Daylight illuminant retrieval using redundant image elements

Joanna Marguier and Sabine Süsstrunk

School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne, Switzerland

*Corresponding author: joanna.marguier@a3.epfl.ch

We present a method for retrieving illuminant spectra from a set of images taken with a fixed location camera, such as a surveillance or panoramic one. In these images, there will be significant changes in lighting conditions and scene content, but there will also be static elements in the background. As color constancy is an under-determined problem, we propose to exploit the redundancy and constancy offered by the static image elements to reduce the dimensionality of the problem. Specifically, we assume that the reflectance properties of these objects remain constant across the images taken with a given fixed camera.

We demonstrate that we can retrieve illuminant and reflectance spectra in this framework by modeling the redundant image elements as a set of synthetic RGB patches. We define an error function that takes the RGB patches and a set of test illuminants as input and returns a similarity measure of the redundant surfaces reflectances. The test illuminants are then varied until the error function is minimized, returning the illuminants under which each image in the set was captured. This is achieved by gradient descent, providing an optimization method that is robust to shot noise.

OCIS codes: 330.0330, 330.1690, 330.1715.

1. Introduction

We propose to solve for color constancy for a set of images containing a small number of redundant elements. More precisely, we are considering N images captured with a fixed location camera under N daylight illuminants. These images exhibit changes in lighting and

content, but also contain a number of static background elements, as illustrated in Figure 1. These objects are unknown, but their reflectance remains unchanged across images. We thus propose to use this additional information to reduce the dimension of the underdetermined problem of illuminant retrieval. Formally, we express all illuminant $E(\lambda)$ and reflectance $S(\lambda)$ spectra by linear models, invert a series of images formation models in parallel, and force their output reflectances to match, allowing to deduce the illuminant in each image. This is achieved through the minimization of an error function, which is a measure of the surfaces' similarities as a function of the illuminant descriptors. Each illuminant is defined by three scalars, the error function thus depends on $3N$ variables and is minimized by gradient descent.

We demonstrate the validity of this approach by applying our algorithm on a set of synthetic RGB values generated for different types of reflectances and daylight illuminants and representing the redundant image elements. We observe that the illuminant estimation becomes more robust when considering an increasing number of images and returns a median angular error under 3° [1] when taking into account sets of four, or more, images. We also show that the results returned by this method are only slightly reduced in the presence of shot noise. Our algorithm generally outperforms the single light method *Shades of Gray* [2], especially in the presence of shot noise.

Our algorithm presents most similarities with the two-stage linear recovery for illuminant and reflectance descriptors developed by D'Zmura and Iverson [3–5]. The authors express all illuminants and reflectances by linear models of fixed dimensions and create a single system of equations from the image formation models for a small number of redundant surfaces viewed under N illuminants. Their approach ensures a perfect recovery of the descriptors when the considered spectra can be exactly represented by linear models, but it starts to fail when only small deviations from the ideal conditions are introduced. Contrarily, our approach only returns an approximation of the illuminants, but remains much more stable in the presence of noise.

This article is structured as follows: Section 2 presents linear models for reflectance and illuminant spectra and several relevant color constancy algorithms. Section 3 details our algorithm, while Section 4 reports the experiments and results. Section 5 concludes the paper.

2. Background

2.A. Linear models of illuminants and reflectances

A way of reducing the dimension of the illuminant retrieval problem is to express the illuminant spectrum $E(\lambda)$ and the reflectance spectrum $S(\lambda)$ as linear sums of a limited number of basis functions. Determining their spectra then reduces to determining their weights, which

we call illuminant and reflectance *descriptors*.

Most natural and man-made surfaces reflectance spectra are smooth enough to be represented by a linear combination of a low number of basis functions. Cohen [6] demonstrated that Munsell surface reflectances could be approximated by the average reflectance plus two additional components, i.e., by a sum of three basis functions. However, later work showed that five to seven [7] or even eight [8] basis functions were necessary to accurately represent real surface spectra.

These basis functions can be computed by principal component analysis on a large set of reflectance spectra [9, 10] and expressed as

$$S(\lambda) = \sum_{i=1}^{N_\sigma} \sigma_i \mathcal{S}_i(\lambda)\,, \tag{1}$$

where σ_i are the weights or reflectance descriptors, $\mathcal{S}_i(\lambda)$ the surface reflectance basis functions, and N_σ the number of basis functions.

Similarly, Judd et al. [11] decomposed daylight illuminant spectra as sums of three basis functions $\mathcal{E}_i$, i.e.,

$$E(\lambda) = \sum_{i=1}^{N_\varepsilon} \varepsilon_i \mathcal{E}_i(\lambda)\,, \tag{2}$$

where ε_i are the illuminant descriptors, $\mathcal{E}_1(\lambda)$ is the average daylight, and $N_\varepsilon = 3$ is the number of basis functions. The authors computed these basis functions by principal component analysis on a total of 622 spectra from three sets of daylight measurements. Using only three basis functions, i.e., the average daylight $\mathcal{E}_1$ plus two orthogonal characteristic vectors, they could reconstruct the illuminant spectra dataset with good accuracy, measured as the variance between real and reconstructed spectra.

We start from the image formation model, which describes the red, green, and blue sensor responses $\rho_k(\mathbf{x})$ as

$$\rho_k(\mathbf{x}) = \int_\lambda E(\mathbf{x},\lambda) S(\mathbf{x},\lambda) R_k(\lambda) d\lambda, \quad k = 1,2,3\,, \tag{3}$$

where $R_k(\lambda)$ are the sensor sensitivities and λ ranges from 400 nm to 700 nm. We express the spectra as (1) and (2) and consider that all quantities are spatially constant, i.e., independent of $\mathbf{x}$, the three equations (3) are then rewritten as

$$\rho_k = \int_\lambda \sum_{i=1}^{N_\varepsilon} \varepsilon_i \mathcal{E}_i(\lambda) \sum_{j=1}^{N_\sigma} \sigma_j \mathcal{S}_j(\lambda) R_k(\lambda) d\lambda = \sum_{i=1}^{N_\varepsilon} \varepsilon_i \sum_{j=1}^{N_\sigma} \sigma_j \int_\lambda \mathcal{E}_i(\lambda) \mathcal{S}_j(\lambda) R_k(\lambda) d\lambda \tag{4}$$

We define the $N_\varepsilon \times N_\sigma$ matrix $\mathbf{\Lambda}_k$ by its elements

$$(\Lambda_k)_{i,j} = \int_\lambda \mathcal{E}_i(\lambda) \mathcal{S}_j(\lambda) R_k(\lambda) d\lambda\,. \tag{5}$$

We represent the quantities $S(\lambda)$, $E(\lambda)$, $R_k(\lambda)$, $\mathcal{E}_i(\lambda)$, and $\mathcal{S}_j(\lambda)$ by their spectra sampled at 10 nm intervals and denote them, respectively, by the vectors $\mathbf{s}$, $\mathbf{e}$, $\mathbf{r}_k$, $\boldsymbol{\mathcal{E}}_i$, and $\boldsymbol{\mathcal{S}}_j$. Equation (3) becomes

$$\rho_k = \mathbf{s}^T \mathrm{diag}(\mathbf{e}) \mathbf{r}_k \,, \tag{6}$$

where $\mathrm{diag}(\mathbf{e})$ is a 31×31 matrix with the vector entries of $\mathbf{e}$ on its diagonal. It can be rewritten with linear models as

$$\rho_k = \boldsymbol{\varepsilon}^T \boldsymbol{\Lambda}_k \boldsymbol{\sigma} = \boldsymbol{\sigma}^T \boldsymbol{\Lambda}_k^T \boldsymbol{\varepsilon} \,, \tag{7}$$

where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\sigma}$ are $N_\varepsilon \times 1$ and $N_\sigma \times 1$ vectors containing the illuminant and reflectance descriptors.

This image formation model using linear models for the illuminant and reflectance spectra is called *bilinear*. It is illustrated by the form of (7) in which $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ have symmetrical roles. If we fix $\boldsymbol{\varepsilon}$, the system becomes linear in $\boldsymbol{\sigma}$, and vice versa. If the illuminant is fixed, (7) can be written as

$$\rho_k = \boldsymbol{\Lambda}_{\varepsilon,k} \boldsymbol{\sigma} \,, \tag{8}$$

where $\boldsymbol{\Lambda}_{\varepsilon,k} = \boldsymbol{\Lambda}_k \boldsymbol{\varepsilon}$ is the *lighting matrix*. It can also be expressed in the form

$$\boldsymbol{\rho} = \boldsymbol{\Lambda}_\varepsilon \boldsymbol{\sigma} \,, \tag{9}$$

with the $3 \times N_\varepsilon$ matrix $\boldsymbol{\Lambda}_\varepsilon = \mathbf{R}\mathrm{diag}(\mathbf{e})\mathbf{B}_s$, where $\mathbf{R}$ is a 3×31 matrix containing the sensor sensitivities and $\mathbf{B_s}$ is a $31 \times N_\sigma$ matrix containing reflectance basis functions. $\boldsymbol{\sigma}$ and $\boldsymbol{\rho}$ are $N_\varepsilon \times 1$ and 3×1 vectors, respectively. For a sensor response corresponding to a given surface, we compute its reflectance descriptors for any illuminant characterized by $\boldsymbol{\varepsilon}$ by inverting (9) as

$$\boldsymbol{\sigma} = \boldsymbol{\Lambda}_\varepsilon^+ \boldsymbol{\rho} \,, \tag{10}$$

where $^+$ represents the Moore-Penrose pseudo-inverse, which can be constructed directly as

$$\boldsymbol{\Lambda}_\varepsilon^+ = \boldsymbol{\Lambda}_\varepsilon^T (\boldsymbol{\Lambda}_\varepsilon \boldsymbol{\Lambda}_\varepsilon^T)^{-1} \,, \tag{11}$$

whenever the matrix $\boldsymbol{\Lambda}_\varepsilon$ is of full rank [12], which should be the case. In practice, it is computed by singular values decomposition. In the expression (10), $\boldsymbol{\sigma}$ is an $N_\sigma \times 1$ vector, $\boldsymbol{\Lambda}_\varepsilon^+$ is an $N_\sigma \times 3$ matrix, and $\boldsymbol{\rho}$ is a 3×1 vector. Finally, the image formation (4) reduces to the expression (9) and can be easily inverted for any given test illuminant.

2.B. Color constancy algorithms

Color constancy algorithms aim at retrieving illuminant descriptors or illuminant independent reflectance descriptors from RGB images. The dimension the problem can be reduced by introducing assumptions on the image or scene content. These algorithms can be classified

in two categories: *Single* light methods exploit the content of one scene or image to retrieve illuminant information, whereas *multiple* lights methods retrieve illuminant descriptors from a set of two, or more, images containing redundant information.

The illuminant can be estimated by introducing simple assumptions on the scene content. The *max-RGB* [13] algorithm relies on the presence of a white patch in the scene to retrieve the illuminant descriptors. More precisely, it assumes the scene white-point to correspond to the maximum sensor response in each channel of an image. The *gray world* algorithm estimates the illuminant by assuming that the average of an image is gray, i.e., that the average reflectance of the objects present in the scene is constant [14]. The scene white-point is estimated by calculating the average value of each color channel. Variations of the gray world algorithm include the *weighted gray world* [15], *gray edge hypothesis* [16], and *shades of gray* [2].

A series of algorithms [17–20] resulted from the observation that not all colors can arise under any illuminant. For example, a bright blue cannot been seen under an incandescent light. The underlying idea of these algorithms is to compare the gamut spanned by the colors in an image with the gamuts corresponding to the colors that can arise under a set of reference illuminants. The image illuminant is then chosen as the reference illuminant for which the respective gamuts are the closest.

Maloney and Wandel [21] express illuminants and reflectances by linear models and assume the number of sensors to be superior to the number of degrees of freedom of the reflectance descriptors. In this framework, they retrieve the illuminant by inversion of a linear system. If we apply this approach to the case of an RGB camera, we have three sensors and can thus approximate the reflectances by a sum of only two functions, which is not accurate enough [6–8].

There are other color constancy approaches that consider *multiple* lights or images. Two images taken under different illuminants or several surfaces viewed under different lights provide redundant information that can be exploited to solve for color constancy. The *flash/no flash* method takes two images of one scene captured with and without flash [22, 23]. Knowing the flash properties, information on the illuminant can be retrieved. The authors take an image pair to compute an estimate of the scene lit by the flash only and, knowing the flash's spectral power distribution, use this last image to retrieve objects reflectance functions. They finally employ these reflectances to estimate the illuminant in the image taken without flash. The *chromagenic* approach takes two images of one scene captured with and without a colored filter placed in front of the camera [24, 25]. The filter is chosen such that the relationship between the filtered and unfiltered RGB values depends strongly on the illumination. The knowledge of its transmittance is used to solve for color constancy. The algorithm is tested on a set of precomputed illuminants. These methods take advantage of the

difference between images pairs (the flash spectra and the filter transmittance, respectively) to retrieve the illuminant in one image.

Finlayson [26] presents a method to retrieve illuminant descriptors taking two sets of three, or more, surfaces viewed under two different illuminants. The illuminant and reflectance spectra are modeled by tri-dimensional linear models. He considers the transforms $\boldsymbol{\mathcal{M}}^{1,2}$ mapping the sensor responses under the first illuminant to the ones under the second illuminant. He shows that the illuminant descriptors $\boldsymbol{\varepsilon}^1 = (\varepsilon_1^1, \varepsilon_2^1, \varepsilon_3^1)$ and $\boldsymbol{\varepsilon}^2 = (\varepsilon_1^2, \varepsilon_2^2, \varepsilon_3^2)$ can be found by solving

$$\boldsymbol{\mathcal{Q}}\boldsymbol{\varepsilon}^1 = \boldsymbol{\mathcal{P}}\boldsymbol{\varepsilon}^2\,, \tag{12}$$

where $\boldsymbol{\mathcal{P}}$ is a 9×3 matrix containing the stretched out 3×3 basis lighting matrices $\boldsymbol{\Lambda}^i$, $i = 1, 2, 3$. The three columns of $\boldsymbol{\mathcal{Q}}$ are the matrices $\boldsymbol{\mathcal{M}}^{1,2}\boldsymbol{\Lambda}^i$ in the form of 9×1 vectors. The solution of (12) is the intersection of the two tri-dimensional spaces spanned by the columns of $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ and is found by the method of the principal angles, which chooses $\boldsymbol{\varepsilon}^1$ and $\boldsymbol{\varepsilon}^2$ such that the angle between $\boldsymbol{\mathcal{Q}}\boldsymbol{\varepsilon}^1$ and $\boldsymbol{\mathcal{P}}\boldsymbol{\varepsilon}^2$ is minimized, returning the descriptors for both illuminants. In [27], the authors note that all the possible transforms mapping the chromaticity coordinates of a surface viewed under a unknown illuminant to a canonical one roughly lie on a straight line and that these transforms for different unknown illuminants lie on different lines. By intersecting them, they can recover illuminant independent chromaticities for the surface.

D'Zmura and Iverson [3–5] present a solution for illuminant and reflectance retrieval when N_p surfaces are viewed under N illuminants. Illuminant and reflectance spectra are expressed by linear models and the image formation is expressed as (4). The authors write (4) for all the surfaces and illuminants and combine the resulting equations in the compact form

$$\mathbf{F}\tilde{\boldsymbol{\sigma}} = 0\,, \tag{13}$$

where $\tilde{\boldsymbol{\sigma}}$ is built using exclusively the reflectance descriptors σ_j and $\mathbf{F}$ depends on the illuminant descriptors ε_i, the illuminant and reflectance basis functions $\mathcal{E}_i(\lambda)$ and $\mathcal{S}_j(\lambda)$, the sensors responses $R_k(\lambda)$, and the color responses $\boldsymbol{\rho}$. The first step of the method is to solve (13) by singular value decomposition and retrieve the reflectance descriptors σ_j from $\tilde{\boldsymbol{\sigma}}$. At this stage, the reflectances are known and the illuminant descriptors can be retrieved in a second step. Indeed, the problem is now reduced to a simple linear system in ε_i. The case $N_\sigma > N_\rho$, with N_ρ the number of channels of the imaging device, is not generally solvable. The authors [5] report the conditions on the number of color channels N_ρ, the number of surfaces N_p, the number of illuminant descriptors N_ε, the number of reflectance descriptors $N_\sigma > N_\rho$, and the number of illuminants N under which the illuminants and reflectances can be simultaneously retrieved. The conditions are of the form $N_p = N_\varepsilon = N_\sigma = N > N_\rho$. The authors also suggest that conditions of the form $N = N_p > N_\varepsilon = N_\sigma > N_\rho$ may lead to

perfect recovery. Moreover, the illuminant and reflectance models need to have exactly N_ε and N_σ degrees of freedom, respectively, to ensure a good result.

3. Our approach

At each pixel location, the image formation model gives a system of three linear equations of the form (7). In the present framework, we know the color responses ρ_k for each surface and the camera sensitivities $R_k(\lambda)$, while both the illuminant and reflectance spectra are unknown. Their spectra are represented by, respectively, $N_\varepsilon = 3$ and $N_\sigma = 7$ descriptors, i.e., by $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3)$ and $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_7)$.

We consider N unknown scenes taken with fixed location cameras and that contain a few static background elements. They are viewed under N unknown illuminants, which we assume to be uniform. We extract N_p patches of uniform color from the objects present in every image and average their pixel values, which we denote $\mathbf{p}$. For each of the NN_p surfaces, we solve a system of three linear equations (9): each surface's mean RGB value $\mathbf{p}$ allows computing the corresponding reflectance descriptors $\boldsymbol{\sigma}$ for any illuminant. We thus compute NN_p systems (10) in parallel for a series of N test illuminants and obtain a set of NN_σ reflectance descriptors. By forcing the reflectances of each surface to match across the N images, we can deduce the N illuminants.

Rather than matching reflectances descriptors $\boldsymbol{\sigma}$ directly, we compute and compare, for each patch, the corresponding color responses as viewed under the standard daylight illuminant D65, which we denote $\mathbf{p}^{D65}$. The argument in favor of this approach is the following: the inverse image formation model (10) returns only one set of reflectance descriptors for each test illuminant, while there always exists a set of metameric surface reflectances that yield to the same RGB values and that we cannot discriminate using this approach. There is thus no intrinsic reason to match reflectance spectra directly. Instead, we are computing a Euclidian distance between the points in $\mathbb{R}^3$ representing the color responses $\mathbf{p}^{D65}$ for each illuminant and each surface. This distance is more meaningful than a Euclidian distance between reflectance descriptors in $\mathbb{R}^{N_\sigma}$. Indeed, the three RGB components contribute equally to the resulting color, as opposed to the reflectance descriptors, which correspond to basis functions having different contributions to the reflectance spectra.

The color responses $\mathbf{p}^{D65} = (p_1^{D65}, p_2^{D65}, p_3^{D65})$ are computed as

$$p_k^{D65} = \Big[\sum_{i=1}^{N_\sigma} \sigma_i \boldsymbol{\mathcal{S}}_i \Big]^T \mathrm{diag}(\mathbf{e}_{D65}) \mathbf{r}_k, \quad k = 1, 2, 3\,, \tag{14}$$

where σ_i are related to a test illuminant $\mathbf{e}_{\mathrm{test}}(\lambda)$ via $\boldsymbol{\Lambda}_\varepsilon$ (10). For any combination of N illuminants we can compute a set of NN_p color responses $\mathbf{p}^{D65}$, which allows defining an error function reaching its minimum when the sensor responses under illuminant D65 match

or, indirectly, when the reflectance spectra match. The illuminant spectra are represented by linear sums of three basis functions (2) and f_e is defined as a function of the illuminant descriptors. The error function is the Euclidean distance between the NN_p sensor responses $\mathbf{p}^{D65}$ and to a combination of N illuminants represented by $\mathbf{E}_k = (\mathbf{e}_1(\boldsymbol{\varepsilon}_k), ..., \mathbf{e}_N(\boldsymbol{\varepsilon}_k))$,

$$f_e(\boldsymbol{\varepsilon}_k) = \frac{1}{N_p}\frac{1}{N}\Big[\sum_{n_p=1}^{N_p}\sum_{j=2,j>i}^{N}\sum_{i=1}^{N}[\mathbf{p}_{i,n_p}^{D65}(\mathbf{E}_k) - \mathbf{p}_{j,n_p}^{D65}(\mathbf{E}_k)]^2\Big]^{\frac{1}{2}}, \tag{15}$$

where the indices i and j run over the images and n_p runs over the N_p patches. The double sum on i and $j > i$ indicates that the distance is summed over the $\binom{N}{2}$ possible image pairs; for example, if we consider three images, Equation (15) becomes

$$f_e(\boldsymbol{\varepsilon}_k) = \frac{1}{N_p}\frac{1}{3}\Big[\sum_{n_p=1}^{N_p}[(\mathbf{p}_{1,n_p}^{D65} - \mathbf{p}_{2,n_p}^{D65})^2 + (\mathbf{p}_{2,n_p}^{D65} - \mathbf{p}_{3,n_p}^{D65})^2 + (\mathbf{p}_{1,n_p}^{D65} - \mathbf{p}_{3,n_p}^{D65})]^2\Big]^{\frac{1}{2}}. \tag{16}$$

A least square error function is used for computational simplicity.

The N scene illuminants corresponding to the N images satisfy

$$\widehat{\mathbf{E}}_k = \arg\min_{\boldsymbol{\varepsilon}_k} f_e(\boldsymbol{\varepsilon}_k)\,. \tag{17}$$

We find the solution of (17) by performing a gradient descent on f_e, which is minimized with the following the gradient estimate

$$\nabla_{\boldsymbol{\varepsilon}_k} f_e = f_e(\boldsymbol{\varepsilon}_k + \boldsymbol{\epsilon}) - f_e(\boldsymbol{\varepsilon}_k - \boldsymbol{\epsilon}) \propto \frac{f_e(\boldsymbol{\varepsilon}_k + \boldsymbol{\epsilon}) - f_e(\boldsymbol{\varepsilon}_k - \boldsymbol{\epsilon})}{\boldsymbol{\epsilon}}\,, \tag{18}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \epsilon_3)$. The step size ϵ_i is fixed. If ϵ_i is too small, the convergence may be slow, whereas a large ϵ_i may lead to erroneous results.

At each step, the illuminant descriptors $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3)$ are updated according to

$$\boldsymbol{\varepsilon}_{k+1} = \boldsymbol{\varepsilon}_k - \alpha[f_e(\boldsymbol{\varepsilon}_k + \boldsymbol{\epsilon}) - f_e(\boldsymbol{\varepsilon}_k - \boldsymbol{\epsilon})]\,. \tag{19}$$

We use fixed α and ϵ_i. In cases where ϵ_i is too large, it can generate oscillations. To overcome this issue, the sign of the gradient is stored at each step and compared at successive steps. Changes of sign are interpreted as oscillations and, if they occur, the amplitude of ϵ_i is reduced.

4. Experiments

4.A. Gradient descent on $f_e = f_e(\varepsilon_1, \varepsilon_2, \varepsilon_3)$

We created sets of RGB values for a series of illuminants and reflectances using the images formation model (9). The illuminants database consists of a total of 173 spectral power

distributions containing both computed and measured illuminant spectra, including the 89 measurements from [31], from which we created a subset of real daylight illuminants (45 spectra). Examples of illuminant spectra are represented in Figure 2. The reflectance database contains 24 MacBeth, 462 Munsell, and 219 Natural spectra. We computed RGB patches values for each of the $24+462+219=705$ reflectance spectra and the 45 illuminant spectra. The sensor sensitivities are the ones of a Canon 350D camera. RGB values were normalized such that the overall maximum is equal to one for each different set of reflectances and each of the 45 illuminants. In other words, the normalization factor varies among illuminants such that it is representative of how RGB images are generally normalized, i.e., such that their maximum value is one. We generated series of 10^3 random combinations of N illuminants and N_p reflectances for each of the three illuminant and three reflectance types and test our algorithm on the resulting RGB values. N takes the values 2, 3, 4, 6, and 10 and N_p takes the values 6 and 10.

The error is reported as the angle between the vectors $\boldsymbol{\rho^E}$ and $\hat{\boldsymbol{\rho}}^{\boldsymbol{E}}$ formed by the real and retrieved white-points in sRGB and is called the *angular error* [28–30]

$$e_{ang} = \arccos\left(\frac{\boldsymbol{\rho^E} \circ \hat{\boldsymbol{\rho}}^{\boldsymbol{E}}}{\| \boldsymbol{\rho^E} \| \| \hat{\boldsymbol{\rho}}^{\boldsymbol{E}} \|}\right). \tag{20}$$

An error under 3° is generally considered as acceptable [1].

Table 1 reports the median and maximum angular errors obtained from the gradient descent algorithm for the MacBeth, Munsell, and Natural reflectances. Figure 3 shows examples of real and retrieved illuminant spectra. The results become more robust with an increasing number of illuminants. When considering four illuminants, the gradient descent method returns a median angular error under 3°. The influence of the number of illuminants N is also visible on the maximum error, ranging from 18.04° when $N=2$ down to 8.37° when $N=10$.

4.B. Noisy sensor responses

We now run our algorithm for $N=4$ illuminants and $N_p=6$ reference surfaces represented by the same RGB sensor responses, but with added shot noise. The noisy sensor responses are modeled as

$$\rho_i = \rho_{i,o} \pm \sigma_{\text{shot noise}}\,, \tag{21}$$

where $\rho_{i,o}$ is the ideal sensor response and $\sigma_{\text{shot noise}}$ is modeled by a Poisson process of parameter $\lambda_{\text{p}} = \sqrt{\rho_{i,o}}$, i.e., centered around $\sqrt{\rho_{i,o}}$. The index i runs over the red, green, and blue sensor responses. In order to control the amount of noise, we introduce a parameter $A_{\text{noise}} \in [0,1]$ scaling $\sigma_{\text{shot noise}}$. To create a model representative of an actual imaging device, the RGB are converted into integer digital values in $[0,255]$ and calculated as

$$\rho_i^{[0,255]} = \rho_{i,o}^{[0,255]} \pm \text{round}(A_{\text{noise}}\sigma_{\text{shot noise}})\,. \tag{22}$$

Figure 5 shows the median and maximum angular errors in sRGB as a function of A_{noise}. The corresponding numerical values can be found in Table 2. Figure 4 shows examples of real and retrieved illuminants spectra for an increasing amount of shot noise.

The minimization on $f_e(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ is not sensitive to the quantization error and quite robust to shot noise with only a moderate change in performance: The median (resp. maximum) angular error increases from $e_{ang} = 2.96°$ (resp. $e_{ang} = 13.00°$) for $A_{\text{noise}} = 0$ to $e_{ang} = 3.42°$ (resp. $e_{ang} = 18.64°$) for $A_{\text{noise}} = 1$.

4.C. Comparison with the Shades of Gray algorithm

We compare the performance of our algorithm with a single image one. The *gray edge* algorithm [16] relies on the gray world assumption, yet not applied on the color image, but on its derivative. While this algorithm shows good performance, it can only be applied to real images and is thus not adapted to a simulation on synthetic data. Instead, we use the *shades of gray* (SoG) method [2], which also relies on the gray world assumption, but employs the Minkowsky p-norm

$$\| \mathbf{X} \|_p = \Big\{ \sum_{i=1}^{N} |X_i|^p \Big\}^{1/p}, \tag{23}$$

where $\mathbf{X} = [X_1, ..., X_N]$ is a vector in $\mathbb{R}^N$. More precisely, rather than estimating the white-point from the scene average, it applies the gray world algorithm on the p-norm of the image.

The gray world and the max-RGB algorithms are two particular cases of SoG with, respectively, $p = 1$ and $p = \infty$. In [2], the authors showed that the best results are obtained for $p = 6$. In [16], the authors report that, while the $p = 6-$norm gray edge performs better than the $p = 6-$norm SoG, both return good results, with angular errors computed on a set of real images of 5.7° and 6.3°, respectively.

Table 3 reports the resulting median and maximum error for the SoG algorithm for all reflectance types and varying N and N_p and should be compared with Table 1 reporting the equivalent results for the gradient descent method. Table 2 reports the results for SoG in the case $N = 4$ and $N_p = 6$ for an increasing amount of shot noise.

The results obtained by gradient descent are generally better than the the ones obtained with SoG, except when considering a larger number of MacBeth reference surfaces ($N_p = 10$). Indeed, taking $N_p = 10$ surfaces out of the 24 MacBeth reflectances ensures are large variety of colors, resulting in a good performance of SoG. Nevertheless, our algorithm generally outperforms SoG in terms of median angular error, that is, exploiting redundancy across images returns significantly better results.

4.D. Comparison with D'Zmura and Iverson's algorithm

D'Zmura and Iverson's algorithm [3] is extremely sensitive to any deviation from ideal data and only leads to perfect recovery of the illuminants and reflectances if their spectra can be exactly represented by linear models. Moreover, it is only applicable for specific numbers of illuminants, surfaces, and descriptors. We consider now $N_p = 3$ surfaces viewed under $N = 2$ illuminants. The algorithm must then be applied with $N_\varepsilon = 3$ illuminant and $N_\sigma = 3$ reflectance descriptors. In the section, we present comparative results with our algorithm, which we run with the same parameters, i.e., $N_p = 3$, $N = 2$, $N_\varepsilon = 3$, and $N_\sigma = 3$.

To ensure a perfect recovery of the descriptors, we first build a set of RGB sensor values from ideal tri-dimensional spectra. To do so, we compute the three first basis functions for the illuminants and reflectances by principal component analysis over the 13 standard illuminants and the 24 MacBeth reflectances. Then, for each of them, we build back the tri-dimensional spectra, i.e., compute the $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ and the $(\sigma_1, \sigma_2, \sigma_3)$ descriptors minimizing, respectively,

$$\| \varepsilon_1 \mathcal{E}_1(\lambda) + \varepsilon_2 \mathcal{E}_2(\lambda) + \varepsilon_3 \mathcal{E}_3(\lambda) - E(\lambda) \| \tag{24}$$

and

$$\| \sigma_1 \mathcal{S}_1(\lambda) + \sigma_2 \mathcal{S}_2(\lambda) + \sigma_3 \mathcal{S}_3(\lambda) - S(\lambda) \| \tag{25}$$

and then construct the approximated spectra

$$E(\lambda) \simeq \varepsilon_1 \mathcal{E}_1(\lambda) + \varepsilon_2 \mathcal{E}_2(\lambda) + \varepsilon_3 \mathcal{E}_3(\lambda) \tag{26}$$

and

$$S(\lambda) \simeq \sigma_1 \mathcal{S}_1(\lambda) + \sigma_2 \mathcal{S}_2(\lambda) + \sigma_3 \mathcal{S}_3(\lambda)\,. \tag{27}$$

We finally build RGB sensors responses from these approximated illuminants (26) and reflectances (27) with the image formation model (6). We apply both algorithms on these perfectly tri-dimensional RGB sensor responses and observe their outcomes for an increasing amount of shot noise (22).

Figure 6 reports the median and maximum angular errors for both algorithms as a function of the shot noise amplitude. Our algorithm's performance under the influence of shot noise is only slightly decreased with a median (resp. maximum) angular error ranging from 3.3° (resp. 13.9°) for $A_{\text{noise}} = 0$ to 3.9° (resp. 17.9°) for $A_{\text{noise}} = 1$. The median angular error for D'Zmura and Iverson's algorithm [3] is around 25° (resp. 150°) even in the absence of noise. While this result may seem surprising, the instability of the algorithm for $A_{\text{noise}} = 0$ is explained by its high sensitivity to small deviations from ideal data and, in the present case, the error is introduced by the quantization of the digital values. If we run the same experiment for smaller A_{noise} while keeping all RGB in floating point, we see that the algorithm indeed performs perfectly and returns a zero median angular error in the absence of noise. It starts

to diverge for small values of A_{noise} and reaches its maximum median error for $A_{\text{noise}} \simeq 0.08$, as illustrated in the second row of Figure 6. While our algorithm only approximates the illuminants and never guarantees a perfect recovery of the descriptors, it is more stable than D'Zmura and Iverson's in more realistic conditions.

5. Conclusions

We proposed to solve for color constancy using a set of images containing a small number of redundant surfaces. This framework corresponds to fixed location cameras, which take images presenting static background elements. We assume that these elements' reflectances, though unknown, remain constant across images. While this assumption still holds in practice, the application on real images will rise several issues, namely variation of intensity across the images, mixed illuminants, shadows, and specularities, whose influences remain to be investigated. We invert a series of image formation models in parallel for a set of test illuminants and, by forcing the output reflectances to match, deduce the illuminants under which each image was captured. We define an error function reaching its minimum when the output reflectances match. The redundant elements are modeled by synthetic RGB patch values. We demonstrate the validity of our method using such patches generated from real daylight illuminants and three types of reflectances.

We observed that the estimation becomes more robust as the number of illuminants N increases. The minimization by gradient descent on $f_e = f_e(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ showed good performances, including in the presence of shot noise. It also generally outperformed the results obtained with the single image algorithm *shades of gray*, especially in the presence of shot noise.

The principal comparative advantage of our method with D'Zmura and Iverson's algorithm [3] is that it only imposes weak constraints on the number of images N, reflectance descriptors N_σ, and surfaces N_p. While the quality of the illuminant estimation will depend on these parameters, the formalism presented here can be applied to any $N \geqslant 2$, $N_\sigma \geqslant 3$, $N_p \geqslant 1$, and $N_\varepsilon \geqslant 1$. Moreover, it is significantly less sensitive to noise. The main comparative disadvantage is that our method requires many iterations and is thus slower.

It must be emphasized that we did not impose any positivity or shape constraint on the illuminant and reflectance spectra, but only set the initial descriptors to be representative of actual illuminants, such as initializing all illuminants to D65. Moreover, our illuminant estimation method by gradient descent uses linear models to represent illuminant and reflectance spectra, but with the advantage of not imposing upper limits on the number of their descriptors. The framework developed here can thus be to extended linear models of arbitrarily high dimensions; that is, it could be applied to non-daylight illuminants, which require to be represented by more than three basis functions.

6. Acknowledgement

The authors acknowledge Professor Graham D. Finlayson for the helpful suggestions and discussions.

References

1. S. Hordley, "Scene illuminant estimation: Past, present, and future," Color Research and Application **31(4)**, 303-314 (1986).
2. G.D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proceedings of the 12th IS&T/SID Color Imaging Conference*, pp. 37-41, (2004).
3. M. D'Zmura and G. Iverson, "Color constancy. I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces," J. Opt. Soc. Am. A **10(10)**, 2148-2165 (1993).
4. M. D'Zmura and G. Iverson, "Color constancy. II. Results for two-stage linear recovery of spectral descriptions for lights and surfaces," J. Opt. Soc. Am. A **10(10)**, 2166-2180 (1993).
5. M. D'Zmura and G. Iverson, "Color constancy. III. General linear recovery of spectral descriptions for lights and surfaces," J. Opt. Soc. Am. A **11(9)**, 2389-2400 (1994).
6. J. Cohen, "Dependency of spectral reflectance curves of the Munsell color chips," Psychon. Sci. **1**, 369-370 (1964).
7. L.T. Maloney, "Evaluation of linear models of surface spectral reflectance with a small number of parameters," J. Opt. Soc. Am. A **3(10)**, 1673-1683 (1986).
8. J.P.S. Parkkinen, J. Hallikainen, and T. Jaaskelainen, "Characteristic spectra of Munsell colors," J. Opt. Soc. Am. A **6(2)**, 318-322 (1989).
9. D.H. Marimont and B.A. Wandell, "Linear Models of surface and illuminant spectra," J. Opt. Soc. Am. A **9(11)**, 1905-1913 (1992).
10. S. Westland, J. Shaw and H. Owens, "Colour statistics of natural and man-made surfaces," Sensor Review **20(1)**, 50-55 (2000).
11. D.B. Judd and D.L. MacAdam and G. Wyszecki, "Spectral distribution of typical daylight as a function of correlated color temperature," J. Opt. Soc. Am. A **54(8)**, 1031-1036 (1964).
12. A.B. Israel and T.N.E. Greville, *Generalized Inverses: Theory and Applications* (John Willey & Sons, 1974).
13. E.H. Land and J.J. McCann, "Lightness and Retinex theory," J. Opt. Soc. Am. A **61(1)**, 1-11 (1971).
14. G. Buchsbaum, "A spatial processor model for object color perception," Journal of the Franklin Institute **310(1)**, 1-26 (1980).

15. R.G. Gershon, A.D. Jepson, and J.K. Tsotsos, "From [R,G,B] to surface reflectance: Computing color constant descriptors in images," Perception **1**, 755-758 (1988).
16. J. van de Weijer and T. and Gevers, "Color constancy based on the gray-edge hypothesis," in Proceedings of IEEE International Conference on Image Processing, (IEEE, 2005), p. 722.
17. D.A. Forsyth, "A novel algorithm for color constancy," International Journal of Computer Vision **5(1)**, 5-36 (1990).
18. K. Barnard, L. Martin, and B. Funt, "Colour by correlation in a three-dimensional colour space," in *Proceedings of the 6th European Conference on Computer Vision*, pp. 275-289 (2000).
19. G.D. Finlayson and S. Hordley, "Improving gamut mapping color constancy," IEEE Trans. Image Processing **9(10)**, 1057-7149 (2000).
20. G.D. Finlayson, S.D.Hordley, and P.M. Hubel, "Color by correlation: A simple, unifying framework for color constancy," IEEE Pattern Analysis and Machine Intelligence **23(11)**, 1209-1221 (2001).
21. L.T. Maloney and B.A. Wandell, "Color constancy: A method for recovering surface spectral reflectance," J. Opt. Soc. Am. A **3(1)**, 29-33 (1986).
22. C. Lu and M.S. Drew, "Practical scene illuminant estimation via flash/no flash pairs," in *Proceedings of the 14th IS&T/SID Color Imaging Conference*, pp. 84-89 (2006).
23. J.M. DiCarlo F.Xiao and B.A. Wandell, "Illuminating illumination," in *Proceedings of the 9th IS&T/SID Color Imaging Conference*, pp. 27-34 (2001).
24. G.D. Finlayson S.D. Hordley, P. Morovic, "Colour constancy using the chromagenic constraint," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2005), p. 1079.
25. G.D. Finlayson S.D Hordley, and P. Morovic, "CMP-C04-06: Chromagenic Colour Constancy," Technical Report, School of Computing Sciences, University of East Anglia (2006).
26. G.D. Finlayson, "Color constancy and a changing illumination," Human Vision, Visual Processing, and Digital Display **1**, 6-10 (1994).
27. G.D. Finlayson, B.V. Funt, and K. Barnard, "Color constancy under varying illumination," in Proceedings of IEEE International Journal of Computer Vision (IEEE, 1995), p.720.
28. S. Hordley and G.D. Finlayson, "Re-evaluating colour constancy algorithms," in Proceedings of the IEEE International Conference on Pattern Recognition (IEEE, 2004), p. 76.
29. K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms - Part I: Methodology and experiments with synthesized data," IEEE Trans.

Image Processing **11(9)** 972-983 (2002).

30. K. Barnard, L. Martin, A. Coath, and B. Funt, "A comparison of computational color constancy algorithms - Part II: Experiments with image data," IEEE Trans. Image Processing **11(9)**, 985-996 (2002).
31. B. Funt, V. Cardei, and K. Barnard, "Learning color constancy," in *Proceedings of the 4th IS&T/SID Color Imaging Conference*, pp. 58-60 (1995).

Table 1. Median angular errors e_{ang} for the Gradient Descent in absence of noise.

		MacBeth		Munsell		Natural	
N	N_p	med e_{ang}	max e_{ang}	med e_{ang}	max e_{ang}	med e_{ang}	max e_{ang}
2	6	3.96	18.04	3.91	18.20	3.98	17.98
3	6	3.38	16.02	3.33	15.40	3.42	15.62
4	6	2.96	13.00	2.90	12.90	2.95	13.46
6	6	2.52	11.79	2.43	11.47	2.57	11.34
6	10	2.51	11.36	2.44	11.11	2.59	11.33
10	6	2.19	8.37	2.10	8.30	2.28	9.81

The algorithm was run 10^3 times using different combinations of N_p patches and N real daylight illuminants. The illuminants are described by three illuminant descriptors and the error function $f_e = f_e(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ is minimized by gradient descent.

Table 2. Comparative median angular errors e_{ang} in presence of shot noise.

	Gradient descent		Shades of Gray	
A_{noise}	med e_{ang}	max e_{ang}	med e_{ang}	max e_{ang}
0.0	2.96	13.00	4.16	36.03
0.1	2.97	13.08	4.21	36.06
0.2	2.95	12.89	4.20	37.05
0.3	2.99	12.90	4.22	34.90
0.4	3.00	13.22	4.28	35.88
0.5	3.03	14.93	4.29	36.96
0.6	3.12	14.15	4.42	35.89
0.7	3.25	15.23	4.49	33.77
0.8	3.25	15.90	4.79	33.18
0.9	3.34	15.06	4.88	36.36
1.0	3.42	18.64	5.22	33.18

The algorithm was run 10^3 times using $N_p = 6$ patches and $N = 4$ daylight illuminants. This table reports the median and maximum angular errors e_{ang} as a function on the amount of shot noise added to the sensor responses for the gradient descent on $f_e = f_e(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ and for Shades of Gray.

Table 3. Median angular errors e_{ang} for the Shades of Gray in absence of noise.

		MacBeth		Munsell		Natural	
N	N_p	med e_{ang}	max e_{ang}	med e_{ang}	max e_{ang}	med e_{ang}	max e_{ang}
2	6	4.06	34.18	5.09	31.89	7.78	27.19
3	6	4.13	36.03	5.12	32.61	7.75	27.03
4	6	4.16	36.03	5.10	32.74	7.83	26.91
6	6	4.07	35.02	5.09	32.65	7.77	27.19
6	10	1.53	22.26	4.01	22.85	6.30	29.84
10	6	4.13	35.71	5.08	32.86	7.73	27.07

The algorithm was run 10^3 times using different combinations of N_p patches and N real daylight illuminants.

7. List of Figure Captions

Fig.1 Images captured with fixed location cameras exhibit changes in lighting and content, but also contain constant static elements.

Fig.2 Examples of measured daylight illuminants

Fig.3 Examples of real (solid lines) and retrieved (dashed lines) illuminant spectra. Each of the three rows represents a different combination of $N = 4$ images and $N_p = 6$ reference surfaces. The median angular errors for rows 1, 2, and 3 are, respectively, 0.23°, 3.02°, and 4.76°.

Fig.4 Examples of real (solid lines) and retrieved (dashed lines) illuminant spectra when noise is added to the sensor responses. The spectra are the ones of Figure 3, plots (a)-(d), for $A_{\text{noise}} = 0.2$, 0.6, and 1.0. The median angular errors corresponding to the rows 1, 2, and 3 are, respectively, 0.51°, 1.58°, and 2.73°.

Fig.5 The left and right graphs show, respectively, the median and maximum angular errors in sRGB obtained by gradient descent on $f_e = f_e(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ (top row) and for the Shades of Gray algorithm (bottom row) for an increasing amount of shot noise. The simulations were run for $N = 4$ illuminants and $N_p = 6$ reference surfaces.

Fig.6 The left and right graphs show, respectively, the median and maximum angular errors in sRGB obtained for both D'Zmura and Iverson's and our algorithm. The results are represented by, respectively, the blue triangles and the red circles and squares.

Figure 1. Images captured with fixed location cameras exhibit changes in lighting and content, but also contain constant static elements.

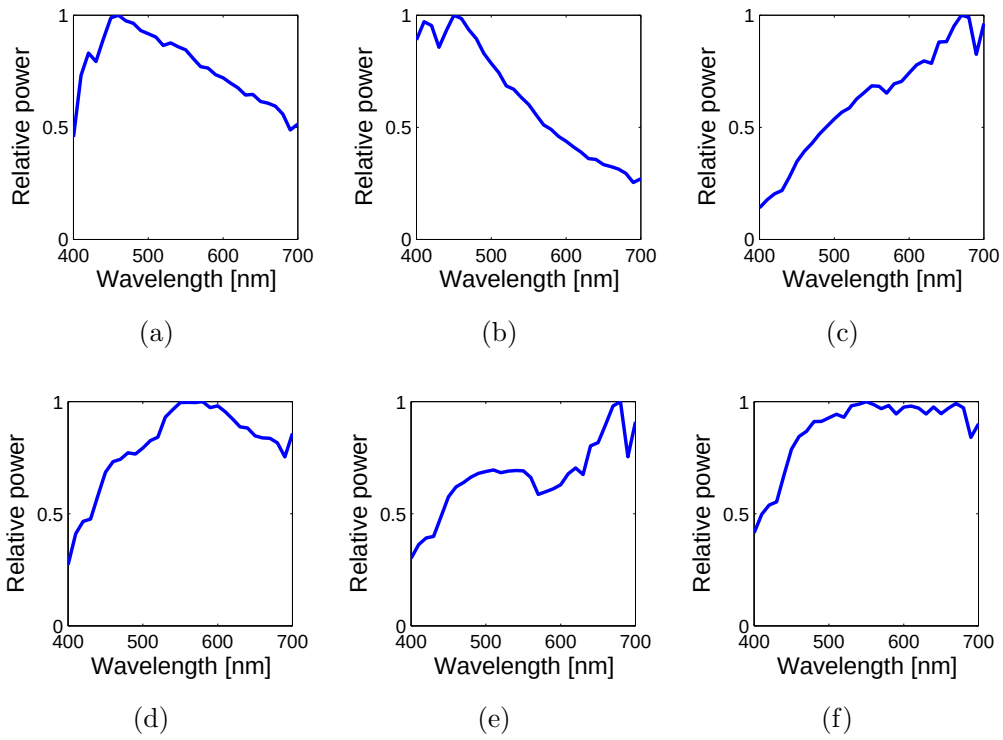

Figure 2. Examples of measured daylight illuminants.

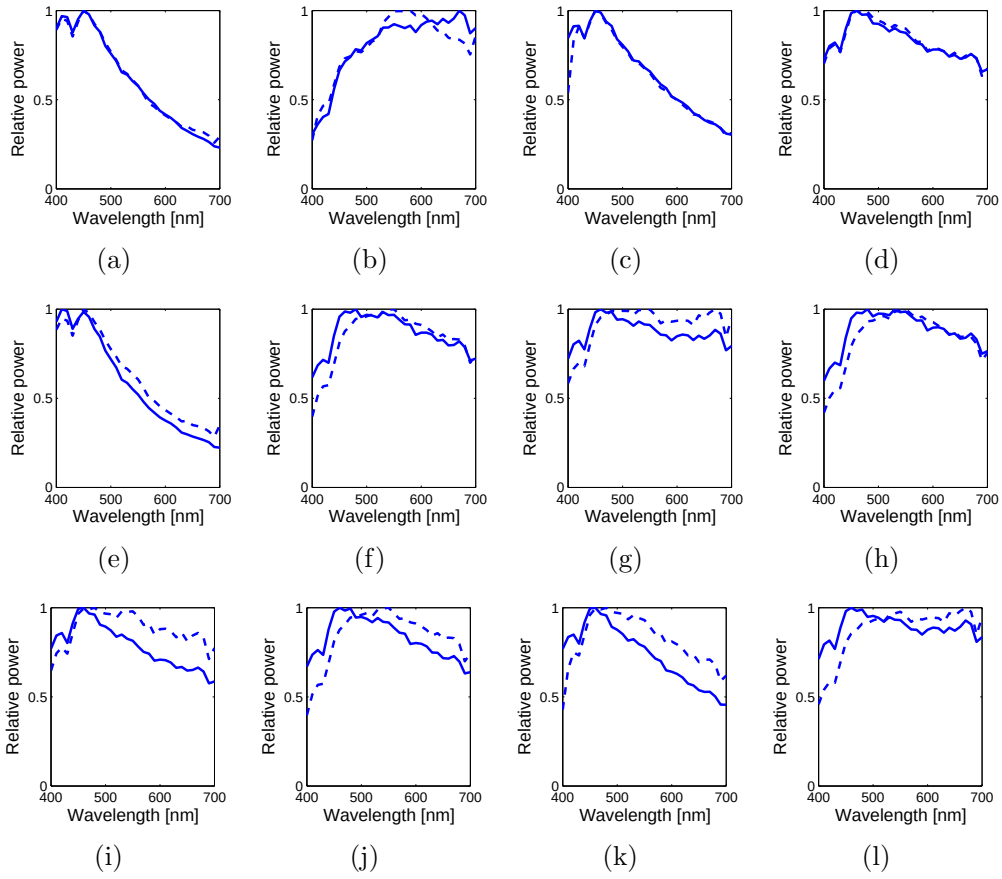

Figure 3. Examples of real (solid lines) and retrieved (dashed lines) illuminant spectra. Each of the three rows represents a different combination of $N = 4$ images and $N_p = 6$ reference surfaces. The median angular errors for rows 1, 2, and 3 are, respectively, 0.23°, 3.02°, and 4.76°.

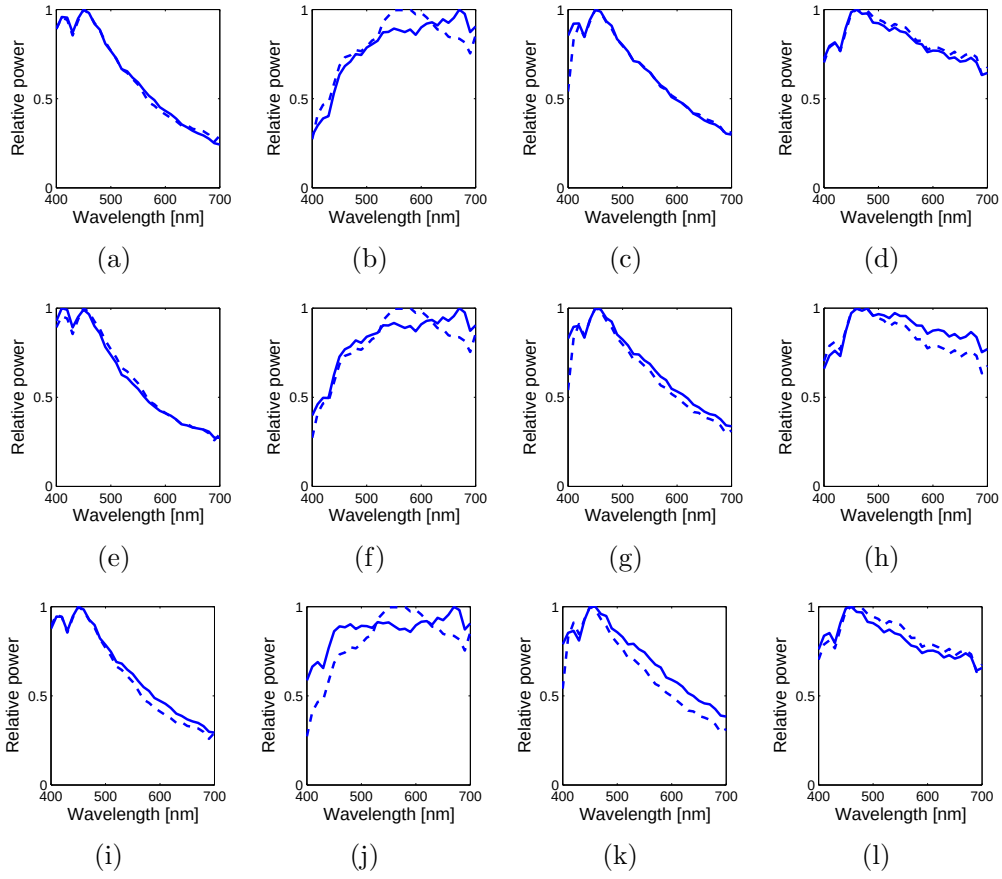

Figure 4. Examples of real (solid lines) and retrieved (dashed lines) illuminant spectra when noise is added to the sensor responses. The spectra are the ones of Figure 3, plots (a)-(d), for $A_{\text{noise}} = 0.2$, 0.6, and 1.0. The median angular errors corresponding to the rows 1, 2, and 3 are, respectively, 0.51°, 1.58°, and 2.73°.

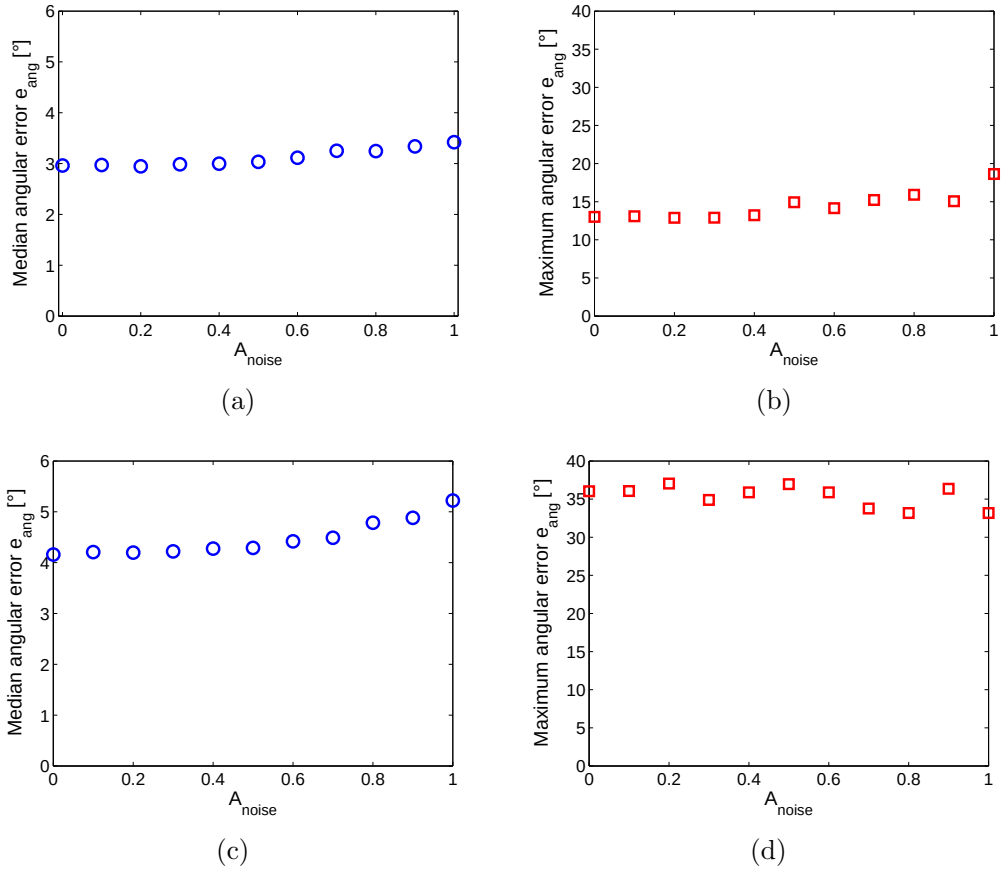

Figure 5. The left and right graphs show, respectively, the median and maximum angular errors in sRGB obtained by gradient descent on $f_e = f_e(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ (top row) and for the Shades of Gray algorithm (bottom row) for an increasing amount of shot noise. The simulations were run for $N = 4$ illuminants and $N_p = 6$ reference surfaces.

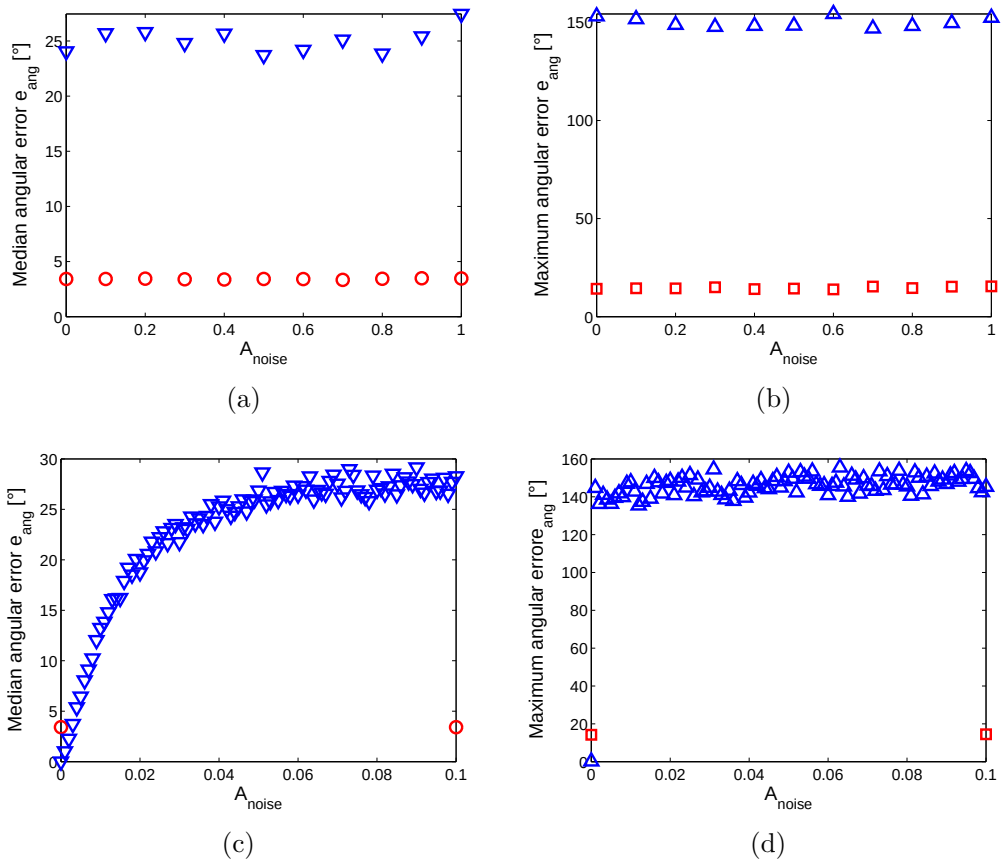

Figure 6. The left and right graphs show, respectively, the median and maximum angular errors in sRGB obtained for both D'Zmura and Iverson's and our algorithm. The results are represented by, respectively, the blue triangles and the red circles and squares.