Color Constancy with Spatio-Spectral Statistics

Ayan Chakrabarti, *Member, IEEE*, Keigo Hirakawa, *Member, IEEE*, and Todd Zickler, *Member, IEEE*

Abstract—We introduce an efficient maximum likelihood approach for one part of the color constancy problem: removing from an image the color cast caused by the spectral distribution of the dominating scene illuminant. We do this by developing a statistical model for the spatial distribution of colors in white balanced images (i.e., those that have no color cast), and then using this model to infer illumination parameters as those being most likely under our model. The key observation is that by applying spatial band-pass filters to color images one unveils color distributions that are unimodal, symmetric, and well represented by a simple parametric form. Once these distributions are fit to training data, they enable efficient maximum likelihood estimation of the dominant illuminant in a new image, and they can be combined with statistical prior information about the illuminant in a very natural manner. Experimental evaluation on standard data sets suggests that the approach performs well.

Index Terms—Color constancy, statistical modeling, spatial correlations, maximum likelihood, illumination statistics.

✦

1 INTRODUCTION

IN addition to its intrinsic reflectance properties, the observed color of a material depends on the spectral and spatial distributions of its surrounding illumination. Thus, in order to use color as a reliable cue for recognition, we must somehow compensate for these extrinsic factors and infer a color descriptor that is stable despite changes in lighting. The ability to make this inference—termed *color constancy*—is exhibited by the human visual system to a certain degree, and there are clear benefits to building it into machines.

One important part of computational color constancy, and the part we consider in this paper, is compensating for the *color cast* that affects a scene as a whole. For this, one ignores spatial variations in lighting spectra and makes the assumption that the spectrum of the illuminant is approximately uniform throughout the scene. The problem is then one of inferring the map $\boldsymbol{M}: \mathbb{R}^3 \to \mathbb{R}^3$ from the space of observed tristimulus vectors (RGB, human cone, etc.) to the space of *canonical colors*—those that would have been obtained for the same scene by a standard observer under a standard illuminant spectral power distribution.

Like most vision tasks, this version of the color constancy problem is ill-posed, so solving it requires a prior model for the set of canonical colors that exist in a scene. Once equipped with such a model, and optionally a prior model for illuminants as well, we can solve for the color cast (the map $\boldsymbol{M}$) that best explains the image under this model.

Traditional approaches to this problem use models based on statistics of per-pixel colors, and the principal challenge they face is how to define useful statistics in a tractable manner. The set of per-pixel colors varies significantly from scene to scene, so methods based on simple first-order statistics like the mean color (*gray world* [1]) or the color of the brightest pixel (*white patch* [2]) can fail quite dramatically. Improving reliability requires more sophisticated models for the set of per-pixel colors in a scene, such as the shape of their convex hull in color space [3] or a nonparametric representation of the empirical probability distribution [4], [5], [6]. These methods improve performance, but at a significant computational cost.

This paper introduces a model that goes beyond statistics of per-pixel colors and leverages joint spatial-color structure. Motivated by the success of filter-based methods [7], [8], we begin by decomposing an input color image into distinct spatial subbands, and then we model the color statistics separately in each subband. We show that unlike per-pixel colors, the empirical probability distributions of the colors in each subband can be well represented using simple parametric (radial exponential) forms, and that these allow inferring the color cast of an image quite effectively. The resulting method outperforms existing approaches on standard databases, operates very efficiently, and allows incorporating a prior model for illuminants in a statistical framework.

1.1 Problem Formulation

Assuming a Lambertian model, we denote the effective spectral reflectance of a surface patch observed at pixel $\boldsymbol{n} \in \mathbb{Z}^2$ by $\kappa(\lambda, \boldsymbol{n})$, where $\lambda \in \mathbb{R}$ denotes wavelength. Here, κ accounts for both the material reflectance and surface orientation with respect to the illuminant. We assume a single dominant scene illuminant and represent its spectral distribution by $\ell(\lambda)$. As depicted in Fig. 1, the spectral distribution of the radiance that is emitted toward the observer is then the product of ℓ and κ.

We assume that we are given three spectral measurements at each pixel $\boldsymbol{n}$:[1]

- *A. Chakrabarti and T. Zickler are with the Harvard School of Engineering and Applied Sciences, Maxwell-Dworkin, 33 Oxford St., Cambridge, MA 02138. E-mail: ayanc@eecs.harvard.edu, zickler@seas.harvard.edu.*
- *K. Hirakawa is with the Intelligent Signal Systems Laboratory, Electrical and Computer Engineering, University of Dayton, 300 College Park, KL241C, Dayton, OH 45469. E-mail: khirakawa1@udayton.edu.*

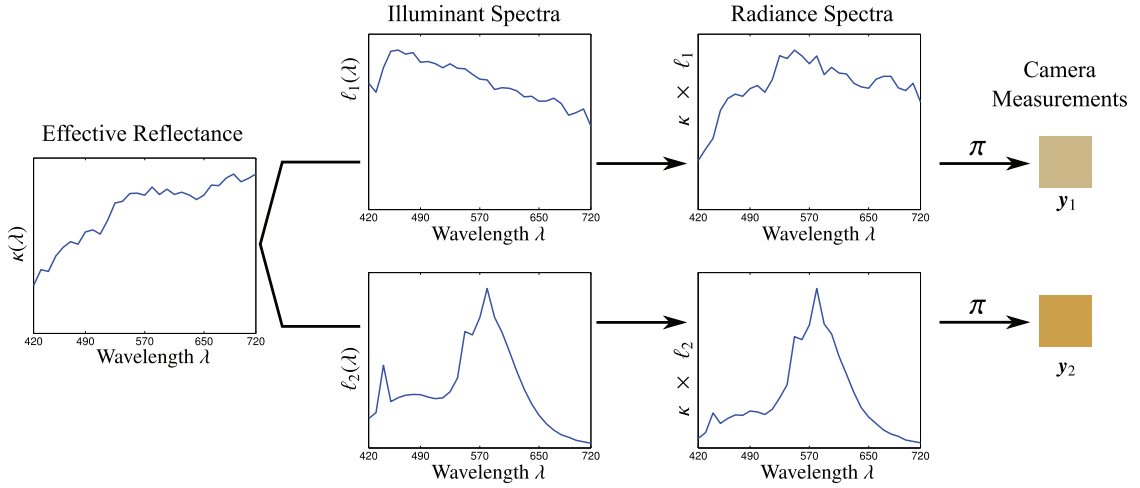1. For typical digital cameras, this assumes idealized demosaicking.

Fig. 1. Camera measurements under different illuminants. The effective surface reflectance (κ) gets shaped by the spectrum (ℓ) of the scene illuminant. Different illuminants can lead to significantly different radiance spectra ($\kappa \times \ell$) being incident on the camera, in turn leading to different trichromatic measurements by the camera's sensors (π).

$$\boldsymbol{y}(\boldsymbol{n}) = \begin{bmatrix} y^{\{1\}}(\boldsymbol{n}) \\ y^{\{2\}}(\boldsymbol{n}) \\ y^{\{3\}}(\boldsymbol{n}) \end{bmatrix} := \int \boldsymbol{\pi}(\lambda)\kappa(\lambda, \boldsymbol{n})\ell(\lambda)\, d\lambda, \qquad (1)$$

where $\boldsymbol{\pi}(\lambda) = [\pi^{\{1\}}(\lambda), \pi^{\{2\}}(\lambda), \pi^{\{3\}}(\lambda)]^T$ are the spectral transmittance distributions of the sensor's color filters. From image $\boldsymbol{y}(\boldsymbol{n})$ our task is to estimate an illuminant-invariant representation, or "canonical color" image, $\boldsymbol{x}(\boldsymbol{n})$ given by

$$\boldsymbol{x}(\boldsymbol{n}) := \int \boldsymbol{\pi}(\lambda)\ell_0(\lambda)\kappa(\lambda, \boldsymbol{n})d\lambda. \qquad (2)$$

We refer to ℓ_0 as our "canonical illuminant," and for the specific choice $\ell_0(\lambda) = 1$, each canonical color $\boldsymbol{x}(\boldsymbol{n})$ can be interpreted as the trichromatic projection (via $\boldsymbol{\pi}$) of the effective reflectance $\kappa(\lambda, \boldsymbol{n})$ at the corresponding surface point.

The colors $\boldsymbol{y}$ and $\boldsymbol{x}$ in (1) and (2) are trichromatic reductions of the full spectral distribution that cannot be exactly reversed, even when illuminants ℓ and ℓ_0 are known. However, prior work [9], [10], [11] suggests that a linear mapping of the form

$$\boldsymbol{x}(\boldsymbol{n}) = \boldsymbol{M}^{-1}\boldsymbol{y}(\boldsymbol{n}), \qquad \boldsymbol{M} \in \mathbb{R}^{3\times 3}, \qquad (3)$$

can achieve accurate chromatic adaptation in many cases, and that it can often be approximated further as a diagonal transformation (i.e., $M_{ij} = 0$, if $i \neq j$). In this paper, we apply such diagonal mappings directly in the color space defined by $\boldsymbol{\pi}$, even though applying them in an optimized (often called "sharpened") color space can improve performance [9], [10], [11]. For this reason, our results may provide a conservative estimate of what can be achieved by our algorithm.

With some abuse of notation, we let $\boldsymbol{M}$ represent a diagonal matrix for the rest of this paper; and with some abuse of terminology, we refer to its recovery as *estimating the illuminant*.

1.2 Related Work

Having settled on a linear diagonal form for the mapping from input color $\boldsymbol{y}$ to canonical color $\boldsymbol{x}$, the color constancy problem reduces to estimating the three diagonal entries of

$\boldsymbol{M}$. Since both the illuminant and scene reflectances are unknown, this is an underdetermined problem, and prior work in this area typically addresses this by introducing a model or statistical prior for surface reflectance κ or canonical color $\boldsymbol{x}$. *White Patch* [2] and *Gray World* [1] are two well-known methods with intuitive interpretations. Based on the observation that color-neutral surfaces (i.e., with constant spectral reflectance) are the most efficient reflectors, the White Patch algorithm posits that pixels observed to have the greatest intensity correspond to a color-neutral surface patch. Similarly, based on the model that the mean surface reflectance in a scene is color neutral, the Gray World method assumes that the sample mean of $\{\boldsymbol{x}(\boldsymbol{n})\}_{\boldsymbol{n}}$ is achromatic or *gray*.[2] The diagonal elements of $\boldsymbol{M}$ are then estimated by mapping $\max_{\boldsymbol{n}}(\boldsymbol{y}(\boldsymbol{n}))$ or $\mathrm{avg}_{\boldsymbol{n}}(\boldsymbol{y}(\boldsymbol{n}))$ to white.

An alternative to these approaches is to represent the set of canonical colors expected in a natural scene by their convex hull, or *gamut* [3]. In a new input image $\boldsymbol{y}(\boldsymbol{n})$, the corresponding corrected colors $\{\boldsymbol{M}^{-1}\boldsymbol{y}(\boldsymbol{n})\}_{\boldsymbol{n}}$ are expected to lie within this gamut, and this induces a set of constraints on the unknown illuminant $\boldsymbol{M}$. In many cases, multiple choices of $\boldsymbol{M}$ may satisfy these constraints for a given image, and different approaches [3], [12] have been proposed to arrive at a unique solution from this feasible set.

Another alternative is to define a probability distribution over the space of observed or canonical colors in a scene. The *Color by Correlation* [13] algorithm begins by discretizing the chromaticity space into a number of bins, and then assigns a probabilities to each bin being observed under every candidate illuminant. The scene illuminant $\boldsymbol{M}$ is then chosen to maximize the likelihood of all observed colors in the input image $\boldsymbol{y}$. Statistical approaches also permit the inclusion of a prior distribution over illuminant spectra. Statistical models for illuminants are combined with those for canonical pixel colors, based on Gaussian [4] or, more effectively, nonparametric [5], [6] distributions, to yield Bayesian estimates of the illuminant $\boldsymbol{M}$.

---

2. In fact, Buchsbaum [1] did not require the average color to be gray—only that it be some known system parameter.

Related to these are a number of leaning-based approaches that effectively use a training database of color images with ground truth illuminant data to learn a function f that maps an observed image $\boldsymbol{y}$ to an estimate of the corresponding illuminant $\boldsymbol{M}$. For example, Shi et al. [14] learn this function f using thin-plate spline interpolation from colors in a training database. Other approaches include learning f in terms of linear filters [15], neural networks [16], and using support-vector regression [17].

All of the above methods can be considered pixel-based because they model the set of individual pixel colors without considering each pixel's spatial context. One can arbitrarily reorder the pixels of any input image, for example, without affecting the resulting estimate of illuminant $\boldsymbol{M}$. Because they ignore spatial structure, these methods can be negatively influenced by scene content that skews the color histogram, such as the presence of large colorful objects.

A variety of work suggests that improvements might be gained by employing spatial image features, such as segmentations or linear filter responses, that incorporate spatial information in a tractable manner. For example, Gershon et al. [18] assume that the average of mean colors of segmented regions of an image, rather than of individual pixels, is color neutral. Alternatively, van de Weijer et al. [7] and Gijsenij et al. [8], respectively, apply the Gray World and gamut constraint procedures described above to the outputs of linear filters instead of the individual pixels they were originally designed for. (The *Gray Edge* method [7] posits that image gradients are on average color neutral, while the generalized gamut mapping algorithm [8] proposes strategies for combining cues from the expected gamuts of various linear filter coefficients.) Finally, Singh et al. [19] reason about the illuminant using a linear spatio-spectral basis for small spatial patches of color images.

The goal of this paper is to leverage joint spatial and spectral structure in a more efficient and effective manner. We achieve this by representing the prior probability distribution over canonical color images in terms of the color coefficients in distinct spatial subbands. We show empirically that, unlike per-pixel colors, the distributions of these subband coefficients can be well represented using convenient parametric forms, and that these parametric forms can be used in a maximum-likelihood framework to estimate the illuminant $\boldsymbol{M}$ quite efficiently and accurately. The paper builds on an earlier version of this work [20] which used Gaussian distributions and tested them on nonlinearly transformed (i.e., gamma-corrected) color images. Here, we work directly on linear data as defined in (1) and (2) and show that distributions with heavier tails are more appropriate.

2 SPATIO-SPECTRAL MODELING

As stated above, we move beyond statistics of individual pixels colors, drawing inspiration from previous studies [7], [8], [19], and seek to exploit information about a pixel's spatial context. We take an approach similar to the gray-edge [7] and generalized-gamut [8] methods by looking at the properties of filter coefficients instead of individual pixels. These methods represent important first steps in looking beyond individual pixels and demonstrate that filter coefficients provide more robust cues for color
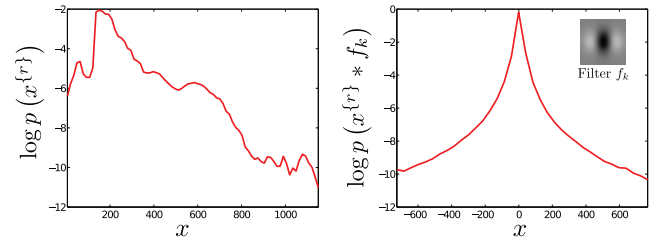

Fig. 2. Distributions of individual pixels versus filter coefficients. Shown are empirical log histograms (over a standard database [6] of 12-bit color images) of the red values of individual pixels (left) and coefficients (right) of a subband filter inset) for images under canonical illumination. The coefficient histogram shows more structure and can be modeled by parametric distributions.

constancy. However, they are based on the direct application of pixel-based color constancy techniques, gray world [1], and gamut mapping [3], respectively, to filter coefficients, so there is reason to believe that a new estimation procedure that is specifically tailored to the statistical behavior of these coefficients, can improve performance.

We begin with the observation that statistics of filter coefficients show far more structure than those of individual pixels. Fig. 2 compares the empirical histograms of red values of individual pixels to those of coefficients for a particular band-pass filter in images under a canonical illuminant. We note that the histogram for the filter coefficients is unimodal and symmetric, while individual pixels show less discernible statistical structure. It is because of this lack of structure that pixel-based color constancy methods have been constrained to either use properties of simple ensemble statistics like the mean (e.g., gray world [1]), employ expensive nonparametric distributions (e.g., Rosenberg [5], [6]), or model only the support, or convex hull, of pixel colors (e.g., gamut mapping [3]).

In contrast, Fig. 2 suggests that filter coefficients have distributions that can be described accurately by parametric models. In this section, we propose such a model and describe a method to learn its parameters from training data. We then present an algorithm that uses this model to estimate the unknown illuminant from an observed image.

2.1 Image Model

We begin by defining a statistical model for an image $\boldsymbol{x}(\boldsymbol{n})$ observed under canonical illumination. Since pixels in a spatial neighborhood are strongly correlated, we first apply a spatially decorrelating transform by filtering the image with a series of spatial subband filters $\{f_k(\boldsymbol{n})\}_k$. While these filters can be chosen to correspond to any standard image decompositions, such as wavelet transforms, steerable pyramids, etc., Gaussian derivative filters have proven successful in edge-based color constancy methods [7], [8]. Therefore, we use a decomposition based on horizontal and vertical second-derivative Gaussian filters at multiple scales, as illustrated in Fig. 3. Each filter is applied separately to all channels of the image and we define the trichromatic color coefficient vector $\boldsymbol{x}_k(\boldsymbol{n})$ as

$$\boldsymbol{x}_k(\boldsymbol{n}) = [(x^{\{1\}} * f_k)(\boldsymbol{n}), (x^{\{2\}} * f_k)(\boldsymbol{n}), (x^{\{3\}} * f_k)(\boldsymbol{n})]^T, \quad (4)$$

where $*$ denotes convolution. We assume that following the decomposition by the filter bank $\{f_k\}$, each subband
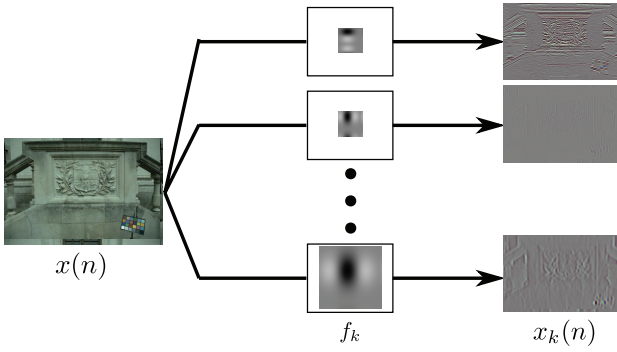
Fig. 3. Subband Decomposition. We use horizontal and vertical second-derivative Gaussian filters at multiple scales to spatially decorrelate the image, and model corresponding coefficients independently.

canonical color image $\boldsymbol{x}_k(\boldsymbol{n})$ can be modeled as being independent from the rest.

We represent the prior probability of each subband canonical color image as the product of independent, identical distributions per trichromatic coefficient vector, $\prod_{\boldsymbol{n}} p(\boldsymbol{x}_k(\boldsymbol{n}))$, where the distribution $p(\boldsymbol{x}_k)$ encodes the joint statistics of the three per-channel coefficients in $\boldsymbol{x}_k$. Specifically, we use the *radial exponential* distribution [21]

$$p(\boldsymbol{x}_k) = \frac{1}{\pi\sqrt{\det(\boldsymbol{\Sigma}_k)}} \exp\left(-2\sqrt{\boldsymbol{x}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{x}_k}\right), \quad (5)$$

where $\boldsymbol{\Sigma}_k$ is a 3×3 positive-definite matrix corresponding to the covariance of $\boldsymbol{x}_k$.

Note that the radial exponential distribution used in (5) is more kurtotic, or heavy-tailed, than a multivariate Gaussian, and can be thought of as a generalization of the Laplace distribution to the multivariate case. It differs, however, from the standard multivariate Laplace distribution [22] in that the multivariate Laplace has equiprobable contours that are L_1-spherical or *diamond* shaped (see Fig. 4 (left)) while those of the radial exponential distribution are ellipsoidal (see Fig. 4 (center)). This implies that the components of $\boldsymbol{x}_k$ along the eigen-vectors of $\boldsymbol{\Sigma}_k$ are uncorrelated but *not* independent under (5). Fig. 4 (right) shows the equiprobable contours of the actual empirical distribution of $\boldsymbol{x}_k$ computed using kernel-density estimation, and we see that these contours are indeed ellipsoidal.

2.2 Learning Model Parameters

Having defined a parametric statistical model for canonical color images, $p(\boldsymbol{x}) = \prod_k \prod_{\boldsymbol{n}} p(\boldsymbol{x}_k(\boldsymbol{n}))$, we are in need of two algorithms. First, we require the ability to fit the model parameters $\boldsymbol{\Sigma}_k$ given a training set of T canonical color images $\{\boldsymbol{x}_t\}_{t=1}^T$ (indexed by t), and second, we require the ability to use this learned prior model to reason about the illuminant $\boldsymbol{M}$ for a given input color image $\boldsymbol{y}(\boldsymbol{n})$. We deal with training in this section, and in the next we address illuminant estimation.

We treat each subband separately during training, so the problem amounts to independently inferring each $\boldsymbol{\Sigma}_k$ given the corresponding subband of the training set, $\{\boldsymbol{x}_{k,t}\}_{t=1}^T$. Although, $\boldsymbol{\Sigma}_k$ corresponds to the covariance matrix of $\boldsymbol{x}_k$, the maximum likelihood (ML) estimate of $\boldsymbol{\Sigma}_k$ is *not* the empirical covariance of $\{\boldsymbol{x}_{k,t}\}$. The log likelihood of the training set as a function of $\boldsymbol{\Sigma}_k$ is
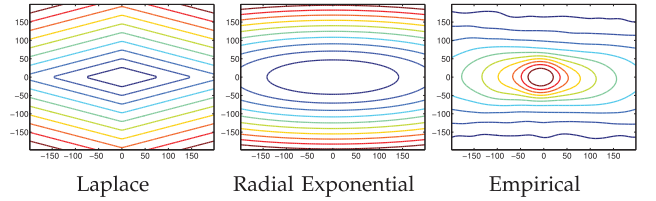

Fig. 4. Equiprobable contours of the distribution of filter coefficient vectors $\boldsymbol{x}_k(\boldsymbol{n})$, corresponding to the filter shown in Fig. 2. Contours are shown along the two major eigen-vectors of the covariance matrix $\boldsymbol{\Sigma}_k$, for a multivariate Laplace distribution (left), A radial exponential distribution (center), and the empirical distribution computed using kernel density estimation (right). The empirical distribution has elliptical contours, making the radial exponential distribution a more appropriate choice for modeling these coefficients.

$$\begin{aligned} J_k(\boldsymbol{\Sigma}) &= \sum_t \log p(\boldsymbol{x}_{k,t}|\boldsymbol{\Sigma}) \\ &= -\frac{T}{2}\log\det(\boldsymbol{\Sigma}) - \sum_t 2\sqrt{\boldsymbol{x}_{k,t}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_{k,t}}. \end{aligned} \quad (6)$$

Unfortunately, the ML estimate of $\boldsymbol{\Sigma}_k$, which is given by $\boldsymbol{\Sigma}_k = \arg\max J_k(\boldsymbol{\Sigma})$, does not have a closed-form solution. Therefore, we propose an iterative estimation procedure.

We initialize $\boldsymbol{\Sigma}_k$ as the identity matrix. Based on the current estimate $\boldsymbol{\Sigma}_k^*$ of $\boldsymbol{\Sigma}_k$ at any iteration, we define an approximation to $J_k(\cdot)$ as

$$J_k^*(\boldsymbol{\Sigma}|\boldsymbol{\Sigma}_k^*) = -\frac{T}{2}\log\det(\boldsymbol{\Sigma}) - \sum_t 2\frac{\boldsymbol{x}_{k,t}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_{k,t}}{\sqrt{\boldsymbol{x}_{k,t}^T \boldsymbol{\Sigma}_k^{*-1} \boldsymbol{x}_{k,t}}}. \quad (7)$$

We then update $\boldsymbol{\Sigma}_k$ in the following iteration as:

$$\boldsymbol{\Sigma}_k = \arg\max_{\boldsymbol{\Sigma}} J_k^*(\boldsymbol{\Sigma}|\boldsymbol{\Sigma}_k^*) = \frac{4}{T}\sum_t \frac{\boldsymbol{x}_{k,t}\boldsymbol{x}_{k,t}^T}{\sqrt{\boldsymbol{x}_{k,t}^T \boldsymbol{\Sigma}_k^{*-1} \boldsymbol{x}_{k,t}}}. \quad (8)$$

The true ML estimate is a fixed point of these iterations, and we find that the procedure converges quickly in practice (usually in less than five iterations).

2.3 Illuminant Estimation

Once we have learned the parameters of the prior model, we are ready to infer the illuminant $\boldsymbol{M}$ for an input color image $\boldsymbol{y}(\boldsymbol{n})$. As per our model in (3), we assume that $\boldsymbol{y}(\boldsymbol{n}) = \boldsymbol{M}\boldsymbol{x}(\boldsymbol{n})$, where $\boldsymbol{M}$ is a diagonal 3×3 matrix with positive entries. Since convolution is a linear operation, it follows that $\boldsymbol{y}_k(\boldsymbol{n}) = \boldsymbol{M}\boldsymbol{x}_k(\boldsymbol{n})$, where $\boldsymbol{y}_k(\boldsymbol{n})$ are the color subband coefficients of $\boldsymbol{y}(\boldsymbol{n})$ for filter f_k as in (4).

We begin by noting that under our prior model for $\boldsymbol{x}_k$, the likelihood of $\boldsymbol{y}_k$ given $\boldsymbol{M}$ is given by

$$\begin{aligned} p(\boldsymbol{y}_k|\boldsymbol{M}) &= \frac{\exp\left(-2\sqrt{\boldsymbol{y}_k^T(\boldsymbol{M}\boldsymbol{\Sigma}_k\boldsymbol{M})^{-1}\boldsymbol{y}_k}\right)}{\pi\sqrt{\det(\boldsymbol{M}\boldsymbol{\Sigma}_k\boldsymbol{M})}} \\ &\propto \frac{1}{\det(\boldsymbol{M})}\exp\left(-2\sqrt{\boldsymbol{y}_k^T(\boldsymbol{M}\boldsymbol{\Sigma}_k\boldsymbol{M})^{-1}\boldsymbol{y}_k}\right). \end{aligned} \quad (9)$$

Therefore, conditioned on $\boldsymbol{M}$, the coefficients $\boldsymbol{y}_k(\boldsymbol{n})$ also have a radial exponential distribution, with the covariance matrix $\boldsymbol{M}\boldsymbol{\Sigma}_k\boldsymbol{M}$. Fig. 5 compares the covariance matrices $\boldsymbol{\Sigma}_k$ and $\boldsymbol{M}\boldsymbol{\Sigma}_k\boldsymbol{M}$ for a typical illuminant $\boldsymbol{M}$. This difference between the kth subband coefficient distributions for the input image $\boldsymbol{y}_k$ and the canonical color image $\boldsymbol{x}_k$, as
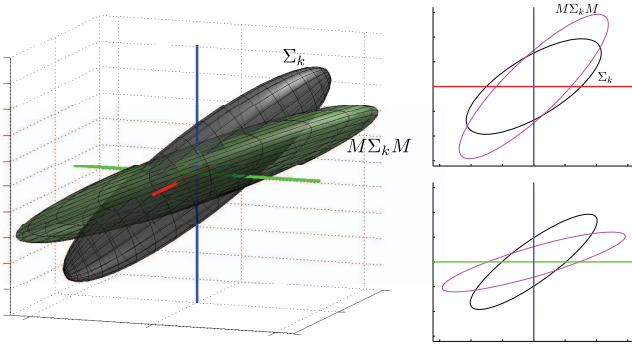
Fig. 5. Covariance matrices Σ_k and $M\Sigma_k M$, for coefficients observed under a canonical illumination and a typical illuminant M. (Left) Ellipsoids corresponding to the covariance matrices, and (right) their projections onto the R-B and G-B planes. The illuminant causes a skew in the shape of the covariance matrix, which serves as a cue for our estimation method.

embodied by these covariance matrices, is the fundamental cue that we will exploit for illuminant estimation.

For notational convenience, we define $\boldsymbol{m} = [m_1, m_2, m_3]$ and $\boldsymbol{w} = [w_1, w_2, w_3]$ to be the diagonal elements of $\boldsymbol{M}$ and $\boldsymbol{M}^{-1}$, respectively, where $m_i = w_i^{-1}$. Then, the ML estimate of $\boldsymbol{M}$ given all observed coefficients $\{\boldsymbol{y}_k(\boldsymbol{n})\}_{k,\boldsymbol{n}}$ is given by

$$
\begin{aligned}
\hat{\boldsymbol{M}}_{ML} &= \arg \max_{\boldsymbol{M} \in \text{diag}(\mathbb{R}^3)} \sum_{k,\boldsymbol{n}} \log p(\boldsymbol{y}_k(\boldsymbol{n})|\boldsymbol{M}) \\
&= \arg \min \sum_{k,\boldsymbol{n}} \log \det(\boldsymbol{M}) + 2\sqrt{\boldsymbol{y}_k^T(\boldsymbol{n})(\boldsymbol{M}\Sigma_k \boldsymbol{M})^{-1}\boldsymbol{y}_k(\boldsymbol{n})} \\
&= \arg \min N \log m_1 m_2 m_3 \\
&\quad + \sum_{k,\boldsymbol{n}} 2\sqrt{\boldsymbol{w}^T\left[(\boldsymbol{y}_k(\boldsymbol{n})\boldsymbol{y}_k^T(\boldsymbol{n})) \circ \Sigma_k^{-1}\right]\boldsymbol{w}},
\end{aligned}
\tag{10}
$$

where $N = \sum_{k,\boldsymbol{n}} 1$ is the total number of coefficients, and $\boldsymbol{A} \circ \boldsymbol{B}$ denotes the Hadamard or entry-wise product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

We propose an iterative algorithm to solve (10) using a similar approach as the training method. Based on the estimate $\boldsymbol{M}^*$ of the illuminant at every iteration, we approximate the cost function in (10) as

$$
\begin{aligned}
L(\boldsymbol{M}|\boldsymbol{M}^*) &= N \log m_1 m_2 m_3 \\
&\quad + \sum_{k,\boldsymbol{n}} 2\frac{\boldsymbol{w}^T[(\boldsymbol{y}_k(\boldsymbol{n})\boldsymbol{y}_k^T(\boldsymbol{n})) \circ \Sigma_k^{-1}]\boldsymbol{w}}{\sqrt{\boldsymbol{y}_k^T(\boldsymbol{n})(\boldsymbol{M}^*\Sigma_k \boldsymbol{M}^*)^{-1}\boldsymbol{y}_k^T(\boldsymbol{n})}} \\
&= N\left[\log m_1 m_2 m_3 + \frac{1}{2}\boldsymbol{w}^T \boldsymbol{A}^* \boldsymbol{w}\right],
\end{aligned}
\tag{11}
$$

where $\boldsymbol{A}^*$ is a 3×3 symmetric matrix given by

$$
\boldsymbol{A}^* = \frac{4}{N}\sum_k \left(\sum_{\boldsymbol{n}} \frac{\boldsymbol{y}_k(\boldsymbol{n})\boldsymbol{y}_k^T(\boldsymbol{n})}{\sqrt{\boldsymbol{y}_k^T(\boldsymbol{n})(\boldsymbol{M}^*\Sigma_k \boldsymbol{M}^*)^{-1}\boldsymbol{y}_k^T(\boldsymbol{n})}}\right) \circ \Sigma_k^{-1}. \tag{12}
$$

In the next iteration, the illuminant estimate $\boldsymbol{M}$ is set to $\arg \min L(\boldsymbol{M}|\boldsymbol{M}^*)$. There is a closed-form solution to this expression for each element of m_i of $\boldsymbol{m}$ when the others elements $m_j, j \neq i$, are fixed. This is given by

$$
\frac{\partial}{\partial m_i}L(\boldsymbol{M}|\boldsymbol{M}^*) = 0 \Rightarrow m_i^2 - \left(\sum_{j \neq i}\frac{A_{ji}^*}{m_j}\right)m_i - A_{ii}^* = 0
$$

$$
\Rightarrow m_i = \frac{1}{2}\left(\sum_{j \neq i}\frac{A_{ji}^*}{m_j} + \sqrt{\left(\sum_{j \neq i}\frac{A_{ji}^*}{m_j}\right)^2 + 4A_{ii}^*}\right). \tag{13}
$$

We therefore apply (13) on each channel i, keeping the others fixed, and iterate until convergence.

Once we have estimated $\hat{\boldsymbol{M}}_{ML}$ through this procedure, the canonical image $\boldsymbol{x}(\boldsymbol{n})$ can be computed simply as $\boldsymbol{x}(\boldsymbol{n}) = \hat{\boldsymbol{M}}_{ML}^{-1}\boldsymbol{y}(\boldsymbol{n})$.

3 ILLUMINANT PRIOR

Implicit in the maximal likelihood estimation of (10) is the notion that all illuminants are equally likely. Drawing from the success of existing color constancy algorithms [4], [6] that leverage the a priori knowledge of the illuminant distribution, however, we develop a strategy to incorporate an illuminant prior. The experimental validation in Section 4 will show that such a *regularized* estimate can improve robustness.

Consider an illuminant prior distribution $p(\boldsymbol{M})$ of the form

$$
p(\boldsymbol{M}) = \left(\frac{2}{\pi}\right)^{3/2}\frac{\det(\boldsymbol{Q})^{-1/2}}{m_1^2 m_2^2 m_3^2}\exp\left(-\frac{1}{2}\boldsymbol{w}^T \boldsymbol{Q}^{-1}\boldsymbol{w}\right), \tag{14}
$$

where $\boldsymbol{Q}$ is a 3×3 positive-definite matrix and $\boldsymbol{m}, \boldsymbol{w} > 0$ are as defined before. It is useful to note that this multivariate Gaussian distribution on $\boldsymbol{w}$ (restricted to positive values) is a conjugate prior to the likelihood expression in (10). We note that (14) may be used to describe the statistics of scene illuminants in general, or those restricted to a category of illuminants (such as indoor versus outdoor).

Ideally, the covariance parameter $\boldsymbol{Q}$ would be computed as a sample covariance of a training set $\{\boldsymbol{w}_t\}$ of illuminant coefficients. However, only the relative spectral radiance of illuminants is usually available, meaning the absolute luminance of the illuminants in the training set is unknown. Therefore, we compute the sample covariance $\boldsymbol{Q}'$ over the set of illuminant vectors normalized such that $\|\boldsymbol{m}_t\| = 1$ during training. During estimation, we first compute the ML estimate $\hat{\boldsymbol{M}}_{ML}$ of (10) (that is, without the illuminant prior), and then reweight the $\boldsymbol{Q}'$ covariance parameter as

$$
\boldsymbol{Q} = \|\boldsymbol{m}_{ML}\|^2 \boldsymbol{Q}'. \tag{15}
$$

This reweighting has the effect of normalizing the training set $\|\boldsymbol{m}_t\|$ to the initial estimate $\|\boldsymbol{m}_{ML}\|$.

Let $\boldsymbol{y}_k$ be the kth subband coefficients of input image as before. The final illuminant estimate incorporating $p(\boldsymbol{M})$ is obtained by solving

$$
\hat{\boldsymbol{M}} = \arg \max\left[\sum_{k,\boldsymbol{n}}\log p(\boldsymbol{y}_k(\boldsymbol{n})|\boldsymbol{M})\right] + \alpha \log p(\boldsymbol{M}). \tag{16}
$$

Here, α is a scalar parameter that weights the relative contribution of the prior to the image evidence, and is learned using cross validation to minimize estimation error on a training set. While, for the special case of $\alpha = 1$, (16) corresponds to maximum a posteriori (MAP) estimation, we

TABLE 1
Angular Error Quantiles for Various Methods on the "Color Checker" Database [6]

	All Images (568)			Indoor (246)			Outdoor (322)		
	Mean	Median	Worst-25%	Mean	Median	Worst-25%	Mean	Median	Worst-25%
Grey World [1]	4.6°	4.1°	8.7°	5.8°	5.5°	9.9°	3.7°	3.1°	7.1°
Grey Edge (1st Order) [7]	4.1°	3.5°	8.0°	4.8°	4.1°	8.8°	3.5°	2.9°	7.1°
Grey Edge (2nd Order) [7]	4.0°	3.4°	7.8°	4.7°	4.0°	8.5°	3.5°	2.8°	7.1°
Gamut Mapping ($\sigma = 5$) [3]	4.1°	2.5°	10.3°	5.5°	4.3°	12.3°	3.1°	1.8°	7.9°
Generalized Gamut ($\sigma = 5,1$-jet) [8]	4.1°	2.5°	10.3°	5.5°	4.4°	12.4°	3.1°	1.8°	7.9°
Rosenberg [6]	4.8°	3.5°	10.5°	6.5°	5.9°	12.0°	3.5°	2.4°	7.9°
Proposed (ML Estimate)	**3.7°**	**3.0°**	**7.6°**	**4.2°**	**3.6°**	**8.1°**	**3.3°**	**2.5°**	**7.1°**
Proposed (with general Prior)	**3.6°**	**3.0°**	**7.4°**	**4.2°**	**3.6°**	**7.9°**	**3.2°**	**2.4°**	**6.9°**
Proposed (with category-wise Prior)	**3.1°**	**2.3°**	**6.5°**	**4.1°**	**3.7°**	**7.8°**	**2.3°**	**1.9°**	**4.6°**

find that a larger weight on the prior yields better estimates in practice.

Solving (16) in a manner similar to (13), we define an approximated cost as

$$L_p(\boldsymbol{M}|\boldsymbol{M}^*) = (N + 2\alpha)\left[\log m_1 m_2 m_3 + \frac{1}{2}\boldsymbol{w}^T \boldsymbol{A}_p^* \boldsymbol{w}\right], \quad (17)$$

where $\boldsymbol{A}_p^*$ is a 3×3 matrix given by

$$\boldsymbol{A}_p^* = \frac{N\boldsymbol{A}^* + \alpha \boldsymbol{Q}}{N + 2\alpha}. \quad (18)$$

Applying (13) iteratively to $\boldsymbol{A}_p^*$ (instead of $\boldsymbol{A}^*$) solves $\boldsymbol{M} = \arg\min L_p(\boldsymbol{M}|\boldsymbol{M}^*)$. If $\boldsymbol{M}^*$ is set to $\boldsymbol{M}_{ML}$, we find that further iterations to update L_p or $\boldsymbol{Q}$ are unnecessary.

4 EXPERIMENTAL EVALUATION

We primarily evaluate the proposed method on the "Color Checker" database collected by the authors of [6], which includes 568 color images, of which 246 were labeled as captured indoors and 322 as captured outdoors. Each image has a color checker chart at manually marked coordinates which serves as ground truth and is masked out during evaluation. We use the version of the database made available by Shi and Funt [23]. It consists of 12-bit linear images in the sensor color space generated directly from the RAW captured data—without using the camera's autowhite balance, gamma-correction, or any demosaicking (the sensor data are subsampled to provide a trichromatic vector at each pixel). While we focus most of our discussion on the performance of various algorithms on this database, we also include results for two other common databases in Section 4.3.

A recent survey paper [24] provides a comprehensive evaluation of a large number of color constancy methods on this database (i.e., the reprocessed version from Shi and Funt [23]), allowing us to compare the proposed algorithm to these methods. Performance is measured in terms of the angular deviation between the estimated and true illuminant, where the latter is computed from the brightest gray patch in the color chart. We use the same error metric and ground truth to evaluate our method and compare performance to the state of the art: gray world [1], gray edge [7], gamut mapping [3], generalized gamut mapping [8], and Rosenberg [5], [6].

4.1 Implementation Details

We implement the proposed method with second-derivative Gaussian filters at three different scales ($\sigma = 1, 2$, and 4). We report performance both with and without the illumination prior. For the case with the prior, we evaluate a general prior over all illuminants, as well as separate priors for the indoor and outdoor images. To train the parameters of our model $\{\Sigma_k\}$, $\boldsymbol{Q}'$, and α, we use three-fold cross validation, where the database is split into three equal folds, and when testing on each fold, we use the remaining images for training. Matlab source code for this implementation and the estimated illuminants for all images in this database are available for download at http://vision.seas.harvard.edu/colorconstancy/.

Since the authors of [24] have made the estimated illuminants and errors for all evaluated methods available, we report performance for most other methods directly from their data. However, we incorporate an important enhancement to the evaluation of gray world and gray edge that improves their performance. While gray world is often interpreted as assuming the mean color of an image to be along $[1, 1, 1]$ (as is the case in [24]), we posit that the mean is along a 3×1 unit vector $\hat{\boldsymbol{g}}$. The illuminant estimate is then given by

$$\hat{\boldsymbol{m}} = [\text{diag}(\hat{\boldsymbol{g}})]^{-1}\text{avg}(\boldsymbol{y}(\boldsymbol{n})), \quad (19)$$

where $\text{avg}(\boldsymbol{y}(\boldsymbol{n}))$ is the trichromatic average of the observed image. This approach is sometimes referred to as *database gray world* [25], and we apply the same interpretation to gray edge, and learn the vector $\hat{\boldsymbol{g}}$ for each algorithm from a training set of images as the mean of the unit vectors corresponding to the pixel or edge averages of each image. We use three-fold cross validation for this training and to pick the optimal smoothing and norm parameters for gray edge as described in [24].

4.2 Results

Table 1 reports the mean and median errors for all algorithms, as well as the *Worst 25 percent* error which is a measure of robustness and refers to the mean of the 25 percent highest error values (note that these may correspond to different images for different methods). These quantiles are reported for the entire database, as well as separately for the indoor and outdoor sets. Figs. 6 and 7 show examples of indoor and
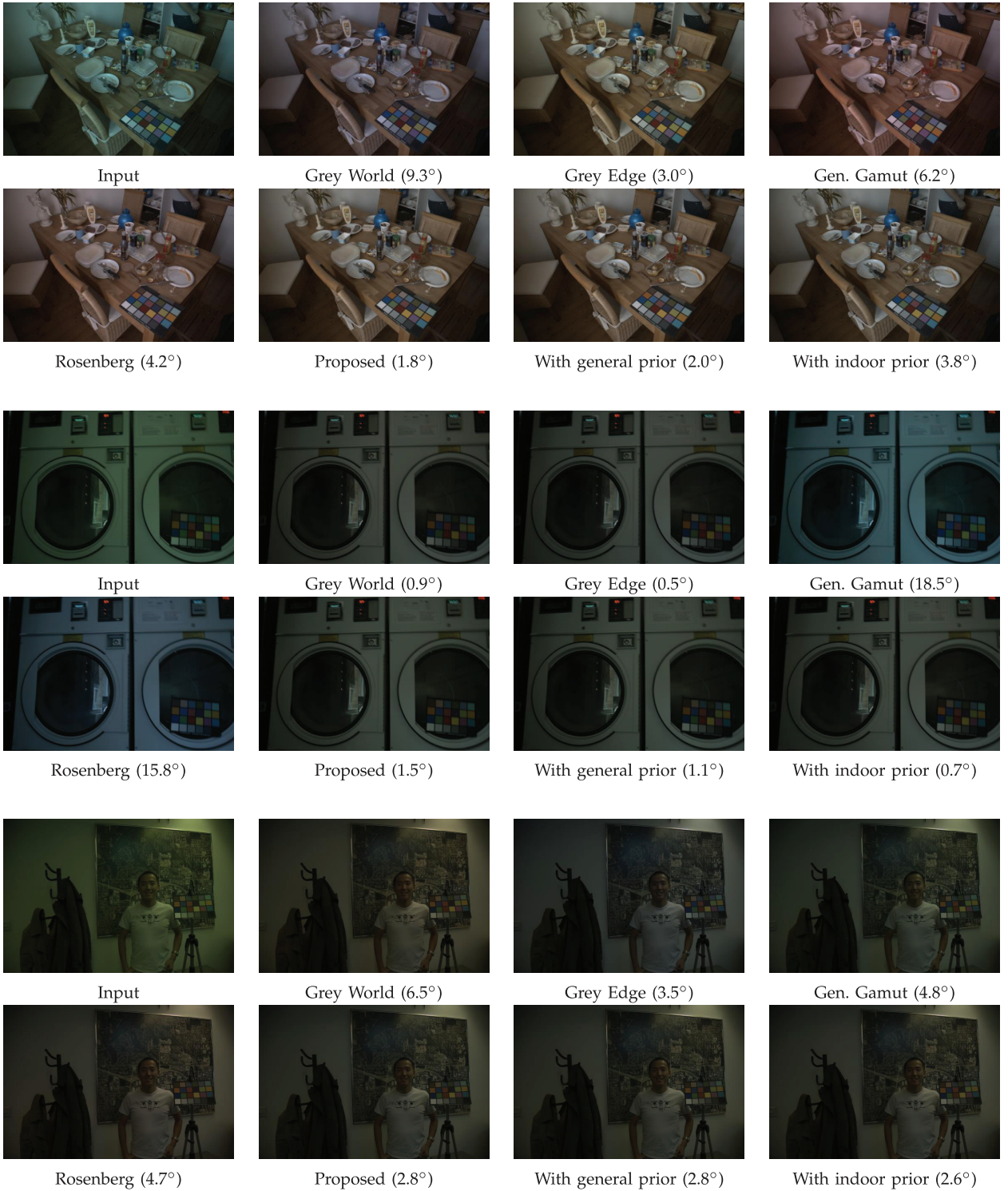
Fig. 6. Indoor images from the color checker database, corrected using different algorithms. Angular errors for estimated illuminant are indicated below each image.

outdoor images, respectively, corrected according to the illuminant estimates of various methods. Table 2 shows results of a *sign test* [26], [27] comparing the performance of every pair of algorithms over the entire database. This test determines whether one algorithm tends to have lower errors compared to another by using significance testing to reject the null hypothesis that $P(\mathrm{err}_i > \mathrm{err}_j) = P(\mathrm{err}_j > \mathrm{err}_i)$, where err_i and err_j denote angular errors from two algorithms i and

j on a random image. Note that the sign test is based on only the sign of the difference in errors for each image, not the magnitude.

We first look at the relative performance of the ML estimator, without any prior information. The sign test results indicate that it typically has lower errors in comparison to most of the state-of-the-art methods over the entire database. The only exception is traditional gamut-mapping, which has

Fig. 7. Outdoor images from the color checker database, corrected using different algorithms. Angular errors for estimated illuminant are indicated below each image.

higher errors than the proposed method on a majority of the images, but by a margin too small to be statistically significant. We also find that the ML estimates have lower mean and worst 25 percent error values among all state-of-the-art methods. Both gamut-based methods have lower median errors, largely due to their superior performance on the outdoor images that were captured in daylight. They have the lowest mean and median errors for those images,

although the proposed ML estimator appears to be more robust, with a lower worst 25 percent error. However, the gamut-based methods perform worse on indoor images and have the highest error quantiles on that set after Rosenberg. In general, all algorithms show poorer performance on the indoor set than on outdoor images, indicating that indoor scenes are more likely to contain regions that act as outliers. However, the ML estimator performs noticeably better than

TABLE 2
Sign Test Results on the "Color Checker" Database [6]

	Grey World — 1	Grey Edge (1st Order) — 2	Grey Edge (2nd Order) — 3	Gamut Map. (σ = 5) — 4	Gen. Gamut (σ = 5,1-jet) — 5	Rosenberg — 6	Proposed (ML) — 7	Proposed (General Prior) — 8	Proposed (categ. Prior) — 9
1		-1	-1	-1	-1	0	-1	-1	-1
2	1		-1	-1	-1	0	-1	-1	-1
3	1	1		-1	-1	1	-1	-1	-1
4	1	1	1		0	0	0	-1	-1
5	1	1	1	0		1	-1	-1	-1
6	0	0	-1	-1	-1		-1	-1	-1
7	1	1	1	0	1	1		-1	-1
8	1	1	1	1	1	1	1		-1
9	1	1	1	1	1	1	1	1	

Value at (row i, column j)	Interpretation at 95% confidence level
1	Algorithm i tends to have lower errors
-1	Algorithm j tends to have lower errors
0	No significant difference

TABLE 3
Performance of Gray Edge and the Proposed Method on the "Color Checker" Database [6], with Both Using the Same Set of Second-Derivative Gaussian Filters with $\sigma = 1$

	Mean	Median	Worst-25%
Grey Edge	4.0°	3.4°	7.8°
Proposed (ML Estimate)	3.9°	3.2°	7.7°
Proposed (with gen. prior)	3.7°	3.0°	7.5°

the state of the art, with the lowest values for all error quantiles on indoor images.

Next, we note that an illuminant prior, even a general one over all images, results in better performance. The sign test shows that the prior leads to a statistically significant improvement in comparison to the ML estimator, and errors are now likely to be lower than for all other methods, including gamut mapping. However, note that while there is an overall improvement in accuracy, estimation errors on individual images may increase as is the case for some examples in Figs. 6 and 7. In Table 1, we see that the most significant improvement is to the worst 25 percent quantile for all sets, indicating that the prior's major contribution is in making the estimation process more robust.

When the illuminant prior is defined separately over sets of indoor and outdoor images (and it is possible that these labels may be obtained during estimation from a light meter on the camera or a scene-classification algorithm), the improvement in overall performance is dramatic. We note that most of this improvement is on the outdoor images, which is likely because of relatively less variability in the set of daylight illuminants.

While the spatio-spectral model allows a natural way to combine cues across subbands at multiple scales, we seek to gauge how well the proposed estimator succeeds at exploiting the information in each subband. Therefore, we perform a more direct comparison of our method to the gray edge algorithm, by restricting the former to work on the same set of coefficients as the latter. Table 3 shows the error quantiles of the two methods (over all images) when both use second-derivative Gaussian filters at scale $\sigma = 1$, which corresponds to the best choice for gray edge. We find that the proposed method yields better estimates even when not using an illuminant prior, and appears to benefit from accurately modeling the statistical structure of subband coefficients.

4.3 Evaluation with Other Databases

Finally, we report performance on two older data sets considered in [24]. The first is a large set of real-world scenes captured by a consumer video camera [28], and the second is a relatively small set of objects captured under various illuminants in laboratory conditions [29]. We use the evaluation results from [24] as an indicator of the performance of state-of-the-art methods, but again, we differ in applying the learning step to gray world and gray edge, as described in Section 4.1.

The "Gray Ball" database [28] is a collection of 11,346 images captured using a video camera. A diffuse gray sphere is attached to the camera and is present at the same location in all images, providing a ground truth estimate of the illuminant. The original database includes gamma-corrected images, and we follow the procedure in [24] by measuring performance after applying inverse gamma-correction to the images and recomputing the ground truth accordingly. We also adopt the approach suggested in [24] for cross validation: Since the database is divided into 15 videos, we carry out estimation on the frames of each video using the remaining videos as training. Table 4 shows error quantiles of the proposed method and compares them to those of other algorithms. Note that we only report errors for the ML estimator since the prior-based estimates have identical error quantiles.

Like with the color checker database [6], we find that the proposed method yields more accurate estimates than other algorithms. However, results on this database should be interpreted with caution as the images are lower in quality, with possible artifacts from compression and the approximate inverse-gamma map. Further, the ground truth information may be unreliable in some cases because the gray ball is close to the camera and can be lit by an illuminant that is different from the dominant illuminant of the scene.

We also report results on the "SFU laboratory" database [29], which consists of 321 images of 31 different objects captured under different illuminants in a laboratory. For training, we take the approach suggested in [24] and fit model parameters to the images of all 31 objects captured under illuminant *syl-50MR16Q* before using these parameters to measure estimation accuracy on the entire

TABLE 4
Performance of Various Methods on the "Gray Ball" Set [28]

	Mean	Median	Worst 25%
Grey World [1]	12.2°	9.7°	25.4°
Grey Edge (1st Order) [7]	10.8°	9.2°	20.9°
Grey Edge (2nd Order) [7]	11.4°	9.9°	21.5°
Gamut Mapping [3]	11.8°	8.9°	24.9°
Generalized Gamut Mapping [8]	11.8°	8.9°	24.9°
Proposed (ML Estimate)	**10.3°**	**8.9°**	**20.3°**

TABLE 5
Performance of Various Methods on the "SFU Lab" Set [29]

	Mean	Median	Worst 25%
Grey World [1]	8.8°	5.6°	20.7°
Grey Edge (1st Order) [7]	6.6°	4.3°	15.6°
Grey Edge (2nd Order) [7]	5.1°	2.5°	13.6°
Gamut Mapping ($\sigma = 4$) [3]	3.7°	2.3°	9.3°
Gen. Gamut ($\sigma = 4$,2-jet) [8]	3.6°	2.1°	9.4°
Proposed (ML Estimate)	**5.6°**	**3.5°**	**12.9°**

database. Note that this means the training and testing images overlap. Further, since the training set contains all objects that appear in the database, albeit under different illuminants, this training scheme may be prone to overfitting. Finally, since the training data contain only one unique illuminant, we can only evaluate the maximum likelihood version of our method, which does not make use of a prior over illuminants.

We report error quantiles for various methods on this database in Table 5. We find that the proposed method performs better than gray world and first-order gray edge. It has a higher mean error and median error than second-order gray edge, but it appears to be more robust with a lower worst 25 percent error.

Interestingly, the gamut mapping methods perform the best, with significantly lower error quantiles. One possible explanation is that object-scale images tend to have a smaller number of independent samples of pixels/coefficients than scene-scale images so that modeling the support (or convex hull) of the observed set is more robust than attempting to describe and exploit its distribution, as done by gray world, gray edge, and the proposed algorithm. However, drawing strong conclusions from images of only 31 distinct objects would be premature. Since they learn a larger number of parameters and are trained on the entire set of reflectances that occur in the database, it is also possible that the gamut mapping methods are simply overfitting to the convex hull of this small object set.

5 CONCLUSION

In this work, we describe an effective way to exploit spatial dependencies among pixels in a color image for the color constancy problem. We model these dependencies by first decomposing the image using a set of spatially decorrelating filters, and then analyzing the statistics of their coefficients. These color coefficients are *better behaved* than colors of individual pixels, and they enable color constancy techniques that are quite accurate and efficient. When no prior information about the illuminant is available, it can be estimated in a maximum likelihood framework, and for cases when one has prior illuminant information, we introduce a prior model that can be incorporated during estimation.

Interesting directions for future work include relaxing the two main assumptions made in this work: spatially constant illuminant spectrum and diagonal transforms for chromatic adaptation. Even when there is a single illuminant in the scene, mutual illumination causes the *effective* illuminant spectrum to vary from point to point. A first step that could lead to improvement would be to remove regions that appear to be outliers to the general statistics of the image, in the hope that these are likely to correspond to areas with significant interreflections. Alternatively, one could explore variants of the proposed method that assume the scene to be lit by a mixture of n illuminants. The problem would then be to estimate which parts of the image are lit by which illuminant, as well the color of each of these illuminants.

It would be interesting to explore avenues to use the spatio spectral model introduced here to go beyond diagonal transforms. One path is to estimate general linear transforms under appropriate constraints. Another is to consider a set of registered training images taken under different illuminants, and then learn the functional form of the map between corresponding pixels. The parameters of this map can then be estimated from a test image to do color constancy.

Another challenge is to adapt this method for applications that need color constancy to work on images captured using consumer digital cameras that do not provide linear (RAW) image data. In this work, we have assumed that training and testing are performed with images taken from the same or similar cameras. This is important because, as shown in [30], the color spaces and nonlinear processing done in cameras can vary significantly, affecting both the subband statistics and the illuminant statistics in Q. Saenko et al. [31] look at the problem of adapting visual category models learned from one camera to another. It would be interesting to investigate whether the same can be done for color constancy in a way that would allow training data from one camera or linear images to be adapted to a new domain, with minimal additional information.

Finally, we note that computational color constancy algorithms are often motivated by observations from the human visual system [2], [32]. The relative success of this method therefore naturally raises questions about whether the human visual system employs related processing for adaptation and color constancy. Indeed, psychophysical experiments have shown a strong interaction between the spatial orientation and frequency of a stimulus and the chromatic adaptation it induces [33], [34]. The mechanisms that govern these interactions are poorly understood [33], and the machinery developed in this paper might be useful in creating experiments to analyze them further. Separate experiments have linked the textures of familiar objects to human color perception [35], [36], and while texture is currently thought to affect color perception through object memory, it is worth exploring the contribution of spatial correlations to this effect.

REFERENCES

[1] G. Buchsbaum, "A Spatial Processor Model for Object Colour Perception," *J. Franklin Inst.,* vol. 310, no. 1, pp. 1-26, 1980.
[2] E. Land, "The Retinex Theory of Color Vision," *Scientific Am.,* vol. 237, no. 6, pp. 108-128, 1977.

[3] D. Forsyth, "A Novel Algorithm for Color Constancy," *Int'l J. Computer Vision,* vol. 5, no. 1, pp. 5-36, 1990.

[4] D. Brainard and W. Freeman, "Bayesian Color Constancy," *J. Optical Soc. of Am. A,* vol. 14, no. 7, pp. 1393-1411, 1993.

[5] C. Rosenberg, T. Minka, and A. Ladsariya, "Bayesian Color Constancy with Non-Gaussian Models," *Proc. Advances in Neural Information Processing Systems,* 2003.

[6] P. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian Color Constancy Revisited," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.

[7] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-Based Color Constancy," *IEEE Trans. Image Processing,* vol. 16, no. 9, pp. 2207-2214, Sept. 2007.

[8] A. Gijsenij, T. Gevers, and J. van de Weijer, "Generalized Gamut Mapping Using Image Derivative Structures for Color Constancy," *Int'l J. Computer Vision,* vol. 86, no. 2, pp. 127-139, 2010.

[9] G. West and M.H. Brill, "Necessary and Sufficient Conditions for Von Kries Chromatic Adaptation to Give Color Constancy," *J. Math. Biology,* vol. 15, no. 2, pp. 249-258, 1982.

[10] G. Finlayson, M. Drew, and B. Funt, "Diagonal Transforms Suffice for Color Constancy," *Proc. IEEE Int'l Conf. Computer Vision,* 1993.

[11] H. Chong, S. Gortler, and T. Zickler, "The Von Kries Hypothesis and a Basis for Color Constancy," *Proc. IEEE Int'l Conf. Computer Vision,* 2007.

[12] K. Barnard, "Improvements to Gamut Mapping Colour Constancy Algorithms," *Proc. European Conf. Computer Vision,* 2000.

[13] G. Finlayson, S. Hordley, and P. Hubel, "Color by Correlation: A Simple, Unifying Framework for Color Constancy," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 11, pp. 1209-1221, Nov. 2001.

[14] L. Shi, W. Xiong, and B. Funt, "Illumination Estimation via Thin-Plate Spline Interpolation," *J. Optical Soc. of Am. A,* vol. 28, no. 5, pp. 940-948, 2011.

[15] A. Hurlbert and T. Poggio, "Synthesizing a Color Algorithm from Examples," *Science,* vol. 239, no. 4839, pp. 482-485, 1988.

[16] V. Cardei, B. Funt, and K. Barnard, "Estimating the Scene Illumination Chromaticity Using a Neural Network," *J. Optical Soc. of Am. A,* vol. 19, no. 12, pp. 2374-2386, 2002.

[17] W. Xiong and B. Funt, "Estimating Illumination Chromaticity via Support Vector Regression," *J. Imaging Science and Technology,* vol. 50, p. 341, 2006.

[18] R. Gershon, A. Jepson, and J. Tsotsos, "From [r, g, b] to Surface Reflectance: Computing Color Constant Descriptors in Images," *Perception,* vol. 17, pp. 755-758, 1988.

[19] B. Singh, W. Freeman, and D. Brainard, "Exploiting Spatial and Spectral Image Regularities for Color Constancy," *Proc. Workshop Statistical and Computational Theories of Vision,* 2003.

[20] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color Constancy Beyond Bags of Pixels," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.

[21] I. Selesnick, "The Estimation of Laplace Random Vectors in Additive White Gaussian Noise," *IEEE Trans. Signal Processing,* vol. 56, no. 8, pp. 3482-3496, Aug. 2008.

[22] F. Sinz, S. Gerwinn, and M. Bethge, "Characterization of the p-Generalized Normal Distribution," *J. Multivariate Analysis,* vol. 100, no. 5, pp. 817-820, 2009.

[23] L. Shi and B. Funt, "Re-Processed Version of the Gehler Color Constancy Dataset of 568 Images," Accessed from http://www.cs.sfu.ca/~colour/data/, 2012.

[24] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational Color Constancy: Survey and Experiments," *IEEE Trans. Image Processing,* vol. 33, no. 9, pp. 2475-2489, Sept. 2011.

[25] K. Barnard, "A Comparison of Computational Color Constancy Algorithm-Part 1: Theory and Experiments with Synthetic Data," *IEEE Trans. Image Processing,* vol. 11, no. 9, pp. 972-984, Sept. 2002.

[26] R. Hogg and E. Tanis, *Probability and Statistical Inference.* Macmillan, 1977.

[27] S. Hordley and G. Finlayson, "Reevaluation of Color Constancy Algorithm Performance," *J. Optical Soc. of Am. A,* vol. 23, no. 5, pp. 1008-1020, 2006.

[28] F. Ciurea and B. Funt, "A Large Image Database for Color Constancy Research," *Proc. IS&T/SID Color Imaging Conf.,* 2003.

[29] K. Barnard, L. Martin, A. Coath, and B. Funt, "A Comparison of Computational Color Constancy Algorithms. II. Experiments with Image Data," *IEEE Trans. Image Processing,* vol. 11, no. 9, pp. 985-996, Sept. 2002.

[30] A. Chakrabarti, D. Scharstein, and T. Zickler, "An Empirical Camera Model for Internet Color Vision," *Proc. British Machine Vision Conf.,* 2009.

[31] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting Visual Category Models to New Domains," *Proc. European Conf. Computer Vision,* 2010.

[32] K. Hirakawa and T. Parks, "Chromatic Adaptation and White-Balance Problem," *Proc. IEEE Conf. Image Processing,* 2005.

[33] A. Werner, "The Spatial Tuning of Chromatic Adaptation," *Vision Research,* vol. 43, no. 15, pp. 1611-1623, 2003.

[34] M. Webster, "Human Colour Perception and Its Adaptation," *Network: Computation in Neural Systems,* vol. 7, no. 4, pp. 587-634, 1996.

[35] T. Hansen, M. Olkkonen, S. Walter, and K. Gegenfurtner, "Memory Modulates Color Appearance," *Nature Neuroscience,* vol. 9, no. 11, pp. 1367-1368, 2006.

[36] M. Olkkonen, T. Hansen, and K. Gegenfurtner, "Color Appearance of Familiar Objects: Effects of Object Shape, Texture, and Illumination Changes," *J. Vision,* vol. 8, no. 5, p. 13, 2008.

Ayan Chakrabarti received the BTech and MTech degrees in electrical engineering from the Indian Institute of Technology Madras, Chennai, India, in 2006, and the SM and PhD degrees in engineering sciences from Harvard University, Cambridge, Massachusetts, in 2008 and 2011, respectively. He is currently a post-doctoral fellow at Harvard University, where he works with Todd Zickler as a member of the Harvard Computer Vision Laboratory, and the Graphics Vision and Interaction group. His research focuses on the development of signal processing and statistical methods for applications in computer vision and computational photography. He is a member of the IEEE.

Keigo Hirakawa received the BS degree in electrical engineering from Princeton University, Princeton, New Jersey, in 2000 with high honors, the MS and PhD degrees in electrical and computer engineering from Cornell University, Ithaca, New York, in 2003 and 2005, respectively, and the MM degree in jazz performance studies from the New England Conservatory of Music, Boston, Massachusetts, in 2006 with high honors. He was a research associate at Harvard University, Cambridge, Massachusetts, from 2006 to 2009. He is currently an assistant professor at the University of Dayton, Dayton, Ohio. He heads Intelligent Signal Systems Laboratory at the University of Dayton, where the group focuses on statistical signal processing, color image processing, and computer vision. He has received honors from the IEEE, the IS&T, and the IAPR. He is a member of the IEEE.

Todd Zickler received the BEng degree in honors electrical engineering from McGill University, Montreal, Quebec, Canada, in 1996 and the PhD degree in electrical engineering from Yale University, New Haven, Connecticut, in 2004, under the direction of P. Belhumeur. He joined the School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, as an assistant professor in 2004 and was appointed the John L. Loeb Associate Professor of the Natural Sciences in 2008. He is the director of the Harvard Computer Vision Laboratory, and his research is focused on modeling the interaction between light and materials and developing algorithms to extract scene information from visual data. His work is motivated by applications in face, object, and scene recognition; image-based rendering; content-based image retrieval; image and video compression; robotics; and human-computer interfaces. He is a recipient of the US National Science Foundation (NSF) Career Award and a Research Fellowship from the Alfred P. Sloan Foundation. His research is funded by the NSF, the US Army Research Office, the US Office of Naval Research, and the Sloan Foundation. He is a member of the IEEE.