

MITSUBISHI ELECTRIC RESEARCH LABORATORIES
<http://www.merl.com>

Bayesian Decision Theory, the Maximum Local Mass Estimate, and Color Constancy

William T. Freeman, David H. Brainard

TR94-23 December 1995

Abstract

Computational vision algorithms are often developed in a Bayesian framework. Two estimators are commonly used: maximum a posteriori (MAP), and minimum mean squared error (MMSE). We argue that neither is appropriate for perception problems. The MAP estimator makes insufficient use of structure in the posterior probability. The squared error penalty of the MMSE estimator does not reflect typical penalties. We apply this new estimator to color constancy. An unknown illuminant falls on surfaces of unknown colors. We seek to estimate both the illuminant spectrum and the surface spectra from photosensor responses which depend on the product of these unknown spectra. In simulations, we show that the MLM method performs better than the MAP estimator, and better than two standard color constancy algorithms. The MLM estimate may prove useful in other vision problems as well.

IEEE Intl. Conf. on Computer Vision, Cambridge, MA, June, 1995

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Copyright © Mitsubishi Electric Research Laboratories, Inc., 1995
201 Broadway, Cambridge, Massachusetts 02139

Bayesian Decision Theory, the Maximum Local Mass Estimate, and Color Constancy

W. T. Freeman
Mitsubishi Electric Research Laboratories
201 Broadway
Cambridge, MA 02139, U.S.A.

D. H. Brainard
Department of Psychology
University of California
Santa Barbara, CA 93106, U.S.A.

In IEEE Intl. Conf. on Computer Vision,
Cambridge, MA, June, 1995.

Abstract

Vision algorithms are often developed in a Bayesian framework. Two estimators are commonly used: maximum a posteriori (MAP), and minimum mean squared error (MMSE). We argue that neither is appropriate for perception problems. The MAP estimator makes insufficient use of structure in the posterior probability. The squared error penalty of the MMSE estimator does not reflect typical penalties.

We describe a new estimator, which we call maximum local mass (MLM) [10, 26, 65], which integrates the local probability density. The MLM method is sensitive to local structure of the posterior probability, which MAP is not. 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 apply this new estimator to color constancy. An unknown illuminant falls on surfaces of unknown colors. We seek to estimate both the illuminant spectrum and the surface spectra from photosensor responses which depend on the product of these unknown spectra. In simulations, we show that the MLM method performs better than the MAP estimator, and better than two standard color constancy algorithms. The MLM method may prove useful in other vision problems as well.

1 The General Problem

The task of perception is to infer properties of the world from sensor measurements. For visual perception, such properties can be object shapes, colors, reflectances, or velocities. A common approach to this problem is Bayesian analysis, which combines the visual data with *prior probabilities* to find the *posterior probability* of properties of the world, which we will call *scene parameters*.

Typically, the goal is to choose a best estimate of the scene parameters. One selects an optimality criterion and uses the posterior probability distribution to find the optimal scene parameter estimate. Two decision rules are almost universally used: maximum a posteriori (MAP) and minimum mean squared error (MMSE). We believe that neither is the best choice for many problems in computational vision.

The MAP rule is to choose the scene parameter values with the highest posterior probability density. This estimator is closely related to maximum likelihood methods, and many computational vision algorithms employ it. Examples include algorithms for shape from shading, stereo, surface reconstruction, color constancy, motion perception, and object inference [29, 1, 61, 54, 16, 38, 63, 32, 57, 2, 39, 20]. Regularization can be interpreted as Bayesian inference using the MAP estimator [56, 62, 61]. Minimum description length analysis [18, 55] also has a Bayesian MAP interpretation [44].

The MMSE rule is to choose the scene parameter values that minimize the average squared distance from the true scene parameter values. It is simple to show that the mean of the posterior distribution is the MMSE estimate. This estimator is also in wide use; Kalman filtering [28], for example, uses it.

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, these estimates can be unsatisfactory. The MAP estimator is insensitive to the detailed structure of the posterior; only the point of maximum probability mass matters. The MMSE estimator can be overly influenced by outlying probability mass. Furthermore, computing the MMSE estimate may require a computationally intensive integration over the entire scene parameter space.

We argue that complicated posterior probability distributions, where these effects matter, occur in real computational vision problems. For example, it is common that several sets of scene parameters can explain the visual data equally well. This effect is given different names: is the *aperture problem* in motion perception [33]; *metamerism* in color perception [21]; the *bas relief* illusion in shape perception [40, 47]. As we will see in an example in Section 2, there can be structure in the posterior probabilities to help disambiguate multiple explanations which the MAP rule ignores. The MMSE rule, on the other hand, may select scene parameters which, while minimizing the squared error, do not explain the visual data at all.

The key to understanding these estimators, and to designing a new one, is the notion of a *loss function*. For a scene parameter \mathbf{x} , a loss function $L(\mathbf{x}, \tilde{\mathbf{x}})$ specifies the penalty for estimating $\tilde{\mathbf{x}}$ when the true scene parameter is \mathbf{x} . A loss function leads to a rule for

estimating the scene parameters from the posterior: choose the scene parameters that minimize the expected loss [8, 3]. The MAP rule assumes the loss function $L(\mathbf{x}, \tilde{\mathbf{x}}) = -\delta(\mathbf{x} - \tilde{\mathbf{x}})$, which we call the minus delta loss function. The MMSE rule assumes the squared error loss function, $L(\mathbf{x}, \tilde{\mathbf{x}}) = (\mathbf{x} - \tilde{\mathbf{x}})^2$.

We believe that neither loss function adequately approximates the loss for incorrect estimates in most perception problems. The minus delta function loss 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 saturates. For example, if we are trying to catch a ball thrown to us, small errors in the perceived size or velocity of the object 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.

The choice of estimation rule can strongly influence algorithm performance. We describe a loss function which is well-suited to perception. It integrates probability density to find the local probability mass in a region. We call it the local mass loss function, and the resulting estimation rule the *maximum local mass* (MLM) estimator. In the context of perception, both we [10, 26] and Yuille and Bulthoff [65] independently suggested this loss function, although neither computed with it. The function rewards approximately correct estimates, and gives a saturating penalty to incorrect estimates. It is related to Shepard’s notion of a “consequential region” [59]. It is in the same spirit as the penalty functions used in robust regression [5, 52, 4] which fit well-parameterized data but ignore outliers. However, the local mass loss function applies to the scene parameters, not to the observations, and affects the equations differently than regression penalty functions.

We illustrate these estimation ideas with a simple problem below. In Section 3, we apply the MLM estimator to a longstanding problem in computational vision: color constancy—how to infer the colors of objects viewed under unknown illumination. We show that the maximum local mass estimate performs better than two existing color constancy algorithms, and significantly better than the MAP estimate. We show how to compute the MLM estimate efficiently. This estimator may improve the performance of other vision algorithms.

2 A Simple Example

We observe a number, y , and are told that it is the product of two other numbers: $y = ab$. What are those two numbers, a and b ?

This is a degenerate case of a perception problem. The datum y corresponds to the visual data, such as an image. The parameters to be estimated, a and b , are the scene parameters. They could generalize to surface and illuminant color, descriptions of shape and reflectance, or estimates of motion. Typical of problems in computational vision, this problem is under-determined and non-linear.

Let us say the datum is $y = 1$, and that we know that $0 < a, b < 4$. From geometric considerations

alone, we can only say that the solution must lie along a segment of the hyperbola $ab = 1$.

A probabilistic analysis yields more. We want to find the posterior probability, $P(\mathbf{x}|\mathbf{y})$, of the parameter vector $\mathbf{x} = (a\ b)^T$, given the observation \mathbf{y} . (Lower-case boldface letters indicate vectors; upper-case boldface letters indicate matrices.) Using Bayes’ theorem, we write the posterior probability as

$$P(\mathbf{x} | \mathbf{y}) = \frac{P(\mathbf{y} | \mathbf{x}) P_{\mathbf{x}}(\mathbf{x})}{P_{\mathbf{y}}(\mathbf{y})} = \frac{[\text{likelihood}] [\text{priors}]}{[\text{normalization}]}. \quad (1)$$

The normalization constant, $P_{\mathbf{y}}(\mathbf{y})$, is independent of the parameters \mathbf{x} that we seek to estimate. (Following convention, we use P to denote the probability density function of its argument. Where possible, we add a subscript to identify the function). $P_{\mathbf{x}}(\mathbf{x})$ is the prior probability of the scene parameter \mathbf{x} . $P(\mathbf{y} | \mathbf{x})$ is called the *likelihood function*. It contains a forward model: given scene parameters, it tells the probability of an observation \mathbf{y} . We allow for observation noise, so the forward model is probabilistic. A deterministic *rendering function*, $\mathbf{f}(\mathbf{x})$, returns the observation associated with scene parameters \mathbf{x} in the absence of noise. Here, the rendering function is $\mathbf{f}(\mathbf{x}) = a\ b = \mathbf{x}_1\ \mathbf{x}_2$. We assume the observation noise to be normally distributed with mean zero and variance σ^2 . Then the likelihood function is

$$P(\mathbf{y} | \mathbf{x}) = \frac{1}{(\sqrt{2\pi}\sigma^2)^N} e^{-\frac{\|\mathbf{y} - \mathbf{f}(\mathbf{x})\|^2}{2\sigma^2}}. \quad (2)$$

We will assume “uniform” priors, $P_{\mathbf{x}}(\mathbf{x}) = \frac{1}{16}$, over the range $[0, 4] \times [0, 4]$, and zero elsewhere. Using this prior and the likelihood of Eq. (2) in Bayes’ theorem, Eq. (1), gives the posterior distribution for our example:

$$P(\mathbf{x} | \mathbf{y} = 1) = \begin{cases} C e^{-\frac{(1-ab)^2}{2\sigma^2}} & \text{if } 0 < a, b < 4 \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

where we have combined constants over the variables of interest into C . This posterior distribution is shown in Fig. 1 (a). Points along the hyperbola $ab = 1$ form a ridge of highest probability. Because of the observation noise, other parameter pairs have a non-zero probability. Note from Fig. 1 (b) that, while the ridge has equal height everywhere, it is wider near $(1, 1)$ than at other points along the hyperbola.

The optimal scene parameters \mathbf{x} minimize the expected loss, $R(\tilde{\mathbf{x}} | \mathbf{y})$, called the *Bayes risk*. Taking the expectation of the loss over the posterior distribution, we have:

$$R(\tilde{\mathbf{x}} | \mathbf{y}) = \int P(\mathbf{x}|\mathbf{y}) L(\mathbf{x}, \tilde{\mathbf{x}}) d\mathbf{x} \quad (4)$$

For the special case of shift invariant loss functions, $L(\mathbf{x}, \tilde{\mathbf{x}}) = l(\mathbf{x} - \tilde{\mathbf{x}})$, the Bayes risk is $l(-\tilde{\mathbf{x}})$ convolved with the posterior $P(\tilde{\mathbf{x}}|\mathbf{y})$.

Part (a) of Fig. 2 shows the loss associated with the MAP rule. Part (d) shows the corresponding Bayes

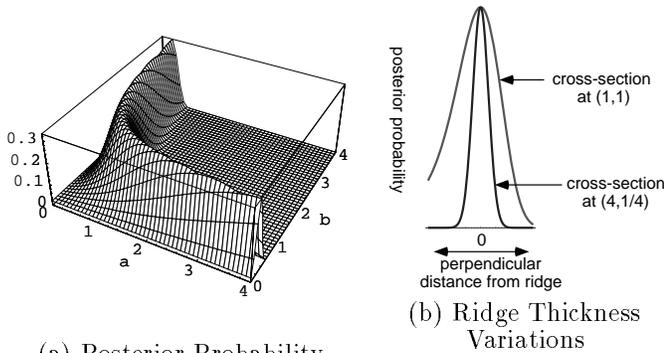


Figure 1: Bayesian analysis of the problem $ab = 1$. Assuming uniform prior probabilities over the graphed region, (a) shows the posterior probability for gaussian observation noise of variance 0.18. The noise broadens the geometric solution into a hyperbola-shaped ridge of maximum probability. (b) Note the different thickness of the ridge; some parts have more local probability mass than others, even though the entire ridge has a constant maximum height.

risk for our $y = ab$ problem with uniform priors. For the MAP estimator, the Bayes risk is minus the posterior probability. (In this and related figures, *increasing* loss is plotted upward to show the extrema more clearly.) Every point along the hyperbolic ridge in the figure has equal probability density, so the MAP rule does not give a unique estimate. Such ridges occur in vision and other estimation problems when many alternatives can account for the observations equally well. MAP estimation ignores all variations in *width* of the ridge, which can be a significant source of information.

Fig. 2 (b) shows the loss function of the MMSE estimator. The Bayes risk is shown in (e). Note that the MMSE estimate is $a = b = 1.3$, 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 which is very unlikely to have generated the observed datum.

The loss function of the maximum local mass (MLM) method addresses deficiencies of both the MAP and MMSE estimators. We define the minus local mass loss function as a gaussian of small covariance:

$$L(\mathbf{x}, \tilde{\mathbf{x}}) = -e^{-|\mathbf{K}_L^{-\frac{1}{2}}(\mathbf{x} - \tilde{\mathbf{x}})|^2}, \quad (5)$$

where we adopt the notation that $|\mathbf{K}^{-\frac{1}{2}} \mathbf{x}|^2 = \mathbf{x}^T \mathbf{K}^{-1} \mathbf{x}$. For matrices \mathbf{K}_L of sufficiently small eigenvalues, this loss function rewards approximately correct answers, yet penalizes all grossly incorrect answers equally, see Fig. 2 (c). Fig. 2 (f) shows the corresponding expected loss for the $ab = 1$ problem. The MLM estimate is $(1.0, 1.0)$, which both accounts well for the datum $y = 1$ and is the estimate that is most probable to be approximately correct, in the sense of maximizing the local probability mass. The ridge of the posterior, Fig. 1 (a), is widest near $(1.0, 1.0)$ (see cross-sections in Fig. 1 (b)). More parameter values near $(1.0, 1.0)$ could have caused the observed datum

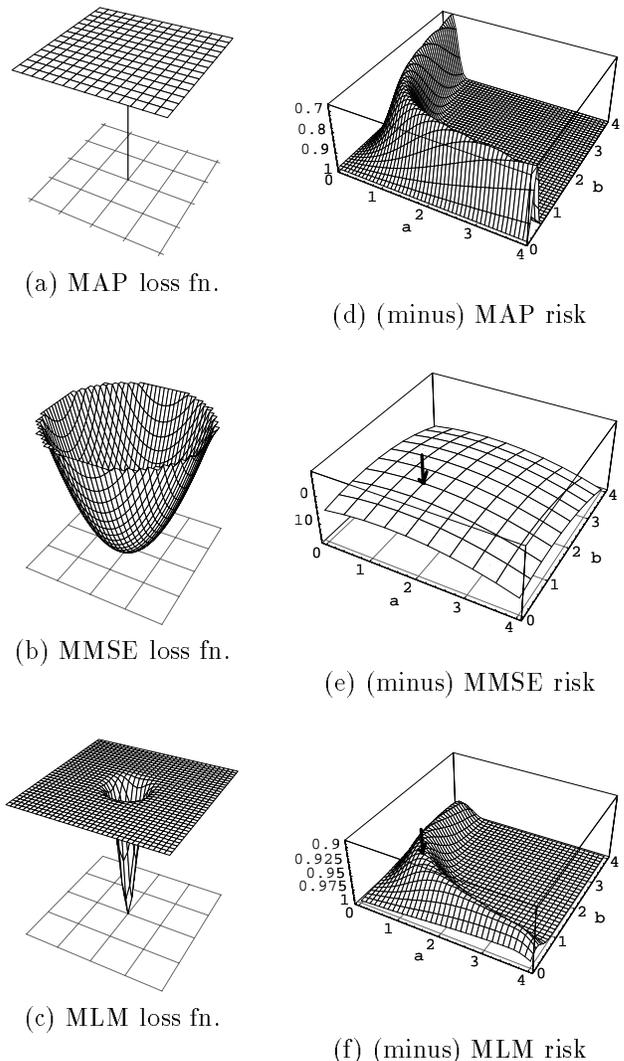


Figure 2: Left column: Three loss functions. Plots show penalty for guessing parameter values offset from the actual value, taken to be the plot center. (a) Minus delta function loss, assumed in MAP estimation. Only *precisely* the correct answer matters. (b) Squared error loss (a parabola), used in MMSE estimation. Very wrong guesses can carry inordinate influence. (c) Minus local mass loss function. Nearly correct answers are rewarded while all others carry nearly equal penalty. Right column: Corresponding expected loss, or Bayes risk, for the $y = ab$ problem. Note: loss *increases* vertically, to show extrema. (d) Expected loss for MAP estimator is minus the posterior probability. There is no unique point of minimum loss. (e) The minimum mean squared error estimate, $(1.3, 1.3)$ (arrow) does not lie along the ridge of solutions to $ab = 1$. (f) The minus local mass loss favors the point $(1.0, 1.0)$ (arrow), where the ridge of high probability is widest. There is the most probability mass in that local neighborhood.

than those near any other estimate.

Our simple product example shows us several things. The loss function is important and can change

the optimal estimate. The MAP solution ignores information from the widths of the ridges of maximum likelihood. The MMSE solution, while optimal in the mean squared error sense, can give solutions which can be non-sensical from a perceptual standpoint ($1.3 \times 1.3 \neq 1.0$). The maximum local mass, MLM, solution integrates probability density to measure the local probability mass. It exploits structure in the posterior distribution and assumes a penalty function which is reasonable for perception problems. We will see that its computation involves local operations in the parameter space. We develop a method to compute the MLM estimate and apply that to a long standing problem in computational vision: color constancy.

3 Application: Color Constancy

The matte reflection from a surface is the product of the incident illumination spectrum and the surface reflectance spectrum. In general, the illumination spectrum is not known, yet human observers do well at estimating colors of surfaces, related to their reflectance spectra, viewed under unknown illumination. The perceived colors are roughly constant [21].

Humans show some color constancy even in a simplified “mondrian world” of matte surfaces, illuminated by a single diffuse illuminant [51, 14]. How color constancy could be achieved in such a simplified world has received much attention [51, 41, 15, 42, 12, 49, 23, 30, 63, 10, 20]. Here we develop the MLM estimate for the illuminant. We expect that insights gained from studying this problem will apply to the more general case of color constancy in the natural image, and may yield improved processing of color images [13]. In addition, our results can serve as an ideal observer benchmark for human performance in the simplified domain.

We consider a collection of N_l matte surfaces. We write the reflectance of the j th surface as a column vector, \mathbf{s}_j . The entries specify the fraction of incident light reflected in N_λ bands throughout the visible spectrum. Similarly, we specify the illuminant spectral power distribution with a column vector \mathbf{e} . The spectral power distribution of the light reaching the imaging device is \mathbf{c}_j , the entry-by-entry product of those two spectra. We write $\mathbf{c}_j = \text{diag}(\mathbf{e}) \mathbf{s}_j$, where $\text{diag}(\mathbf{x})$ places the elements of the vector \mathbf{x} along the diagonal of a matrix, with zeros elsewhere.

A visual system typically samples the reflected spectrum \mathbf{c}_j with N_r classes of linear photoreceptors. We specify the photoreceptor sensitivities with a matrix \mathbf{R} . The pq th element of \mathbf{R} is the sensitivity of the p th sensor class to light in the q th wavelength band. We let the vector \mathbf{r}_j represent the responses of each of the sensor classes to the reflected spectrum \mathbf{c}_j . Without sensor noise, we have the rendering equation for the sensor responses \mathbf{r}_j to the j th colored surface as a function of the scene parameters, \mathbf{e} and \mathbf{s}_j :

$$\mathbf{r}_j = \mathbf{f}(\mathbf{e}, \mathbf{s}_j) = \mathbf{R} \mathbf{c}_j = \mathbf{R} \text{diag}(\mathbf{e}) \mathbf{s}_j. \quad (6)$$

We seek to find the illuminant spectrum \mathbf{e} and the surface spectra \mathbf{s}_j from the photoreceptor responses \mathbf{r}_j to each surface.

3.1 Previous Work

3.1.1 Representation

Researchers have developed finite-dimensional linear models for naturally occurring surface and illuminant spectra. An illuminant spectrum lies within an N_e dimensional linear model \mathbf{B}_e if we can write $\mathbf{e} = \mathbf{B}_e \mathbf{w}_e$, where \mathbf{B}_e is a matrix (N_λ by N_e) and \mathbf{w}_e is an N_e dimensional column vector. The entries of \mathbf{w}_e are called the linear model weights for the illuminant \mathbf{e} . Similarly, the surface reflectance spectra lie within the linear model \mathbf{B}_s if we can write $\mathbf{s}_j = \mathbf{B}_s \mathbf{w}_{sj}$, where \mathbf{B}_s is an N_λ by N_s matrix and \mathbf{w}_{sj} is an N_s dimensional column vector.

Linear models with either three or four basis functions describe a large sample of measured illuminants [36] and the CIE has standardized a three dimensional linear model for daylights [17]. Linear models with dimension as low as three also capture a large percentage of the variance of measured surface reflectance spectra [48, 34]. Brainard [9] provides a handbook-style review of the use of linear models in computational color vision.

In this representation, the rendering equation, Eq. (6), becomes

$$\begin{aligned} \mathbf{r}_j = \mathbf{f}(\mathbf{w}_e, \mathbf{w}_{sj}) &= \mathbf{R} \text{diag}(\mathbf{B}_e \mathbf{w}_e) \mathbf{B}_s \mathbf{w}_{sj} \\ &= \mathbf{R} \text{diag}(\mathbf{B}_s \mathbf{w}_{sj}) \mathbf{B}_e \mathbf{w}_e. \end{aligned} \quad (7)$$

3.1.2 Algorithms

Most modern color constancy algorithms incorporate linear model constraints. Maloney and Wandell [49] showed that when there are N_r classes of photoreceptors, the rendering equation may be inverted if the illuminants lie within an N_r dimensional linear model and that the surfaces lie within an N_{r-1} dimensional linear model. We refer to their algorithm as the subspace algorithm [49, 64]. For human vision there are only three classes of cone photoreceptors, so that Maloney and Wandell’s analytical result applies only if the surface spectra lie within a two-dimensional linear model, which is not the case for natural surfaces.

In Buchsbaum’s algorithm [15], the illuminants and surfaces must lie within N_r dimensional linear models. The algorithm assumes that the spatial mean of the surface reflectances is constant for all images. This is the “gray world” assumption, which implies that color constancy must break down for images where the mean reflectance differs from the assumed mean. However, the human visual system maintains at least partial color constancy in spite of shifts of the mean surface reflectance [27, 11, 50].

Recently, researchers have studied color constancy as a statistical estimation problem. Trussell and Vhrel [63] have used a maximum likelihood approach (closely related to MAP estimation) to estimate the illuminant. D’Zmura and Iverson [20] also use a MAP approach. Brainard and Freeman [10] applied an MMSE criterion to estimate the illuminant.

None of the previous approaches to this problem uses an error criterion which is tailored to the perceptual problem at hand. Here, we show the benefits of the maximum local mass method. To specify our

algorithm we must select the prior probabilities, then minimize the expected loss.

3.2 Priors

For our simulations, we created three dimensional linear models for surfaces and illuminants. We choose a linear model for surfaces by performing principal components analysis on the data of Kelly et al. [37, 53] who measured the reflectance functions of 462 Munsell papers. We then found the best fitting model weights for each individual surface in the data set. We fit these distributions with gaussians of the sample mean and covariance matrix.

We took our three dimensional linear model for illuminants to be the CIE linear model for daylight [17]. To generate a prior distribution on the weights, we generated CIE daylight with correlated color temperatures drawn according to a normal distribution with mean 6500°K and standard deviation 4000°K. (Draws outside the range 3000° to 25000°K were rejected.) We scaled these illuminants by factors drawn uniformly between 1 and 10. We computed the linear model weights \mathbf{w}_e on each function and fit the resulting distribution with a gaussian.

Note that it is not required that the priors be gaussians for the results which follow.

3.3 Approximation for the Local Probability Mass

We want to apply the maximum local mass estimator to the color constancy problem. The Bayes risk, which we seek to minimize, involves an integral over the entire scene parameter space, Eq. (4), yet our loss function, Eq. (5), is only appreciably non-zero near the estimate $\tilde{\mathbf{x}}$. We want to exploit that locality to find an approximation for the Bayes risk which 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.

To examine this limit, we write the observation noise covariance as $\frac{1}{\tau} K_n$, where K_n is fixed, and let τ become large. Similarly, to form a localized loss function, we write its covariance as $\frac{1}{\mu} K_L$, where a fixed K_L defines the shape of the loss covariance, and we let μ become large. For the loss function to sample the full width of the ridges of the posterior distribution, the loss function must be larger than those ridges, so we will study the regime where $\frac{\mu}{\tau}$ is small.

Combining Eqs. (1), (2) and (4) for the posterior, likelihood and risk, we have:

$$\begin{aligned} R(\tilde{\mathbf{x}}|\mathbf{y}) &= C \int [\text{likelihood}] [\text{priors}] [\text{loss function}] d\mathbf{x} \\ &= -C \int \exp[-\frac{\tau}{2}|\mathbf{K}_n^{-\frac{1}{2}}(\mathbf{r} - \mathbf{f}(\mathbf{x}))|^2] \mathbf{P}_{\mathbf{x}}(\mathbf{x}) \\ &\quad \exp[-\frac{\mu}{2}(|\mathbf{K}_L^{-\frac{1}{2}}(\mathbf{x} - \tilde{\mathbf{x}})|^2)] d\mathbf{x} \end{aligned} \quad (8)$$

For an integral of the form

$$I(\tau) = \int \exp[-\tau\phi(\mathbf{x})] g(\mathbf{x}) d\mathbf{x}, \quad (9)$$

the leading order term in an asymptotic expansion for large τ is [6]:

$$I(\tau) \approx \frac{e^{-\tau\phi(\mathbf{x}_0)}}{\sqrt{|\det(\phi_{x_i x_j}(\mathbf{x}_0))|}} \left(\frac{2\pi}{\tau}\right)^{\frac{n}{2}} g(\mathbf{x}_0), \quad (10)$$

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

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

$$\phi(\mathbf{x}) = \frac{1}{2}|\mathbf{K}_n^{-\frac{1}{2}}(\mathbf{y} - \mathbf{f}(\mathbf{x}))|^2 + \frac{\mu}{2\tau}|\mathbf{K}_L^{-\frac{1}{2}}(\mathbf{x} - \tilde{\mathbf{x}})|^2. \quad (11)$$

Then, using the approximation of Eq. (10), we have:

$$\begin{aligned} R(\tilde{\mathbf{x}}|\mathbf{y}) &\approx -C e^{[-\tau(\frac{1}{2}|\mathbf{K}_n^{-\frac{1}{2}}(\mathbf{y} - \mathbf{f}(\mathbf{x}_0))|^2 + \frac{\mu}{2\tau}|\mathbf{K}_L^{-\frac{1}{2}}(\mathbf{x}_0 - \tilde{\mathbf{x}})|^2)]} \times \\ &\quad \frac{\mathbf{P}_{\mathbf{x}}(\mathbf{x}_0)}{\sqrt{|\det(\phi_{x_i x_j}(\mathbf{x}_0))|}}. \end{aligned} \quad (12)$$

Twice differentiating $\phi(\mathbf{x})$ in Eq. (11) gives

$$\phi_{x_i x_j}(\mathbf{x}_0) = \mathbf{f}'_i{}^T \mathbf{K}_n^{-1} \mathbf{f}'_j - (\mathbf{y} - \mathbf{f}(\mathbf{x}_0))^T \mathbf{K}_n^{-1} \mathbf{f}''_{ij} + \frac{\mu}{\tau} [\mathbf{K}_L^{-1}]_{ij}, \quad (13)$$

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

$$\mathbf{f}'_i = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_i} \Big|_{\mathbf{x}=\mathbf{x}_0}, \text{ and } \mathbf{f}''_{ij} = \frac{\partial^2 \mathbf{f}(\mathbf{x})}{\partial x_i \partial x_j} \Big|_{\mathbf{x}=\mathbf{x}_0}. \quad (14)$$

To apply Eq. (12) for the expected loss at $\tilde{\mathbf{x}}$, we need to find an expansion point \mathbf{x}_0 where $\phi(\mathbf{x})$ of Eq. (11) is minimized. If we restrict attention to evaluating the risk 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 Eq. (12) to evaluate the risk at points of maximum likelihood. (We believe setting $\mathbf{x}_0 = \tilde{\mathbf{x}}$ in Eq. (12) will well approximate the risk for other points, as well).

At $\mathbf{x}_0 = \tilde{\mathbf{x}}$, the difference between the risk $R(\tilde{\mathbf{x}}|\mathbf{y})$ in Eq. (12) and the negative posterior, Eq. (1), is the factor, $\frac{1}{\sqrt{|\det(\phi_{x_i x_j}(\tilde{\mathbf{x}}))|}}$. This allows the local mass risk to respond to the width as well as the height of probability ridges such as those shown in Fig. 1.

A related approximation is used in Bayesian statistics, dating to Laplace [43, 22, 35, 8]. The first two terms of $\phi_{x_i x_j}(\mathbf{x}_0)$ form the conditional Fisher information matrix, \mathbf{I}_{ij} [22, 3]. It can be used to marginalize the posterior over nuisance parameters [7, 45, 3], yielding a factor of $\frac{1}{\sqrt{\det(\mathbf{I}_{ij})}}$ after integration (for indices i, j of the marginalized variables). Recent authors have exploited this in parameter estimation [31, 60, 46] and computer vision [24, 25]. However, if one marginalizes over parameters, one can not estimate their optimal values. Furthermore, for underdetermined estimation problems, $\det(\mathbf{I}_{ij})$ can be zero, spoiling the approximation. The loss function approach avoids this singularity, and allows for trading off accuracy requirements among scene parameter components.

4 Results

We simulated scenes consisting of 8 randomly drawn surfaces under a single randomly drawn illuminant. We used the rendering equation Eq. (7) and the Smith-Pokorny estimates of the human cone sensitivities [19] to compute the cone responses, perturbing them with additive gaussian noise.

We searched for the minimum loss illuminant. At a given illuminant value in the search, we solved the rendering equation, Eq. (7), for the surface weight values which accounted for the observations. We then used Eq. (12) to find the local mass loss at these values. We used a BFGS variable metric optimization algorithm [58] (the `constr` optimization routine in Matlab 4.2a) to descend to a locally optimal estimate. A broad range of τ and μ give similar results; we used $\tau = 100$, $\mu = 10000$, and set K_L equal to the covariance matrix of the prior probability of the scene parameters.

For comparison, we estimated the illuminant using the subspace [64], gray world [15], and MAP algorithms. We used the gray world assumption to set the overall scaling of the subspace algorithm, which only recovers relative spectra.

4.1 Comparison among algorithms

The four panels of Figure 3 show the performance of the four algorithms for a single illuminant. The actual illuminant spectrum is shown by the solid line in each panel of the figure. The dotted lines in each panel show individual estimates produced by each algorithm. To produce each estimate, we drew a set of surfaces from the prior distribution and computed the sensor responses. We then added 1% noise to the sensor responses and applied each algorithm. The maximum local mass estimates are grouped closest to the actual illuminant spectrum. The gray world algorithm estimates are correct on average but have wider variability. The MAP estimator ignores relevant information in the posterior distribution, which results in a systematic bias of its estimates. The subspace algorithm is not guaranteed to work under these conditions, and its estimates are extremely noisy [10].

We also investigated how the algorithms performed under some violations of the assumed prior statistics. In one set of simulations, we varied the illuminant while continuing to draw surfaces from the prior distribution for surfaces. We used two illuminants in addition to the 6500°K daylight. One was a 4000°K daylight and the other a 10000°K daylight. We summarize the results by computing the average (over 19 individual runs) fractional root mean squared error (RMSE) between the estimate and true illuminant. Figure 4 (a) shows this error measure for each algorithm and each illuminant. For all three illuminants, the MLM algorithm performs best, and the quality of its estimates is independent of which illuminant is simulated.

In a second set of simulations, we used the 6500°K illuminant but biased the mean of the distribution used to draw the surfaces. We used two biases. For one, we biased the surface mean so that the mean sensor responses under the 6500°K illuminant were the same as for the unbiased surfaces under the 4000°K illuminant. For the other, we biased the surface mean so that the mean sensor responses under the 6500°K illuminant were the same as for the unbiased surfaces

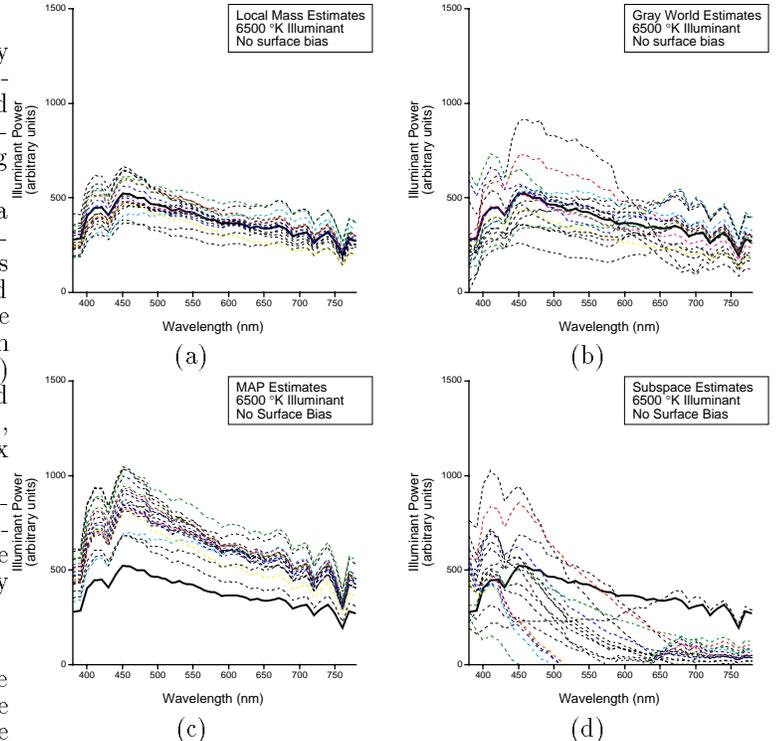


Figure 3: Visual comparison of illumination spectrum estimates for four color constancy algorithms: local mass, gray world, MAP and subspace. For a given illuminant, shown in dark line, a set of surfaces was drawn from the prior distribution 19 times. For each draw, each algorithm estimated the illuminant reflectance spectrum. The maximum local mass estimates, (a), are grouped closest to the actual illumination spectrum. The gray world algorithm estimates, (b), have wider variability. The MAP estimator, (c), ignores relevant information in the posterior distribution, which results in a systematic bias of its estimates. The subspace algorithm, (d), was not designed to work under the tested conditions, and performs poorly.

under the 10000°K illuminant. This manipulation is a test of the robustness of the algorithms with respect to violations of the gray world assumption. Figure 4 (b) shows the fractional RMSE for the unbiased condition and the two biased conditions. Biasing the surfaces impacts the performance of all of the algorithms. However, the MLM algorithm is clearly the most robust with respect to the manipulation, and its error is roughly half that of the next best algorithm (the gray world algorithm) for all three surface draw conditions. The prior built into the MLM algorithm is not so strong as to make it brittle when the particular scene is biased from the “average” scene.

5 Conclusions

We study perception from the point of view of Bayesian decision theory. We argue that neither of the two commonly used estimators, MAP and MMSE, is appropriate for the task of perception. We propose a new estimator, called the maximum local mass, MLM, estimate. This estimate maximizes an integral

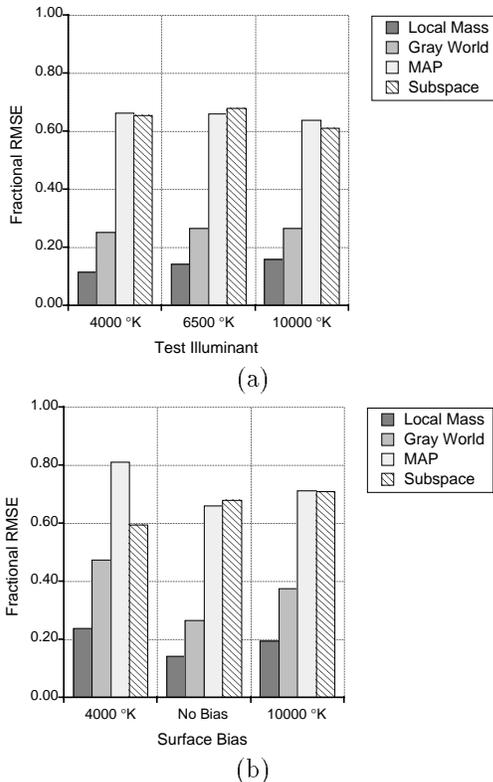


Figure 4: Summary results. (a) shows the performance of all four algorithms for three illuminants. (b) shows the performance of all four algorithms for three surface draw conditions. The performance measure is the average (over 19 individual runs) fractional root mean squared error (RMSE) between the estimate and true illuminant. For all conditions, the MLM estimate performs substantially better than the other algorithms. It is seen to be robust against these violations of its prior assumptions.

over the local posterior probability density. The MLM estimate assumes a loss function appropriate to perception problems: it rewards approximately correct answers, and penalizes all grossly incorrect answers equally. The maximum local mass estimator finds the most probable approximately correct answer.

The MLM estimate may be most useful when a set of competing explanations account for the observed image data. Then the MLM estimate exploits important local structure in the posterior probability that the MAP estimate overlooks. We give an analytic approximation for the expected loss to be minimized, valid for low levels of observation noise. We show with a simple example that the choice of estimator matters, and that the MLM estimate can give a sensible answer where the MAP and MMSE estimates do not.

We apply the MLM method to a long-standing computational vision problem: color constancy. We estimated illuminant and surface spectral compositions from photosensor observations of products of such spectra. In computer simulations, we compared our MLM algorithm with the MAP estimate, and two standard color constancy algorithms, the subspace al-

gorithm and the gray-world algorithm. The MLM algorithm performed better than the gray-world and subspace algorithms, and significantly better than the MAP estimate. We compared the algorithms under some violations of the prior assumptions used by the algorithms. The MLM method still performed well, and better than the other algorithms.

We believe use of the maximum local mass estimate may improve the performance of other Bayesian computational vision algorithms as well.

6 Acknowledgements

The authors acknowledge helpful conversations with E. Adelson, M. D’Zmura, G. Iverson, D. Mumford, D. Pelli, E. Simoncelli, B. Wandell, and A. Yuille. DHB was supported by NEI grant EY 10016.

References

- [1] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9:147–169, 1985.
- [2] P. N. Belhumeur. A computational theory for binocular stereopsis. In D. Knill and W. Richards, editors, *Visual Perception: Computation and Psychophysics*. Cambridge Univ. Press, 1995.
- [3] J. O. Berger. *Statistical decision theory and Bayesian analysis*. Springer, 1985.
- [4] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proc. 4th Intl. Conf. Comp. Vis.*, pages 231–236. IEEE, 1993.
- [5] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.
- [6] N. Bleistein and R. A. Handelsman. *Asymptotic expansions of integrals*. Dover, 1986.
- [7] G. E. P. Box and G. C. Tiao. A Bayesian approach to the importance of assumptions applied to the comparison of variances. *Biometrika*, 51(1 and 2):153–167, 1964.
- [8] G. E. P. Box and G. C. Tiao. *Bayesian inference in statistical analysis*. John Wiley and Sons, Inc., 1973.
- [9] D. H. Brainard. Colorimetry. In M. Bass, editor, *OSA Handbook of Optics*. Opt. Soc. Am., Washington, D.C., 1994.
- [10] D. H. Brainard and W. T. Freeman. Bayesian method for recovering surface and illuminant properties from photosensor responses. In *Proc. SPIE*, volume 2179, San Jose, CA, Feb. 1994. Human Vision, Visual Processing and Digital Display V.
- [11] D. H. Brainard and J. M. Speigle. Achromatic loci measured under realistic viewing conditions. *Invest. Ophthalm. and Visual Sci. Suppl.*, 35:1328, 1994.
- [12] D. H. Brainard and B. A. Wandell. Analysis of the retinex theory of color vision. *J. Opt. Soc. Am. A*, 3:1651–1661, 1986.
- [13] D. H. Brainard and B. A. Wandell. Calibrated processing of image color. *Color Res. Appl.*, 15:266–271, 1990.
- [14] D. H. Brainard, B. A. Wandell, and E. J. Chichilnisky. Color constancy: from physics to appearance. *Current dir. in psychol. science*, 2:165–170, 1993.
- [15] G. Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310:1–26, 1980.
- [16] H. H. Bulthoff. Bayesian models for seeing shapes and depth. *J. Theo. Biol.*, 2(4), 1991.
- [17] CIE. Colorimetry. Bureau Central de la CIE, 1986.
- [18] T. Darrell, S. Sclaroff, and A. Pentland. Segmentation by minimal description. In *Proc. 3rd Intl. Conf. Comp. Vis.*, pages 112–116. IEEE, 1990.

- [19] 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.
- [20] M. D’Zmura, G. Iverson, and B. Singer. Probabilistic color constancy. In R. D. L. et. al., editor, *Geometric representations of perceptual phenomena: papers in honor of Tarow Indow’s 70th birthday*. Lawrence Erlbaum, Hillsdale, NJ, in press.
- [21] R. M. Evans. *The perception of color*. Wiley, New York, 1974.
- [22] R. A. Fisher. *Statistical methods and scientific inference*. Hafner, 1959.
- [23] D. A. Forsyth. A novel approach to colour constancy. In *Proc. 1st Intl. Conf. Comp. Vis.*, pages 9–18. IEEE, 1988.
- [24] W. T. Freeman. Exploiting the generic view assumption to estimate scene parameters. In *Proc. 4th Intl. Conf. Comp. Vis.*, pages 347–356. IEEE, 1993.
- [25] W. T. Freeman. The generic viewpoint assumption in a framework for visual perception. *Nature*, 368(6471):542–545, April 7 1994.
- [26] W. T. Freeman. The generic viewpoint assumption in a Bayesian framework. In D. Knill and W. Richards, editors, *Visual Perception: Computation and Psychophysics*. Cambridge Univ. Press, 1995. in press.
- [27] H. Fuchs. *Eine experimentelle untersuchung zur farbkonstanz*. PhD thesis, University of Regensburg, 1992.
- [28] A. Gelb, editor. *Applied optimal estimation*. MIT Press, 1974.
- [29] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Pat. Anal. Mach. Intell.*, 6:721–741, 1984.
- [30] R. Gershon and A. D. Jepson. The computation of color constant descriptors in chromatic images. *Color Res. Appl.*, 14:325–334, 1989.
- [31] S. F. Gull. Bayesian inductive inference and maximum entropy. In G. J. Erickson and C. R. Smith, editors, *Maximum Entropy and Bayesian Methods in Science and Engineering*, volume 1. Kluwer, 1988.
- [32] D. J. Heeger and E. P. Simoncelli. Model of visual motion sensing. In L. Harris and M. Jenkin, editors, *Spatial Vision in Humans and Robots*. Cambridge Univ. Press, 1992.
- [33] B. K. P. Horn and B. G. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
- [34] T. Jaaskelainen, J. Parkkinen, and S. Toyooka. A vector-subspace model for color representation. *J. Opt. Soc. Am. A*, 7:725–730, 1990.
- [35] H. Jeffreys. *Theory of probability*. Oxford, Clarendon Press, 1961.
- [36] D. B. Judd, D. L. MacAdam, and G. W. Wyszecki. Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.*, page 1031, 1964.
- [37] K. L. Kelly, 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.
- [38] D. Kersten. Transparency and the cooperative computation of scene attributes. In M. S. Landy and J. A. Movshon, editors, *Computational Models of Visual Processing*, chapter 15. MIT Press, Cambridge, MA, 1991.
- [39] D. C. Knill, D. Kersten, and A. Yuille. A Bayesian formulation of visual perception. In D. Knill and W. Richards, editors, *Visual Perception: Computation and Psychophysics*. Cambridge Univ. Press, 1995.
- [40] J. J. Koenderink and A. J. van Doorn. Affine structure from motion. *J. Opt. Soc. Am. A*, 8(2):374–385, 1991.
- [41] E. H. Land. The retinex theory of color vision. *Sci. Am.*, 237(6):108–128, 1977.
- [42] E. H. Land. Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. *Proc. Nat. Acad. Sci. USA*, 80:5163–5169, 1983.
- [43] P. S. Laplace. *Theorie analytique des probabilites*. Courcier, 1812.
- [44] Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *Intl. J. Comp. Vis.*, 3:73–102, 1989.
- [45] D. V. Lindley. *Bayesian statistics, a review*. Society for Industrial and Applied Mathematics (SIAM), 1972.
- [46] D. J. C. MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992.
- [47] J. Malik. Newton Institute, Cambridge, England, July 1993. lecture.
- [48] L. T. Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am. A*, 3(10):1673–1683, 1986.
- [49] L. T. Maloney and B. A. Wandell. Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc. Am. A*, 3(1):29–33, 1986.
- [50] J. J. McCann. Psychophysical experiments in search of adaptation and the gray world. In *Proc. IS&T 47th Ann. Mtg.*, pages 397–401, 1994.
- [51] J. J. McCann, S. P. McKee, and T. H. Taylor. Quantitative studies in retinex theory: a comparison between theoretical predictions and observer responses to the ‘color mondrian’ experiments. *Vision Research*, 16:445–458, 1976.
- [52] P. Meer, D. Mintz, and A. Rosenfeld. Robust regression methods for computer vision: a review. *Intl. J. Comp. Vis.*, 6(1):59–70, 1991.
- [53] D. Nickerson. Spectrophotometric data for a collection of Munsell samples. U.S. Department of Agriculture Report, 1957.
- [54] A. P. Pentland. Local shading analysis. *IEEE Pat. Anal. Mach. Intell.*, 6(2):170–187, 1984.
- [55] A. P. Pentland. Automatic extraction of deformable part models. *Intl. J. Comp. Vis.*, 4:107–126, 1990.
- [56] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(26):314–139, 1985.
- [57] A. R. Pope and D. G. Lowe. Learning object recognition models from images. In *Proc. 4th Intl. Conf. Comp. Vis.*, pages 296–301. IEEE, 1993.
- [58] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge Univ. Press, 1992.
- [59] R. N. Shepard. Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323, 1987.
- [60] J. Skilling. Classic maximum entropy. In J. Skilling, editor, *Maximum Entropy and Bayesian Methods*, Cambridge, pages 45–52. Kluwer, 1989.
- [61] R. Szeliski. *Bayesian Modeling of Uncertainty in Low-level Vision*. Kluwer Academic Publishers, Boston, 1989.
- [62] D. Terzopoulos. Regularization of inverse problems involving discontinuities. *IEEE Pat. Anal. Mach. Intell.*, 8(4):413–424, 1986.
- [63] H. J. Trussell and M. J. Vrhel. Estimation of illumination for color correction. In *Proc. ICASSP*, pages 2513–2516, 1991.
- [64] B. A. Wandell. The synthesis and analysis of color images. *IEEE Pat. Anal. Mach. Intell.*, 9:2–13, 1987.
- [65] A. L. Yuille and H. H. Bulthoff. Bayesian decision theory and psychophysics. In D. Knill and W. Richards, editors, *Visual Perception: Computation and Psychophysics*. Cambridge Univ. Press, 1995. in press.