Computational Color Constancy with Spatial Correlations

Computational Color Constancy with Spatial Correlations

Ayan Chakrabarti
Keigo Hirakawa
and
Todd Zickler

TR-09-10

Computer Science Group
Harvard University
Cambridge, Massachusetts

This report has also been submitted to the IEEE Transactions on Pattern Analysis and Machine Intelligence.

Computational Color Constancy with Spatial Correlations

Ayan Chakrabarti, Keigo Hirakawa, and Todd Zickler

Abstract—The color of a scene recorded by a trichromatic sensor varies with the spectral distribution of the illuminant. For recognition and many other applications, we seek to process these measurements to obtain a color representation that is unaffected by illumination changes. Achieving such *color constancy* is an ill-posed problem because both the spectral distribution of the illuminant and the scene reflectance are unknown. For the most part, methods have approached this problem by leveraging the statistics of individual pixel measurements, independent from their spatial contexts. In this work, we show that the strong spatial correlations that exist between measurements at neighboring image points encode useful information about the illuminant and should not be ignored. We develop a method to encode these correlations in a statistical model and exploit them for color constancy. The method is computationally efficient, allows for the incorporation of prior information about natural illuminants, and performs well when evaluated on a large database of natural images.

Index Terms—Color Constancy, Statistical Modeling, Spatial Correlations, Model Fitting, Illuminant Statistics, Regularization

1 Introduction

A trichromatic sensor, be it retinal or RGB, collects three measurements of light, and these measurements depend on the spectral reflectance of an observed surface as well as the spectral irradiance that is incident upon it. If the incident illumination changes, so do the measurements, and this makes color a difficult cue to exploit for scene understanding. To deal with this inconvenience, we seek to achieve *color constancy*—the ability to maintain a representation of surface color that is unaffected by illumination changes. The human visual system achieves color constancy with reasonable success, and computer vision systems are likely to benefit from emulating this functionality.

Determining an 'illumination-invariant' representation of the colors of an observed scene is an ill-posed problem because both the illumination and reflectance are typically unknown. In fact, even when the spectral distribution of the illuminant is known, the task of mapping a trichromatic image to its 'canonical' illumination-invariant representation is severely under-determined. Many traditional approaches address these issues by exploiting the statistics of individual pixels in an image [1]–[11], and in doing so, they ignore the spatial correlations that exist between neighboring image points. This lies in contrast to the human visual system, which seems to employ joint processing of spatial and spectral signals during chromatic adaptation [12].

In this paper, we explore the utility of combining spatial and spectral cues to strengthen the inference on spectral information. Our goals are to (I) develop a statistical model that accounts for the spatio-spectral redundancies in natural color images; and (II) develop a method to find the mapping from an input image to its illumination-invariant representation by model-fitting. As suggested by an earlier version of this work [13] and further demonstrated in the sequel, incorporating a set of spatial basis vectors, like those typically used for spatial decompositions of greyscale images, can greatly enhance our ability to exploit the intrinsic structure of color images and enable more robust color constancy. Conveniently, the estimation algorithm has low computational cost, and admits incorporating prior knowledge about illuminant statistics.

• *AC and TZ are with the Harvard School of Engineering and Applied Sciences, Cambridge, MA 02138. KH is with the University of Dayton, Dayton, OH 45469. E-mails: {ayanc@eecs.harvard.edu, k.hirakawa@notes.udayton.edu, zickler@eecs.harvard.edu}.*

1.1 Problem Formulation

We assume a Lambertian model in which spectral reflectance reduces from a function of wavelength and surface orientation to one of wavelength alone. Let $\kappa : \mathbb{R} \times \mathbb{Z}^2 \to [0, 1]$ be the spectral reflectance as a function of wavelength $\lambda \in \mathbb{R}$ at a surface patch observed by pixel $\boldsymbol{n} \in \mathbb{Z}^2$. We represent the spectral irradiance that is incident at each scene point by $\ell : \mathbb{R} \times \mathbb{Z}^2 \to \mathbb{R}_+$; it is also a function of λ and $\boldsymbol{n}$. As depicted in Fig. 1, the spectral distribution of the radiance that is emitted toward the camera is then the product of ℓ and κ.

We further assume that a linear trichromatic sensor provides three spectrally-distinct measurements of the radiance along each observed ray.[1] We denote the three-vector of measurements obtained at each pixel by:

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

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

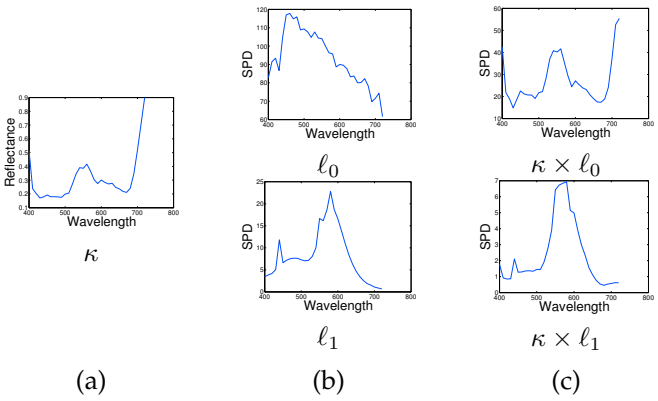


Fig. 1. SPD of light (plotted vs wavelength λ on the x axis) reflected by the same surface under two different illuminants. (a) Diffuse reflectivity of the surface. (b) Spectrum of the two illuminants. (c) Spectrum of light incident on camera.

where $\boldsymbol{f}(\lambda) = [f^{\{1\}}(\lambda), f^{\{2\}}(\lambda), f^{\{3\}}(\lambda)]^T$ are the spectral transmittance distributions of three color filters (e.g., red, green and blue). Without loss of generality, measurements $\boldsymbol{y}(\boldsymbol{n})$ can be taken as coordinates in a standard color space, such as CIE XYZ, because a camera's color filters are usually designed to (approximately) span the same spectral subspace as the standard color matching functions. Finally, in many cases, the recorded tristimulus vectors undergo an enhancement step referred to as *gamma correction* before being stored or transmitted. This is a compressive nonlinearity of the form:

$$\gamma\left(y^{\{i\}}(\boldsymbol{n})\right) = \left(y^{\{i\}}(\boldsymbol{n})\right)^{1/g}, \tag{2}$$

where g is typically 2.2.

In computational color constancy, we are given the observations $\boldsymbol{y}(\boldsymbol{n})$ (or their gamma-corrected versions $\gamma(\boldsymbol{y}(\boldsymbol{n}))$) and we seek an illumination-invariant representation $\boldsymbol{x}(\boldsymbol{n})$ given by

$$\boldsymbol{x}(\boldsymbol{n}) := \int \boldsymbol{f}(\lambda)\kappa(\lambda, \boldsymbol{n})\ell_0(\lambda)\, d\lambda. \tag{3}$$

Here ℓ_0 is understood to be a known "canonical illuminant" (D65, E, etc.) that is spatially-uniform. Transforming from $\boldsymbol{y}$ to $\boldsymbol{x}$ can be viewed as a *white balance* operation and is an example of *chromatic adaptation*.

1.2 Related Work

The observation models in (1) and (3) are trichromatic reductions of the spectral distribution κ that cannot be reversed, even when ℓ and ℓ_0 are known. Thus, this formulation of the computational color constancy problem is severely under-determined. Two standard techniques for addressing such ill-posed and under-determined inverse problems are (I) parameterized estimation (*i.e.*, an assumed parametric form for the mapping from $\boldsymbol{y}$ to $\boldsymbol{x}$) and (II) the introduction of a "model" (sometimes a prior) for κ or $\boldsymbol{x}$. Once the parametric form of the mapping is chosen, its parameters for a given input image can be determined according to the assumed model for κ or $\boldsymbol{x}$.

Prior work in the area of color constancy follows this standard practice. Under spatially-uniform illumination (*i.e.*, $\ell(\lambda, \boldsymbol{n}) = \ell(\lambda)$), a *linear* mapping of the form

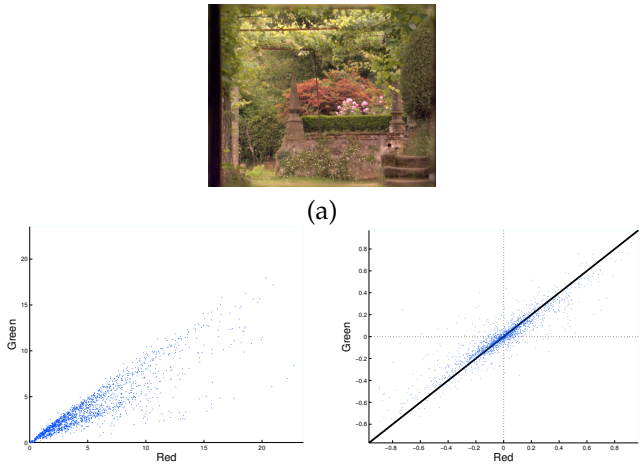
$$\hat{\boldsymbol{x}}(\boldsymbol{n}) = \boldsymbol{M}\boldsymbol{y}(\boldsymbol{n}), \qquad \boldsymbol{M} \in \mathbb{R}^{3\times 3} \tag{4}$$

can achieve exact chromatic adaptation in many cases [14], and this can even be simplified to a diagonal transformation (in an optimized linear color space) when the 'world' of illuminant spectra and spectral reflectances satisfy certain joint constraints [14]–[16]. Under these conditions, mapping an image to its canonical, illumination-invariant representation reduces to the problem of estimating three diagonal coefficients. With some abuse of notation, we let $\boldsymbol{M}$ denote a diagonal matrix for the remainder of this paper.

Having settled on a linear diagonal form for the mapping from $\boldsymbol{y}$ to $\boldsymbol{x}$, another question is how to estimate the three parameters of this mapping for an input image that is captured under unknown illumination. *White Patch* [1] and *Grey World* [2] 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 Grey World method assumes that the sample mean of $\boldsymbol{x}(\boldsymbol{n})$ is 'grey', meaning that it is proportional to $[1, 1, 1]^T$. The diagonal elements of $\boldsymbol{M}$ are then estimated by mapping $\max(\boldsymbol{y})$ or $\text{avg}(\boldsymbol{y})$ to white.

Alternatives to White Patch and Grey World include gamut constraint methods. Based on the observation that the gamut (the complete set of tri-stimulus vectors observed in an image) is skewed by a change in illumination, these methods find the diagonal matrix $\boldsymbol{M}$ that maps all observed tri-stimulus vectors inside the convex hull of the "expected gamut" under canonical illumination [3]. Other alternatives are those based on the dichromatic model [17], according to which the observed tri-stimulus vectors from a curved, spatially-uniform, dielectric surface lie in a plane containing the illuminant color. When many such regions can be identified in an image, the resulting color histograms can be analyzed to estimate the illuminant color, and thus the diagonal entries of $\boldsymbol{M}$ (e.g., [4]–[6]). Finally, a number of learning-based methods have been proposed that use databases of labeled data to train linear filters [18], neural networks [10], and Bayesian prior distributions of surface reflectances and illuminant spectra [7]–[9]. The resulting models can then be used to infer the entries of $\boldsymbol{M}$ for a novel input image.

These methods lack robustness to phenomenon such as noise and the presence of large colorful objects—that skew the color histogram of $\boldsymbol{y}$, and a variety of



(a) (b) (c)

Fig. 2. Statistics of pixel colors individually and with spatial context. (a) Synthetic gamma-corrected RGB image generated from hyper-spectral reflectance data image [23] and the flat illuminant ℓ_0. Scatter plots for intensity values of red and green channels of (b) individual pixels; and (c) image patches projected onto a particular spatial basis vector, that are found to be largely distributed along the achromatic $[1\ 1]$ axis represented by the black line.

work suggests that improvements might be gained by employing spatial image features that are less sensitive to such effects. For example, Gershon et al. [19] assume that the average of mean colors of segmented regions of an image, rather than of individual pixels, is color neutral. van de Weijer et al. [20] and Gijsenij et al. [21] propose applying Grey World and gamut constraint approaches respectively to filtered versions of an image rather than to the image itself. The *Grey Edge* method [20] posits that image gradients are on average color neutral, while the generalized gamut mapping algorithm [21] proposes strategies for combining cues from the "expected gamuts" of different linear filter coefficients. Finally, Singh et al. [22] reason about the illuminant using linear bases for small spatial windows of spectral images. Broadly, our paper falls into this category as well, since we too seek to exploit combinations of spatial and spectral information that provide more direct access to scene illumination.

2 Spatio-Spectral Modeling

Instead of relying on statistics of individual pixels colors, we draw inspiration from previous studies [19]–[22] and seek to exploit information about a pixel's spatial context. We aim to build an explicit statistical model of the spatial dependencies between nearby image points, and thereby exploit the high degrees of correlation that exist among neighboring pixels in a typical scene.

2.1 Spatial Correlations

As motivation, consider Fig. 2(b), which illustrates the scatter plot of pixel colors contained in a trichromatic image that has been rendered synthetically from hyper-spectral image data [23]. This has been rendered using equal-energy illumination (illuminant E), so the three-vector for the illuminant color is proportional to $[1, 1, 1]^T$. This plot, which has been projected onto the red/green plane for visualization, demonstrates the difficulty of extracting information about the illuminant color when spatial correlations are ignored. Since the distribution of tri-stimulus vectors is strongly dependent on scene content, it exhibits high variability from scene to scene, even when these scenes are observed under the same illuminant. It is very difficult to distinguish the illuminant effects from this natural scene variability.

As an alternative, we can investigate the benefits of transforming the image through standard techniques—used in image coding, denoising, and so on—to account for correlations between neighboring pixels. Figure 2(c) depicts a scatter plot of the three-vectors resulting from projecting each 8×8 image patch of each color channel of the gamma-corrected[2] trichromatic image onto a particular spatial basis vector (corresponding to a "band-pass" filter). In this case, we see that the scatter is highly concentrated along the achromatic direction, which is the direction of the illuminant color. Unlike the scatter of individual pixels, this concentration occurs in spite of the presence of dominant image colors. This is similar to the observation made in [20] that image gradients tend to be largely achromatic.

More insight into this phenomenon can be gained by looking at the distribution of energy in the Fourier domain. Figure 3 shows the decomposition of the same gamma-corrected trichromatic image into luminance and chrominance components, along with magnitude plots of their respective Fourier transforms. Here, luminance is defined as the component of the color image along the direction of the illuminant, and chrominance is the residual. Specifically, for this gamma-corrected image, which is rendered under canonical illuminant ℓ_0 (illuminant E) the two components are given by

$$x_{\text{lum}}(\boldsymbol{n}) = \frac{1}{3}\left\langle \gamma(\boldsymbol{x}(\boldsymbol{n})), [1\ 1\ 1]^T \right\rangle,$$
$$x_{\text{chr}}^{\{i\}}(\boldsymbol{n}) = \gamma(x^{\{i\}}(\boldsymbol{n})) - x_{\text{lum}}(\boldsymbol{n}), \qquad i \in \{1, 2, 3\}. \quad (5)$$

Figure 3 clearly shows that the energy of the chrominance signal (*i.e.*, $\sum_i |x_{\text{chr}}^{\{i\}}|^2$) in the Fourier domain decays more rapidly than that of the luminance signal (*i.e.*, $|3x_{\text{lum}}|^2$), which in turn reflects the fact that that the correlations between spectral distributions at distinct image points are spatial-frequency dependent. This dominance of the luminance signal at high spatial frequencies is in direct agreement with Fig. 2(c), which also shows the contributions from the chrominance signal (the component orthogonal to $[1\ 1\ 1]$) in a high-frequency sub-band to be limited. That the energy of the chrominance signal

2. Ones sees similar effects using the linear image, but empirically, we find that gamma correction improves energy compaction in this plot. Further discussion can be found in Sect. 5.

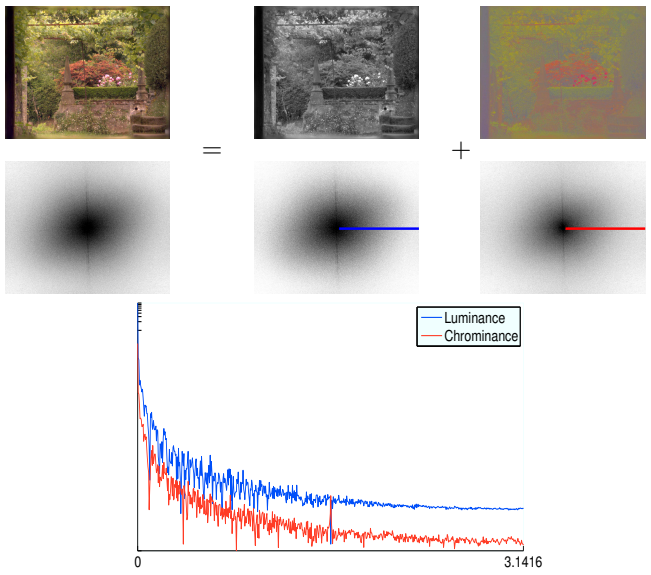


Fig. 3. Decomposition of (top row) a color image into luminance (middle column) and (right column) chrominance components and (middle row) log-magnitudes of the corresponding Fourier spectra. The bottom row shows the plots of the log-magnitudes for the luminance and chrominance images along a line in frequency space (as labeled in the middle row).

is concentrated at lower spatial frequencies agrees with the intuition that the luminance signal captures most inter-object boundaries and intra-object textures, while the residual chrominance signal captures the deviation from the baseline intensity image and is largely void of content at high spatial frequencies. These observations are also consistent with literature in color image modeling(e.g., [24]) and the contrast sensitivity functions of the human visual system [25].

2.2 Statistical Model

The observations of the previous section suggest that image components across spatial frequencies provide more direct information about the scene illuminant than individual pixels, and that a joint spatio-spectral image model that incorporates a notion of spatial frequency will be an effective tool for color constancy. This section develops a spatio-spectral model for gamma corrected images that has two key properties: 1) its parameters can be learned efficiently from a database of natural images captured under canonical illumination; and 2) once trained, it can be used to systematically infer the "white-balance transform" $\boldsymbol{M}$ from an input image $\gamma(\boldsymbol{y})$ captured under unknown illumination.

Since pixels in local neighborhoods (patches) in an image are strongly correlated with each other, the first step is to apply a decomposition that breaks an image into un-correlated components. Canonical linear transforms such as the discrete cosine transform (DCT) [26] and wavelet decompositions [27] have been used to model the local regularity of image data effectively in image denoising [28], compression [29], and other applications. These transforms have attractive joint spatial and frequency localization properties and we propose using them in the color constancy context to decompose a color image into spatial frequency "sub-bands" that can be modeled separately. A comparison of performance for the different transforms is provided in Sec. 4.1.

Formally, $\boldsymbol{X} = \left[\boldsymbol{X}^{\{1\}^T}, \boldsymbol{X}^{\{2\}^T}, \boldsymbol{X}^{\{3\}^T}\right]^T$ is defined to be a random vector corresponding to a patch of color pixels cropped from $\gamma(\boldsymbol{x})$,

$$\boldsymbol{X}^{\{i\}} := \left[\left\{\gamma\left(x^{\{i\}}(\boldsymbol{n})\right) : \boldsymbol{n} \in \eta\right\}\right], \tag{6}$$

where η is a local neighborhood of pixel locations. Let $\{\boldsymbol{D}\}_{k=0,\ldots,K-1}$ be a set of decorrelating filters that serves as a basis set for $\boldsymbol{X}$ such that $\boldsymbol{D}_k^T \boldsymbol{X}^{\{i\}} \in \mathbb{R}$ denotes the k^{th} sub-band coefficient for the i^{th} color channel. $\boldsymbol{\mathcal{D}}_k^T \boldsymbol{X} \in \mathbb{R}^3$ is defined to apply a separable application of the transform in $\boldsymbol{D}_k$ to each color channel as

$$\boldsymbol{\mathcal{D}}_k^T \boldsymbol{X} = \begin{bmatrix} \boldsymbol{D}_k^T & & \\ & \boldsymbol{D}_k^T & \\ & & \boldsymbol{D}_k^T \end{bmatrix} \begin{bmatrix} \boldsymbol{X}^{\{1\}} \\ \boldsymbol{X}^{\{2\}} \\ \boldsymbol{X}^{\{3\}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{D}_k^T \boldsymbol{X}^{\{1\}} \\ \boldsymbol{D}_k^T \boldsymbol{X}^{\{2\}} \\ \boldsymbol{D}_k^T \boldsymbol{X}^{\{3\}} \end{bmatrix}. \tag{7}$$

Now we model $\boldsymbol{X}$ by assigning distributions to every $\boldsymbol{\mathcal{D}}_k^T \boldsymbol{X}$.

Without loss of generality, $\boldsymbol{\mathcal{D}}_0^T \boldsymbol{X} \in \mathbb{R}^3$ may be assumed to be the scaling coefficients/DC component of $\boldsymbol{X}$ (the lowest spatial frequency sub-band). Empirically, one finds $\boldsymbol{\mathcal{D}}_0^T \boldsymbol{X}$ to have the largest variance in natural images, and there is little or no discernible structure in the distribution of these coefficients. We therefore model them as uniform:

$$\boldsymbol{\mathcal{D}}_0^T \boldsymbol{X} \overset{\text{i.i.d.}}{\sim} \mathcal{U}(\nu_{\min}, \nu_{\max}). \tag{8}$$

This assumption is common in many applications and ensures that the DC component does not contribute during inference.

Empirically, we also find that the distributions of the higher-order coefficients $\boldsymbol{\mathcal{D}}_k^T \boldsymbol{X}$ are zero-mean and symmetric. Though more sophisticated distribution models exist [28], [30]–[32], we model their distributions as multivariate normal:

$$\boldsymbol{\mathcal{D}}_k^T \boldsymbol{X} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \boldsymbol{\Lambda}_k), \qquad k > 0. \tag{9}$$

In this model, the image correlations that vary with spatial frequency are captured via the per-sub-band covariance matrices, $\boldsymbol{\Lambda}_k = E[\boldsymbol{\mathcal{D}}_k^T \boldsymbol{X} \boldsymbol{X}^T \boldsymbol{\mathcal{D}}_k^T] \in \mathbb{R}^{3\times 3}$.

Training this model requires assigning values to the covariance matrices, $\boldsymbol{\Lambda}_k$, using a database of natural images with canonical illumination. Figure 4 visualizes the covariance matrices (for a few sub-bands k) that were learned from a database of such images (see Sect. 4). The covariance matrices are depicted using iso-density ellipsoids, and one immediately sees that the largest variance exists along the achromatic axis (*i.e.*, along $[1,1,1]^T$). This evidence further supports the observations made

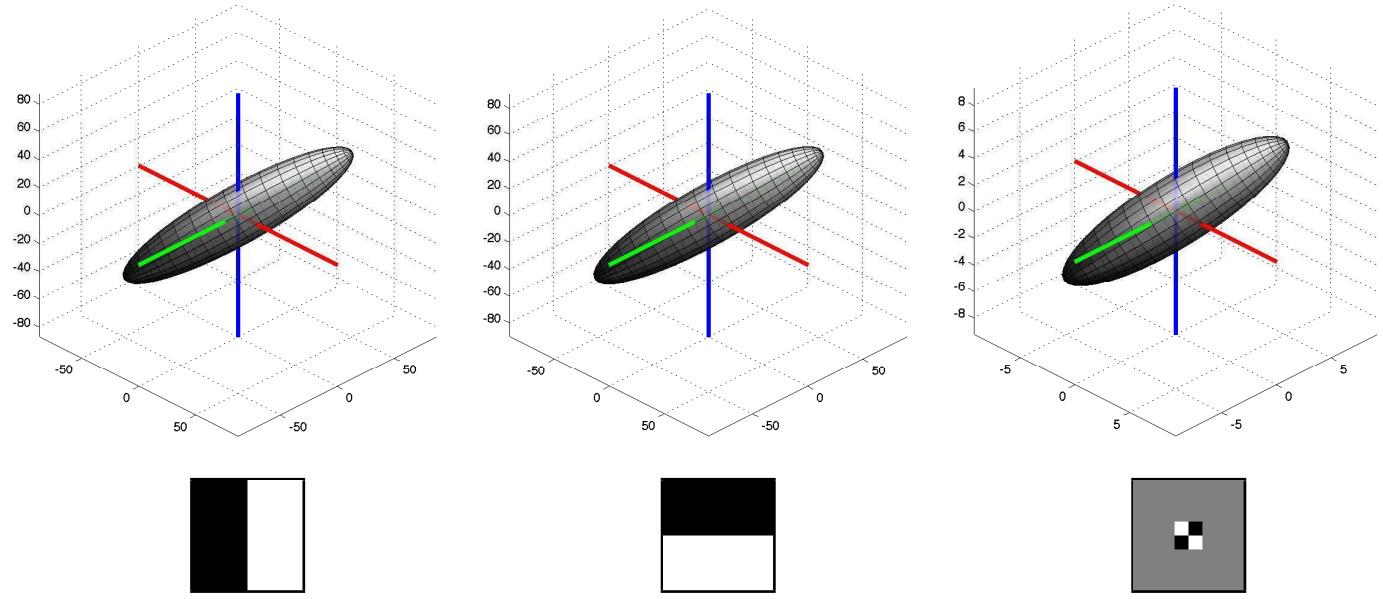


Fig. 4. Ellipsoids corresponding to the covariance matrices $\mathbf{\Lambda_k}$ for three values of $k > 0$. In all cases, the major eigenvector is along $[1\ 1\ 1]$ and the corresponding eigenvalue is much larger than the other eigenvalues. The eccentricities of these ellipsoids,*i.e.*, the ratios of the eigenvalues, determine the expected concentration of variance along the grey direction during estimation.

in Figs. 2 and 3 that luminance dominates for non-DC spatial components, and thus that higher spatial frequencies (*i.e.*, $k > 0$) provide more direct information about the scene illuminant than individual pixels.

Natural comparisons to this approach are the techniques described in [20] and [21] that also seek to characterize the properties of filter coefficients of natural images. Grey edge [20] may be thought of as using a statistic of only a "single" sub-band corresponding to the choice of edge filter. Generalized gamut mapping [21], on the other hand, considers the output of multiple linear filters (which we interpret as multiple sub-bands). It determines the "gamut" for these coefficients, *i.e.*, a valid range of "colors" that they may have. In contrast, we focus on a statistical model that exploits the phenomenon observed in Fig. 3. These dissimilarities lead to differences in estimation methodology (see Sec. 3) and performance (see Tables 2 and 4).

3 Estimation

Given an image $\boldsymbol{y}$ captured under an unknown illuminant, the spatio-spectral model described in the previous section can be leveraged to efficiently estimate the matrix $\boldsymbol{M}$ in (4), and therefore to generate the corresponding illuminant-invariant image $\hat{\boldsymbol{x}}$. The optimal $\hat{\boldsymbol{M}}$ is computed as

$$\hat{\boldsymbol{M}} = \arg\min_{\boldsymbol{M}} R(\gamma(\hat{\boldsymbol{x}})) = \arg\min_{\boldsymbol{M}} R\left(\gamma(\boldsymbol{M}\boldsymbol{y})\right), \qquad (10)$$

where the cost function $R(\cdot)$ is defined with respect to the model defined in (9). With the restriction that $\boldsymbol{M}$ is diagonal, $\gamma(\cdot)$ commutes with matrix multiplication and (10) simplifies to

$$\hat{\boldsymbol{M}} = \underset{\boldsymbol{M}\in\mathrm{diag}(\mathbb{R}^3)}{\arg\min}\ R\left(\gamma(\boldsymbol{M})\gamma\left(\boldsymbol{y}\right)\right). \qquad (11)$$

While $R(\cdot)$ may take various forms—based on maximal likelihood, maximum a posteriori, and posterior mean estimates—this paper focuses on one that admits a closed-form solution.

3.1 Model Fitting Approach

The cost function $R(\cdot)$ is formulated such that the $\hat{\boldsymbol{x}}$ corresponding to $\hat{\boldsymbol{M}}$ has a high probability of occurring under our model. For a patch $\hat{\boldsymbol{X}}$ cropped from $\hat{\boldsymbol{x}}$, a set of functions $R_k(\cdot)$ is defined as

$$\begin{aligned} R_k(\boldsymbol{\mathcal{D}}_k^T\hat{\boldsymbol{X}}) &= -\log\Pr\left(\boldsymbol{\mathcal{D}}_k^T\boldsymbol{X} = \boldsymbol{\mathcal{D}}_k^T\hat{\boldsymbol{X}}\right) \\ &= \frac{1}{2}\left(\boldsymbol{\mathcal{D}}_k^T\hat{\boldsymbol{X}}\right)^T\boldsymbol{\Lambda}_k^{-1}\left(\boldsymbol{\mathcal{D}}_k^T\hat{\boldsymbol{X}}\right) + \text{const.}\,, \end{aligned} \qquad (12)$$

where $R_k(\cdot)$ measures the fit of the k^{th} spatial sub-band component of the patch $\hat{\boldsymbol{X}}$ to the model in (9).

Let $\{\boldsymbol{Y}_j\}$ be the set of patches cropped from $\boldsymbol{y}$ as per (6). The cost function $R(\cdot)$ is then defined to aggregate the model-fit measure across all sub-band components and patches as

$$\begin{aligned} R(\gamma(\boldsymbol{M})\gamma(\boldsymbol{y})) &= \sum_{j,k>0} z_k R_k\left(\boldsymbol{\mathcal{D}}_k^T\hat{\boldsymbol{X}}_j\right) \\ &= \sum_{j,k>0} z_k R_k\left(\gamma(\boldsymbol{M})\boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y}_j\right), \end{aligned} \qquad (13)$$

where the scalars z_k weigh the relative contributions of each sub-band to $R(\cdot)$. Substituting (13) in (11), we have

$$\hat{\boldsymbol{M}} = \underset{\boldsymbol{M}\in\text{diag}(\mathbb{R}^3)}{\arg\min} \sum_{j,k>0} z_k(\gamma(\boldsymbol{M})\boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y_j})^T\boldsymbol{\Lambda}_k^{-1}(\gamma(\boldsymbol{M})\boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y_j}). \tag{14}$$

Defining $\hat{\boldsymbol{m}} = \text{diag}(\gamma(\hat{\boldsymbol{M}}))$, and using the identity

$$\text{diag}(\boldsymbol{m})\boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y}_j = \text{diag}(\boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y}_j)\boldsymbol{m}, \tag{15}$$

the optimization problem in (14) is equivalent to solving

$$\hat{\boldsymbol{m}} = \underset{\boldsymbol{m}\in\mathbb{R}^3}{\arg\min}\ \boldsymbol{m}^T\boldsymbol{A}\boldsymbol{m}, \tag{16}$$

where $\boldsymbol{A}\in\mathbb{R}^{3\times3}$ is defined as

$$\boldsymbol{A} = \sum_{j,k>0} z_k\ \text{diag}(\boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y_j})\boldsymbol{\Lambda}_k^{-1}\text{diag}(\boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y_j}). \tag{17}$$

Without any additional constraints, the solution to (16) is trivially $\hat{\boldsymbol{m}} = 0$. Therefore, an additional condition, $\|\hat{\boldsymbol{m}}\|^2 = c,\ c > 0$, is added. It is easy to show that the solution to this constrained optimization problem is $\hat{\boldsymbol{m}} = \sqrt{c}\boldsymbol{e}$, where $\boldsymbol{e}$ is the eigen-vector of $\boldsymbol{A}$ corresponding to the smallest eigen-value. It is worth noting that the *color* of the optimal coefficients $\hat{\boldsymbol{m}}$ does not depend on the constant $c > 0$.

3.2 Robustness to Noise

So far, it has been assumed that the image $\boldsymbol{y}$ is observed without noise. It is well accepted that pixel sensor measurements are Poisson counts that result from spatio-spectral-temporal integration of photons that arrive at the sensor. The stochastic nature of the photon arrival process is therefore well understood to be the predominant source of the heteroscedastic noise in low light conditions.

Let $\tilde{\boldsymbol{y}}$ be the noisy version of $\boldsymbol{y}$. Through a technique called "variance stabilization" [33], [34], it has been shown that the noise after gamma correction may be assumed to independent of the signal and Gaussian, *i.e.*,

$$\tilde{\boldsymbol{y}} \sim \mathcal{P}(\boldsymbol{y}) \quad\Leftrightarrow\quad \gamma(\tilde{\boldsymbol{y}}) \sim \mathcal{N}(\gamma(\boldsymbol{y}), \sigma^2), \tag{18}$$

under the approximation $g = 2$ in (2). It therefore follows that

$$\boldsymbol{\mathcal{D}}_k^T\tilde{\boldsymbol{Y}}_j = \boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y}_j + \boldsymbol{w}_{jk}, \tag{19}$$

where $\boldsymbol{w}_{jk} \overset{\text{i.i.d}}{\sim} \mathcal{N}\left(0, \sigma^2\|\boldsymbol{D}_k\|^2\boldsymbol{I}\right)$.

In light of this, the matrix $\boldsymbol{A}$ in (17) can be related to the noisy version $\tilde{\boldsymbol{A}}$ as

$$\begin{aligned}
\mathbb{E}[\tilde{\boldsymbol{A}}] &= \mathbb{E}\left[\sum_{k>0,j} z_k\text{diag}(\boldsymbol{\mathcal{D}}_k^T\tilde{\boldsymbol{Y}}_j)\boldsymbol{\Lambda}_k^{-1}\text{diag}(\boldsymbol{\mathcal{D}}_k^T\tilde{\boldsymbol{Y}}_j)\right] \\
&= \mathbb{E}\left[\sum_{k>0,j} z_k\text{diag}(\boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y_j} + \boldsymbol{w}_{jk})^T\boldsymbol{\Lambda}_k^{-1}\text{diag}(\boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y_j} + \boldsymbol{w}_{jk})\right] \\
&= \sum_{k>0,j} z_k\text{diag}(\boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y}_j)\boldsymbol{\Lambda}_k^{-1}\text{diag}(\boldsymbol{\mathcal{D}}_k^T\boldsymbol{Y}_j) + \sigma^2\boldsymbol{\Sigma} \\
&= \boldsymbol{A} + \sigma^2\boldsymbol{\Sigma},
\end{aligned} \tag{20}$$

where $\boldsymbol{\Sigma}\in\mathbb{R}^{3\times3}$ is a diagonal matrix whose entries are the diagonal entries of $\sum_{k,j} z_k\|\boldsymbol{D}_k\|^2\boldsymbol{\Lambda}_k^{-1}$. $\boldsymbol{A}$ can therefore be easily computed from $\tilde{\boldsymbol{A}}$ and σ^2.

In practice, the proposed color constancy algorithm will work correctly without any special considerations for the noise, or any need to estimate the noise variance σ^2. Due to symmetry in color channels, the diagonal entries of $\boldsymbol{\Lambda}_k$ and consequently of $\boldsymbol{\Sigma}$ are all equal. Therefore, the estimate of $\hat{\boldsymbol{M}}$ based on $\boldsymbol{A}$ and the observed noisy version, $\tilde{\boldsymbol{A}}$, will be identical as the addition of $\sigma^2\boldsymbol{\Sigma}$ leaves the eigen-vectors unchanged.

3.3 Regularization with Illuminant Statistics

So far, the cost function $R(\cdot)$ used in (12) leverages the spatio-spectral statistics of natural scenes, but assumes that all illuminants—and consequently all estimates of $\hat{\boldsymbol{m}}$—are equally likely. This section shows that the definition of $R(\cdot)$ in (13) can be modified to incorporate available prior knowledge about the statistics of illuminant spectra in natural environments [7]–[9].

Let us define a new cost function R_{reg} as

$$\begin{aligned}
R_{\text{reg}}\left(\text{diag}(\boldsymbol{m})\gamma(\boldsymbol{y});\Gamma\right) \quad &= \quad (1-\Gamma)\ R\left(\text{diag}(\boldsymbol{m})\gamma(\boldsymbol{y})\right) \\
&\quad + \Gamma\ \boldsymbol{m}^T\boldsymbol{Q}^{-1}\boldsymbol{m}.
\end{aligned} \tag{21}$$

Here, we let $\boldsymbol{m}^T\boldsymbol{Q}^{-1}\boldsymbol{m}$ be a *regularization* term that is independent of the observed image $\boldsymbol{y}$ and depends only on illuminant statistics. It assigns a relative weighting to the different values of $\boldsymbol{m}$ based on the entries of the 3×3 matrix $\boldsymbol{Q}$. During estimation, $\Gamma\in[0,1]$ controls the contribution of the regularization term relative to $R(\cdot)$.

The regularization matrix $\boldsymbol{Q}$ and the scalar Γ can be learned from a set of training images $\{\boldsymbol{y}_t\}$ captured under a representative set of known illuminants. This set can be a general set of natural images, or restricted to match any available prior information, such as whether the images are captured indoors or outdoors. Let $\{\hat{\boldsymbol{m}}_t\}$ be the "ground-truth" white balance transforms corresponding to the known illuminants for the training images $\{\boldsymbol{y}_t\}$. Then the matrix $\boldsymbol{Q}$ is computed as

$$\boldsymbol{Q} = \sum_t \hat{\boldsymbol{m}}_t\hat{\boldsymbol{m}}_t^T, \tag{22}$$

and the parameter Γ is computed by cross-validation as

$$\Gamma = \underset{\Gamma'\in[0,1]}{\arg\min} \sum_t \text{err}\left(\hat{\boldsymbol{m}}_t, \underset{\boldsymbol{m}}{\arg\min}\ R_{\text{reg}}\left(\text{diag}(\boldsymbol{m})\gamma(\boldsymbol{y}_t);\Gamma'\right)\right), \tag{23}$$

where $\text{err}(\cdot,\cdot)$ is a measure of the error between the estimated and "true" $\boldsymbol{m}$. In general, (23) does not have a closed-form solution, so Γ must be computed using an iterative search algorithm.

Once $\boldsymbol{Q}$ and Γ are trained, incorporating the regularization term adds almost no computational complexity to the inference of $\hat{\boldsymbol{M}}$. The regularized estimate is given by $\hat{\boldsymbol{m}} = \sqrt{c}\boldsymbol{e}'$, where $\boldsymbol{e}'$ is the smallest eigen-vector of $\left[(1-\Gamma)\boldsymbol{A} + \Gamma\boldsymbol{Q}^{-1}\right]$, and $\boldsymbol{A}$ is as defined in (17).

4 Experimental Results

We show results on the 568 color images collected by the authors of [9], of which 246 were labeled as captured indoors, and 322 as captured outdoors under daylight. In addition, we captured a new database of 64 images under various artificial illuminants. RAW images from the cameras (Canon EOS 40D, 1DS and 5D) were used to generate gamma-corrected RGB images using a standard pipeline, *leaving out the camera's auto-white balance step*. The images were then scaled down by a factor of 5 (using an anti-aliasing filter followed by sub-sampling) to remove potential demosaicking artifacts. Every image contained a Macbeth color checker chart at a known location. The grey squares on the chart served as ground truth. The chart was masked out during evaluation.

We evaluate the proposed algorithm on this dataset and also provide comparisons to other computational color constancy methods in the literature. We gauge performance with the standard "angular error" metric—the angular deviation in degrees between the estimated and true "white point" of the input image, *i.e.*, the recorded color of an achromatic surface.

With methods that require training data, we use three fold cross validation, *i.e.*, we divide the database in to three roughly equal folds, and perform estimation on one fold using the other two for training. We first evaluated the proposed method across multiple trials with different random divisions of the database and found the results to be roughly stable across 10 trials— the mean and median error values varied from 3.836° to 3.847° and 2.763° to 2.817° respectively. This was expected since the images in the database are generally uncorrelated, and therefore for all remaining experiments we use a fixed division that is kept constant for all methods.

As noted above, every image in the database is labeled as *"daylight"*, *"artificial illumination"* or *"indoor"*, and we use these labels while evaluating methods that use prior illuminant statistics. In addition to reporting results for all images corrected using illuminant and reflectance statistics learned over the whole database, we also report performance for the case where each image is corrected using illuminant statistics trained only on images with the same label. Note that reflectance statistics are still trained on the whole database.

4.1 Implementation Details

As mentioned in Sect 2.2, the covariance matrices $\{\boldsymbol{\Lambda}_k\}$ need to be estimated from a set of images captured under canonical illumination. Given a set of images for training, we select those images for which the angular deviation of the true illuminant from "grey" is less than 10°. In addition, a corrective diagonal transform was applied to the selected images based on the ground truth to remove any residual effects of illuminant color, and these corrected images were used to compute $\{\boldsymbol{\Lambda}_k\}$.

The weights z_k in (13) were chosen to be equal to the sum of the eigen-values of the corresponding $\boldsymbol{\Lambda}_k$. For regularization, the white point of each image, computed from the gray patches, was used as illuminant color to estimate $\boldsymbol{Q}$ and the parameter Γ was computed as per (23) using the angular error metric.

TABLE 1
Comparison of different spatial decompositions for $\{\boldsymbol{D}_k\}$

	Mean Error	MATLAB Run-time for 584×878 image
Harr Wavelet	3.8°	4.6 secs
Daubechies-4 Wavelet	3.8°	5.6 secs
Daubechies-6 Wavelet	3.9°	6.5 secs
Bi-orthogonal 3,3 Wavelet	4.0°	7.4 secs
DCT 8×8	4.0°	10.1 secs

We next evaluate different spatial decompositions as choices for $\{\boldsymbol{D}\}_k$. Table 1 shows a comparison of running times for estimation on a single image as well as mean angular error values across all images, for the proposed method (without regularization) using the DCT and separable wavelet decompositions. The DCT was computed using 8×8 overlapping patches, while the wavelet results correspond to a 3-level un-decimated separable decomposition for each case.

The Harr wavelet decomposition performs the best and also has the lowest running time. It should be noted however that in a camera pipeline or when working with JPEG images, the DCT coefficients may already be computed and available, rendering the computational cost for DCT negligible. The remaining results use the Harr wavelet decomposition.

4.2 Comparison with Other Methods

We next compare the proposed method (with and without regularization) to the Grey World [2], Grey Edge [20], Generalized Gamut [21] and Rosenberg [9] algorithms. We use the code provided by the authors of [20] for evaluating both Grey World and Grey Edge. For Grey World, the illuminant color was estimated by averaging in the "linear" RGB color space assuming the gamma correction parameter g was 2.2. This choice corresponded to the traditional interpretation of Grey World and performed better than simply computing the mean of the gamma corrected pixels.

For Grey Edge, the parameters were fixed to using first order derivatives, a Minkowski norm of 1.2 and a smoothing parameter of 4. These were determined to be optimal using an exhaustive search over the parameter space to minimize the median error on randomly selected subset of 100 images. The optimal parameters were found to be the same when using different randomly chosen subsets.

The Generalized Gamut method was tested using code made available by the authors. In addition to the automated training step for estimating the "canonical gamut", the experimental evaluation in [21] found different parameters to be optimal for different databases.

TABLE 2
Angular error quantiles for various methods on "Color Checker" database [9]

	All Images (632)		Daylight (322)		Artificial Illum. (64)		Indoor (246)	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Grey World	4.7°	3.6°	5.5°	4.3°	2.9°	2.0°	4.2°	3.3°
Generalized Gamut	5.2°	3.1°	2.5°	1.8°	4.7°	3.7°	8.9°	6.7°
Grey Edge	3.9°	3.0°	3.7°	3.0°	3.2°	2.5°	4.3°	3.2°
Proposed	3.8°	2.8°	3.6°	2.6°	3.0°	2.1°	4.3°	3.1°
Illuminant Statistics from All Images								
Rosenberg	3.7°	2.2°	2.1°	1.3°	3.0°	1.8°	6.0°	3.9°
Proposed with Reg.	3.7°	2.8°	3.4°	2.5°	3.0°	2.2°	4.4°	3.2°
Illuminant Statistics from Respective Label								
Rosenberg	3.5°	2.0°	2.0°	1.3°	2.3°	1.4°	5.8°	4.0°
Proposed with Reg.	2.7°	1.9°	1.6°	1.4°	2.4°	1.8°	4.3°	3.0°

TABLE 3
Running times for different methods

	Running Time
Proposed (C)	2.5 secs
Proposed (MATLAB)	4.6 secs
Grey World (MATLAB)	1.8 secs
Grey Edge (MATLAB)	3.1 secs
Generalized Gamut (MATLAB)	23.2 secs
Rosenberg (MATLAB & C++)	104.8 secs

Therefore, we evaluated various choices for these— the filters to consider, the combination strategy to use, and the smoothing parameter— on one fold of the database, and report performance with settings that yielded minimum mean error. We found that using intersection of the complete "*1-jet*" outperformed intersection of the complete "*2-jet*", intersection of only 1st order filters, and computing the mean over the different algorithm outputs. Further, we found that choosing a smoothing parameter of 5 lead to the lowest error.

The implementation provided by the authors of [9], which includes modifications and enhancements to the original method, was used to evaluate the Rosenberg algorithm. We show results using recommended settings found to perform best in [9].

Table 2 reports the mean and median angular error values for all methods and test cases. In addition to reporting error over all images, we also show results for individual sections of the database. Figures 5,6 and 7 show examples of images from the daylight, artificial illuminant and indoor sub-sets of the database, corrected by the different methods. In Table 3, we report the computational times for the different methods. All results are for a 584×878 image, on a Pentium D 2.8 GHz machine with 1 GB of RAM.

4.3 Discussion

From Table 2 we note that, amongst methods that do not rely on illuminant statistics, the proposed method has the lowest mean and median error values over all images. Grey Edge outperforms the related Generalized Gamut method which also uses cues from image derivatives, and has lower mean and median error values than Grey World as well. Looking at the break down of these results across different sections of the database, we see that Generalized Gamut mapping has impressive performance on the "Daylight" images, but has higher errors for the other sets. The simple Grey World algorithm performs surprisingly well on the "artificial illuminant" and "indoor" sets. This is possibly because these scenes often feature white or grey walls that provide a stable cue to the algorithm. The proposed algorithm's performance is relatively consistent across the different labels.

We note that over the whole database, the prior-based methods have lower errors and that performance improves as expected when using illuminant statistics from images in the same sub-section of the database. Here too, looking at the individual break down of results proves insightful.

We find that the Rosenberg algorithm performs significantly worse on the "indoor" images (consistent with the evaluation in [9]) even when using specific illuminant statistics. This is in contrast to its performance on the "daylight" and "artificial illuminant" set, which is responsible for its overall lower error values on the whole database. The difference in performance is not unexpected since the indoor images have the largest variance in illuminants as these images contain scenes lit by a complex set of illuminants (both artificial and natural), whereas the "daylight" and "artificial illuminant" images have comparatively less variation in illuminant color. The prior for this set is therefore least informative. The significantly higher error values however signify that the Rosenberg method depends heavily on an informative prior for good performance.

Adding regularization to the proposed method improves performance. With a label-specific prior, the effect on "daylight" images is most significant. As with the Rosenberg method, the benefit from illuminant statistics

Fig. 5. Images from the "Daylight" set, corrected with different algorithms. Angular error values for all corrected images are indicated below each image.

is least for the "indoor" images. However, the robust cues from the spatio-spectral modeling ensure that the method remains competitive for this case as well.

The example images also provide some intuition about the performance of the different methods. For example, in Fig. 5(ii), we see that the proposed method, Generalized Gamut and Grey Edge, which use cues from spatial components do better than the Rosenberg method and Grey World which rely on pixel domain statistics. Grey World and Rosenberg correct the wall to white, even though in the actual scene it has a reddish tinge. We also note that although the regularization term leads to better estimates overall as seen in Table 2, for individual images it may lead to higher errors.

In summary, the proposed method appears to benefit from cues from the the statistical characterization of spatial sub-band coefficients leading to consistent and robust performance. It is also able to exploit prior knowledge about illuminant statistics effectively when this knowledge is informative. Further, from Table 3, it is clear that the algorithm is competitive in terms of computational cost.

In addition to the comparisons in this section, we also report results (for non prior-based methods) in the appendix on an older database [35], and compare to other algorithms whose performance on that database is documented in the literature. Source code for the proposed method and the new set of artificial illuminant images, are available at **http://www.eecs.harvard.edu/~ayanc/color-constancy/**.

5 Conclusion

In this work, we noted the existence of spatial dependencies among pixels in a color image and demonstrated their utility for the color constancy problem. We developed a statistical model to quantify these dependencies and an estimation procedure, with a closed form solution, to leverage the model for color correction. We also described the inclusion of a regularization term to

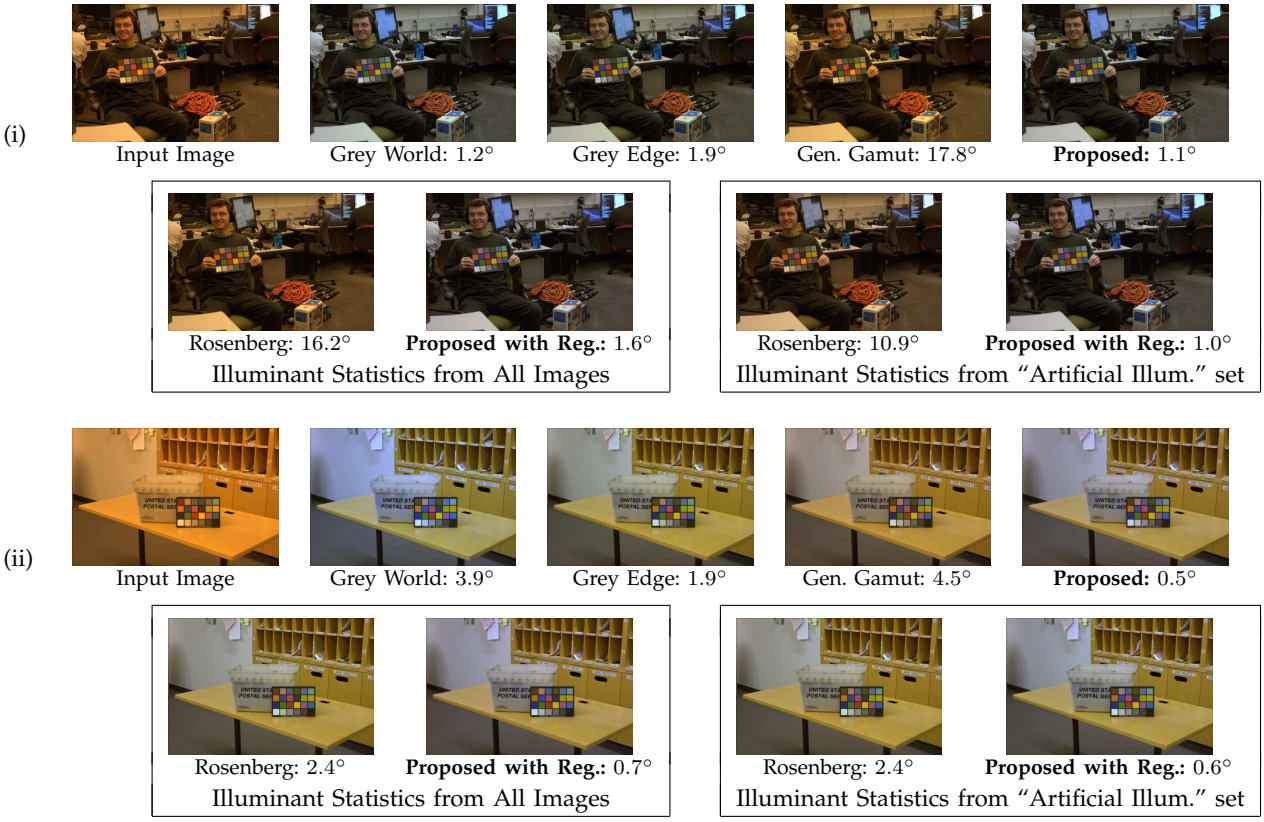


Fig. 6. Images from the "Artificial Illuminants" set, corrected with different algorithms. Angular error values for all corrected images are indicated below each image.

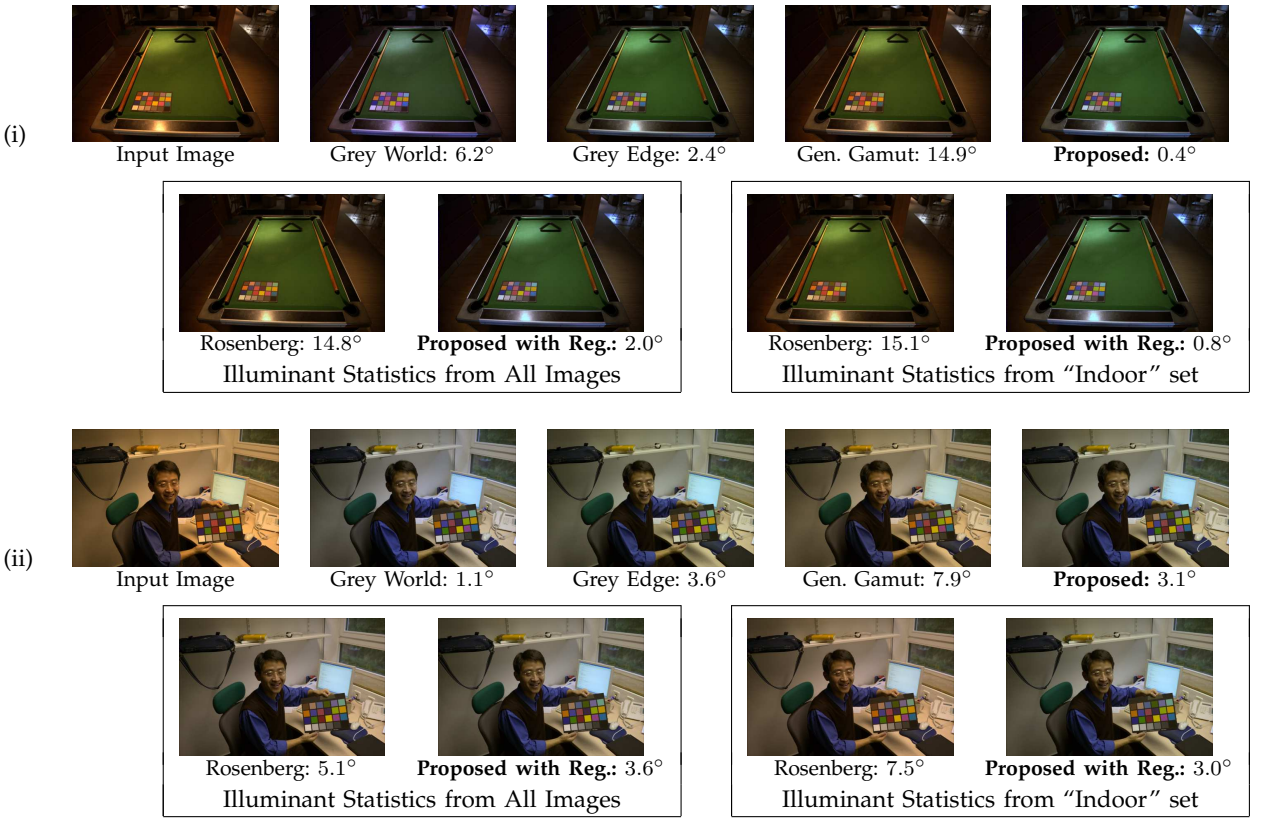


Fig. 7. Images from the "Indoor" set, corrected with different algorithms. Angular error values for all corrected images are indicated below each image.

take into account prior information about the statistics of illuminants. A database of digital color images with ground truth was used to evaluate the effectiveness of our approach.

One of the questions that future work will address is building statistical models in linear RGB space. While the current algorithm is found to perform better using a Gaussian distribution on the sub-band coefficients post-gamma correction, a model on the gamma corrected image implies a corresponding distribution on the linear RGB values, albeit one that is heavier-tailed than the Gaussian. This could possibly be due to inhomogeneity of large valued sub-band coefficients corresponding to boundaries that need more careful treatment.

At the same time, many applications will need color constancy to work on images captured using consumer digital cameras, where RAW image data is not available. In this work, we have assumed that training and testing will be done on images taken from the same or similar cameras. This is important, because as shown in [36], the color spaces and non-linear processing done in cameras can vary significantly— affecting both the sub-band statistics in $\mathbf{\Lambda}_k$ and the illuminant statistics in $\boldsymbol{Q}$. Saenko et al. [37] 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.

The other avenues worth exploring involve relaxing the assumptions made in this work — spatially invariant illumination and diagonal transforms. Even if a scene is lit by a single illuminant, mutual illumination causes the "effective illuminant" spectrum to vary across the scene. A first step that could lead to improved estimates would be to remove regions that appear to be outliers to the general statistics of the image, as these are likely to correspond to areas with significant inter-reflections. Furthermore, variants of the proposed method could be developed that assume the scene is 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.

The problem of moving beyond diagonal transforms is interesting as well. 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 "learning" 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.

Computational color constancy algorithms are often motivated by observations from the human visual system [38], [39]. 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 [12], [40]. The mechanisms that govern these interactions are poorly understood [12], 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 [41], [42], 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.

Finally, it would be interesting to consider similar joint spatio-spectral models for hyper-spectral images, in which we have dense spectral measurements at each pixel. Analyzing a large database of hyper-spectral natural images would likely yield a deeper understanding of the patterns that we have observed empirically in this paper, and it would likely enable more powerful computational tools.

APPENDIX
EVALUATION WITH "GREY BALL" DATABASE

The "Grey Ball" database [35] is a collection of 11355 images captured using a video camera. A grey ball was attached to the camera and is present at the same location in all images, serving as ground truth. It is pioneering in being the first large database of real world images captured to evaluate color constancy methods, and performance measures for a large number of algorithms on this database are available in the literature.

It should be noted here that the images in this database, being frames sampled from a video sequence, are lower in resolution/quality and highly correlated. In fact, the illuminants are virtually identical within many of the sets since they correspond to videos captured at the same location. Training illuminant statistics from these sets therefore leads to over-fitting, causing these statistics to generalize poorly to images in other sets. Indeed, adding illuminant prior-based regularization to our method degrades performance on this database (mean and median angular errors increase from 6.0° and 4.4° to 6.4° and 4.6° respectively). Therefore, we do not include results for illuminant prior-based methods here.

Furthermore, the ground truth information is unreliable in some cases since the grey ball is close to the camera and may be lit by a different illuminant than the scene being captured (the color charts in the database in [9] are placed within the scene to avoid this). As such, the error values on this database need to be interpreted with caution.

We evaluate the proposed method (without regularization) using three fold cross validation. However, because of the correlated nature of the data, care needs to be taken to ensure that frames of the same scene are not used for training and testing simultaneously. The database is divided in to 12 sets of different scenes (there are actually 15 sets corresponding to 15 videos from which the frames are sampled, but we fuse multiple

TABLE 4
Angular error quantiles for various methods on "Grey Ball" database [35]

	Mean	Median
Proposed	$\mathbf{6.0}^\circ$	$\mathbf{4.4}^\circ$
Grey World [2]	6.9°	6.0°
Grey Edge [20]	6.4°	5.3°
Generalized Gamut [21]*	6.5°	5.5°
White Patch [1]*	7.1°	6.7°
Regular Gamut [3]*	7.2°	5.7°
Color by Correlation [11]*	8.1°	6.5°

* Values reported from [21].

sets whose names indicated they were taken at the same venue— such as *CICC2002, CICC2002_2, CICC2002_3*). Three fold cross validation is then done at the level of these sets, ensuring that when testing on an image, frames from the same location are not used for training.

Table 4 reports mean and median error values for various algorithms. We ran our own experiments for the proposed method, Grey World and Grey Edge, while the remaining values are taken from [21]. For Grey Edge, we used parameters suggested in [20] for the database and for Grey World, we compute the mean in linear space as described in Sec. 4.2— this again leads to better results than computing the mean directly. The proposed method has the lowest error values amongst all algorithms in terms of both mean and median.

REFERENCES

[1] V. Cardei and B. Funt, "Committee-based color constancy," in *Proc. IS&T/SID Color Imaging Conf.*, 1999, pp. 311–313.

[2] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst*, vol. 310, no. 1, pp. 1–26, 1980.

[3] D. Forsyth, "A novel algorithm for color constancy," *Intl. J. of Comp. Vis.*, vol. 5, no. 1, 1990.

[4] G. Finlayson and G. Schaefer, "Convex and non-convex illuminant constraints for dichromatic colour constancy," in *Proc. CVPR*, 2001.

[5] H.-C. Lee, "Estimating the illuminant color from the shading of a smooth surface," in *A.I. Memo 1068, MIT Artificial Intelligence Laboratory*, 1988.

[6] ——, "Method for computing the scene-illuminant chromaticity from specular highlights," *J. of Opt. Soc. Am. A*, vol. 3, pp. 3–10, 1986.

[7] D. Brainard and W. Freeman, "Bayesian color constancy," *J. of Opt. Soc. Am. A*, vol. 14, no. 7, pp. 1393–1411, 1993.

[8] C. Rosenberg, T. Minka, and A. Ladsariya, "Bayesian Color Constancy with Non-Gaussian Models," in *Proc. NIPS*, 2003.

[9] P. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian Color Constancy Revisited," in *Proc. CVPR*, 2008.

[10] V. Cardei, B. Funt, and K. Barnard, "Estimating the scene illumination chromaticity using a neural network," *J. of Opt. Soc. Am. A*, vol. 19, no. 12, pp. 2374–2386, 2002.

[11] G. Finlayson, S. Hordley, and P. Hubel, "Color by correlation: A simple, unifying framework for color constancy," *IEEE Trans. Pattern Anal. and Mach. Intell.*, 2001.

[12] A. Werner, "The spatial tuning of chromatic adaptation," *Vision Research*, vol. 43, no. 15, pp. 1611–1623, 2003.

[13] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy beyond bags of pixels," in *Proc. CVPR*, 2008.

[14] H. Chong, S. Gortler, and T. Zickler, "The von Kries hypothesis and a basis for color constancy," in *Proc. ICCV*, 2007.

[15] G. West and M. H. Brill, "Necessary and sufficient conditions for von kries chromatic adaptation to give color constancy," *J. of Math. Bio.*, vol. 15, no. 2, pp. 249–258, 1982.

[16] G. Finlayson, M. Drew, and B. Funt, "Diagonal transforms suffice for color constancy," in *Proc. ICCV*, 1993.

[17] S. A. Shafer, "Using color to separate reflection components," *Color Research and Application*, vol. 10, pp. 210–218, 1985.

[18] A. Hurlbert and T. Poggio, "Synthesizing a color algorithm from examples," *Science*, vol. 239, no. 4839, pp. 482–485, 1988.

[19] 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.

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

[21] A. Gijsenij, T. Gevers, and J. van de Weijer, "Generalized gamut mapping using image derivative structures for color constancy," *Intl. J. of Comp. Vis.*, vol. 86, no. 2, pp. 127–139, 2010.

[22] B. Singh, W. Freeman, and D. Brainard, "Exploiting spatial and spectral image regularities for color constancy," in *Proc. Workshop on Statistical and Computational Theories of Vis.*, 2003.

[23] D. Foster, S. Nascimento, and K. Amano, "Information limits on neural identification of colored surfaces in natural scenes," *Visual Neuroscience*, vol. 21, no. 03, pp. 331–336, 2005.

[24] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Color plane interpolation using alternating projections," *IEEE Trans. Image Processing*, vol. 11, no. 9, pp. 997–1013, 2002.

[25] G. Wyzecki and W. Stiles, "Color Science," *Concepts and Methods, Quantitative Data and Formulae*, 1982.

[26] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transfom," *IEEE Transactions on Computers*, vol. 100, no. 23, pp. 90–93, 1974.

[27] S. Mallat, *A wavelet tour of signal processing*. Academic Pr, 1999.

[28] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Proc.*, vol. 12, no. 11, pp. 1338–1351, 2003.

[29] D. Taubman, M. Marcellin, and M. Rabbani, "JPEG2000: Image compression fundamentals, standards and practice," *J. of Electronic Imaging*, vol. 11, p. 286, 2002.

[30] L. Sendur and I. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Sig. Proc.*, vol. 50, no. 11, pp. 2744–2756, 2002.

[31] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter-and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans. Image Proc.*, vol. 11, no. 5, pp. 545–557, 2002.

[32] K. Hirakawa and X. L. Meng, "An empirical Bayes EM-wavelet unification for simultaneous denoising, interpolation, and/or demosaicing," in *Proc. ICIP*, 2006.

[33] M. Bartlett, "The use of transformations," *Biometrics*, vol. 3, no. 1, pp. 39–52, 1947.

[34] F. Anscombe, "The Transformation of Poisson, Binomial and Negative-Binomial Data," *Biometrika*, vol. 35, no. 3-4, pp. 246–254, 1948.

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

[36] A. Chakrabarti, D. Scharstein, and T. Zickler, "An empirical camera model for internet color vision," in *Proc. BMVC*, 2009.

[37] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Proc. ECCV*, 2010.

[38] E. Land and J. McCann, "Lightness and Retinex Theory," *J. of Opt. Soc. Am.*, vol. 61, no. 1, p. 1, 1971.

[39] K. Hirakawa and T. Parks, "Chromatic adaptation and white-balance problem," in *Proc. ICIP*, 2005.

[40] M. Webster, "Human colour perception and its adaptation," *Network: Computation in Neural Systems*, vol. 7, no. 4, pp. 587–634, 1996.

[41] T. Hansen, M. Olkkonen, S. Walter, and K. Gegenfurtner, "Memory modulates color appearance," *Nature Neuroscience*, vol. 9, no. 11, pp. 1367–1368, 2006.

[42] M. Olkkonen, T. Hansen, and K. Gegenfurtner, "Color appearance of familiar objects: Effects of object shape, texture, and illumination changes," *Journal of Vision*, vol. 8, no. 5, p. 13, 2008.