Computer Vision and Image Understanding xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

A sequential Bayesian approach to color constancy using non-uniform filters

Sandra Skaff *, Tal Arbel, James J. Clark

Centre for Intelligent Machines, McGill University, Montréal, Que., Canada H3A 2A7

ARTICLE INFO

Article history:
Received 20 June 2008
Accepted 25 March 2009
Available online xxxx

Keywords:
Color constancy
Bayesian
Sequential: chaining
Posterior
Prior
Filters: binary
Multiple

ABSTRACT

This paper introduces a non-uniform filter formulation into the Brainard and Freeman Bayesian color constancy technique. The formulation comprises sensor measurements taken through a non-uniform filter, of spatially-varying spectral sensitivity, placed on the camera lens. The main goal of this paper is two-fold. First, it presents a framework in which sensor measurements obtained through a non-uniform filter can be sequentially incorporated into the Bayesian probabilistic formulation. Second, it shows that such additional measurements obtained reduce the effect of the prior in Bayesian color constancy. For the purposes of testing the proposed framework, we use a filter formulation of two portions of different spectral sensitivities. We show through experiments on real data that improvement in the parameter estimation can be obtained inexpensively by sequentially incorporating additional information obtained from the sensor through the different portions of a filter by Bayesian chaining. We also show that our approach outperforms previous approaches in the literature.

1. Introduction

Color constancy is the ability of a vision system to compute a measure of a surface's color that is independent of the spectrum of the light incident on the surface. The human visual system has an effective, if imperfect, color constancy mechanism. Studies of the human color constancy system have led to the development of many color constancy methods for machine vision applications. None of the existing color constancy techniques function perfectly in all situations, and most fail outside of a limited domain of applicability. The difficulty in solving the color constancy problem lies in the bilinear nature of the underlying relation between the illuminant and surface reflectance spectra and the measurements provided by the photoreceptors. Perhaps the most successful methods of dealing with the inherent ambiguity induced by the bilinearity of the problem employ regularization techniques. A notable example of this can be found in Brainard and Freeman's Bayesian technique [1].

One of the advantages of the Bayesian approach to solving the color constancy problem is that it permits a straightforward integration of multiple sources of information. A drawback of any Bayesian method is the need to specify a prior distribution on the solution space, and the bias induced by the particular choice of prior. If more information, in the form of additional independent measurements, is provided, the influence of the prior is reduced. Studies of the human visual system reveal a possible source for this extra information. In looking at the structure of the retina, it is clear that the spectral sensitivity of the photoreceptors is not uniform across the retina. Because of the effects of the macular pigment, as well as the varying path length of light rays through the lens material, the foveal photoreceptors are less sensitive to blue wavelengths than the photoreceptors in the periphery. Despite this, the human perception of color is invariant to eye position. This means that we perceive the same color of an item both peripherally and foveally. One explanation for this invariance of color perception across the retina is that the visual system is applying a perceptual-stability constraint. A theory of how this could be accomplished was proposed by Clark and O'Regan [2]. They suggested that moving the eye results in a variation in the retinal signal. This variation, under the assumption that the world itself is not changing, allows for the generation of a signal for adaptation, which can be used to drive learning of invariance. Once invariance is learned, the system will produce the same percept of color no matter on which part of the retina the stimulus falls.

We incorporate the intuition provided by the Clark–O'Regan color stability theory into the Brainard and Freeman Bayesian technique. A non-uniform filter with spatially-varying spectral sensitivity can be used to represent the photoreceptors at different retinal locations. Measurements can be obtained sequentially by moving the gaze across a surface in the scene. These measurements can be used to update the inference resulting from our Bayesian formulation. Moreover, we propose to acquire multiple images of the same scene by moving the camera. This camera motion also allows for acquiring additional measurements which can be used to reduce the effect of the prior in the Bayesian formulation. A major advantage of our approach is that it does not require all surface

* Corresponding author.
E-mail addresses: sandra@cim.mcgill.ca (S. Skaff), arbel@cim.mcgill.ca (T. Arbel), clark@cim.mcgill.ca (J.J. Clark).

1077-3142/$ - see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.cviu.2009.03.014

patches in the scene to be viewed through the whole filter. In other words, some patches can be viewed through one portion of the filter but not another.

All color constancy approaches, including the proposed one, have a limited domain of applicability which is usually defined by a set of assumptions. In this paper, we also have a set of assumptions that are stated up front to make it simpler for the reader to follow through the derivations later on in this paper. First, it is assumed that the surfaces are Mondrian [3]. A Mondrian is a planar surface composed of several, overlapping, matte, patches named after the style of painting produced by the artist Piet Mondrian. Also, the surface patches are uniform and are uncorrelated from each other. Next, it is assumed that the sensor measurements obtained through the different portions of the filter are statistically independent from each other. In the case of obtaining measurements through different portions of the filter for the same surface, the independence assumption is still applied. In this case, one can think of the different parts of the surface viewed through different portions of the filter as different surfaces. The fact that they have the same underlying spectra does not necessarily imply that they are parts of the same surface. Moreover, it is assumed that each sensor measurement comprises three statistically independent measurements for the three sensor types: Red (R), Green (G), and Blue (B). Furthermore, the light illuminating a Mondrian scene is assumed to be locally constant; that is the spectral characteristics of the light vary slowly. Therefore, the spectrum of light falling on the lens does not vary with viewpoint for one surface patch, and any change in measurements after moving the sensors would be due to noise. However, it is assumed that these measurements are constant. Note that assuming constant illumination across a scene is a common practice in color constancy approaches [4]. Moreover, since the surface patches are flat in nature and they are placed on a flat surface, there are no interreflections between them. This means that a surface spectral reflectance and the illuminant spectrum vectors are sufficient statistics for the measurements of the corresponding single surface patch. Furthermore, it is assumed that segmentation and correspondence tasks have already been run on the images, and therefore we know which surface patches are viewed through which portions of the filter. Finally, note that while the assumptions stated above may impose practical limitations, we show in this paper that our approach can still be applied successfully to real image data.

This paper is organized as follows. First, color constancy approaches which use linear model representations for surface and illuminant spectra are summarized in Section 2. The Bayesian approach is detailed in this section as well. Next, the proposed approach is explained and derived in Section 3. For the purposes of testing the proposed approach on real data, a filter formulation of two spectral sensitivities is used. Section 4 shows experimental results on real data where improvement in illuminant and surface spectral estimation is attained by incorporating additional sensor measurements through filter portions of different spectral sensitivities. Section 5 compares the proposed approach to state-of-the-art approaches in the literature. The paper is concluded in Section 6.

2. Linear color constancy algorithms

Many color constancy algorithms lie in computing surface and illuminant spectra, which are represented by linear models. Such representations appear in Brill and West's work [5] on von Kries adaptation [6], which is the oldest algorithm for color constancy. These representations also appear in Buchsbaum's [7] and Gershon et al.'s work [8], who proposed variants of the gray world algorithm, which assumed that the average of the surface reflectances in the scene is gray. Since this assumption can be easily violated, these approaches do not perform well in practice. Linear model representations also appear in Maloney and Wandell [9,10] and Yuille's work [11], in which the assumption of having more photoreceptor responses than the number of surface and illuminant spectra basis functions is required. This is a serious limitation, especially in estimating surface spectra. It implies that no more than three basis functions can be used in the event that the rod photoreceptor response is used in addition to those of the cones. D'Zmura and Lennie [12] and D'Zmura and Iverson [13,14] also used linear model representations for surface and illuminant spectra. The authors in [13,14] proposed a model check algorithm to determine necessary and sufficient conditions for unique recovery of surface and illuminant spectra. Another approach is Ho et al.'s [15], which estimates surface and illuminant spectra using linear model representations from the color signal as opposed to the photoreceptor response. This signal, which is the product of the illuminant and surface spectra, needs to be measured using a spectroradiometer if the approach were to be applied on real data.

The issues in all the approaches mentioned above led us to base our work on Brainard and Freeman's approach [1], which estimates surface and illuminant spectral basis function weights from photoreceptor responses. This approach, explained in Section 2.1, imposes no restrictions on the number of surface spectra basis functions compared to the number of photoreceptor responses. The novelty of our approach lies in showing how surface and illuminant spectra can be used to inexpensively accumulate information obtained from the sensor through filter portions, of different spectral sensitivities, using the Bayesian framework.

2.1. Bayesian color constancy

Brainard and Freeman used a regularization technique to solve the problem of computing the values of the illuminant and surface reflectance spectra parameters [1]. The regularization technique used is based on Bayesian inference [16,17]. The advantage of using the Bayesian model is that it embeds all possible uncertainties as probability distributions, and that it performs inference based on all the information available in the data. For a given problem, the Bayesian approach makes use of *a priori* information about likely physical configurations of the solution. It is this prior information that helps in resolving ambiguities in the problem. Therefore, in this sense, the Bayesian approach can be thought of as a regularizer of the problem to obtain the solution in question. On the other hand, the drawbacks of using the Bayesian method should be taken into consideration. First, it may be difficult to specify a prior distribution. Also, in the event of maximizing the posterior, it may be difficult to specify a suitable cost function for the optimization process and, furthermore, this function may be difficult to maximize.

The surface and illuminant spectra in the Bayesian color constancy approach are parametrized using Maloney and Wandell's bilinear model [10]. The authors in [10] proposed an algorithm for computing the model parameters from sensor measurements of the light reflected from a set of surface patches. They start by describing the spectrum of light arriving at location x on an array of sensors by the function:

$$I^x(\lambda) = E(\lambda)S^x(\lambda), \tag{1}$$

where $E(\lambda)$ is the spectral power distribution of the ambient light in the scene, $S^x(\lambda)$ is the surface spectral reflectance, and λ denotes the wavelength. Assuming that there are p sensors at each location x, and the relative wavelength sensitivity of the k^{th} sensor is $R_k(\lambda)$, the response recorded at each location is given by

$$\rho_k^x = \sum_{\lambda=1}^{M} E(\lambda)S^x(\lambda)R_k(\lambda)d\lambda, \quad k = 1, 2, \ldots, p, \tag{2}$$

Please cite this article in press as: S. Skaff et al., A sequential Bayesian approach to color constancy using non-uniform filters, Comput. Vis. Image Understand. (2009), doi:10.1016/j.cviu.2009.03.014

where M denotes the dimensionality of the spectra. S and E are represented as linear models with n and m basis functions, respectively:

$$S^x(\lambda) = \sum_{j=1}^{n} \sigma_j^x S_j(\lambda), \tag{3}$$

$$E(\lambda) = \sum_{i=1}^{m} \epsilon_i E_i(\lambda). \tag{4}$$

In both cases, the basis functions are fixed and are assumed to be known. The basis functions for the surface reflectances, $S_j(\lambda)$, are computed using principal components analysis (PCA) on a set of 150 Munsell color chips. Munsell patches can be found in the Munsell Book of Color [18]. The basis functions of the light spectra $E_i(\lambda)$ are computed using PCA on a set of 622 different spectra of natural daylight. Therefore, to find the surface reflectance and ambient light, the basis function weights σ_j^x and ϵ_i should be computed. The sensor responses can be viewed as bilinear functions of the unknown basis function weights. This bilinearity means that various choices of ϵ and σ_x can produce identical sensor measurements and, therefore, that the problem of finding the weights is ill-posed.

Substituting Eqs. (3) and (4) into Eq. (2) yields the following equation:

$$\rho_k^x = \sum_{j=1}^{n} \sigma_j^x S_j(\lambda) \sum_{i=1}^{m} \epsilon_i E_i(\lambda) R_k(\lambda) d\lambda, \quad k = 1, 2, \ldots, p. \tag{5}$$

Maloney and Wandell assumed that a surface spectrum can be modeled by only two basis functions although this is a severe constraint. The authors require such an assumption because their algorithm requires more sensors than surface basis functions, and there are only three sensors: red, green, and blue. This assumption limits the results when the algorithm is presented with surfaces that cannot be represented accurately by only two basis functions. For example, Parkkinen et al. showed that Munsell patch spectra need as many as eight basis functions to achieve an accurate representation [19]. However, taking into account the fact that the first few principal components contain the most information in general, this assumption might not pose a problem in some cases.

Bayesian inference theory is comprised of three probability density functions: the prior, the posterior, and the likelihood. We denote the vector of the surface and illumination spectra model weights as $\vec{w}$, and the sensor responses as $\vec{y}$. We can obtain a statistical model for $\vec{w}$ by the conditional *posterior* density function of $\vec{w}$ given the measurement $\vec{y}$ as $p(\vec{w}|\vec{y})$ such that

$$p(\vec{w}|\vec{y}) = \frac{p(\vec{y}|\vec{w})p(\vec{w})}{p(\vec{y})} \propto p(\vec{y}|\vec{w})p(\vec{w}). \tag{6}$$

$p(\vec{y}|\vec{w})$ is the *likelihood* which models the relationship between the illuminant and surface spectra model weights and the sensor responses. $p(\vec{w})$ represents the *prior* information on the model parameters. $p(\vec{y})$ represents the probability of the sensor responses or the measurements and is a normalization term that does not affect the shape of the posterior distribution. In the Brainard–Freeman formulation, the prior is represented by a normal distribution on the weights of the illuminant and surface reflectance spectra. These weights are computed by projecting these illuminant and surface reflectance spectra vectors on the illuminant and surface reflectance spectra basis functions. The illuminant data set is composed of daylight spectra with temperatures ranging from 3000 to 25,000 K, while the surface data set is composed of Munsell reflectance spectra. The likelihood $p(\vec{y}|\vec{w})$ is also represented by a normal distribution. Given the posterior $p(\vec{w}|\vec{y})$, they compute a loss function which they call the Bayesian expected loss:

$$\bar{L}(\tilde{\vec{w}}|\vec{y}) = \int_{\vec{w}} \bar{L}(\tilde{\vec{w}}|\vec{w}) p(\vec{w}|\vec{y}) dw. \tag{7}$$

This function computes the penalty for choosing a single estimate $\tilde{\vec{w}}$ when the actual parameters are $\vec{w}$ [16]. Brainard and Freeman choose an estimate for $\vec{w}$ such that the loss is minimal. They discuss three types of loss functions: the Maximum A Posteriori (MAP), the Minimum Mean Squared Error (MMSE), and the Maximum Local Mass (MLM). For simplicity in this paper, we choose to use the MAP function.

3. The sequential non-uniform filter formulation

We introduce a technique that builds upon Brainard and Freeman's Bayesian approach in which only one sensor response is acquired. Our technique acquires sensor measurements through different filter portions, each having its own spectral sensitivity. Therefore, the inherent ill-posedness of the problem is addressed through the introduction of more sources of information. This acquisition of measurements is similar to when a person moves his/her gaze across a scene and is inspired by Clark and O'Regan's theory of *color stability* [2]. This theory states that the human's perception of color is invariant to eye position. The acquisition of information is sequential and can be modeled through a sequential Bayesian estimation process. Note that this process is computationally feasible as its complexity lies in the multiple computations of the probability density functions of the Bayesian formulation.

We had shown in simulation in [20] that there is improvement in the surface patch spectra estimates with the placement of a non-uniform filter on the camera lens. Two types of filters were considered: one with two portions, each having a different spectral sensitivity characteristic, and one with more than two portions. The latter formulation was chosen to be Gaussian as it mimics the variation in the spectral sensitivity of the photoreceptors across the retina. This Gaussian formulation could be constructed with a drop of food coloring placed on transparent glass. Therefore, this filter would have spectral transmittance that is maximum in the center and that decreases gradually toward the peripheries. However, constructing such a filter is beyond the scope of this paper. Therefore, in order to show how acquiring additional measurements improves on spectral estimation, we use a filter of two portions with two corresponding spectral sensitivities here.

As in the Brainard and Freeman approach, evidence for the lighting and surface color parameters of either one or many surface patches in a scene is represented by a conditional probability density function given the sensor measurements. Since there are two distinct portions of the filter, X and Y, this probabilistic evidence is then accumulated sequentially over X and Y. In practice, this formulation can be modeled by placing a filter with the appropriate absorption characteristic onto half of the camera lens. We shall refer to this model as a binary filter, which has one transparent part and one filter part. Therefore, evidence can be accumulated over the pixels of the image of the scene. The three sensor responses obtained through one portion of the filter will be denoted as *RGB*. In practice, the *RGB* response (or measurement) for a single patch would be the average of all its pixel sensor responses. We shall start with the derivations for the single surface patch case before developing the case for more surface patches. Finally, we discuss how the case of a two portion filter, or binary filter, can be straight-forwardly generalized to the case of a multiple portion filter.

3.1. A single patch

Let us consider a simple case where there is only one surface patch in the scene, illuminated by a single light source. For this surface patch, the *RGB* sensor responses or measurements through filter portion X will be denoted as RGB_{X1}, and those through filter

portion Y will be denoted as RGB_{Y1}. The surface spectral model weights vector is denoted by a_1 and the illuminant spectral model weights vector by b. Suppose that the patch is visible through both filter portions as shown in Fig. 1 (a_1). Note that the mentioned figure contains three-surface patches; however, it is assumed that only patch 1 is in the scene for now for illustration purposes.

To accumulate probabilistic evidence of the scene patch and illuminant spectra over different filter portions, we derive the conditional posterior density function, $p(a_1, b|\{RGB\})$, for the parameters a_1 and b given the set of measurements of the entire scene, $\{RGB_1\} \triangleq RGB_{X1}, RGB_{Y1}$. Using Bayes' rule:

$$p(a_1, b|\{RGB_1\}) \propto p(\{RGB_1\}|a_1, b)p(a_1, b). \quad (8)$$

Therefore,

$$p(\{RGB_1\}|a_1, b) = p(RGB_{X1}, RGB_{Y1}|a_1, b) = p(RGB_{X1}|a_1, b)p(RGB_{Y1}|a_1, b), \quad (9)$$

by the statistical independence of the measurements assumption (explained in Section 1). Substituting $p(\{RGB_1\}|a_1, b)$ into Eq. (8), we obtain

$$p(a_1, b|\{RGB_1\}) \propto p(RGB_{X1}|a_1, b)p(RGB_{Y1}|a_1, b)p(a_1, b). \quad (10)$$

By Bayes' rule, we can state

$$p(a_1, b|RGB_{X1}) \propto p(RGB_{X1}|a_1, b)p(a_1, b). \quad (11)$$

Therefore, by substituting $p(a_1, b|RGB_{X1})$ from Eq. (11) into Eq. (10), we get

$$p(a_1, b|\{RGB_1\}) \propto p(a_1, b|RGB_{X1})p(RGB_{Y1}|a_1, b). \quad (12)$$

From Eq. (12) we can conclude that the posterior of the parameters a_1 and b is the product of the posterior for one filter portion (in this case X), $p(a_1, b|RGB_{X1})$, and the likelihood of the RGB measurement from the other filter portion (in this case Y), $p(RGB_{Y1}|a_1, b)$. Therefore, the extension to Brainard and Freeman's approach is the additional likelihood term (in this case $p(RGB_{Y1}|a_1, b)$) resulting from the measurement (in this case RGB_{Y1}) from the tinted portion of the filter. This gives more information used to compute the unknown basis function weights thus resulting in better estimates. Note that this formulation assumes X and Y are interchangeable.

3.2. Multiple patches

We now wish to describe the more complex case where it is assumed that there is more than one surface patch in the scene, illuminated by a single light source. RGB measurements for surface patch number n obtained from filter portions X and Y will be denoted as RGB_{Xn} and RGB_{Yn}, respectively. The surface spectral model weights vector for patch n is denoted by a_n and the illuminant spectral model weights vector by b. Two cases are considered in this situation. The first one is when the image is acquired once, that is when the camera is static. The second one is when the image of the scene is acquired several times, from different positions, that is when the camera is moved.

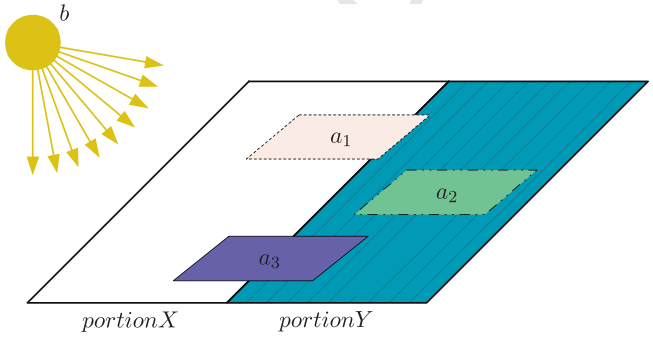


Fig. 1. *Static camera.* Three patches in the scene, illuminated by a single light source viewed by a sensor through a two portion or binary filter, which has one transparent part and one part tinted blue–green (represented with the hashed part), for example. a_1, a_2, a_3 and b are the weights in the spectral linear models for patches 1, 2, 3 and the illuminant, respectively. (For interpretation of references to color in this figure legend, the reader is referred to the web version of this article.)

3.2.1. Static camera

Consider a scene with three-surface patches. Suppose that surface patches 1 and 3 are visible through both portions of the filter, while surface patch 2 is visible through filter portion Y alone, as shown in Fig. 1.

Let $\{RGB\}$ denote the total set of measurements of the scene:

$$\{RGB\} \triangleq RGB_{X1}, RGB_{X3}, RGB_{Y1}, RGB_{Y2}, RGB_{Y3}. \quad (13)$$

The conditional posterior density function for the parameters, a_1, a_2, a_3 and b, given the set of measurements of the scene, $\{RGB\}$, is denoted by $p(a_1, a_2, a_3, b|\{RGB\})$. First, using Bayes' rule:

$$p(a_1, a_2, a_3, b|\{RGB\}) \propto p(\{RGB\}|a_1, a_2, a_3, b)p(a_1, a_2, a_3, b). \quad (14)$$

The no interreflection assumption implies that the prior probabilities for the surface reflectance weights a_n are statistically independent of each other and of the spectral function weights of the illuminant, b:

$$p(a_1, a_2, a_3, b) = p(a_1, a_3, b)p(a_2). \quad (15)$$

Consequently, substituting $p(a_1, a_2, a_3, b)$ from Eq. (15) into Eq. (14), it becomes

$$p(a_1, a_2, a_3, b|\{RGB\}) \propto p(\{RGB\}|a_1, a_2, a_3, b)p(a_1, a_3, b)p(a_2). \quad (16)$$

The statistical independence of the measurements assumption (explained in Section 1) implies

$$p(\{RGB\}) = p(RGB_{X1}, RGB_{X3}, RGB_{Y1}, RGB_{Y2}, RGB_{Y3}) = p(RGB_{X1}, RGB_{X3})p(RGB_{Y1})p(RGB_{Y2})p(RGB_{Y3}).$$

Therefore, the likelihood in Eq. (14) can be expressed as

$$p(\{RGB\}|a_1, a_2, a_3, b) = p(RGB_{X1}, RGB_{X3}, RGB_{Y1}, RGB_{Y2}, RGB_{Y3}|a_1, a_2, a_3, b) = p(RGB_{X1}, RGB_{X3}|a_1, a_2, a_3, b)p(RGB_{Y1}|a_1, a_2, a_3, b) \times p(RGB_{Y2}|a_1, a_2, a_3, b)p(RGB_{Y3}|a_1, a_2, a_3, b).$$

By the no interreflection assumption, the surface spectra model weights of patches 1 and 3, for example, do not affect the sensor measurements of surface patch 2. Therefore,

$$p(RGB_{Y2}|a_1, a_2, a_3, b) = p(RGB_{Y2}|a_2, b). \quad (17)$$

Applying the same assumption to the scene in Fig. 1, we can write

$$p(RGB_{X1}, RGB_{X3}|a_1, a_2, a_3, b) = p(RGB_{X1}, RGB_{X3}|a_1, a_3, b),$$
$$p(RGB_{Y1}|a_1, a_2, a_3, b) = p(RGB_{Y1}|a_1, b),$$
$$p(RGB_{Y3}|a_1, a_2, a_3, b) = p(RGB_{Y3}|a_3, b).$$

Consequently,

$$p(\{RGB\}|a_1, a_2, a_3, b) = p(RGB_{X1}, RGB_{X3}|a_1, a_3, b)p(RGB_{Y1}|a_1, b)p(RGB_{Y2}|a_2, b)p(RGB_{Y3}|a_3, b).$$

Substituting the latter equation of $p(\{RGB\}|a_1, a_2, a_3, b)$ into Eq. (16) we get

$$p(a_1, a_2, a_3, b|\{RGB\}) \propto p(RGB_{X1}, RGB_{X3}|a_1, a_3, b)p(RGB_{Y1}|a_1, b) \times p(RGB_{Y2}|a_2, b)p(RGB_{Y3}|a_3, b)p(a_1, a_3, b)p(a_2).$$

Applying Bayes' rule for the first term and the second to last we get

$$p(a_1, a_3, b|RGB_{X1}, RGB_{X3}) \propto p(RGB_{X1}, RGB_{X3}|a_1, a_3, b)p(a_1, a_3, b). \quad (18)$$

Therefore,

$$p(a_1,a_2,a_3,b|\{RGB\}) \propto p(a_1,a_3,b|RGB_{X1},RGB_{X3})p(RGB_{Y1}|a_1,b) \times p(RGB_{Y2}|a_2,b)p(RGB_{Y3}|a_3,b)p(a_2). \quad (19)$$

From Eq. (19), we can conclude that the posterior for the entire scene $p(a_1,a_2,a_3,b|\{RGB\})$ is a function of the posterior for one filter portion (X), the likelihood of the RGB measurements taken through the other filter portion (Y) and the prior of the spectral function weights of the surface patches viewed through only the latter portion ($p(a_2)$). In turn, the posterior for filter portion X can be expressed as

$$\begin{aligned} &p(a_1,a_3,b|RGB_{X1},RGB_{X3}) \\ &\propto p(RGB_{X1},RGB_{X3}|a_1,a_3,b)p(a_1,a_3,b)(\text{by Bayes' rule}) \\ &= p(RGB_{X1},RGB_{X3}|a_1,a_3,b)p(a_1,b)p(a_3) \\ &\quad \times (\text{by the no interreflection assumption}) \\ &= p(RGB_{X1}|a_1,a_3,b)p(RGB_{X3}|a_1,a_3,b)p(a_1,b)p(a_3) \\ &\quad \times (\text{by the statistical independence of the measurements assumption}) \\ &= p(RGB_{X1}|a_1,b)p(RGB_{X3}|a_3,b)p(a_1,b)p(a_3) \\ &\quad \times (\text{by the no interreflection assumption}) \\ &\propto p(a_1,b|RGB_{X1})p(RGB_{X3}|a_3,b)p(a_3)(\text{by Bayes' rule}). \end{aligned} \quad (20)$$

Note that this posterior is a function of the posterior for surface patch 1 and the likelihood and prior for surface patch 3. The sequential nature of the strategy can once again be observed here as for each filter portion: the posterior for each surface patch acts as a prior for the next surface patch. We would like to add that there are a number of different ways to write the posterior, depending on the order of filter portions considered. However, all these ways are equivalent as they give rise to the same value of the posterior function. Once again, note in Eq. (19) the likelihoods resulting from the additional measurements for surface patches 1 and 3 (RGB_{Y1} and RGB_{Y3}) obtained from the tinted filter portion Y. This addition to the formulation results in an improvement in the estimates of the parameters a_1, a_3, and b.

3.2.2. Moving camera

Humans constantly move their eyes thus accumulating evidence to acquire more knowledge of the world. This fact motivated us to move the camera to acquire multiple images of the same scene, thus placing the surface patches of the scene in different positions with respect to the portions of the filter. This allows for accumulating more measurements for some surface patches in the case of a binary filter. In the following sections, we study two cases: one in which the camera is moved once and one in which the camera is moved twice.

3.2.2.1. One move. We begin by examining the case where the camera is moved once such that two images of the same scene are taken from two different positions. This places the surface patches in different locations with respect to the filter as shown in the first move (equivalently the first two snapshots) in Fig. 2. The posteriors, two in this case, are derived in a similar way as in the previous section. The first posterior is P_1, which is that obtained from the first image after gathering the first set of measurements. The second is P_2, which is that obtained from the second image after the camera is moved and thus more measurements are gathered. We will demonstrate the sequential nature of this probabilistic inference once again in this section.

At the initial stage, that is before the camera is moved, the total set of measurements denoted by $\{RGB\}$ is given by

$$\{RGB\} \triangleq RGB_{X1},RGB_{X2},RGB_{X3},RGB_{Y1}. \quad (21)$$

The posterior can be derived given the set of measurements $\{RGB\}$ from the scene in a similar way to that of Eq. (19) (in Section 3.2.1):

$$\begin{aligned} P_1 &= p(a_1,a_2,a_3,b|\{RGB\}) \\ &\propto p(a_1,a_2,a_3,b|RGB_{X1},RGB_{X2},RGB_{X3})p(RGB_{Y1}|a_1,b). \end{aligned} \quad (22)$$

The next step involves the movement of the camera and gathering a new set of RGB measurements which is given by

$$\{RGB\} \triangleq RGB_{X1},RGB_{X2},RGB_{X3},RGB_{Y1},RGB_{Y2}. \quad (23)$$

We will show how P_1 acts as a prior for P_2, which can now be computed given the new RGB measurements:

$$\begin{aligned} P_2 &= p(a_1,a_2,a_3,b|\{RGB\}) \\ &\propto p(\{RGB\}|a_1,a_2,a_3,b)p(a_1,a_2,a_3,b) \\ &= p(RGB_{X1},RGB_{X2},RGB_{X3},RGB_{Y1},RGB_{Y2}|a_1,a_2,a_3,b)p(a_1,a_2,a_3,b) \\ &\quad \times (\text{by Bayes' rule}) \\ &= p(RGB_{X1},RGB_{X2},RGB_{X3}|a_1,a_2,a_3,b)p(RGB_{Y1},RGB_{Y2}|a_1,a_2,a_3,b) \\ &\quad \times p(a_1,a_2,a_3,b)(\text{by the statistical independence of the measurements assumption}) \\ &= p(RGB_{X1},RGB_{X2},RGB_{X3}|a_1,a_2,a_3,b)p(RGB_{Y1}|a_1,a_2,a_3,b) \\ &\quad \times p(RGB_{Y2}|a_1,a_2,a_3,b)p(a_1,a_2,a_3,b)(\text{by the statistical independence of the measurements assumption}) \\ &= p(RGB_{X1},RGB_{X2},RGB_{X3}|a_1,a_2,a_3,b)p(RGB_{Y1}|a_1,b)p(RGB_{Y2}|a_2,b) \\ &\quad \times p(a_1,a_2,a_3,b)(\text{by the no interreflection assumption}) \\ &\propto p(a_1,a_2,a_3,b|RGB_{X1},RGB_{X2},RGB_{X3})p(RGB_{Y1}|a_1,b)p(RGB_{Y2}|a_2,b) \\ &\quad \times (\text{by Bayes' rule}) = P_1 p(RGB_{Y2}|a_2,b). \end{aligned} \quad (24)$$

Therefore, we conclude once again that the posterior computation is sequential and that it is a function of the posterior obtained from the first image and the likelihood obtained from the second image.

3.2.2.2. Two moves. We then consider the case where the camera is moved twice such that three images of the scene are taken from three different positions (Fig. 2). This places the surface patches in three different locations, as opposed to two in Section 3.2.2.1, with respect to the filter. Therefore, more measurements for a patch may be acquired depending on the filter positions. The measurements in this case are the ones used for computing P_2 as well as an additional measurement obtained for surface patch 3 when viewed through portion Y of the filter. In this case, three posteriors are obtained: one from the set of data of each image.

For simplicity, we assume that the first move is similar to the first move in the previous section. This means that the first two images captured in this case are the same as the two images captured in the previous case. Therefore, P_1 and P_2 can be derived in a similar way and using the same assumptions as in Eqs. (22) and (24). As explained in the previous section, P_1 acts as a prior for P_2. In this section the camera is moved a second time, and a new set of data is gathered to obtain the posterior P_3. Therefore, a new set of RGB measurements is obtained in addition to the old ones that were obtained from the two initial movements. This new set of RGB measurements is given by the following equation:

$$\{RGB\} \triangleq RGB_{X1},RGB_{X2},RGB_{X3},RGB_{Y1},RGB_{Y2},RGB_{Y3}. \quad (25)$$

P_2 acts as a prior for P_3:

$$P_3 = p(a_1,a_2,a_3,b|\{RGB\}) \propto P_2 p(RGB_{Y3}|a_3,b). \quad (26)$$

In the same way as information is accumulated over different filter portions, information is accumulated over scenes, such that the posterior for each image acts as a prior for the set of data of the next image. The posteriors obtained from images due to additional moves can be derived in a similar way.

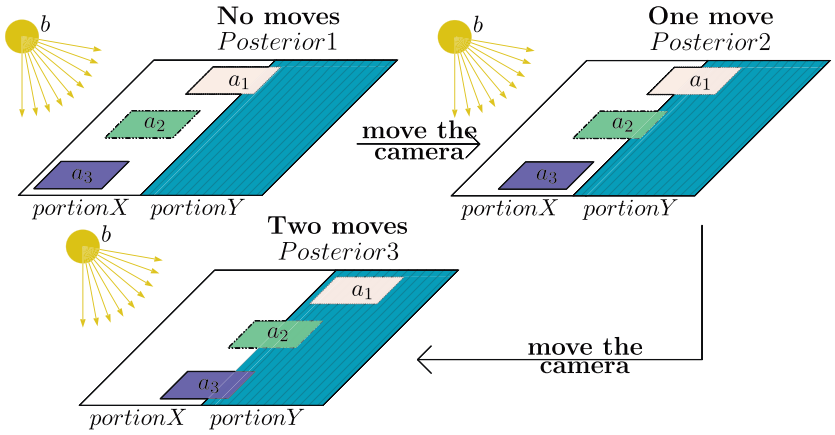


Fig. 2. *Moving camera.* Three patches in the scene, illuminated by a single light source viewed by a sensor through a two portion or binary filter, which has one transparent part and one part tinted blue-green (represented with the hashed part), for example. The camera is moved twice. a_1, a_2, a_3 and b are the weights in the spectral linear models for patches 1, 2, 3 and the illuminant, respectively. (For interpretation of the references to colours in this figure legend, the reader is referred to the web version of this paper.)

In short, the active Bayesian formulation is, in fact, *sequential*, where the posterior for each image acts as a prior for the next image, and the posterior for each filter portion acts as a prior for the next filter portion. This is similar to the sequential nature of the "Static Camera" case (in Section 3.2.1), where the posterior for each surface patch acts as a prior for the next surface patch. The addition of more surface patches in the scene provides more information regarding the illumination and thus, in turn, about the surface patch spectra themselves. In addition, moving the camera provides more information about the scene as the surface patches are viewed from different locations with respect to the filter, and thus many images of the same scene are obtained. This allows for acquiring multiple measurements for the same surface patch and thus improves on parameter estimation.

3.3. Generalization: two to multiple portion filters

We illustrated the sequential nature of the proposed active Bayesian formulation for the case of two portion filters. As shown above, the posterior for each filter portion acts as a prior for the next filter portion. This hypothesis can be generalized to the case of more than two portion filters in the scene. Consider, for example, the "Static Camera" case when there is a filter of three portions, taken in the order X, Y, Z, and when there are three-surface patches in the scene. The scene in this example looks similar to that in Fig. 1 except that patch 2 is viewed through filter portion *Y and Z* instead of only filter portion *Y*. Patches 1 and 3 are viewed through filter portions *X* and *Y* as in Fig. 1. Call P_3 the conditional posterior function for all three portions of the multiple filter for the parameters of the surfaces and the light source given the set of measurements of the scene. P_3 would be a function of the posterior for the first two portions *X* and *Y* of the filter, (call it P_2), the likelihood of the measurements taken through the third portion of the filter (Z), and the prior of the spectral function weights of the surface patches viewed through only portion *Z* of the filter (none in this case):

$$P_3 = P_2 p(RGB_{Z2}|a_2, b). \tag{27}$$

In turn, P_2 would be a function of the posterior for the first filter portion (X), the likelihood of the measurements taken through the second filter portion (Y), and the prior of the spectral function weights viewed through only filter portion *Y*:

$$P_2 = P_1 p(RGB_{Y1}|a_1, b) p(RGB_{Y2}|a_2, b) p(RGB_{Y3}|a_3, b) p(a_2). \tag{28}$$

Therefore the derivation of the posterior can be straight-forwardly extended from incorporating measurements taken through two filter portions to incorporating measurements taken through three filter portions. This extension can also be applied to the "Moving Camera" case (Section 3.2.2).

4. Experiments

The main purpose of the experiments performed is to show that, despite the assumptions imposed, additional measurements obtained for a patch viewed through a filter do improve surface and illuminant spectral estimates. Constructing a filter with multiple portions of different spectral sensitivities is beyond the scope of this work as explained in Section 3. Therefore we choose to illustrate the performance of the approach for the case of two portion filters. Note that we had shown in simulation in [20] that improvement in spectral estimation can be obtained with the introduction of filters with more than two portions into the scene as well.

In this work, we refer to a measurement system with two portions, transparent and filtered, as a binary filter. The optical filter chosen for the filtered portion should be such that there are significant differences between the responses from the transparent portion and the filtered portion in the long-, medium-, and short-wavelength range channels. In addition, the filter should not completely block the light from the wavelengths that a given photoreceptor is sensitive to. For example, a "brick-wall" yellow optical filter should not be employed, as it would completely block the portion of the spectrum responded to by the short-wavelength channel while passing through unchanged the portions of the spectrum responded to by the long- and medium-wavelength channels. Taking into account these issues, we choose to use an Edmund Optics BG 18 bandpass filter of spectral transmittance peaking in the blue-green wavelength region. This filter shifts the spectral sensitivity of the long- and short-wavelength sensors towards medium wavelengths, and narrows the bandwidth of the medium-wavelength sensor. Thus, it is expected that the binary filter measurement system should provide better spectral modeling at medium wavelengths than the non-filter measurement system.

Before moving on to describing the implementation and testing specifics of the algorithm and the results obtained, we would like to note that some previous color constancy approaches investigated the idea of using multiple measurements for the same surface. For example, Tsukada and Ohta [21], who employed linear model representations for surface and illuminant spectra, used two illuminants to capture two images of the same scene. While

their approach may seem similar to the proposed one, the authors in [21] employed more unknowns as the number of basis function illuminant spectra is doubled with the use of two illuminants. Also, D'Zmura and Iverson [13,14,22] studied the case where multiple light sources were used to capture multiple images of the same scene. However their approach suffers from many limitations as is discussed in Section 2.

Within the color correction approaches that use 3D color space representations such as *RGB*, it has been shown how illumination variation in an image can be used as an additional constraint to improve on color constancy [23,24]. Finlayson et al. introduced the Chromagenic color constancy approach [25,26], where a filtered and a non-filtered image are obtained for each scene. As mentioned in [25], the additional measurements are not used to increment the number of knowns to facilitate the estimation of the model parameters as is the case in the proposed algorithm. These measurements are used to compute the relationship between the filtered and non-filtered *RGB*'s. This relationship is used to select the scene illuminant, and given the illuminant, color correction can be performed. The Chromagenic color constancy approach is described for the case of using two measurements for each surface. However, in the case of three or more measurements per surface, the calculations can become tedious, and it would be more difficult to compute the relationship between the non-filtered and the filtered *RGB*'s at once. On the other hand, the proposed spectral based approach can handle the incorporation of additional measurements in a more natural and inexpensive way using Bayesian techniques as described in the previous section. Moreover, our approach allows for flexibility in the sense that if only non-filtered *RGB* measurements are available for some surface patches, these can still be used in the posterior distribution. This is not the case in Chromagenic color constancy, where each surface patch should have two sets of measurements in order for the formulation of the problem to be complete.

4.1. Experimental setup

We captured images of the pages in the Munsell Book of Color [18] with a Panasonic WV-CP410 camera. All the images were obtained in our laboratory at night to ensure that the only illumination was that of the scene light source. We obtained a 50×50 pixel sample from each of the Munsell patches using a semi-automated segmentation algorithm. Since we assume Mondrian scenes where the illumination is locally constant, we averaged the *RGB* responses of all the pixels in a segmented patch to obtain one *RGB* response per patch. The same steps were carried out in the case of the filtered portion of the camera lens to obtain *RGB* responses for the patches viewed through a filter. Since we assume flat scenes where there are no interreflections, the average responses corresponding to the patches in the scene, in the filter and no filter cases, were inputted into our algorithm.

We start by explaining how the scenes were constructed. Ten surface patches, which we shall refer to as main patches, were chosen at random. Ten groups of two patches each were then selected randomly, and these groups were used to construct 10 three-surface patch scenes with the main patch repeated in all 10 scenes. Since there were 10 main patches, a total of 100 scenes was constructed. The selection of scenes was done in this fashion so as to allow more robust evaluation of spectral estimation for the main patch which is repeated in 10 scenes. All the constructed scenes comprised of three randomly chosen *matte* Munsell surface patches [18] illuminated by a single 60 W tungsten light source at temperature 2800 K. The spectra of these patches were measured by Parkkinen and Silfsten [27]. The database of light source spectra comprised of a set of tungsten light spectra at temperatures ranging from 2600 to 3500 K, in steps of 100 K, and was obtained from the IES lighting handbook [28]. The spectral sensitivities of the camera sensor were obtained from the manufacturer. The wavelengths considered are over the visible range, 430–700 nm, which is discretized into 5 nm intervals. Note that the choice of surface and illuminant spectral databases would not make a difference in the performance of the approach as the basis functions can be updated accordingly to be the principal components of the new spectral databases. The performance of the approach, however, depends on how well the basis functions represent the spectral databases for each of the surfaces and the light source.

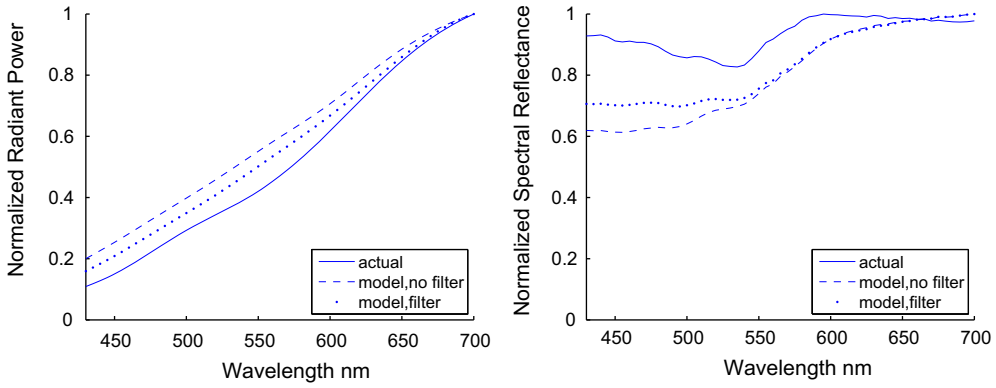
For the Bayesian formulations, the likelihood functions $p(\{RGB\}|a_1, a_2, a_3, b)$ are computed using the sensor measurements as well as the corresponding model predictions, which are given in Eq. (5). We describe what the sensor measurements are for each case studied in the next section. The standard deviation of the likelihood is an approximation of the standard deviation of the measurement noise. The latter standard deviation is computed by capturing 50 snapshots of each of the Munsell patches with a 100 ms delay between snapshots. Then, the standard deviation of the noise corresponding to each patch sensor response is calculated using the 50 snapshots of the patch. Finally, the standard deviations of the sensor response noise of all the patches is averaged to obtain the likelihood standard deviation. While this may be an oversimplified way of computing the standard deviation, it serves the purpose and can be reproduced fairly easily upon changing the database of surface patches. Note that the same standard deviation as computed above is used when modeling the likelihood in the case of a filter. This means that it is assumed that the standard deviation of the noise is approximately the same in this case. This assumption is made for simplification purposes as it avoids the need of estimating standard deviations for different filters. The illuminant spectrum $E(\lambda)$ is modeled with five basis functions, while the surface spectrum $S^x(\lambda)$ is modeled with eight basis functions. The weights for these spectra are assumed to have Gaussian distributions. The basis functions for the surface spectrum model are represented by the principal components of the spectra of the 1269 Munsell patches provided by Parkkinen and Silfsten [27]. The basis functions for the light source spectrum model are represented by the principal components of the spectra of the 10 tungsten light spectra [28].

The prior distributions for the spectral model parameters $p(a_1, a_2, a_3, b)$ are assumed to be independent and are modeled by Gaussian distributions. The means and variances of the prior functions are computed from the distribution of weights corresponding to the 1269 Munsell patch spectra and the 10 tungsten light spectra used. These weights are obtained by projecting the measured spectra onto the basis function sets.

The location of the mode of the posterior distribution is estimated by a standard MATLAB optimization package, *lsqnonlin*, resulting in a set of estimated surface and light source spectra weight vectors. Note that any package can be used provided that the negative of the cost function can be minimized. In order to eliminate the chances of obtaining a set of weights corresponding to a local minimum, the optimization is run with 10 different randomly chosen starting points. In most cases, the optimization terminates at the same cost function value. In the occasional cases when it does not, we choose the solution weight vector corresponding to the smallest value.

4.2. Experimental results

First, we compare the spectral estimates obtained in the binary filter case to those obtained in the no filter case, which is equivalent to Brainard and Freeman's approach. Next, we show how moving the camera, once and then twice, to acquire additional



(a) Tungsten Illuminant at 2800K (b) Munsell Patch 1203

Fig. 3. The estimated and actual spectra when there are multiple surface patches in the scene in the cases of no filter and binary filter for each of the (a) tungsten illuminant at 2800 K, RMSE = 0.0930 (no filter), RMSE = 0.0533 (filter) and (b) Munsell patch 1203, RMSE = 0.1705 (no filter), RMSE = 0.1282 (filter).

measurements for a surface patch improves spectral estimation. Experiments with three-surface patches in the scene, illuminated by a single light source (as described in Section 3) are performed. The resulting estimates for the model and the actual spectra for the surface patches and the light sources are plotted and the root mean square (RMS) errors are computed in each of the two cases below. We would like to note that, to the best of our knowledge, this is the first time that Brainard and Freeman's approach is applied on real data.

4.2.1. Static camera

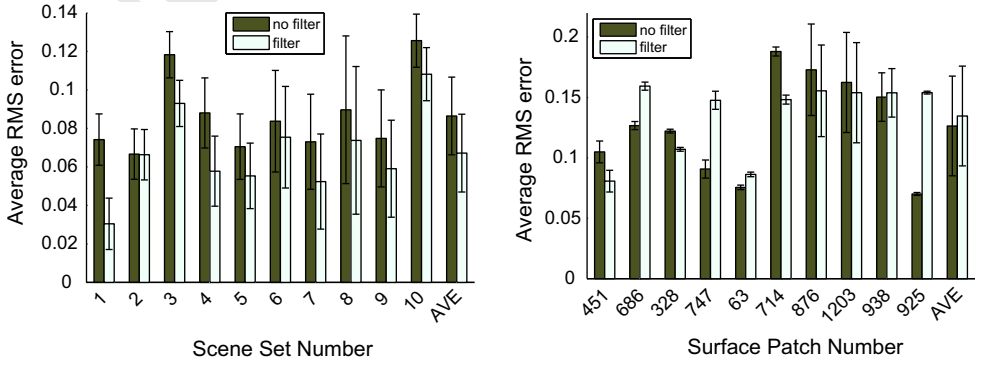
We compare the performance of the approach in the no filter case to that of the binary filter case. In the no filter case, one *RGB* response is obtained for each surface patch in the scene. In the binary filter case, two *RGB* responses corresponding to the different portions of the filter are obtained for each of surface patches 1 and 3. One *RGB* response corresponding to filter portion *Y* is obtained for surface patch 2 (see Fig. 1). These sensor responses are used in computing the likelihood functions, which in turn are used in the computation of the posteriors as described in Section 3.2.1. The resulting estimates for the model and the actual spectra for the tungsten light source and a surface patch from one of the scenes can be found in Fig. 3. Note that RMS error is indicated as RMSE in the corresponding plots. The plots show that the model spectra are closer to the actual spectra in the binary filter case than in the no filter case.

Moreover, we depict in Fig. 4 the average over 10 scenes of the RMS errors between the model and actual spectra for the illuminant and the main surface patches, which are repeated in a set of 10 scenes. We also depict the average of all the average RMS errors in the last two bars for each plot. AVE denotes this average. From these plots we observe that there is always an improvement in the illuminant spectral estimates. Theoretically, this should lead to an improvement in the surface spectral estimates. Despite the fact that this is not always the case as we can see in Fig. 4b, there is either no improvement or an improvement in 50% of the cases.

To get a better understanding of the comparative performance of the approach in the no filter and filter cases, the paired *T*-test [29] is run on the average RMS errors for the 10 scene sets (depicted in Fig. 4) for a 95% confidence level. Note that this test assumes that the differences between the average RMS errors have a normal distribution. As a result of the test, the improvement in performance for the binary filter case over the no filter case for the illuminant spectral estimates is statistically significant ($t = 4.848$, $p < 0.05$). As for the surface spectral estimates, the difference in performance between the no filter and the binary filter case is not statistically significant ($t = -0.639$, $p < 0.05$). Therefore we cannot conclude from the latter results that the average performance for the no filter case is better than that for the binary filter case as implied by the last two bars in Fig. 4b.

4.2.2. Moving camera

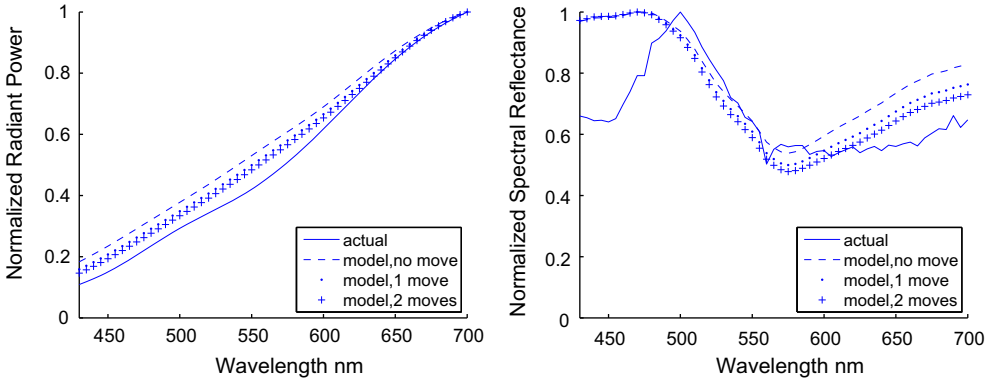
We show results for the moving camera case, as described in Section 3.2.2. There are three-surface patches in the scene with a filter of two portions, represented by a binary filter on the lens having a transparent part and a tinted blue–green part. The camera is moved: once and then twice. Before moving the camera, the sensor responses obtained are two for surface patch 1, corresponding to



(a) Tungsten Illuminant at 2800K (b) Munsell Patches

Fig. 4. Average RMS errors for the illuminant and 10 surface patch spectra for the no filter and binary filter cases when there are multiple surface patches in the scene. Each error is the average RMS error over 10 scenes chosen at random for the spectrum of the (a) tungsten illuminant at 2800 K and (b) repeated Munsell patch in each 10 scenes.

Please cite this article in press as: S. Skaff et al., A sequential Bayesian approach to color constancy using non-uniform filters, Comput. Vis. Image Understand. (2009), doi:10.1016/j.cviu.2009.03.014



(a) Tungsten Illuminant at 2800K (b) Munsell Patch 686

Fig. 5. The estimated and actual spectra when there are multiple surface patches in the scene in the binary filter case when the camera is moved twice for each of the (a) tungsten illuminant at 2800 K, RMSE = 0.0762 (no motion), RMSE = 0.0519 (1 move), RMSE = 0.0399 (2 moves) and (b) Munsell patch 686, RMSE = 0.1621 (no motion), RMSE = 0.1464 (1 move), RMSE = 0.1428 (2 moves).

each of the filter portions X and Y, and one for each of surface patches 2 and 3, corresponding to filter portion Y. The sensor responses obtained after the camera is moved the *first* time are the same as the previous ones, with an additional sensor response obtained for surface patch 2 corresponding to filter portion Y (see Fig. 2). The sensor responses obtained after the camera is moved the *second* time are the same as the previous ones, with an additional sensor response obtained for surface patch 3 corresponding to filter portion Y. These sensor responses are used in computing the likelihood functions corresponding to each of the one move and two moves cases. The posterior functions are then computed as described in Section 3.2.2. The hypothesis is that the estimates will improve with each move for the patches for which additional measurements are obtained upon moving. Fig. 5 shows the model and the actual spectra as well as the RMS errors for both the light source and one of the three-surface patches in a scene respectively. Notice how the RMS errors in this figure are reduced by moving.

Moreover, we depict the average over 10 scenes of the RMS errors between the model and actual spectra for the illuminant and the main surface patches, which are repeated in a set of 10 scenes as seen in Fig. 6. We also depict the average of all the average RMS errors in the last three bars for each plot. We can see that the average RMS errors are reduced with every camera move for the illuminant. This should imply decreasing RMS errors with every move for the surface patches as well, which is evident in 80% of the cases.

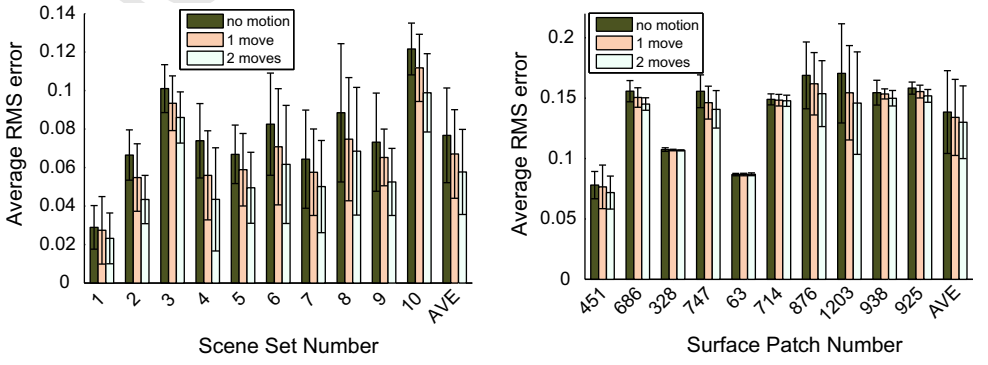
To get a better understanding of the comparative performance of the approach in the no move, one move, and two moves cases, the paired T-test [29] is run on the average RMS errors for the 10 scene sets (depicted in Fig. 6) for a 95% confidence level. The two pairs for which the average RMS errors are employed are: (i) the no motion and one move case and (ii) the one move and two moves case. As a result of this test, we conclude that the improvement in the illuminant spectral estimates is statistically significant upon moving the camera once versus no motion ($t = 6.532$, $p < 0.05$), and upon moving the camera twice versus moving once ($t = 9.271$, $p < 0.05$). The same is true for the surface spectral estimates which are statistically significantly better for the one move case versus the no motion case ($t = 2.557$, $p < 0.05$), and for the two moves case versus the one move case as well ($t = 3.867$, $p < 0.05$). Therefore we can conclude that these results confirm our hypothesis that when the effect of the prior is reduced, better surface and illuminant spectral estimates can be obtained.

5. Comparisons to previous work

We compare the proposed approach to five of the prominent color correction algorithms following the framework proposed by Barnard et al. [30]. We start by discussing the different approaches before moving on to the numerical comparisons.

5.1. Color correction approaches

In color correction algorithms, the illuminant estimate is typically referred to as the sensor response to a white surface under



(a) Tungsten Illuminant at 2800K (b) Munsell Patches

Fig. 6. Average RMS errors for the illuminant and 10 surface patch spectra for the binary filter case when there are multiple surface patches in the scene when the camera is moved twice. Each error is the average RMS error over 10 scenes chosen at random for the spectrum of (a) the tungsten illuminant at 2800 K and (b) one Munsell patch.

Please cite this article in press as: S. Skaff et al., A sequential Bayesian approach to color constancy using non-uniform filters, Comput. Vis. Image Understand. (2009), doi:10.1016/j.cviu.2009.03.014

a canonical illuminant. The canonical illuminant is the one selected from the database of illuminant spectra for which the camera is most balanced. The first algorithm is the *Gray World*, GW [7]. It assumes that the scene average is identical to the camera response to a chosen gray surface under the canonical illuminant. The illuminant *RGB*, which is equivalent to the camera response to a white surface, is taken to be double the response to a gray surface, again under the canonical illuminant. The second algorithm, *Scale-by-Max*, takes the illuminant *RGB* estimate to be the maximum in each channel (red, green, blue) over all the camera responses to all the surfaces in a scene [31]. The third algorithm is *Color by Correlation* proposed by Finlayson et al. [32,33] as an improvement to the Color in Perspective algorithm [34]. In the Color by Correlation algorithm, a correlation matrix in which each column corresponds to a different training illuminant and each row corresponds to a different image chromaticity is built. The chromaticities are computed from the surface reflectances in the training set. The convex hull of the chromaticities corresponding to one illuminant is computed. All the entries of the chromaticity bins falling in the convex hull are set to one, while all other elements are set to zero. The correlation matrix is multiplied by a vector containing information about the image chromaticities. This vector has the same discretization as the rows in the correlation matrix. An entry of this vector is set to one if the chromaticity occurs in the image and to zero otherwise. The scene illuminant selected is the one corresponding to the maximum correlation between the correlation matrix and the vector of chromaticities. This algorithm is labeled as C-by-C-01.

The last two algorithms which we compare our approach to are based on Forsyth's *gamut mapping* approach [35], which generally comprises two steps. The first step is to form two convex sets of *RGB*'s represented by their respective convex hulls. One set comprises all possible surface *RGB*'s when illuminated under a canonical illuminant. The other set comprises all surface *RGB*'s under the unknown illuminant, and therefore the convex hull is the set of observed *RGB*'s in this case. The two hulls are a unique diagonal mapping of each other under the diagonal assumption of illumination change. The set of possible maps is computed and the solution diagonal map is sought. This map corresponds to the solution set, which corresponds to the set of the surface *RGB*'s under the canonical illuminant. This is the second step of the algorithm. Finlayson added two ideas to Forsyth's approach to introduce the Color in Perspective (*CIP*) approach [34]. First, he showed how to perform gamut mapping in 2D (chromaticity) instead of 3D (*RGB*) space as this simplifies gamut mapping and renders it usable in the real world, where specularities and varying illumination are commonly-arising problems. Second, he suggested that the diagonal maps can be restricted to only those corresponding to expected illuminants. As the illumination constraint is non-convex, Barnard [36] introduced the assumption of the convex hull for this constraint (*ECRULE*). As for obtaining the solution set, which is the second step of a gamut mapping algorithm, there are three methods. The first one is choosing the solution set with the maximum volume [34,35]. The second one is taking the average of the constraint set [36]. The third one is averaging over the non-convex constraint set using Monte Carlo integration [37]. Following Barnard et al.'s framework [30], here the solution is selected by taking the center of mass of the constraint set, referred to as ICA (Illumination Constrained Average).

5.2. Numerical results

Our goal is to compare our approach to the algorithms discussed in the previous section. This implies that the chosen data sets and priors needed for these algorithms are chosen in ways that yield them comparable to the method proposed in this paper. First, the scenes used for testing are the same as the ones used in the previous section. They therefore consist of three-surface patches illuminated by a single light source. Second, the database of light source spectra used for training in the Color by Correlation approach is the same as that used for computing the prior functions in our approach. This is a set of tungsten light spectra at temperatures ranging from 2600 to 3500 K, in steps of 100 K. Next, the canonical illuminant selected from this set is such that the camera is most balanced, and is therefore tungsten light at 3500 K. Note that the canonical illuminant choice can be arbitrary according to Barnard et al. [30]. In addition, the set of *RGB* measurements used to obtain the convex hull in the 2D and 3D Gamut Mapping approaches are the camera sensor responses of all 1269 patches from the Munsell Book of Color [18]. In addition, the reflectance spectra of these patches are used in the Gray World approach. All the spectra employed in the previous approaches are 55D and over the same wavelength range as in our approach.

The approach proposed in this paper estimates both illuminant and surface spectra. However, the described algorithms perform color correction after carrying out an illumination estimation step. Therefore, we focus on comparing the performance of illuminant estimation of the different algorithms. Since we are interested in the chromaticity of the illuminant, we choose the second error measure in Barnard's framework [30] for comparison. To obtain this error measure, the chromaticities for each of the actual and model illuminants given their *RGB*'s is computed:

$$(r_a, g_a) = (R_a/(R_a + G_a + B_a), G_a/(R_a + G_a + B_a)),$$
$$(r_m, g_m) = (R_m/(R_m + G_m + B_m), G_m/(R_m + G_m + B_m)), \quad (29)$$

where (R_a, G_a, B_a) is the *RGB* of the actual illuminant and (R_m, G_m, B_m) is the *RGB* of the model illuminant. The error E is the vector distance between the two chromaticities:

$$E = \sqrt{(r_m - r_a)^2 + (g_m - g_a)^2}. \quad (30)$$

Our approach estimates an illuminant spectral model, which is projected onto the three sensitivity curves of the camera sensor to obtain the equivalent three-dimensional *RGB* measurement. From this *RGB*, the corresponding chromaticity coordinates and then E are computed for each scene. The median RMS error is computed to measure the performance of the approach in estimating illuminant spectra for all 100 scenes [38].

Next, the median RMS error is computed for the illuminant chromaticity estimate of the same 100 scenes as given by the five previous algorithms. These errors are shown in Table 1. As there are many variants for each of the Color by Correlation, 2D Gamut Mapping, and 3D Gamut Mapping algorithms, the variant which

Table 1
The median RMS errors for the illuminant estimates of 100 scenes (except for 2D Gamut Mapping) given by five previous approaches, Brainard and Freeman's approach, and the different variants of the proposed approach. For the scene setup in the case of Sequential Bayesian – Binary Filter, Static Camera, see Fig. 1. For the scene setup in the case of Sequential Bayesian – Binary Filter: No moves, One move, and Two moves, see Fig. 2. N denotes the number of gathered sensor responses or measurements.

Algorithm	Median RMS error
Gray World	0.2528
Scale-by-Max	0.2520
C-by-C-01	0.0366
2D Gamut Mapping (CIP) – ICA	0.0619
3D Gamut Mapping (ECRULE) – ICA	0.0698
Brainard–Freeman $N = 3$	0.0454
Sequential Bayesian – Binary Filter, No Moves $N = 4$	0.0427
Sequential Bayesian – Binary Filter, One Move $N = 5$	0.0355
Sequential Bayesian – Binary Filter, Static Camera $N = 5$	0.0350
Sequential Bayesian – Binary Filter, Two Moves $N = 6$	0.0310

gives rise to the minimal median RMS error over the 100 illuminants is selected. In addition, the median RMS errors of Brainard and Freeman's approach as well as the different variants of our approach are shown in the table in decreasing order. As expected, the order of median RMS errors corresponds to an increasing number of sensor measurements (denoted by N in the table) in each of the scenarios. However, in the Sequential Bayesian – Binary Filter, One Move and the Sequential Bayesian – Binary Filter, Static Camera cases, the number of gathered *RGB* measurements is the same ($N = 5$). This reflects in the very small difference in the corresponding median RMS errors.

First, we would like to note that the 2D Gamut Mapping algorithm did not find a solution for 28 out of the 100 scenes. This may be due to the fact that the canonical gamut, formed by the *RGB*'s of the surfaces viewed under the canonical illuminant, is limited as it does not take into account all possible surfaces. Only the surface patches from our data set are considered. The median RMS error given in the table for 2D Gamut Mapping is, therefore, over 72 scenes. From the table, it can be concluded that the Gray World and the Scale-by-Max algorithms perform poorly as expected since the assumptions imposed by these algorithms are strong and are not always valid for real world images. The Color by Correlation algorithm performs relatively well compared to the other four previous approaches as it does not impose as strong a set of assumptions. This explains its low median RMS error. Finally, we conclude that the proposed Sequential Bayesian approach yields the best illuminant spectral estimates when at least five sensor measurements are employed when there are three-surface patches in the scene, as in the Binary Filter – Static Camera, the Binary Filter – One Move, and the Binary Filter – Two Moves cases. Our approach has an advantage over the gamut mapping approaches in that it does not require a canonical gamut, which should be formed by taking into account as many surface patches as possible. Moreover, our approach does not impose any assumptions on the scene under study, which is the case in the Gray World and Scale-by-Max algorithms. Therefore, our approach demonstrates flexibility of use.

6. Conclusions

In this work, we have proposed a *sequential Bayesian* strategy that builds on the Brainard–Freeman Bayesian approach to color constancy [1]. The recursion helps in gathering information across different scenes and filter portions of different spectral sensitivities. Brainard and Freeman used a Bayesian technique to regularize the problem of computing the values of the illuminant and surface reflectance spectra parameters [39]. This technique uses the bilinear model of Maloney and Wandell [10] to provide a parametrization of the illuminant and surface patch spectra.

Our approach enhances Brainard and Freeman's model and is based on insight from the characteristics of human vision, and the human eye, in particular. In our approach, measurements are acquired sequentially from each portion of a filter, similar to when a person moves his/her gaze across a surface in the scene. In this paper we modeled the case of a non-uniform filter, which has two portions of different spectral sensitivities. We also extended this case to moving the camera across the scene.

The experimental results in Section 4 indicate that there is considerable improvement in the illuminant spectra estimates with the introduction of the binary filter method compared to the no filter method, which is equivalent to that of Brainard and Freeman. Another result shown in Section 4 is that of the moving camera: the more the number of moves of the camera, the more measurements for surface patches obtained. This motion results in an improvement in the estimates of the illuminant and surface spectra. Next, the numerical comparisons of illuminant estimation in Section 5 show that the proposed algorithm, in the case of a certain number of gathered sensor responses from a scene, can outperform state-of-the-art color correction approaches.

In our work, we rely heavily on the Bayesian probabilistic formulation. We believe that this formulation is a suitable approach because it models the uncertainties in all parts of a system, on one hand, and because it employs spectral models, on the other hand. Spectral models make it easier to further extend this Bayesian approach to incorporate measurements from different scenes. In addition, spectral models as used in our approach do not require all surface patches of the scene to be viewed under all portions of a filter. Finally, the computational complexity of the Bayesian formulation is tolerable, which makes it feasible to implement and run our sequential formulation.

Acknowledgements

We would like to thank Kobus Barnard for making his code publicly available. This research was funded by a grant from the Canadian Natural Sciences and Engineering Research Council (NSERC).

References

[1] D.H. Brainard, W.T. Freeman, Bayesian color constancy, Journal of the Optical Society of America A 32 (7) (1997) 1393–1411.
[2] J.J. Clark, J.K. O'Regan, A temporal difference learning model for perceptual stability in color vision, in: IEEE International Conference on Pattern Recognition, Barcelona, Spain, 2000, pp. 503–506.
[3] J.J. McCann, S.P. McKee, T.H. Taylor, Quantitative studies in retinex theory – a comparison between theoretical predictions and observer responses to the 'color mondrian' experiments, Vision Research 16 (5) (1976) 445–448.
[4] K. Barnard, L. Martin, A. Coath, B. Funt, A comparison of computational color constancy algorithms. II. Experiments with image data, IEEE Transactions on Image Processing 11 (9) (2002) 985–996.
[5] M. Brill, G. West, Contributions to the theory of invariance of colour under the condition of varying illumination, Journal of Mathematical Biology 11 (3) (1981) 337–350.
[6] J.v. Kries, Beitrag zur physiologie der gesichtsempfinding, Arch. Anat. Physiol. 2 (1878) 5050–5524.
[7] G. Buchsbaum, A spatial processor model for object colour perception, Journal of the Franklin Institute 310 (1980) 1–26.
[8] R. Gershon, A.D. Jepson, J.K. Tsotsos, From [r, g, b] to surface reflectance: computing color constant descriptors in images, in: International Joint Conference on Artifical Intelligence, 1987, pp. 755–758.
[9] L.T. Maloney, Computational approaches to color constancy, Ph.D. Thesis, Stanford University, Stanford, CA, 1985.
[10] L.T. Maloney, B.A. Wandell, Color constancy: a method for recovering surface spectral reflectance, Journal of the Optical Society of America A 3 (1) (1986) 29–33.
[11] A. Yuille, A method for computing spectral reflectance, Biological Cybernetics 56 (1987) 195–201.
[12] M.D. D'Zmura, P. Lennie, Mechanisms of color constancy, Optical Society of America A 10 (3) (1986) 1662–1672.
[13] M.D. D'Zmura, G. Iverson, Color constancy I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces, Journal of the Optical Society of America A 10 (10) (1993) 2148–2165.
[14] M.D. D'Zmura, G. Iverson, Color constancy II. Results for two-stage linear recovery of spectral descriptions for lights and surfaces, Journal of the Optical Society of America A 10 (10) (1993) 2166–2180.
[15] J. Ho, B.V. Funt, M.S. Drew, Separating a color signal into illumination and surface reflectance components: theory and applications, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (10) (1990) 966–977.
[16] T.O. Berger, Statistical Decision Theory and Bayesian Analysis, Springer, New York, NY, 1985.
[17] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, Stanford, CA, 1973.
[18] Munsell Book of Color-Matte Finish Collection, Munsell Color, Baltimore, MD, 1976.
[19] J.P.S. Parkkinen, J. Hallikainen, T. Jaaskelainen, Characteristic spectra of Munsell colors, Journal of the Optical Society of America A 6 (2) (1989) 318–322.
[20] S. Skaff, T. Arbel, J.J. Clark, Active Bayesian color constancy with non-uniform sensors, in: IEEE International Conference on Pattern Recognition, vol. 2, Quebec City, QC, 2002, pp. 681–684.
[21] M. Tsukada, Y. Ohta, An approach to color constancy using multiple images, in: Proceedings of the 3rd IEEE International Conference on Computer Vision, Osaka, Japan, 1990, pp. 385–389.
[22] M.D. D'Zmura, G. Iverson, Color constancy III. General linear recovery of spectral descriptions for lights and surfaces, Journal of the Optical Society of America A 11 (9) (1994) 2389–2400.

[23] G.D. Finlayson, B.V. Funt, K. Barnard, Color constancy under varying illumination, in: IEEE International Conference on Computer Vision, Cambridge, MA, 1995, pp. 720–725.
[24] K. Barnard, G. Finlayson, B. Funt, Color constancy for scenes with varying illumination, Computer Vision and Image Understanding 65 (2) (1997) 311–321.
[25] G.D. Finlayson, S.D. Hordley, P. Morovic, Colour constancy using the chromagenic constraint, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 1079–1086.
[26] G.D. Finlayson, P.M. Morovic, S.D. Hordley, Chromagenic colour constancy, in: Proceedings of the 10th Congress of the International Colour Association, 2005.
[27] J.P. Parkkinen, P. Silfsten, Database of spectra, University of Joensuu, Finland, Available from: <http://spectral.joensuu.fi/index.php?page=database&info=munsell_matt/>.
[28] I.E.S. of North America, IES lighting handbook, Illuminating Engineering Society of North America, New York, NY, 1981.
[29] D.J. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, Chapman & Hall/CRC Press, New York, NY, 2000.
[30] K. Barnard, V. Cardei, B. Funt, A comparison of computational color constancy algorithms. I. Methodology and experiments with synthesized data, IEEE Transactions on Image Processing 11 (9) (2002) 972–984.
[31] E.H. Land, The retinex theory of color vision, Scientific American 237 (6) (1977) 108–128.
[32] G.D. Finlayson, P.M. Hubel, S. Hordley, Color by correlation, in: Proceedings of the Fifth Color Imaging Conference, 1997, pp. 6–11.
[33] G.D. Finlayson, S.D. Hordley, P.M. Hubel, Color by correlation: a simple unifying framework for color constancy, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (11) (2001) 1209–1221.
[34] G.D. Finlayson, Color in perspective, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (10) (1996) 1034–1038.
[35] D.A. Forsyth, A novel algorithm for color constancy, International Journal of Computer Vision 5 (1) (1990) 5–36.
[36] K. Barnard, Computational colour constancy: taking theory into practice, Master's Thesis, Simon Fraser University, School of Computing (August 1995).
[37] G. Finlayson, S. Hordley, Selection for gamut mapping colour constancy, Image and Vision Computing 17 (8) (1999) 597–604.
[38] S.D. Hordley, G.D. Finlayson, Reevaluation of color constancy algorithm performance, Journal of the Optical Society of America A 23 (5) (2006) 1008–1020.
[39] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-Posed Problems, V.H. Winston, Washington, DC, 1977.

Please cite this article in press as: S. Skaff et al., A sequential Bayesian approach to color constancy using non-uniform filters, Comput. Vis. Image Understand. (2009), doi:10.1016/j.cviu.2009.03.014