0.35	0.36	6.11×10^6
3		0.31	0.33	6.23×10^6
4	6500	0.31	0.32	5.96×10^6
5		0.29	0.31	5.91×10^6
6		0.28	0.29	5.46×10^6
7	15000	0.26	0.27	4.76×10^6

[a] Each of the seven illuminants is a CIE daylight. The table provides the CIE $x-y$ chromaticities and luminance (in arbitrary units) of these illuminants. We also provide the correlated color temperature (CCT) for illuminants 1, 4, and 7.

Table 2. Mean Simulated Surfaces[a]

Surface Index	Label	x	y	Luminance Factor
1	"Yellow"	0.43	0.40	41.82
2		0.42	0.39	38.84
3		0.39	0.37	39.04
4		0.38	0.37	39.13
5		0.36	0.35	39.31
6	None	0.36	0.35	35.83
7		0.35	0.34	35.67
8		0.33	0.33	34.25
9		0.32	0.32	31.73
10		0.30	0.30	30.86
11	"Blue"	0.29	0.29	31.00

[a] Each of the 11 surface distributions is obtained by shifting the mean of the normal distribution underlying our standard prior surface distribution. We provide the CIE $x-y$ chromaticity and luminance factor of the mean of each surface distribution when it is rendered under an equal-energy illuminant. Since the distributions are truncated normals, we provide sample means. Surface distributions 1, 6, and 11 carry the labels "yellow," "none," and "blue," respectively.

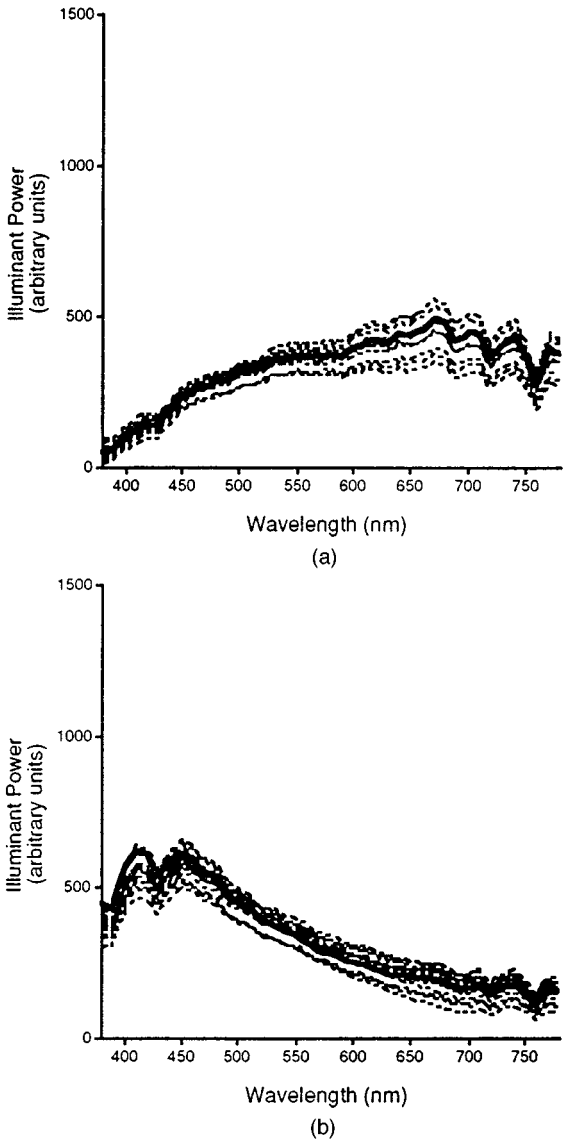


Fig. 6. Performance of the MLM algorithm for two illuminants that differ from the prior mode. To generate each panel, we fixed the simulated illuminant and repeated the simulation for 15 different sets of eight randomly drawn surfaces (no surface bias). The solid curve in each plot shows the simulated illuminant. The dashed curves show the individual estimates of this illuminant produced by the MLM algorithm. (a) Results for the simulated 4000-K daylight, (b) results for the simulated 15,000-K daylight.

We did confirm that our implementation of the Subspace algorithm performs well when run for simulations for which the two-dimensional linear model constraint holds.

Figure 5 shows results from the case in which the illuminant was equal to the mode of the illuminant prior and in which the surfaces were drawn at random from the surface prior. That is, the simulation conditions match those for which the algorithms were designed. An important part of algorithm evaluation is to consider what happens when the algorithm must deal with conditions that differ from those for which it was designed. For this reason we evaluated the performance of the algorithms by using simulations for which we biased either the illumi-

nant or the distribution from which the collection of surfaces was drawn, without adjusting the algorithm design to match. To bias the distribution from which the surface collection was drawn, we added a constant offset to the prior distribution for surfaces. Since we enforced a physical realizability constraint when we drew the surfaces, there is no simple analytic expression for the resulting means of the truncated distributions. Tables 1 and 2 specify the illuminants and the sample means of the surface collection distributions, respectively, that we used in our simulations.

Figure 6 shows the performance of the MLM algorithm for two illuminants that differ considerably from the mode of the illuminant prior. We see that in each case the algorithm returns accurate estimates. Thus we conclude that the MLM algorithm is not sensitive only to its illuminant prior: It is able to use the sensor data to track illumination changes.

When the collection of surfaces in the scene is in fact drawn from our surface prior, the color-constancy problem may not be too difficult. Indeed, under these conditions, the Gray World algorithm performs reasonably well. More challenging is to estimate the illuminant correctly in the face of biases in the collection of surfaces in the scene. To evaluate how well the algorithms handled this problem, we simulated conditions with the 6500-K il-

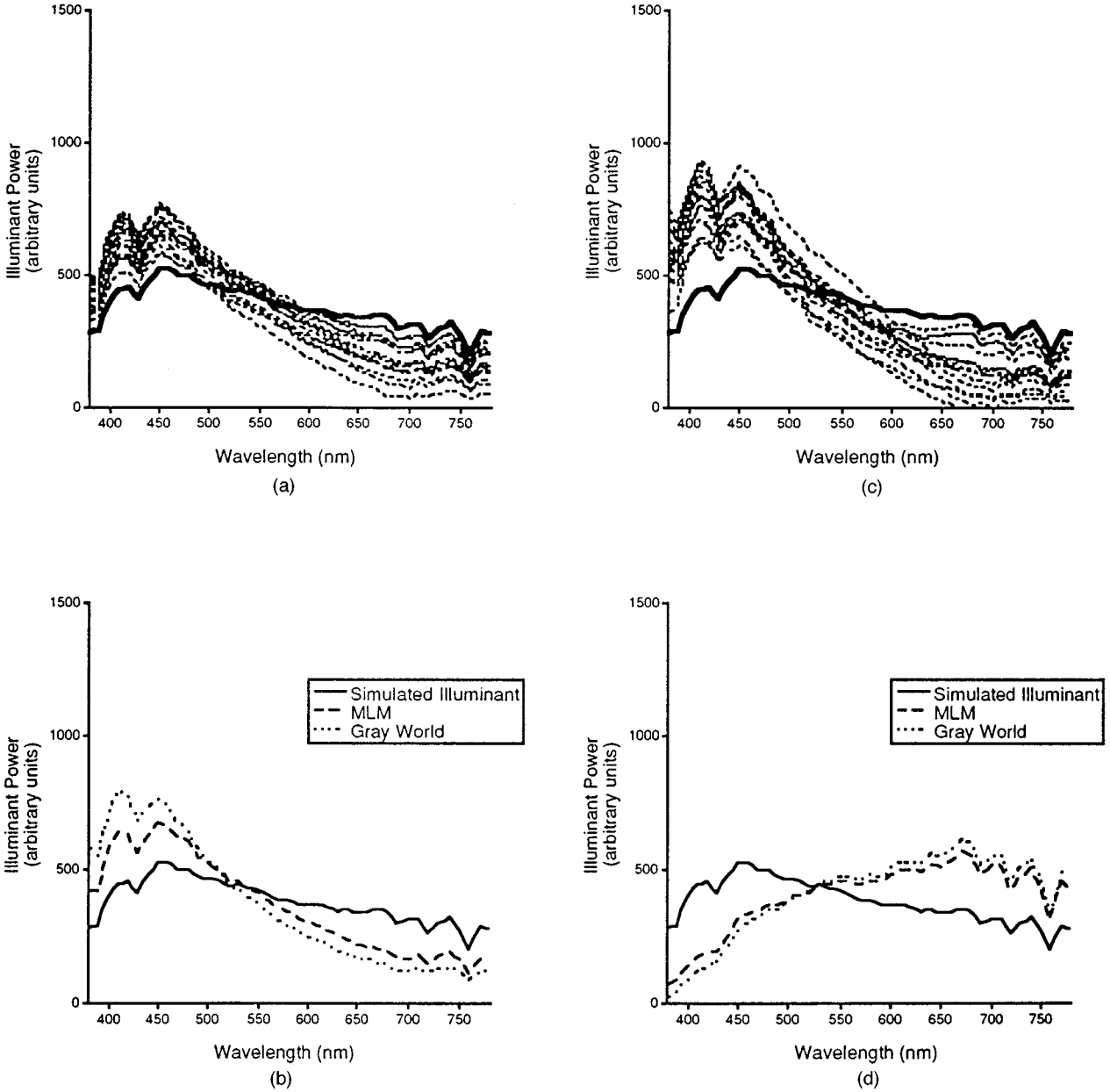


Fig. 7. Performance of MLM and Gray World algorithms when the surfaces are drawn from biased distributions. (a) Performance of the MLM algorithm for the simulated 6500-K daylight when the surfaces were drawn from the "blue" distribution specified by Table 2, (c) performance of the Gray World algorithm for the same conditions. In each case we fixed the simulated illuminant and repeated the simulation for 15 different sets of eight randomly drawn surfaces. Plots (b) and (d) highlight the difference between the performances of the two algorithms. Plot (b) compares the mean estimate of the MLM algorithm for the simulated 6500-K daylight when the surfaces were drawn from the blue distribution with the mean estimate of the Gray World algorithm for the same conditions. Plot (d) shows the same comparison when the surfaces were drawn from the "yellow" distribution.

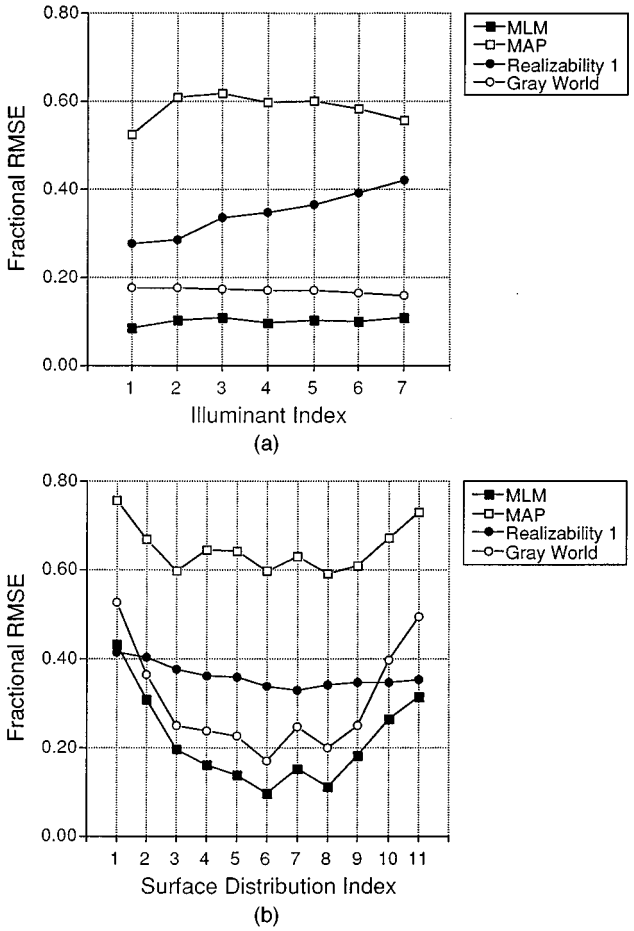


Fig. 8. Summary of algorithm performance for a variety of simulation conditions. The figure shows the mean fractional RMSE computed in the spectral domain for each algorithm for a variety of simulation conditions. (a) Error plotted as a function of the seven simulated illuminants listed in Table 1. For these simulations there was no bias in the distribution of surfaces. (b) Error plotted as a function of the 11 surface distributions listed in Table 2. For these simulations the 6500-K simulated illuminant was used.

luminant and biased the distribution from which we drew the surfaces. Figure 7 shows the performance of the MLM algorithm for the 6500-K daylight when the surfaces are drawn from biased distributions. Figure 7(a) shows the performance of the MLM algorithm for the simulated 6500-K daylight when the surfaces were drawn from the "blue" distribution specified by Table 2. Here we see that the bias in the surface distribution has a substantial effect on the illuminant estimate. Figure 7(c) shows the performance of the Gray World algorithm for the same conditions. By comparing these two plots, we see that although the MLM algorithm is influenced by the bias in the surface distribution, it is not as severely affected as the Gray World algorithm. This fact is seen more clearly in Fig. 7(b), where the mean estimates of each algorithm are plotted together. Figure 7(d) shows a similar comparison for the "yellow" surface distribution. For this surface draw, the difference between the two algorithms is less pronounced.

2. Quantitative Analysis

The qualitative analysis presented in Subsection 3.B.1 leads to two main conclusions. First, of the algorithms that we have considered, the MLM algorithm provides the best illuminant estimates. Second, the performance of all the algorithms is influenced when the distribution of surfaces is heavily biased. We can confirm these conclusions with a more quantitative analysis of our simulation results.

Quantifying the performance of the various algorithms involves many issues. Ideally, there would be a single performance metric appropriate for all applications of color-constancy algorithms. As our discussion above makes clear, however, different loss functions may be appropriate for different applications. If we take the Bayesian view and specify a loss function, then the expected loss provides a natural metric for algorithm comparison. This metric may not appeal to non-Bayesians, however. In addition, the expected loss can be difficult to compute numerically. As an extreme example, consider the minus delta loss function of the MAP estimate. Since this loss function is constant for any but exactly the correct estimate, simulations of any algorithm are highly likely to give the same estimate of expected loss. Similar considerations apply to any localized loss function.

Here we have chosen to use a standard metric for comparing the algorithms. We compute the fractional root-mean-squared error (RMSE) between the estimates of the illuminant spectral power distribution returned by the algorithms and the simulated illuminant spectral power distribution. This error measure, $|\mathbf{B}_e\tilde{\mathbf{w}}_e - \mathbf{B}_e\mathbf{w}_e|^2/|\mathbf{B}_e\mathbf{w}_e|^2$, captures roughly how well the algorithms recover the physical properties of the illuminant. In spite of the difficulties with using a squared-error metric as a loss function for designing algorithms (see Subsection 2.B), our choice has the didactic advantage of familiarity. In addition, since it is not highly localized, we can estimate its expected value by using simulations.

For each illuminant–surface prior pair, we computed the mean fractional RMSE for each algorithm (over the 15 repetitions of the simulation; rare convergence failures excluded). Figure 8 summarizes the performance of the MLM, MAP, Realizability 1, and Gray World algorithms for a variety of simulation conditions. Figure 8(a) shows the error for different simulated illuminants when the surfaces are drawn from an unbiased distribution. As with our qualitative analysis, we see that the MLM algorithm does well at tracking changes in the illuminant: The estimation error is roughly constant as the simulated illuminant is varied. In addition, Fig. 8(a) shows that the MLM algorithm performs better than the other algorithms according to the fractional RMSE metric.

Figure 8(b) shows the error for different surface distributions when the 6500-K illuminant is simulated. Again, the MLM algorithm performs best. It is also clear that the algorithm is not immune to shifts in the surface distribution. As the surface distribution is biased, the error for the MLM algorithm increases, although not as rapidly as that for the Gray World algorithm. Note that the performance of the algorithm degrades gracefully in the face of biases in the surface distribution.

4. DISCUSSION

A. Algorithm Performance

Although the MLM method is more robust against changes in the collection of surfaces than the Gray World algorithm, it is still influenced substantially by changes in the collection of surfaces in the scene. For our viewing geometry, the separation of changes in illuminant from changes in surface collection is a hard problem, even with the use of all the statistical information available. This separation has been achieved only when the surface reflectances are severely restricted to lie within a two-dimensional linear model. In this case both the MLM and Subspace algorithms perform well (simulations not presented).

Little is known about the actual distributions of illuminants and surfaces in natural scenes. It seems unlikely that the priors that we used adequately capture all of the regularities of these distributions. One attractive feature of our MLM algorithm is that it depends only locally on the prior. Although we have used truncated normals, our expression for the expected loss may be implemented for arbitrary priors without any increase in computational complexity. Richer priors may improve algorithm performance.

As we noted in Section 1, there is good evidence that human vision successfully separates changes in illumination from changes in surface collection.[15,22,46,47] These studies, however, employed stimuli composed of real illuminated surfaces. Such stimuli are not characterized by the simple model of image formation that we analyzed. Only a few studies[28,29,83] have measured how well human vision separates illuminant changes from changes in the surface collection by using stimuli that conform to our imaging model. These studies are not conclusive, but they leave open the possibility that the human visual system is unable to separate the two physical processes effectively for simple viewing geometries. Thus it is possible that the problem of separating illuminant changes from changes in the surface collection is not soluble for the simple image model that we have considered, given reasonable assumptions about the distribution of illuminant spectral power distributions and surface reflectance functions. Presumably guided by this view, a number of authors have investigated the computational color-constancy problem for richer imaging models and concluded that information about the illuminant is carried by interreflection between surfaces,[84] specular highlights,[85,86] and variation in the illumination.[32,33] We believe that extending the Bayesian approach to these richer viewing geometries may provide further insight as to what information is most useful.

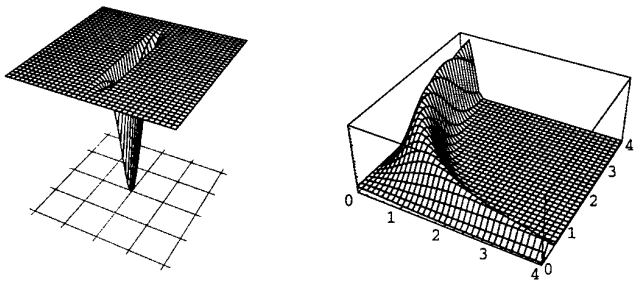
The Gray World algorithm performs reasonably across our simulation conditions, although uniformly worse than the MLM algorithm. This is not surprising, as the MLM algorithm uses all of the information contained in the sensor responses while the Gray World algorithm uses only the mean of the sensor responses. The fact that the Gray World algorithm performs reasonably suggests that the mean of the sensor responses carries the bulk of the information about the illuminant available for our simple viewing geometry. Maloney[87] has pointed out that the Bayesian approach can be used to test this idea more precisely, since we can compare the performance of different Bayesian algorithms when each is given access to a different summary of the sensor data. For the local mass loss function, such algorithms may be implemented by substituting different rendering functions into the formulas developed in Appendix A.

B. Local Mass Loss Function

Other research relates to the local mass loss function. Yuille and Bulthoff[88] independently suggested such a loss function for perception problems, although they did not use it in an algorithm. The local mass loss function is similar in spirit to Shepard's notion of a consequential region.[89] It also looks similar to the penalty functions used in robust regression,[90–92] which fit well-parametrized data but ignore outliers. However, the local mass loss function applies to the scene parameters, not to the data, and thus plays a different role from that of regression penalty functions.

Freeman proposed a Bayesian interpretation of the generic viewpoint assumption[93,94] that involved marginalizing the posterior probability over particular variables, such as viewpoint, which are not to be estimated. Improved estimates of the scene parameters can result from using the marginalized posteriors, thus exploiting the assumption of generic viewpoint. The local mass loss function offers a more general computation with a similar interpretation. The loss function multiplies the posterior probability by a Gaussian of a particular width in each parameter direction before integration to find the expected loss. The loss function shape specifies how much uncertainty we tolerate in each scene parameter dimension. Marginalization over a generic variable may be understood as using an infinitely wide loss function in that parameter dimension. The local mass loss function approach, however, offers two advantages over marginalization. First, marginalizing over a generic variable precludes estimating an optimal value for that parameter, while the loss function approach allows us to specify that we care more about the precision of some parameters than others but still to obtain a best estimate for all of them. Second, the asymptotic approximation for the marginalized posterior probability[88] may become singular if the likelihood function is flat along the generic variable. The Gaussian window of the local mass loss function avoids that singularity.

An asymmetric loss function is appropriate in situations in which 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. Whether using an asymmetric loss function affects the estimates of the MLM algorithm, however, depends on the particular shape of the posterior. Figure 9 shows the expected loss for the $y = ab$ product example for an asymmetric loss function. Here the posterior ridge curves sharply in the parameter space, and using an asymmetric loss function has a substantial effect. For the color-constancy problem of our simulations, we have found only a moderate sensitivity to varying the shape of



(a) Asymmetric loss function

(b) (minus) expected loss

Fig. 9. Product example with an asymmetric loss function. (a) Loss function; (b) corresponding expected loss, computed according to expression (A5). As in Fig. 3, decreasing loss is plotted upward for visual clarity. The loss function aspect ratio is 0.14 to 1.0. The optimal parameter estimate is (0.37, 2.7), which differs from the results obtained for the symmetric local mass loss function (Fig. 3).

the loss function. We also found that the MLM algorithm performed similarly to an algorithm that used MAP estimation on the marginal posterior for the illuminant parameters.[5]

C. Bayesian Approach

The Bayesian approach has three explicit components. To calculate the likelihood, we must model image formation. To specify the prior, we must quantify our assumptions about images. And to minimize the expected loss, we must state the cost of estimation errors. Within the Bayesian framework, each of these components may be considered separately: We need not confound our assumptions about the world (the prior) with how we will use our estimate (the loss function).

The Bayesian approach is sometimes criticized exactly because it requires specification of priors and loss functions. This is regarded as undesirable because of uncertainty about what prior or loss is appropriate. We disagree. While specifying the prior and the loss may be difficult, that is not the fault of the approach. When the result of a Bayesian algorithm depends strongly on either the prior or the loss, what we learn is that the estimation problem itself is difficult and that we must carefully consider what assumptions we make to solve it. Such assumptions are implicit in any algorithm that solves the problem; the Bayesian approach makes the assumptions explicit.[1]

Bayesian methods may be used to understand non-Bayesian algorithms. The idea is to find a prior and a loss function for which the non-Bayesian algorithm mimics the Bayesian solution. As an example, Appendix C analyzes a special case of the color-constancy problem, lightness constancy in a monochromatic world. We show that Land and McCann's[18] normalization rule of assigning the percept of white to the brightest patch follows from a Bayesian analysis. Although not stated in terms of illuminant estimation, their rule is equivalent to assuming that the actual illuminant has the minimum power consistent with physical realizability. For a class of prior probability distributions, our analysis in Appendix C shows that their heuristic implements the MLM estimate. This provides a principled derivation of a heuristic rule. Note that our Realizability 1 algorithm represents one possible way of generalizing the minimum power heuristic from lightness to color. Although the Realizability 1 algorithm performed reasonably, it was not as effective as the MLM algorithm derived from a full analysis of the color-constancy problem.

D. Relation of Estimation to Perception

We close with a few remarks about the relation among Bayesian estimation, loss functions, and perception. In this paper we have focused on estimation of the illuminant as the key to color constancy. We designed our Bayesian algorithm to minimize error in a parameter space that specified physical illuminants and surfaces, and we evaluated the algorithms by assessing how well they estimated the simulated illuminants. As emphasized by Forsyth[16] (see also Brainard *et al.*,[17] McCann *et al.*,[11] Land and McCann,[18] Land,[19,20] Marimont and Wandell,[56] and Finlayson[34]), an explicit estimate of the illuminant is not necessary for color constancy. Rather, all that is required is that the visual system construct a representation of surfaces that is constant across changes of viewing context. Given that estimating the physical illuminant and surfaces has proved to be a difficult problem, it seems worth considering whether the problem of estimating constant surface descriptors is more tractable. Note that there is no guarantee that the algorithm that is best at estimating physical spectra will also be best at returning stable surface descriptors.

For human vision it is clear that a great deal of information about the physical properties of surfaces and illuminants is lost at the initial stage of light encoding, since the cone photosensors collectively provide only three pieces of information about the incident spectrum at each location. Certain variations in spectra are completely invisible to the human visual system (see Brainard *et al.*[17]; also, Brainard[41] and Nielsen and Wandell[95]). Rather than estimating the full illuminant or surface spectra, it might be advantageous to estimate surface descriptors that are weaker yet carry enough information to discriminate between surfaces in an invariant manner across illuminants. In this spirit Marimont and Wandell[56] showed how the design of linear models to represent spectra could take advantage of the spectral sensitivities of the sensors. In preliminary analyses of our current simulations, we found some combinations of illuminant and surface bias for which an algorithm that performed poorly according to our spectral error metric performed better when error was evaluated in terms of chromaticity.

In the context of Bayesian estimation, these ideas may be incorporated into the loss function. Suppose, for example, that we consider a surface descriptor consisting of the cone responses to the surface rendered under a canonical illuminant. Let $\mathbf{g}(\mathbf{x})$ be a function that maps a vector $\mathbf{x}$ in our full illuminant–surface parameter space to a vector of the cone coordinates of each surface described by $\mathbf{x}$ when rendered under the canonical illuminant. Given this function $\mathbf{g}(\mathbf{x})$, we can construct a loss function of the form

$$L(\tilde{\mathbf{x}}, \mathbf{x}) = -\exp\{-|\mathbf{K}_L^{-1/2}[\mathbf{g}(\tilde{\mathbf{x}}) - \mathbf{g}(\mathbf{x})]|^2\}. \quad (16)$$

Unlike the loss function that we used in developing our MLM algorithm, this loss function penalizes differences not in the physical parameter space but rather in the surface descriptor space induced by $\mathbf{g}(\mathbf{x})$. Extending the analysis of Appendix A to handle loss functions of the form of Eq. (16) is straightforward as long as the function $\mathbf{g}(\mathbf{x})$ is twice differentiable in $\mathbf{x}$. We hope to learn whether the surface descriptors produced by this approach are stable.

APPENDIX A

In this appendix we derive an approximation to the expected loss for the local mass loss function. We consider the general problem of estimating a parameter vector $\mathbf{x}$ from an observation $\mathbf{y}$. We write the rendering function as $\mathbf{y} = \mathbf{f}(\mathbf{x})$ and assume that there is additive Gaussian observation noise with zero mean and known covariance. The prior probability is given by the density function $p(\mathbf{x})$. We use an asymptotic approximation to calculate the expected loss. In Appendix B we specialize to the color-constancy problem, and in Appendix C we address monochromatic lightness constancy.

Computing the expected loss, which we seek to minimize, involves an integral over the entire scene parameter space, Eq. (3), yet our loss function, Eq. (11), is appreciably nonzero only near the estimate $\tilde{\mathbf{x}}$. We want to exploit that locality to find an approximation for the expected loss that depends only on measurements at $\tilde{\mathbf{x}}$. This simplification applies generally to scene parameter estimation by the MLM method in the limit of low observation noise.

Let $\mathbf{K}_n$ and $\mathbf{K}_L$ be matrices that describe the shape and the relative sizes of the noise and loss function covariances. We can compute the expected loss in the low-noise limit by scaling each matrix by a factor $1/\tau$ and examining the limit as τ becomes large. (This approximation neglects any effects that depend on the absolute scale of $\mathbf{K}_L$.)

Combining Eq. (2) (Bayes rule) with the multivariate generalization of Eq. (7) (for the likelihood function) in Eq. (3) gives the expected loss:

$$\begin{aligned}\bar{L}(\tilde{\mathbf{x}}|\mathbf{y}) &= C \int (\text{likelihood})(\text{priors})(\text{loss function})\, d\mathbf{x} \\ &= -C \int \exp\left\{-\frac{\tau}{2}|\mathbf{K}_n^{-1/2}[\mathbf{y}-\mathbf{f}(\mathbf{x})]|^2\right\} p(\mathbf{x}) \\ &\quad \times \exp\left[-\frac{\tau}{2}|\mathbf{K}_L^{-1/2}(\mathbf{x}-\tilde{\mathbf{x}})|^2\right] d\mathbf{x}. \end{aligned} \tag{A1}$$

For an integral of the form

$$I(\tau) = \int \exp[-\tau\phi(\mathbf{x})] g(\mathbf{x}) d\mathbf{x}, \tag{A2}$$

the leading-order term in an asymptotic expansion for large τ is[96]

$$I(\tau) \approx \frac{\exp[-\tau\phi(\mathbf{x}_0)]}{\{|\det[\phi_{\mathbf{xx}}(\mathbf{x}_0)]|\}^{1/2}} \left(\frac{2\pi}{\tau}\right)^{n/2} g(\mathbf{x}_0), \tag{A3}$$

where $\mathbf{x}_0$ minimizes $\phi(\mathbf{x})$.

We identify $g(\mathbf{x}) = p(\mathbf{x})$ and

$$\phi(\mathbf{x}) = \tfrac{1}{2}|\mathbf{K}_n^{-1/2}[\mathbf{y}-\mathbf{f}(\mathbf{x})]|^2 + \tfrac{1}{2}|\mathbf{K}_L^{-1/2}(\mathbf{x}-\tilde{\mathbf{x}})|^2. \tag{A4}$$

Then, using approximation (A3), we have

$$\bar{L}(\tilde{\mathbf{x}}|\mathbf{y}) \approx -C \exp(-\tau\{\tfrac{1}{2}|\mathbf{K}_n^{-1/2}[\mathbf{y}-\mathbf{f}(\mathbf{x}_0)]|^2 + \tfrac{1}{2}|\mathbf{K}_L^{-1/2}(\mathbf{x}_0-\tilde{\mathbf{x}})|^2\}) \frac{p(\mathbf{x}_0)}{\{|\det[\phi_{\mathbf{xx}}(\mathbf{x}_0)]|\}^{1/2}}. \tag{A5}$$

Twice differentiating $\phi(\mathbf{x})$ in Eq. (A4) gives the following for the (i,j)th element of the matrix $\phi_{\mathbf{xx}}$:

$$[\phi_{\mathbf{xx}}(\mathbf{x}_0)]_{ij} = \mathbf{f}_i'^{\mathrm{T}}\mathbf{K}_n^{-1}\mathbf{f}_j' - [\mathbf{y}-\mathbf{f}(\mathbf{x}_0)]^{\mathrm{T}}\mathbf{K}_n^{-1}\mathbf{f}_{ij}'' + [\mathbf{K}_L^{-1}]_{ij}, \tag{A6}$$

where $[\,\cdot\,]_{ij}$ means the (i,j)th array element,

$$\mathbf{f}_i' = \left.\frac{\partial\mathbf{f}(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\mathbf{x}_0} \quad \text{and} \quad \mathbf{f}_{ij}'' = \left.\frac{\partial^2\mathbf{f}(\mathbf{x})}{\partial x_i \partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0}. \tag{A7}$$

To apply expression (A5) for the expected loss at $\tilde{\mathbf{x}}$, we need to find an expansion point $\mathbf{x}_0$ where $\phi(\mathbf{x})$ of Eq. (A4) is minimized. If we restrict attention to evaluating the expected loss for estimates $\tilde{\mathbf{x}}$ at local maxima or ridges of the likelihood function, then both terms of $\phi(\mathbf{x})$ are minimized locally by the choice $\mathbf{x}_0 = \tilde{\mathbf{x}}$. Thus we can set $\mathbf{x}_0 = \tilde{\mathbf{x}}$ in expression (A5) to evaluate the expected loss at points of maximum likelihood. [We believe that setting $\mathbf{x}_0 = \tilde{\mathbf{x}}$ in expression (A5) provides an upper bound for the expected loss for other points as well.]

At $\mathbf{x}_0 = \tilde{\mathbf{x}}$ the difference between the expected loss $\bar{L}(\tilde{\mathbf{x}}|\mathbf{y})$ in expression (A5) and the negative posterior is the factor $1/\{|\det[\phi_{\mathbf{xx}}(\tilde{\mathbf{x}})]|\}^{1/2}$. This term allows the local mass loss to respond to the width as well as the height of probability ridges such as those shown in Fig. 2(a). A related approximation is used in Bayesian statistics, dating to Laplace.[2,97–99] The first two terms of $\phi_{\mathbf{xx}}(\mathbf{x}_0)$ form the conditional Fisher information matrix $\mathbf{I}$.[1,98] It can be used to marginalize the posterior over nuisance parameters,[1,100,101] yielding a factor of $1/\sqrt{\det(\mathbf{I})}$ after integration. Recent authors have exploited this in parameter estimation[102–104] and computer vision.[93,94] However, if one marginalizes over parameters, one cannot estimate their optimal values. Furthermore, for underdetermined estimation problems, $\det(\mathbf{I})$ can be 0, spoiling the approximation. The loss function approach used here avoids this singularity and allows for trading off accuracy requirements among scene parameter components.

APPENDIX B

To apply the general results of Appendix A to the color-constancy problem, we need to identify $\mathbf{x}$ and evaluate the derivatives of Eqs. (A7). From our representation of the spectral variables that we want to estimate, we have

$$\mathbf{x} = \begin{pmatrix}\mathbf{w}_s \\ \mathbf{w}_e\end{pmatrix}, \qquad \mathbf{x}_0 = \begin{pmatrix}\mathbf{w}_{s0} \\ \mathbf{w}_{e0}\end{pmatrix}. \tag{B1}$$

To evaluate $\mathbf{f}_i'$ of Eqs. (A7), we use the rendering equation as written in Eq. (6) to write

$$\partial \mathbf{f}(\mathbf{w}_s, \mathbf{w}_e)/\partial w_{si} = \mathbf{N}(\mathbf{w}_e)_i, \tag{B2}$$

$$\partial \mathbf{f}(\mathbf{w}_s, \mathbf{w}_e)/\partial w_{ei} = \mathbf{N}(\mathbf{w}_s)_i, \tag{B3}$$

where $\mathbf{N}(\mathbf{w}_e)_i$ is the ith column of the matrix $\mathbf{N}(\mathbf{w}_e)$.

For this example it is simpler to write the full matrix of Eq. (A6), rather than the (i,j)th element of it. We have

$$\phi_{\mathbf{xx}}(\mathbf{x}_0) = \begin{pmatrix} \mathbf{N}^{\mathrm{T}}(\mathbf{w}_{e0}) \\ \mathbf{N}^{\mathrm{T}}(\mathbf{w}_{s0}) \end{pmatrix} \mathbf{K}_n{}^{-1} (\mathbf{N}(\mathbf{w}_{e0}) \quad \mathbf{N}(\mathbf{w}_{s0})) + \mathbf{K}_L{}^{-1}, \tag{B4}$$

where we examine only feasible solutions, so $\mathbf{y} - \mathbf{f}(\mathbf{x}_0) = 0$. This term appears in the desired expression for the expected loss, expression (A5). The exponential in that expression is 1, leaving the prior term, discussed in Subsection 3.A.1.

APPENDIX C

We specialize Appendix A to the case of lightness constancy for a monochromatic visual system in a monochromatic world. The scene parameter $\mathbf{x}$ is

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \\ x_I \end{pmatrix}, \tag{C1}$$

where x_i, $i = 1, \ldots, N$, is the reflectance of the ith surface and x_I is the illuminant strength. The rendering equation is

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} x_I. \tag{C2}$$

Using the above and Eqs. (A7), it follows that

$$\mathbf{f}'^{\mathrm{T}} \mathbf{K}_n{}^{-1} \mathbf{f}' = \frac{1}{\sigma_n^2} \begin{bmatrix} {x_I}^2 & 0 & \ldots & 0 & x_I x_1 \\ 0 & {x_I}^2 & \ldots & 0 & x_I x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & {x_I}^2 & x_I x_N \\ x_I x_1 & x_I x_2 & \ldots & x_I x_N & \sum_{i=1}^N {x_1}^2 \end{bmatrix}, \tag{C3}$$

written for the case of identically distributed observation noise of variance σ_n. We also assume an isotropic loss function covariance $\mathbf{K}_L{}^{-1} = \mathbf{I}/\sigma_L^2$, where $\mathbf{I}$ is the identity matrix.

Finding the expected loss for a value of the scene parameter $\mathbf{x}$ that accounts for the observed data [i.e., $\mathbf{y} = \mathbf{f}(\mathbf{x})$] involves evaluating $\det(\phi_{\mathbf{xx}}) = \det(\mathbf{f}'^{\mathrm{T}} \mathbf{K}_n{}^{-1} \mathbf{f}' + \mathbf{K}_L{}^{-1})$ from Eq. (A6). Substituting from above, we have

$$\phi_{\mathbf{xx}} = \frac{1}{\sigma_n^2} \begin{bmatrix} {x_I}^2 + \sigma_n^2/\sigma_L^2 & 0 & \ldots & 0 & x_I x_1 \\ 0 & {x_I}^2 + \sigma_n^2/\sigma_L^2 & \ldots & 0 & x_I x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & {x_I}^2 + \sigma_n^2/\sigma_L^2 & x_I x_N \\ x_I x_1 & x_I x_2 & \ldots & x_I x_N & \sum_{i=1}^N {x_i}^2 + \sigma_n^2/\sigma_L^2 \end{bmatrix}. \tag{C4}$$

We expand the determinant by minors for each element of the bottom row. The submatrix of the minor for the bottom right element is diagonal. By inspection, that bottom right element times its cofactor is $(\sum_{i=1}^N {x_i}^2 + \sigma_n^2/\sigma_L^2)({x_I}^2 + \sigma_n^2/\sigma_L^2)^N$. The minors of the other elements of the bottom row are also straightforward. For each submatrix of a minor, there is one row that has only one nonzero element, and the submatrix associated with that nonzero element is diagonal. The jth element of the bottom row times its cofactor then works out to be $-(x_I x_j)^2({x_I}^2 + \sigma_n^2/\sigma_L^2)^{N-1}$. Thus we have

$$\begin{aligned} \det[\phi_{\mathbf{xx}}(\mathbf{x})] &= \left(\sum_{i=1}^N {x_i}^2 + \frac{\sigma_n^2}{\sigma_L^2} \right) \left({x_I}^2 + \frac{\sigma_n^2}{\sigma_L^2} \right)^N \\ &\quad - \sum_{j=1}^N (x_I x_j)^2 \left({x_I}^2 + \frac{\sigma_n^2}{\sigma_L^2} \right)^{N-1} \\ &= \frac{\sigma_n^2}{\sigma_L^2} \left({x_I}^2 + \frac{\sigma_n^2}{\sigma_L^2} \right)^{N-1} \left(\sum_{i=1}^N {x_i}^2 + {x_I}^2 + \frac{\sigma_n^2}{\sigma_L^2} \right). \end{aligned} \tag{C5}$$

Under the low-noise conditions, where σ_L is sufficiently larger than σ_n, we have $\sigma_n^2/\sigma_L^2 \ll x_i$, and the expression for $\det[\phi_{\mathbf{xx}}(\mathbf{x})]$ simplifies to

$$\begin{aligned} \det[\phi_{\mathbf{xx}}(\mathbf{x})] &\approx \frac{\sigma_n^2}{\sigma_L^2} ({x_I}^2)^{N-1} \left(\sum_{i=1}^N {x_i}^2 + {x_I}^2 \right) \\ &\approx \frac{\sigma_n^2}{\sigma_L^2} ({x_I}^2)^{N-2} \left(\sum_{i=1}^N {y_i}^2 + {x_I}^4 \right). \end{aligned} \tag{C6}$$

Using the above and evaluating, expression (A5) at feasible solutions [for which $\mathbf{y} = \mathbf{f}(\mathbf{x}_0)$ and $\mathbf{x}_0 = \tilde{\mathbf{x}}$], we have, for the expected loss,

$$\bar{L}(\mathbf{x}|\mathbf{y}) \approx -\frac{Cp(\mathbf{x})}{{x_I}^{N-2} \left(\sum_{i=1}^N {y_i}^2 + {x_I}^4 \right)^{1/2}}, \tag{C7}$$

where we have absorbed constants over $\mathbf{x}$ into C. We see that the integration of the local probability mass has introduced a bias favoring low illuminant values.

For the case of uniform priors over surface reflectances (in a linear parameterization), we have

$$p(x_i) = \begin{cases} 1 & \text{if } x_i \text{ is between 0 and 1} \\ 0 & \text{otherwise} \end{cases} \tag{C8}$$

for $1 \leq i \leq N$. If we assume that the prior probability of illuminants is uniform over the range of interest, then, for scene parameters that account for the data, we have

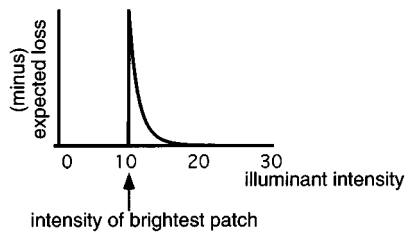


Fig. 10. Expected utility (minus expected loss) derived for the case of monochromatic lightness constancy. The surfaces and the illuminant were assumed to have a uniform prior probability. To make the plot, we calculated with eight surfaces, the maximum observation set to 10, and the mean squared observation set to 33. Land and McCann's white hypothesis (call the brightest surface white or equivalently choose the dimmest possible illuminant consistent with physical realizability) is the MLM estimate here.

$$p\left(\frac{y_i}{x_I}\right) = \begin{cases} 1 & \text{if } y_i/x_I \text{ is between 0 and 1} \\ 0 & \text{otherwise} \end{cases}. \quad \text{(C9)}$$

Expression (C7) then has a simple interpretation: $\bar{L}(x_I|y_1, \ldots, y_N)$ is 0 unless the lighting strength x_I is such that all surface reflectances would be between 0 and 10. Within that range there is a bias favoring the dimmest illuminant value consistent with the observations. This equation interprets Land's brightest patch = white heuristic[12] as the MLM estimator for the patch reflectance. Figure 10 shows a plot of the resulting expected utility (minus expected loss) for an example set of conditions.

The bias will favor the lowest feasible illuminant for the uniform priors used in the example above and for many other prior distributions as well. The expected loss for the MAP estimator does not contain this term, which suggests why the MAP estimate for the full color problem consistently overestimated spectral intensities.

For the priors of Eq. (C8), with $N > 2$, one can also motivate the brightest patch = white heuristic in another way. If one calculates the marginal posterior distribution for the illuminant intensity (by integrating the full posterior over all possible surface reflectance values), then the MAP estimate for this marginalized posterior is the same as the MLM estimate derived above.

ACKNOWLEDGMENTS

The authors acknowledge helpful conversations with E. Adelson, M. D'Zmura, G. Iverson, L. Maloney, D. Pelli, E. Simoncelli, B. Wandell, and A. Yuille. D. H. Brainard was supported by National Eye Institute grant EY10016.

REFERENCES AND NOTES

1. T. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York, 1985).
2. G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis* (Wiley, New York, 1973).
3. D. Knill and W. Richards, eds., *Perception as Bayesian Inference* (Cambridge U. Press, Cambridge, 1996).
4. Preliminary reports of our work may be found in Brainard and Freeman[5] and Freeman and Brainard.[6] See Trussell and Vrhel[7,8] and D'Zmura *et al.*[9] for related statistical approaches to color constancy.
5. D. H. Brainard and W. T. Freeman, "Bayesian method for recovering surface and illuminant properties from photoreceptor responses," in *Human Vision, Visual Processing, and Display V*, B. E. Rogowitz and J. P. Allebach, eds., Proc. SPIE **2179**, 364–376 (1994).
6. W. T. Freeman and D. H. Brainard, "Bayesian decision theory, the maximum local mass estimate, and color constancy," in *Proceedings of the 5th International Conference on Computer Vision* (IEEE Computer Society Press, Los Alamitos, Calif., 1995), pp. 210–217.
7. H. J. Trussell and M. J. Vrhel, "Estimation of illumination for color correction," *Proceedings of the International Conference in Acoustics, Speech, and Signal Processing* (IEEE, New York, 1991), pp. 2513–2516.
8. M. J. Vrhel and H. J. Trussell, "Color correction using principal components," Color Res. Appl. **17**, 329–337 (1992).
9. M. D'Zmura, G. Iverson, and B. Singer, "Probabilistic color constancy," in *Geometric Representations of Perceptual Phenomena: Papers in Honor of Tarow Indow's 70th Birthday*, R. D. Luce, M. D'Zmura, D. Hoffman, G. Iverson, and A. K. Romney, eds. (Erlbaum, Hillsdale, N. J., 1995), pp. 187–202.
10. R. M. Evans, *The Perception of Color* (Wiley, New York, 1974).
11. 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 Res. **16**, 445–458 (1976).
12. M. D. Fairchild and P. Lennie, "Chromatic adaptation to natural and incandescent illuminants," Vision Res. **32**, 2077–2085 (1992).
13. D. H. Brainard, B. A. Wandell, and E.-J. Chichilnisky, "Color constancy: from physics to appearance," Curr. Dir. Psychol. Sci. **2**, 165–170 (1993).
14. L. E. Arend, "How much does illuminant color affect unattributed colors?" J. Opt. Soc. Am. A **10**, 2134–2147 (1993).
15. D. H. Brainard and J. M. Speigle, "Achromatic loci measured under realistic viewing conditions," Invest. Ophthalmol. Visual Sci. Suppl. **35**, 1328 (1994).
16. D. A. Forsyth, "A novel algorithm for color constancy," Int. J. Comput. Vision **5**, 5–36 (1990).
17. D. H. Brainard, B. A. Wandell, and W. B. Cowan, "Black light: how sensors filter spectral variation of the illuminant," IEEE Trans. Biomed. Eng. **36**, 140–149 (1989).
18. E. H. Land and J. J. McCann, "Lightness and retinex theory," J. Opt. Soc. Am. **61**, 1–11 (1971).
19. E. H. Land, "Recent advances in retinex theory," Vision Res. **26**, 7–21 (1986).
20. E. H. Land, "Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image," Proc. Natl. Acad. Sci. USA **80**, 5163–5169 (1983).
21. W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory* (Academic, San Diego, 1989).
22. J. J. McCann, "Psychophysical experiments in search of adaptation and the gray world," in *Proceedings of the 47th Annual Conference on Imaging Science and Technology* (The Society for Imaging Science and Technology, Springfield, Va., 1994), pp. 397–401.
23. J. J. McCann, J. A. Hall, and E. H. Land, "Color Mondrian experiments: the study of average spectral distributions," J. Opt. Soc. Am. **67**, 1380 (1977).
24. L. E. Arend and A. Reeves, "Simultaneous color constancy," J. Opt. Soc. Am. A **3**, 1743–1751 (1986).
25. L. E. Arend, A. Reeves, J. Schirillo, and R. Goldstein, "Simultaneous color constancy: papers with diverse Munsell values," J. Opt. Soc. Am. A **8**, 661–672 (1991).
26. A. Valberg and B. Lange-Malecki, "Mondrian complexity does not improve 'color constancy'," Invest. Ophthalmol. Visual Sci. Suppl. **28**, 92 (1987).
27. D. H. Brainard and B. A. Wandell, "Asymmetric color-

matching: how color appearance depends on the illuminant," J. Opt. Soc. Am. A **9**, 1433–1448 (1992).
28. K. H. Bauml, "Illuminant changes under different surface collections: examining some principles of color appearance," J. Opt. Soc. Am. A **12**, 261–271 (1995).
29. K. H. Bauml, "Color appearance: effects of illuminant changes under different surface collections," J. Opt. Soc. Am. A **11**, 531–542 (1994).
30. G. Buchsbaum, "A spatial processor model for object colour perception," J. Franklin Inst. **310**, 1–26 (1980).
31. L. T. Maloney and B. A. Wandell, "Color constancy: a method for recovering surface spectral reflectances," J. Opt. Soc. Am. A **3**, 29–33 (1986).
32. M. D'Zmura and G. Iverson, "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).
33. M. D'Zmura and G. Iverson, "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).
34. G. D. Finlayson, "Color constancy in diagonal chromaticity space," in *Proceedings of the 5th International Conference on Computer Vision* (IEEE Computer Society Press, Los Alamitos, Calif., 1995), pp. 218–223.
35. B. K. P. Horn and B. G. Schunk, "Determining optical flow," Artif. Intell. **17**, 185–203 (1981).
36. J. J. Koenderink and A. J. van Doorn, "Affine structure from motion," J. Opt. Soc. Am. A **8**, 377–385 (1991).
37. T. Marill, "Emulating the human interpretation of line-drawings as three-dimensional objects," Int. J. Comput. Vision **6**, 147–161 (1991).
38. Y. G. Leclerc and M. A. Fischler, "Line drawings as 3D wire frames," Int. J. Comput. Vision **9**, 113–136 (1992).
39. P. Sinha and E. H. Adelson, "Recovering reflectance and illumination in a world of painted polyhedra," in *Proceedings of the 4th International Conference on Computer Vision* (IEEE Computer Society Press, Los Alamitos, Calif., 1993), pp. 156–163.
40. E. Saund and T. P. Moran, "Perceptual organization in an interactive sketch editing application," in *Proceedings of the 5th International Conference on Computer Vision* (IEEE Computer Society Press, Los Alamitos, Calif., 1995), pp. 597–604.
41. D. H. Brainard, "Colorimetry," in *Handbook of Optics: Volume 1. Fundamentals, Techniques, and Design*, M. Bass, ed. (McGraw-Hill, New York, 1995), pp. 26.1–26.54.
42. 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. **54**, 1031–1040 (1964).
43. L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," J. Opt. Soc. Am. A **3**, 1673–1683 (1986).
44. T. Jaaskelainen, J. Parkkinen, and S. Toyooka, "A vector-subspace model for color representation," J. Opt. Soc. Am. A **7**, 725–730 (1990).
45. D. H. Brainard and B. A. Wandell, "Analysis of the retinex theory of color vision," J. Opt. Soc. Am. A **3**, 1651–1661 (1986).
46. A. L. Gilchrist, "When does perceived lightness depend on perceived spatial arrangements?" Percept. Psychophys. **28**, 527–538 (1980).
47. A. L. Gilchrist, "Lightness contrast and failures of constancy: a common explanation," Percept. Psychophys. **43**, 415–424 (1988).
48. R. O. Brown, "Saturation and color constancy," in *Advances in Color Vision*, Vol. 4 of 1992 OSA Technical Digest Series (Optical Society of America, 1992), pp. 110–111.
49. A. B. Poirson and B. A. Wandell, "Appearance of colored patterns—pattern color separability," J. Opt. Soc. Am. A **10**, 2458–2470 (1993).
50. B. Singer and M. D'Zmura, "Color contrast induction," Vision Res. **34**, 3111–3126 (1994).
51. B. Singer and M. D'Zmura, "Contrast gain control—a bilinear model for chromatic selectivity," J. Opt. Soc. Am. A **12**, 667–685 (1995).
52. J. W. Jenness and S. K. Shevell, "Color appearance with sparse chromatic context," Vision Res. **35**, 797–805 (1995).
53. B. A. Wandell, "The synthesis and analysis of color images," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-9**, 2–13 (1987).
54. We use the notation $p(\ )$ to denote different probability-density functions. The particular function in any context is indicated by the argument.
55. T. W. Hungerford, *Algebra* (Springer-Verlag, New York, 1974).
56. D. H. Marimont and B. A. Wandell, "Linear models of surface and illuminant spectra," J. Opt. Soc. Am. A **9**, 1905–1913 (1992).
57. J. B. Tenenbaum and W. T. Freeman, "Separable mixture models: separating style and content," in *Advances in Neural Information Processing Systems 9*, M. C. Mozer, M. I. Jordan, and T. Petsche, eds. (MIT Press, Cambridge, Mass., to be published).
58. J. J. Koenderink and A. J. van Doorn, "The generic bilinear calibration–estimation problem," Int. J. Comput. Vision (to be published).
59. M. S. Landy and J. A. Movshon, eds., *Computational Models of Visual Processing* (MIT Press, Cambridge, Mass., 1991).
60. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**, 721–741 (1984).
61. R. Szeliski, *Bayesian Modeling of Uncertainty in Low-Level Vision* (Kluwer, Boston, 1989).
62. T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," Nature (London) **317**, 314–319 (1985).
63. D. Terzopoulos, "Regularization of inverse problems involving discontinuities," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-8**, 413–424 (1986).
64. A. P. Pentland, "Automatic extraction of deformable part models," Int. J. Comput. Vision **4**, 107–126 (1990).
65. Y. G. Leclerc, "Constructing simple stable descriptions for image partitioning," Int. J. Comput. Vision **3**, 73–102 (1989).
66. A. Gelb, *Applied Optimal Estimation* (MIT Press, Cambridge, Mass., 1974).
67. A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1984).
68. K. L. Kelley, K. S. Gibson, and D. Nickerson, "Tristimulus specification of the Munsell Book of Color from spectrophotometric measurements," J. Opt. Soc. Am. **33**, 355–376 (1943).
69. D. Nickerson, "Spectrophotometric data for a collection of Munsell samples," (U.S. Department of Agriculture, Washington, D.C., 1957; available from Munsell Color Company, Baltimore, Md.).
70. Although the Munsell papers are a man-made collection of surfaces, analyses of natural surfaces[43,44] suggest that these are described by similar linear models.
71. E. Schrodinger, "Theorie der pigmente von grosster leuchtkraft," Ann. Phys. (Leipzig) **62**, 603–622 (1920), as discussed in Ref. 72.
72. G. Wyszecki, "Color appearance," in *Handbook of Perception and Human Performance,* K. R. Boff, L. Kaufman, and J. P. Thomas, eds. (Wiley, New York, 1986), pp. 9.1–9.56.
73. S. Rosch, "Die Kennzeichnung der Farben," Phys. Z. **29**, 83–91 (1928), as discussed in Ref. 72.
74. D. L. MacAdam, "The theory of the maximum visual efficiency of colored materials," J. Opt. Soc. Am. **25**, 249–252 (1935).
75. M. R. Pointer, "The gamut of real surface colours," Color Res. Appl. **5**, 145–155 (1980).
76. M. Richter and K. Witt, "The story of the DIN color system," Color Res. Appl. **11**, 138–145 (1986).
77. G. Wyszecki and W. S. Stiles, *Color Science—Concepts*

and Methods, Quantitative Data and Formulae, 2nd ed. (Wiley, New York, 1982).
78. J. M. Speigle and D. H. Brainard, "Luminosity thresholds: effects of test chromaticity and ambient illumination," J. Opt. Soc. Am. A **13**, 436–451 (1996).
79. D. A. Forsyth, "A novel approach to colour constancy," in *Proceedings of the International Conference on Computer Vision* (IEEE Computer Society Press, Los Alamitos, Calif., 1988), pp. 9–18.
80. CIE, *Colorimetry*, 2nd ed. (Bureau Central de la CIE, Paris, 1986).
81. P. DeMarco, J. Pokorny, and V. C. Smith, "Full-spectrum cone sensitivity functions for X-chromosome-linked anomalous trichromats", J. Opt. Soc. Am. A **9**, 1465–1476 (1992).
82. A. Grace, *Optimization Toolbox for Use with MATLAB—User's Guide* (MathWorks, Natick, Mass., 1990).
83. H. Fuchs, "Eine experimentelle undersuchung zur farbkonstanz," unpublished Ph.D. dissertation (University of Regensburg, Regensburg, Germany, 1992).
84. M. S. Drew and B. V. Funt, "Variational approach to interreflection in color images", J. Opt. Soc. Am. A **9**, 1255–1265 (1992).
85. H. Lee, "Method for computing the scene-illuminant chromaticity from specular highlights," J. Opt. Soc. Am. A **3**, 1694–1699 (1986).
86. M. D'Zmura and P. Lennie, "Mechanisms of color constancy," J. Opt. Soc. Am. A **3**, 1662–1672 (1986).
87. L. T. Moloney, Department of Psychology, New York University, New York, N.Y. (personal communication, 1995).
88. A. L. Yuille and H. H. Bulthoff, "Bayesian decision theory and psychophysics," in *Visual Perception: Computation and Psychophysics*, D. Knill and W. Richards, eds. (Cambridge U. Press, Cambridge, 1995), pp. 123–161.
89. R. N. Shepard, "Toward a universal law of generalization for psychological science," Science **237**, 1317–1323 (1987).
90. A. Blake and A. Zisserman, *Visual Reconstruction* (MIT Press, Cambridge, Mass., 1987).
91. P. Meer, D. Mintz, and A. Rosenfeld, "Robust regression methods for computer vision: a review," Int. J. Comput. Vision **6**, 59–70 (1991).
92. M. J. Black and P. Anandan, "A framework for the robust estimation of optical flow," in *Proceedings of the 4th International Conference on Computer Vision* (IEEE Computer Society Press, Los Alamitos, Calif., 1993), pp. 231–236.
93. W. T. Freeman, "The generic viewpoint assumption in a framework for visual perception," Nature (London) **368**, 542–545 (1994).
94. W. T. Freeman, "Exploiting the generic viewpoint assumption," Int. J. Comput. Vision **20**, 243–261 (1996).
95. K. R. K. Nielsen and B. A. Wandell, "Discrete analysis of spatial sensitivity models," J. Opt. Soc. Am. A **5**, 743–755 (1988).
96. N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals* (Dover, New York, 1986).
97. P. S. Laplace, *Theorie Analytique des Probabilities* (Courcier, Paris, 1812).
98. R. A. Fisher, *Statistical Methods and Scientific Inference*, 2nd ed. (Hafner, Oliver and Boyd, Edinburgh, 1959).
99. H. Jeffreys, *Theory of Probability* (Clarendon, Oxford, 1961).
100. G. E. P. Box and G. C. Tiao, "A Bayesian approach to the importance of assumptions applied to the comparison of variances," Biometrika **51**, 153–167 (1964).
101. D. V. Lindley, *Bayesian Statistics, a Review* (Society for Industrial and Applied Mathematics, Philadelphia, 1971).
102. S. F. Gull, "Bayesian inductive inference and maximum entropy," in *Maximum-Entropy and Bayesian Methods in Science and Engineering*, G. J. Erickson and C. R. Smith, eds. (Kluwer, Boston, 1988), pp. 53–74.
103. J. Skilling, "Classic maximum entropy," in *Maximum Entropy and Bayesian Methods, Cambridge*, J. Skilling, ed. (Kluwer, Dordrecht, The Netherlands, 1989), pp. 45–52.
104. D. J. C. MacKay, "Bayesian interpolation," Neural Comput. **4**, 415–447 (1992).