



**HAL**  
open science

## Perceptual Color Characterization of Cameras

Javier Vazquez-Corral, David Connah, Marcelo Bertalmío

► **To cite this version:**

Javier Vazquez-Corral, David Connah, Marcelo Bertalmío. Perceptual Color Characterization of Cameras. 2014. hal-01015537v1

**HAL Id: hal-01015537**

**<https://hal.science/hal-01015537v1>**

Preprint submitted on 27 Jun 2014 (v1), last revised 29 Nov 2014 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Perceptual Color Characterization of Cameras

Javier Vazquez-Corral, David Connah, and Marcelo Bertalmío

**Abstract**—Color camera characterization, mapping outputs from the camera sensors to an independent color space such as  $XYZ$ , is an important step in the camera processing pipeline. Until now, this procedure has been primarily solved by using a  $3 \times 3$  matrix obtained via a least-squares optimization. In this paper, we propose to use the spherical sampling method, recently published by Finlayson *et al.* to perform a perceptual color characterization. In particular, we search for the  $3 \times 3$  matrix that minimizes three different perceptual errors, one pixel-based, and two spatial-based. For the pixel-based case, we minimize the CIE Delta E error, while for the spatial-based case we minimize both the SCIELAB error, and the CID error measure. Our results demonstrate an improvement of approximately 3% for the Delta E error and 15% for the two other error measures.

**Index Terms**—Color characterization, perceptual correction, camera sensor response.

## I. INTRODUCTION

AT first glance, it would seem that for a camera to accurately capture colors matching our perception, the color triplets obtained by the camera sensor(s) should correspond to the cone responses of the human visual system. But this is never the case, as Figure 1 shows, and since the spectral sensitivities of sensor and cones are different, the responses must also be different.

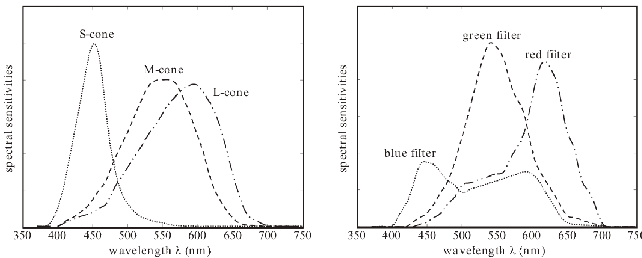


Fig. 1. Spectral sensitivities of: (a) the three types of cones in a human eye, and (b) a typical digital camera. Images from [1].

There are several reasons for this difference, like the fact that it is difficult to tune the spectral response of the pigments or dyes of the color filter arrays, and that having spectral sensitivities with a large amount of overlap (as in the responses of medium and long wavelength cones) would not be practical from a signal-to-noise point of view [2]. But while emulating cone responses is not practical for image capture, it is essential in the subsequent processing of the image signal: the stimulus the scene would have produced in the human visual system must be estimated as accurately as possible [2]. This is why

Javier Vazquez-Corral and Marcelo Bertalmío are with the Department of Information and Communications Technologies, Universitat Pompeu Fabra. David Connah is with the Centre for Visual Computing, University of Bradford.

Manuscript received XXX; revised YYY

we must be able to transform the  $(R, G, B)$  values of the sensor into  $(X, Y, Z)$  tristimulus values, i.e. go from  $RGB$  into CIE  $XYZ$ , which we recall is a color space that uses the color-matching functions  $\bar{x}, \bar{y}, \bar{z}$  of a standard observer, obtained from perceptual color-matching experiments.

We can transform  $(R, G, B)$  into  $(X, Y, Z)$  by imposing the Luther-Ives condition [3]: that the sensor response curves are a linear combination of the color matching functions. Manufacturing processes and the properties of the materials used make it difficult to adjust at will the sensor response curves, and the Luther-Ives condition is usually not met in practice [4]. Despite this fact a three-channel camera with three arbitrary sensor response curves is able to estimate the tristimulus values of an object as long as the object's spectral reflections are always composed of three principal components and they don't change steeply with respect to wavelength [4]. This implies that with a linear transformation we can go from the observed  $(R, G, B)$  triplet to its corresponding  $(X, Y, Z)$  tristimulus value, and this process is called *color characterization*. We must point out that the transform itself could also be a non-linear mapping [5], [6]. While the non-linear approach generally uses more parameters, and hence has greater potential to reduce error, the linear approach remains very popular. This is due to two key advantages: 1) with linear methods scaling the  $RGB$  responses results in a direct scaling of  $XYZ$  values, and 2) linear methods preserve the hue-planes of the input  $RGB$  image [7].

A standard way for camera manufacturers to perform color characterization is the following [4]:

- 1) Build a set of  $n$  test patches of representative or important colors.
- 2) Under controlled conditions, with a known illuminant (e.g. D65), measure the tristimulus values of the patches with a tristimulus colorimeter obtaining  $(X_i, Y_i, Z_i)$ ,  $1 \leq i \leq n$ .
- 3) Under the same conditions, use the camera to measure the  $(R, G, B)$  values of the patches, obtaining  $(R_i, G_i, B_i)$ ,  $1 \leq i \leq n$ .
- 4) A linear transformation, a  $3 \times 3$  matrix called the *colorimetric matrix*, gives an estimated tristimulus value  $(\hat{X}_i, \hat{Y}_i, \hat{Z}_i)$  from  $(R_i, G_i, B_i)$ . This matrix is computed so as to minimize the total visual color difference  $J$ , which is a weighted sum of the color differences  $\Delta E$  (computed for instance in a CIE perceptual uniform color space) between the target tristimulus  $(X_i, Y_i, Z_i)$  and its estimate  $(\hat{X}_i, \hat{Y}_i, \hat{Z}_i)$ , for each patch  $i$ ,  $1 \leq i \leq n$ :

$$J = \sum_{i=1}^n w_i \Delta E(X_i, Y_i, Z_i, \hat{X}_i, \hat{Y}_i, \hat{Z}_i),$$

where  $w_i$  are the weights for the different patches. The colorimetric matrix is usually obtained through least-squares minimization.

In [8] it is noted that the above procedure has the problem that the white point is not preserved, i.e. white in  $RGB$  is not mapped to white in the CIE  $XYZ$  color space where white in  $XYZ$  represents the  $XYZ$  response to a perfect reflecting diffuser under a specified illuminant; an additional term can be added to  $J$  in order to prevent this [9], and more accurate and robust techniques have also been proposed [10].

Some cameras come with several pre-set matrices computed under different illuminations. For instance, using the matrix for fluorescent lighting removes a noticeable green cast that would otherwise be present if we used a matrix computed with a standard illuminant like D65 or D50; other pre-sets may correspond for instance to a “film look” (with de-saturated colors), or may give a very vivid color palette. These pre-set matrices can also be adjusted manually so as to achieve a certain image look, since changing the colorimetric matrix affects hue and saturation (the white point is preserved, though, and color matrix adjustment must not be confused with white balance). Video cameras were the first to incorporate the possibility of modifying the colorimetric matrix, so that multiple cameras in live broadcasts could be color matched and there appeared no color jumps when switching from one camera to another [11].

In this paper we propose an approach to characterization which retains the advantages of linear transforms, while explicitly minimizing perceived differences in perceptual color spaces. The approach works in two stages: in the first stage an  $RGB$  to  $XYZ$  transform is computed that minimizes the least-squares error in the usual way; in the second stage this transform is adjusted to reduce a defined perceptual measure. This adjustment is done using the method of spherical sampling of Finlayson et al. [12], which generates a set of discrete putative transforms that deviate systematically from the first transform. We will show that choosing the best of these transforms to minimize the perceptual error results in an improved characterization performance. Let us stress here that, to the authors’ knowledge, this is the first time spherical sampling is used out of the spectral sharpening domain. This may open the door to the application of spherical sampling to further applications in different domains.

This is valid for any kind of digital camera, be it for video or still photography, using CCD or CMOS sensors. Our method provides camera manufacturers and also color researchers and professional photographers and cinematographers with a simple way to record images whose colors are more faithful, in a perceptual sense, to those in the real world scene where the images were taken.

## II. BACKGROUND

### A. Sources of Error

The error in the characterization - i.e. the difference between the target and estimated  $XYZ$  values - is dependent upon the sensor sensitivity of the camera, the reflectance spectra of the objects in the scene, and the illuminant of the scene. In the case that the  $RGB$  sensor sensitivities are within an exact  $3 \times 3$  transform of the  $XYZ$  color matching functions (CMFs), or equivalently, the  $RGB$  bases and  $XYZ$  CMFs span the

same linear subspace, then the error in the characterization is guaranteed to be zero [13].

In practice this condition is rarely met: as we mentioned, there are inherent variations in the sensor manufacturing process, as well as conflicting design considerations such as robustness to noise.

One approach to solving this problem would be to map  $RGB$  sensors direct to  $XYZ$  CMFs using the linear transform  $T$  minimizing:

$$\arg \min_T \|S_{xyz} - S_{rgb}T\|_2 \quad (1)$$

Where  $S_{xyz}$  denotes an  $m \times 3$  matrix of  $XYZ$  color matching functions sampled at  $m$  discrete wavelengths,  $S_{rgb}$  is an  $m \times 3$  matrix of camera  $RGB$  responses, and  $T$  is a  $3 \times 3$  linear transform. This can be shown to be the best approach when no other information about spectra is known (the maximum ignorance assumption) [14]; i.e. when spectra are pure noise stimuli. However this is not the case for real spectra, which are both positive and smoothly varying functions of wavelength [15], [16]. As such, the structure of reflectances becomes an important factor in defining the optimal transform. For this reason, the minimization is usually performed as

$$\arg \min_T \|M_{xyz} - M_{rgb}T\|_2 \quad (2)$$

Where  $M_{xyz}$  denotes an  $n \times 3$  matrix representing  $XYZ$  values for a set of illuminants and reflectances, and  $M_{rgb}$  is an  $n \times 3$  matrix representing the  $RGB$  values for the same set of illuminants and reflectances.

In a world where reflectances can be described by three, or fewer, parameters in a linear model, the characterization process is again guaranteed to result in zero error [17]. In practice this condition is not met: real reflectances are described by more than 3 parameters [18], and the characterization error will be non-zero.

In addition to these fundamental sources of error, the imaging process will contain various sources of noise. These factors are not considered in our coverage here.

### B. Perceptual Error

The goal of characterization is to ensure that colors remain stable regardless of the imaging device used. Ultimately, when considered with a display device as part of the color management pipeline, the captured colors should appear to be as similar to the original scene as possible. The implication for the characterization error is that errors should reflect the perceived difference between colors.

The light entering the human eye is firstly captured by the photoreceptors, and is immediately subject to neural processing; this comes in the form of adaptation, spatial channel mixing, and both pooling and contrasting responses across receptors, even before the visual signal has left the retina. Interactions between neurons in the visual cortex, and at higher processing sites, mean that relationships between perceived colors are complex, and the color of a stimulus cannot be defined without taking into account its spatial surround and context.

Some of the behavior of the human visual system can be predicted by ‘color appearance models’. These perceptual models approximate the transformation from raw cone-responses (which can be derived from  $XYZ$  values) to perceptual attributes such as lightness, colorfulness and hue. An early color appearance model, the CIELab color space [19], was derived specifically to map colors into a space where color differences correlate closely with Euclidean distances measured in CIELab. Other more recent models, such as CIECAM02 [20], provide predictors of the perceptual attributes of the colors, but CIELab color difference formulae (which measure Euclidean distance in CIELab space) are still used extensively for color difference calculations, and are being incrementally improved alongside the color appearance models, e.g. CIE  $\Delta E$  2000 [21].

Color appearance models may take into account the non-linear response of the visual system to light transduced by the photoreceptors, including effects such as adaptation and opponent color mechanisms, but they have a relatively basic notion of spatial relationships between colors. To model these effects multiple image-difference predictors, and image-appearance models, have been proposed. These include the relatively simple S-CIELAB model, which incorporates spatial filtering into the CIELab calculation [22], up to more complex models such as iCAM [23], which includes spatial filtering and local adaptation in a more complete color appearance model.

Recently, some perceptually-based image quality metrics have been presented. In particular, the CID quality measure [24] computes a color difference metric for a complete image taking into account multiple imaging factors in a perceptually-based color space, in particular hue, chroma, lightness, lightness-contrast and lightness-structure.

### C. Linear vs. Non-linear

The characterization process involves a mapping from one 3-vector,  $RGB$ , to a second 3-vector,  $XYZ$ . This can be done through either linear or non-linear mapping. Non-linear approaches such as polynomials [5], neural networks [6], or geometric approaches [25] have been favored in some applications, as they introduce a greater number of parameters, and hence can reduce characterization error. Nonetheless, linear approaches are still popular, and are widely employed in default conversions used by color camera manufacturers. Linear transforms, while potentially less accurate, preserve two important properties: scalability, and the preservation of hue planes, which we will briefly describe now.

The scalability property means that the characterization is equally valid when the exposure duration of the camera is changed - this is not true for most non-linear characterizations, where a change in exposure duration will scale  $RGB$  values, but may result in angular shifts in the  $XYZ$  vectors, which in turn correlate to visible color shifts. In general, when applying non-linear methods the transformation matrix needs to be recalculated each time exposure is altered, although recently some solutions have been proposed for retaining scalability while increasing the number of free parameters [7], [26], [27].

Similarly, the planes of constant hue in an  $RGB$  image, i.e. the plane of  $RGB$  responses to a matte surface with some

elements in light and some in shadow, will be preserved by a linear transformation, but with a non-linear transform there may be induced color shifts. This issue is covered in detail in [7] where a method is proposed to circumvent this problem.

We note here that the scalability and constant hue planes properties follow directly from the linear response to light of CCD / CMOS sensors [28]. Any non-linearity is added in post-processing on the camera, such as a display gamma or the result of noise suppression. It is therefore more practical to directly capture the images in RAW mode, thus circumventing the in-camera image-processing.

## III. SPHERICAL SAMPLING FOR CAMERA CHARACTERIZATION

The goal of the present paper is to build a  $3 \times 3$  matrix transform that can minimise perceptual errors, rather than errors in  $XYZ$  space. Human color perception is a non-linear phenomenon, and therefore, methods based on the minimization of a linear measure, such as least-squares with the  $L_2$  norm, are not well adapted to this goal. To achieve this we need a method for performing a constrained search of the space of possible sensors.

Spherical sampling [29], [12] provides a means for discretely sampling points on a sphere and relating them to sensors. In particular, given a set of sensors, spherical sampling converts each sensor into a point in the sphere and, from there, samples discrete points on the sphere within a pre-specified angular distance from the starting point. Each of these sample points represents a new sensor. So, for example, given a set of three  $RGB$  sensors, these can be mapped to three separate points on a sphere. Multiple points, say 10000, can then be sampled close to each sensor on the sphere. In this example this results in 10000 new R sensors, 10000 new G sensors and 10000 new B sensors. A discrete set of plausible sensors can then be generated by enumerating all possible combinations of the R, G, and B sensors. These potential sensors can then be compared using an error metric.

There are alternative methods to exploring the sensor space. For example, one could sample the 9 coefficients of the transform matrix  $T$  directly. However, spherical sampling has several practical and theoretical advantages.

First, the problem of uniform sampling on a sphere is well understood: we can be sure that sampled points are equally distributed in all different directions, i.e. without over- or under-sampling in any given direction. Sampling the matrix  $T$  directly does not ensure this, which can slow the convergence and is less likely to find the local error minimum.

Secondly, using spherical sampling it is possible to define the extent over which the sampling will take place in the sensor space, since the number of points that are sampled in the sphere is directly related to the mean distance between any two solutions. Once again, this information is not available by directly sampling on the matrix, since the same increment in different coefficients of the matrix might represent very different distances in terms of the resulting sensors.

Finally, spherical sampling allows us to define the distance from the starting point with a simple and intuitive measure:

the angular distance. Using this distance it is possible to avoid solutions that are too far away from the starting point. This has been shown to be important in other applications such as color ratio stabilisations [12], where it was shown that failing to enforce a distance constraint could result in extreme sensor shifts such as a green sensor being modified to a red one.

### A. Details of the method

The problem of camera characterization can be thought as the conversion from an original set of  $RGB$  sensors to a different set,  $XYZ$ . In particular, in the domain of camera characterization the problem is to map each camera response  $RGB$  to an  $XYZ$  tristimulus value. When using a  $3 \times 3$  linear transform each new coordinate is a linear combination of  $RGB$ , with weights given by a column of the transform matrix, e.g.

$$X = t_{11}R + t_{21}G + t_{31}B \quad (3)$$

With the appropriate processing (that we discuss in the next paragraph), the vector  $[t_{11}t_{21}t_{31}]$  (along with the other columns of the transform  $T$ ) can be considered as a point on a sphere. Spherical sampling then provides a method for systematically exploring sensor sets by choosing three points on the sphere as the three columns of the transform matrix.

Mathematically, let us represent our original camera sensors  $S$  as an  $m \times 3$  matrix where  $m$  denotes the number of discrete wavelengths at which the sensor functions are sampled (this value is typically 31, which comes from sampling the visible spectrum 400nm to 700nm at 10nm intervals) and 3 is the number of sensors, which for most applications corresponds to the red, green and blue color channels. We perform the reduced singular value decomposition (SVD) of these sensors in order to obtain a basis.

$$S = U \cdot \Sigma \cdot V^t \quad (4)$$

where  $U$  is an orthogonal matrix with dimension  $m \times 3$ ,  $\Sigma$  is a diagonal  $3 \times 3$  matrix containing the singular values of matrix  $S$ ,  $V$  is an orthogonal  $3 \times 3$  matrix and  $^t$  is the transpose operator. Then,  $U$  is the basis we seek.

From this basis  $U$ , we can define a new set of sensors  $\tilde{S}$  ( $m \times 3$ ), different from the original sensors  $S$ , by multiplying the basis by any linear transformation  $P$  ( $3 \times 3$ ), where  $P$  consists of 3 column vectors,  $\underline{p}_1, \underline{p}_2, \underline{p}_3$  that are sampled over the 2-sphere. Then,

$$\tilde{S} = UP, \quad P = [\underline{p}_1, \underline{p}_2, \underline{p}_3]; \quad (5)$$

We are interested in the relation between the original sensors  $S$  and the newly defined sensors  $\tilde{S}$ .

Using Equations 4 and 5 we have:

$$\tilde{S} = UP = U\Sigma V^t(\Sigma V^t)^{-1}P = S(\Sigma V^t)^{-1}P. \quad (6)$$

Now, replacing this into Equation 2 we obtain

$$T = (\Sigma V^t)^{-1}P. \quad (7)$$

We can also rearrange this equation in order to relate a transformation matrix  $T$  with a set of points  $P$  over the sphere.

$$P = \Sigma V^t T. \quad (8)$$

A visual explanation of this procedure can be seen in Figure 2.

To find a global error minimum it is possible, in principle, to sample sensors across the whole surface of the sphere. However, the combination of three separate sensors means that the number of possible 3-sensor combinations for  $N$  samples would be  $O(N^3)$ , which would be prohibitive to compute for large  $N$ . For example, if we select  $N$  to be 25000, which represents an angular distance between points in the sphere of  $1^\circ$ , we will have  $25000^3 = 2.6$  trillion sensor sets. Furthermore, it is clear that some sensor choices would be clearly unsuitable, both in terms of generating high error, and in some degenerate cases where the transform matrix becomes rank-reduced and maps  $RGB$  values to a plane or vector, rather than a complete 3D color space.

For these reasons, instead of sampling all the sphere, we look at the two last equations for a simplified solution. We start by considering  $T_{LS}$ , that is the solution to Equation 2 that minimizes the least-squares characterization error in  $XYZ$  space, which is computed by solving the normal equation. Once this matrix is found, we apply Equation 8 to obtain the  $T_{LS}$  representatives in the sphere and we sample in their surrounds.

## IV. EXPERIMENTS AND RESULTS

We have performed three different experiments to demonstrate the capability of our method, each using a different error metric. The first experiment deals with a pixel-wise measure, the CIE  $\Delta_E$ . This is, to the best of our knowledge, the only perceptual error metric that has been used, until now, to optimise camera characterization [30]. The other two experiments that we perform, which are based on minimizing a color appearance model (S-CIELAB) and an image quality metric (CID), are novel. They characterize the camera not only in terms of isolated pixel values, but also considering the context; a strategy that is more consistent with human vision.

Our experiments have been performed using simulated camera systems, whereby the camera sensitivities have been measured and recorded in advance, and the illumination and reflectance spectra of the scene are taken from existing databases of spectra and multispectral images.

The simulation works as follows: for each image (or set of pixels) we know the reflectance spectrum at each pixel, and from this we compute both the  $XYZ$  value at each pixel, and the  $RGB$  response of the camera using the standard image formation equations. Then, the  $RGB$  camera responses are converted to estimated  $XYZ$  by the different methods. Therefore, we have the real and the estimated  $XYZ$  images and we calculate the different metrics from these.

To evaluate the device dependency of our results we use a total of 37 cameras, including 28 used by Jiang *et al.* [31], and 9 from the image engineering webpage [32]. While we use these cameras to simulate the capture of still images,

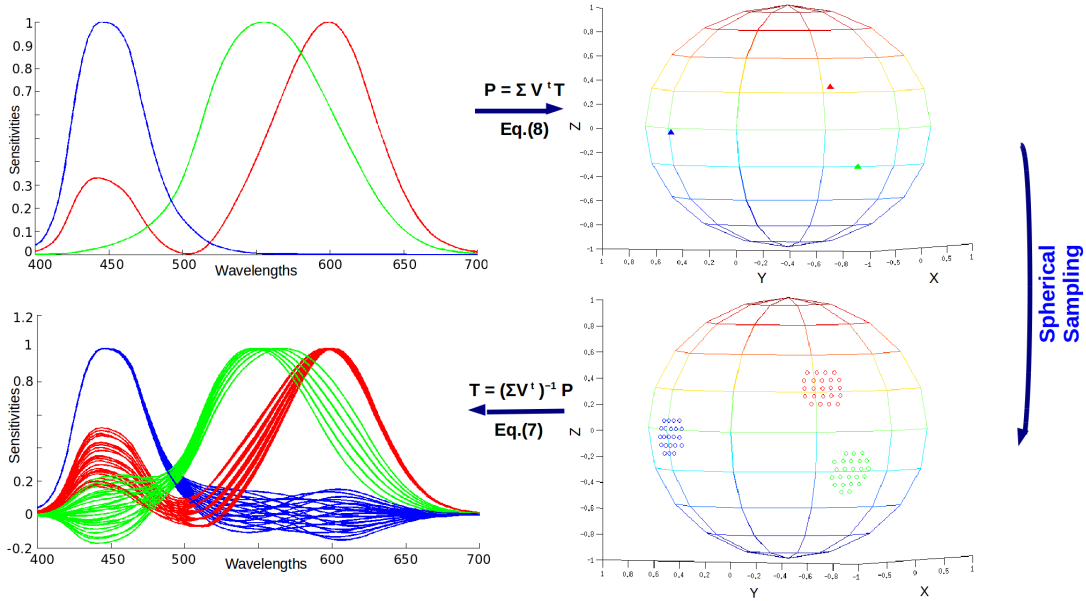


Fig. 2. Pipeline of the spherical sampling method. Adapted from [12]. From the original set of sensors, representatives in the sphere are obtained. Then, a sampling over those points in the sphere is performed. Finally, all the points added are transformed back to the sensor domain, representing new set of sensors.

the characterisation process applies equally to video images. For all three experiments we have computed the matrix  $T_{LS}$  which minimizes Equation 2 in the least-squares sense, using 102 illuminants and 1995 reflectances from the Simon Fraser University dataset [33]. The spherical sampling procedure considers points within a distance of  $3.3^\circ$  from those representing the starting sensors in the first experiment, and  $2.5^\circ$  for the second and third; we sample 30000 points in the sphere in all the cases except in experiment 1b, where we sample 50000 points. Let us note again here that with a bigger angular distance, and a larger spherical sampling resolution, better results might still be obtained at the cost of extra computations.

#### A. Experiment 1a: CIE $\Delta_E$

As we mentioned earlier,  $XYZ$  is not a perceptual space, i.e. the same distance between two points in different regions of the space represent different perceived distances by a human observer. As a result, minimizing error in  $XYZ$  space does not necessarily correspond to a minimized perceptual error. In order to avoid this issue different perceptual color spaces have been defined, most notably: CIELab, CIELuv [34], and CIECAM [20]. The most commonly used space is the CIELab space, as in this space Euclidean distance correlates with perceptual distance, and the color conversion from  $XYZ$  to Lab is straightforward. The Euclidean distance in CIELab space is referred to as the CIE  $\Delta_E$  and is defined mathematically as:

$$\Delta_E(p_1, p_2) = \|p_1 - p_2\|_2 \quad (9)$$

where  $p_1$  and  $p_2$  are two points in the Lab space with dimensions  $3 \times 1$ .

More recently some improvements on this metric have been proposed that are better correlated with perceptual differences,

notably CIE  $\Delta_{E00}$  [21], which is known to give more accurate results for blue colors. The formula for this calculation can be found in [21], and we denote the result as  $\Delta_{E00}$  [35].

In this section our goal is to minimize, for each particular camera, the average perceptual error for a set of illuminants and reflectances, and we report results for both the  $\Delta_E$  and of the  $\Delta_{E00}$  metrics.

Mathematically, for a particular camera, we call  $p_{i,j}$  (of dimension  $3 \times 1$ ) the real  $XYZ$  value of the pixel for the reflectance  $i \in I$  under illuminant  $j \in J$ , and  $q_{i,j}$  ( $3 \times 1$ ) the  $RGB$  camera value of the pixel. We look for the matrix  $T$  that minimizes

$$\arg \min_T \frac{\sum_{i \in I} \sum_{j \in J} \Delta_E(\text{Lab}(p_{i,j}^t), \text{Lab}(q_{i,j}^t \cdot T))}{\#I \cdot \#J} \quad (10)$$

where  $\text{Lab}()$  represents the transformation from  $XYZ$  to CIE Lab. We follow the same procedure to calculate  $\Delta_{E00}$ , by replacing Equation 9 with the standard formula in [21].

We note here that a similar experiment regarding color constancy using sharp sensors was performed in [12].

We have performed our experiment using the 1995 reflectances and 102 illuminants described above. Computing all possible surfaces under all lights induces a total of 203490 color signals. We separate these into training and test sets by assigning 90% of the color signals at random to the training set, and the remaining 10% to the test set. We repeat this procedure 100 times to avoid any bias caused by the random selection.

The results for all 37 cameras are reported in Figures 3 and 4 for the  $\Delta_E$  and  $\Delta_{E00}$  metrics respectively. Tables I and II show error statistics relating to the figures, presenting the improvement obtained in the minimum, maximum, mean and median statistics. In the tables the error for both the

least-squares method and our spherical sampling approach are averaged over all the cameras.

TABLE I  
 $\Delta_E$  RESULTS: AVERAGE FOR 37 CAMERAS

	Lst Sqr.	Spher. Samp.
Min	<b>1.7764</b>	<b>1.7764</b>
Max	10.9341	<b>10.5822</b>
Mean	3.0581	<b>2.9536</b>
Median	2.5827	<b>2.4903</b>

TABLE II  
 $\Delta_{E00}$  RESULTS: AVERAGE FOR 37 CAMERAS

	Lst Sqr.	Spher. Samp.
Min	<b>1.1178</b>	<b>1.1178</b>
Max	7.3300	<b>7.1277</b>
Mean	1.8991	<b>1.8438</b>
Median	1.5049	<b>1.4950</b>

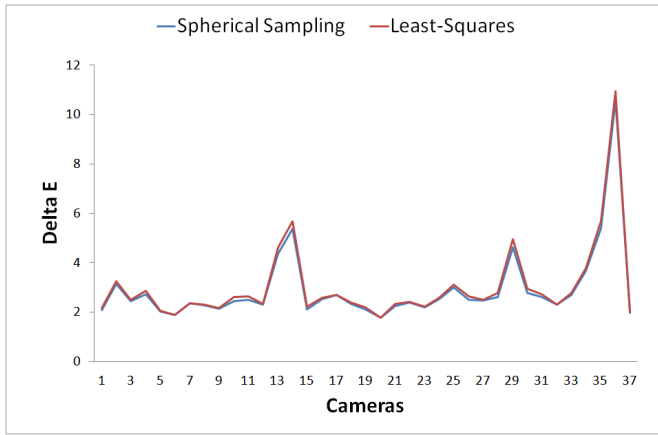


Fig. 3.  $\Delta_E$  error for the different cameras. In blue:  $\Delta_E$  for spherical sampling. In red: error for the least-squares minimization.

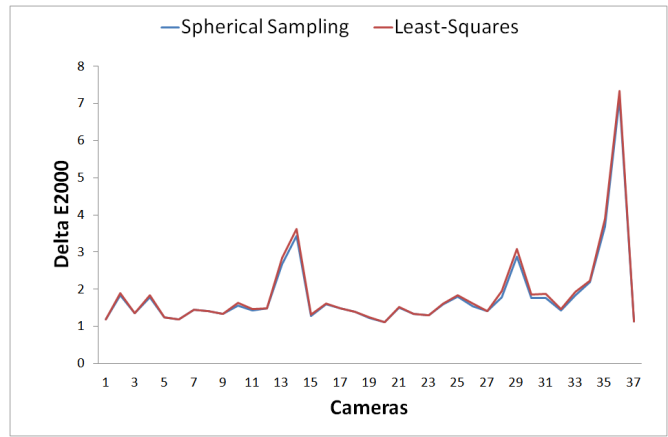


Fig. 4.  $\Delta_E$  error for the different cameras. In blue:  $\Delta_E$  for spherical sampling. In red: error for the least-squares minimization.

process 100 times. Results of this procedure, measured using  $\Delta_E$  and  $\Delta_{E00}$  error, are shown in Tables III and IV.

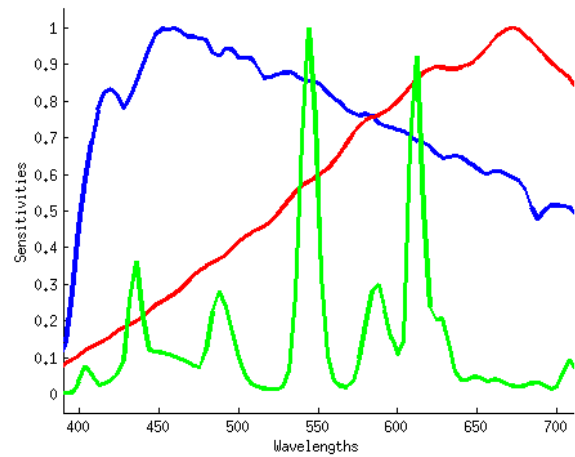


Fig. 5. Spectral sensitivities of the three illuminants used in Experiment 1b.

*B. Experiment 1b: Different illuminants*

As we point out in the introduction, some cameras calculate pre-set color correction matrices under different illuminations, which gives better color correction for each individual illuminant. In this section we look at how well our method works in this situation by deriving separate color correction matrices under different illuminants, and comparing the performance of our technique with the least-squares approach, where the least squares solution is derived using the same illuminant.

Using the same set of 1995 reflectances we generated color signals for each of three illuminants: one representing daylight, one a fluorescent illuminant, and the third an incandescent illuminant. The spectral sensitivities of these illuminants can be found in Figure 5. For each illuminant we separate the data into training and test sets, with a proportion of 90% for training and 10% for the test. We then use the spherical sampling procedure to generate a color correction matrix for the training set and test the performance on the test set, and repeat this

TABLE III  
 $\Delta_E$  ERROR FOR THREE DIFFERENT ILLUMINANTS ON THE SIMON FRASER DATASET.

	Daylight		Incandescent		Fluorescent	
	LS	SS	LS	SS	LS	SS
Min	<b>1.0581</b>	<b>1.0581</b>	<b>0.9804</b>	<b>0.9804</b>	<b>1.0092</b>	<b>1.0092</b>
Max	6.9781	<b>6.8166</b>	6.3126	<b>6.1703</b>	4.5314	<b>4.3278</b>
Mean	1.8353	<b>1.8153</b>	1.6491	<b>1.6297</b>	1.7122	<b>1.6690</b>

TABLE IV  
 $\Delta_{E00}$  ERROR FOR THREE DIFFERENT ILLUMINANTS ON THE SIMON FRASER DATASET.

	Daylight		Incandescent		Fluorescent	
	LS	SS	LS	SS	LS	SS
Min	<b>0.6729</b>	<b>0.6729</b>	<b>0.5527</b>	<b>0.5527</b>	<b>0.4915</b>	<b>0.4915</b>
Max	4.5663	<b>4.4134</b>	4.3101	<b>4.2190</b>	2.5441	<b>2.4224</b>
Mean	1.1240	<b>1.1116</b>	0.9465	<b>0.9330</b>	0.8057	<b>0.7967</b>



C. Experiment 2: S-CIELAB

Our previous experiments, which can be considered the de-facto experiments for perceptual minimization, have one main drawback: they do not take into account the image context. For example, when looking at a yellow pixel, the distance between the real and approximated value will be the same both if it has a pink neighborhood or an orange one. This is known to be false, since human perception relies deeply on the context of the scene [36]. It is for this reason that S-CIELAB was proposed [22]. Basically, S-CIELAB computes the  $\Delta_E$  error measure after applying spatial pre-processing to account for the spatial-color sensitivity of human vision.

Given that SCIELAB is applied to images, for the purposes of our experiments it is important to use meaningful images and meaningful illumination. To this end, we have used the 16 hyperspectral images obtained by Foster *et al.* [37], which contain a range of man-made objects, natural landscapes and both indoor and outdoor scenes, and the D65 illuminant. The 16 images are presented in Figure 6.



Fig. 6. Foster *et al.* [37] hyperspectral dataset used for experiments 2 and 3.

In this experiment, for each camera, we use spherical sampling to calculate a color correction matrix that minimizes S-CIELAB error. Mathematically, if we define  $I_{h,D65}$  (dimension  $N \times 3$ ) as the real  $XYZ$  values of image  $h$  of the database under the D65 illuminant, and  $J_{h,D65}$  (dimension  $N \times 3$ ) the image obtained by the camera for the same image and illuminant, we search the matrix  $T$  that minimizes:

$$\arg \min_T \frac{\sum_{h=1}^M SCIELAB(I_{h,D65}, J_{h,D65} \cdot T)}{M} \quad (11)$$

To separate the training and test sets we use a *leave-one-out* procedure whereby we use 15 out of the 16 images to build the transform matrix, and test the method on the remaining image. We repeat this procedure 16 times, leaving out each image in turn, and then calculate the mean error over all 16 images.

Results for all the different cameras are shown in Figure 7. A statistical analysis of the results is shown in Table V.

TABLE V  
SCIELAB RESULTS: AVERAGE RESULT FOR 37 CAMERAS USING A LEAVE-ONE-OUT PARADIGM.

	Lst Sqr.	Spher. Samp.
Min	0.7139	<b>0.5173</b>
Max	4.4245	<b>3.8121</b>
Mean	1.2100	<b>1.0056</b>
Median	1.0529	<b>0.8701</b>

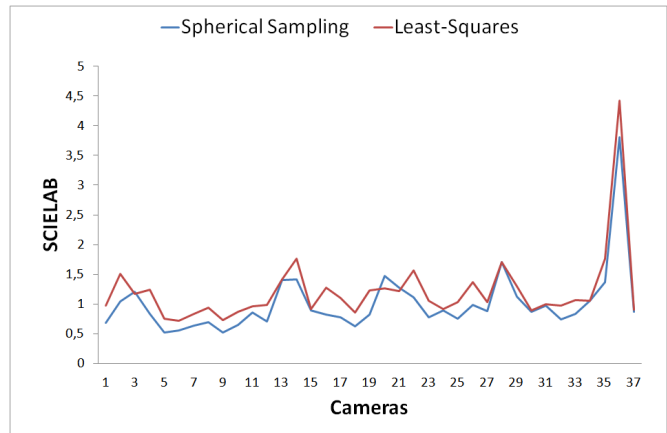


Fig. 7. SCIELAB error for the different cameras. In blue: S-CIELAB error for spherical sampling. In red: error for the least-squares minimization.

D. Experiment 3: CID measure

Lissner *et al.* have recently proposed a new perceptually-based color image metric, which shows a good correlation with human evaluations of gamut mapping algorithms [24]. The method is originally based on the intensity SSIM image quality metric of Wang *et al.* [38].

In this experiment we have followed the same procedure as section IV-C, and have used the leave-one-out paradigm to minimize the mean CID error for a subset of the hyperspectral images, and to test on a novel image for which the method wasn't trained.

Defining  $I_{h,D65}$  as the real  $XYZ$  image  $h$  of the database under the D65 illuminant, and  $J_{h,D65}$  as the image obtained by the camera for the same image and illuminant, we search for the matrix  $T$  minimizing:

$$\arg \min_T \sum_{h=1}^M CID(I_{h,D65}, J_{h,D65} \cdot T). \quad (12)$$

Results are shown in Figure 8, and, once again spherical sampling outperforms the least-squares optimization. A statistical analysis is shown in Table VI.

E. Qualitative evaluation

Figure 9 presents a qualitative evaluation of color correction matrices that minimize both the SCIELAB and the CID measures. The images are presented in *sRGB* using the standard



TABLE VI  
CID RESULTS: AVERAGE FOR 37 CAMERAS.

	Lst Sqr.	Spher. Samp.
Min	0.0024	<b>0.0018</b>
Max	0.1144	<b>0.1017</b>
Mean	0.0092	<b>0.0076</b>
Median	0.0045	<b>0.0040</b>

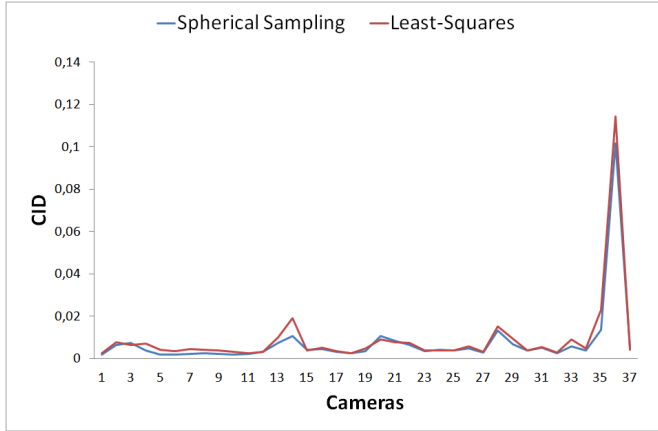


Fig. 8. CID error for the different cameras. In blue: CID error for spherical sampling. In red: error for the least-squares minimization.

transform to convert from  $XYZ$  to  $sRGB$ . In the top row, i.e. the garden image, we present color correction results that minimize SCIELAB error. The image on the left, which is the least-squares minimization, has a mean SCIELAB difference from the middle image (the real image) of 8.74. The image created by spherical sampling, shown on the right, gives a mean difference of 4.86. The middle row, i.e. the flower image, shows the color correction results from minimising the CID measure. The image on the left (the least-squares solution) has a mean difference from the central image (the original) of 0.0569, while the image generated from spherical sampling, shown on the right, gives an error of 0.0232. A cropped version of this comparison is presented in the last row. We note here the yellowish cast in the least-squares image.

## V. DISCUSSION

In the previous section we have minimized three different error metrics independently. In Table VII we compare the results between these metrics using the mean percentage improvement for the three different cases. As expected, our method shows an improvement when minimizing  $\Delta_E$  and  $\Delta_{E00}$ , but less of an improvement than it shows for the image-based metrics. Even in this case, our improvement is approximately 3%. More impressive are the results for the other two metrics, where our spherical sampling procedure reduces the error by more than 15%.

Another way to represent the results is to analyse what is happening in the sensor space. As our method maps  $XYZ$  to  $RGB$  responses, one can also conceptualise this as mapping the  $RGB$  sensor sensitivities directly to the  $XYZ$  color matching functions. The question then is: are the approximate  $XYZ$  color matching functions found by spherical sampling

TABLE VII  
PERCENTAGE IMPROVEMENT IN ERROR-METRIC FOR SPHERICAL SAMPLING WHEN COMPARED TO LEAST-SQUARES.

	$\Delta_E$	$\Delta_{E2000}$	SCIELAB	CID
Mean percentage	3.42%	2.91%	16.89%	17.11%

closer to the real  $XYZ$  sensors than those obtained via the least-squares optimization? To address this question we have defined a new measure.

Let us define  $X, Y$  and  $Z$  as  $m \times 1$  vectors, which are discrete representations of the color matching functions  $\bar{x}, \bar{y}$  and  $\bar{z}$  respectively, sampled at  $m$  wavelengths spanning the domain  $\omega$ . Let us also define  $\hat{X}^{LS}, \hat{Y}^{LS}$ , and  $\hat{Z}^{LS}$  as the approximation derived by least-squares (LS) and  $\hat{X}^{SS}, \hat{Y}^{SS}$ , and  $\hat{Z}^{SS}$  as the approximation derived by spherical sampling. We also define  $e_i^{LS}$  and  $e_i^{SS}, i \in \{X, Y, Z\}$ , as the error obtained for both methods, which we compute for each sensor separately; e.g. the error for the  $X$  sensor is calculated as:

$$e_X^{LS} = \sum_{m \in \omega} \|X_m - \hat{X}_m^{LS}\|_q \quad (13)$$

where  $q$  represents the chosen norm. In this paper we will use  $q = 1$  and  $q = 2$ . We choose these values since  $q = 1$  represents the response to a flat white surface under a white illuminant, and  $q = 2$  is the typical Euclidean norm. From the previous equation the difference between the two approaches can be computed as:

$$diff_X(LS, SS) = e_X^{LS} - e_X^{SS}, \quad (14)$$

therefore, when this difference is smaller than zero the least-squares sensors are closer to the real sensors than spherical sampling ones and vice-versa.

We note here that while Equation 13 could be minimized more directly by computing a matrix  $T$  from Equation 1, the purpose here is to compare the transforms derived using real spectra and illuminants for their ability to approximate the  $XYZ$  color matching functions. This metric provides an indirect measure of how well the methods will generalise to unseen data, and we use it as a ‘‘sanity check’’ to ensure that overtraining is not occurring.

In Table VIII we show the results averaged over all cameras for all the three sensors in the three different experiments. In this case, we have performed the minimization of the three measures on the full set of images, i.e. we do not separate training and test data, since the metric we are computing is different from the metric used to learn the transforms. We can see that the spherical sampling technique results in lower error than the least-squares method in 16 out of 18 cases, which reinforces the adequacy of our method.

## VI. CONCLUSION

The present work investigates a novel technique for mapping camera  $RGB$  responses to device independent, CIE $XYZ$ , color co-ordinates. The method works by discretely sampling sensor sets that are close to an initial ‘‘best-guess’’ solution derived by minimizing the least-squares error. The discrete

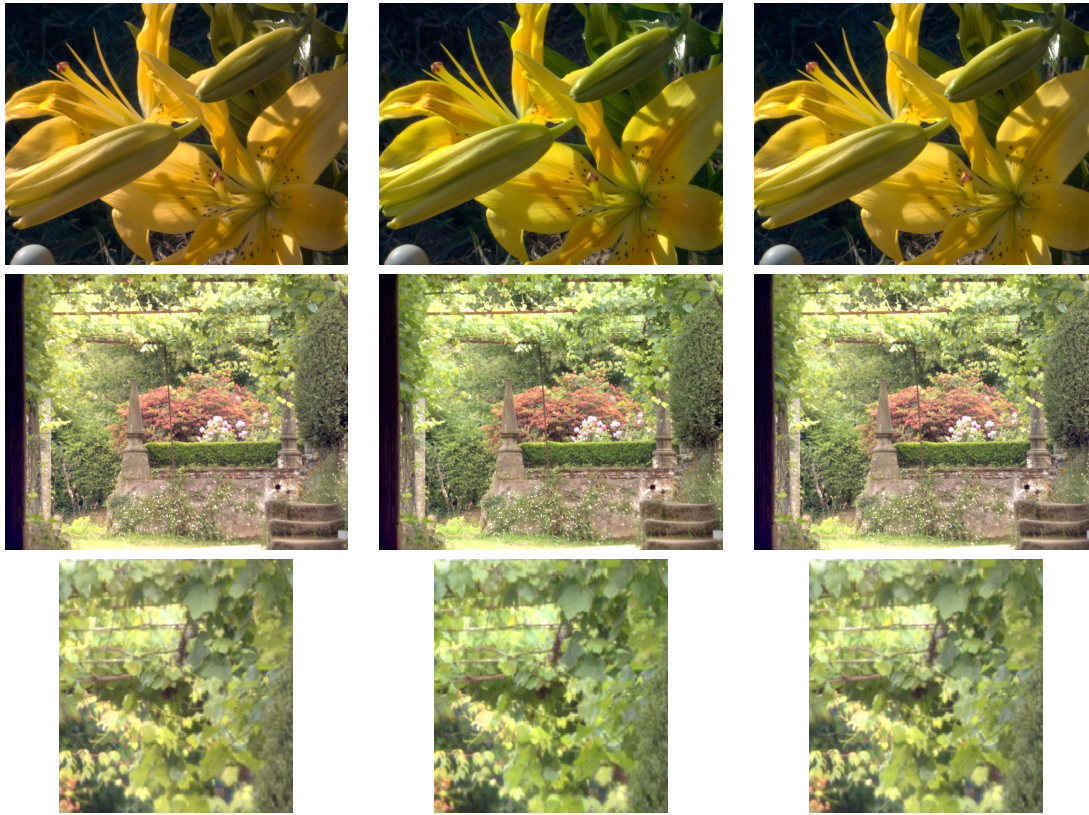


Fig. 9. Qualitative evaluation of the approach. Top row: minimising the SCIELAB measure; middle and bottom rows: minimising the CID measure. For each column the images are: Original (left), least-squares (centre), spherical sampling (right)

TABLE VIII

ERROR IN A LEAST-SQUARES OPTIMIZATION MINUS ERROR USING A SPHERICAL SAMPLING PROCEDURE COMPUTED IN THE SENSOR SPACE.

	$\Delta_E$			SCIELAB			CID		
	X	Y	Z	X	Y	Z	X	Y	Z
L1	0.71	0.59	2.05	-0.43	0.30	2.15	0.30	0.38	1.28
L2	1.02	0.85	3.37	-0.08	0.58	2.87	1.16	0.79	1.55

nature of the algorithm means that error can be minimized in any chosen color space, which allows the optimization to take into account spatial characteristics of colors in images, and takes a step towards characterizing perceptual, as well as physical, aspects of the color.

The method was tested on a range of different images, and simulated camera response curves. The results show that for over 90% of image and sensor combinations the initial least-squares solution can be improved upon by spherical sampling. This improvement ranges from approximately 3% for CIELAB  $\Delta E$  error and 15% for the SCIELAB and CID error metrics.

This method would be of direct and immediate applicability for camera manufacturers, who would be able to appreciably increase the accuracy of the color characterization of their cameras, without any need to change the experimental set-up that they use. Furthermore, the proposed method is of interest for color researchers and professional photographers/cinematographers, who may use an affordable calibration device to estimate the *RGB* response curves of the camera

and then apply our method to properly correct the RAW data.

#### ACKNOWLEDGMENTS

This work was supported by European Research Council, Starting Grant ref. 306337, and by Spanish grants AACC, ref. TIN2011-15954-E, and Plan Nacional, ref. TIN2012-38112.

#### REFERENCES

- [1] E. Lam and G. Fung, "Automatic white balancing in digital photography," *Single-sensor imaging: Methods and applications for digital cameras*, pp. 267–294, 2008.
- [2] P. Hubel, J. Holm, G. Finlayson, M. Drew *et al.*, "Matrix calculations for digital photography," in *The Fifth Color Imaging Conference: Color, Science, Systems and Applications*, 1997.
- [3] G. Sharma, "Color fundamentals for digital imaging," *Digital Color Imaging Handbook*, vol. 20, 2003.
- [4] P. Hung, "Color theory and its application to digital still cameras," *Image Sensors and Signal Processing for Digital Still Cameras*, ed. J. Nakamura, pp. 205–221, 2005.
- [5] G. Hong, M. R. Luo, and P. A. Rhodes, "A study of digital camera colorimetric characterization based on polynomial modeling," *Color Research and Application*, vol. 26, pp. 76–84, 2001.
- [6] T. L. V. Cheung, S. Westland, D. R. Connah, and C. Ripamonti, "A comparative study of the characterization of colour cameras by means of neural networks and polynomial transforms," *Journal of Coloration Technology*, vol. 120, pp. 19–25, 2004.
- [7] C. Find-Andersen and J. Y. Hardeberg, "Colorimetric characterization of digital cameras preserving hue planes," in *IS&T/SID 13th Color Imaging Conference: Color, Science, Systems and Applications*, 2005, pp. 141–146.
- [8] R. Ramanath, W. Snyder, Y. Yoo, and M. Drew, "Color image processing pipeline," *Signal Processing Magazine, IEEE*, vol. 22, no. 1, pp. 34–43, 2005.

- [9] G. Finlayson and M. Drew, "Constrained least-squares regression in color spaces," *Journal of Electronic Imaging*, vol. 6, no. 4, pp. 484–493, 1997.
- [10] S. Bianco, F. Gasparini, A. Russo, and R. Schettini, "A new method for RGB to XYZ transformation based on pattern search optimization," *Consumer Electronics, IEEE Transactions on*, vol. 53, no. 3, pp. 1020–1028, 2007.
- [11] G. Adcock, "Charting your camera," *Creative COW Magazine*, November 2011.
- [12] G. D. Finlayson, J. Vazquez-Corral, S. Süsstrunk, and M. Vanrell, "Spectral sharpening by spherical sampling," *Journal of the Optical Society of America A*, vol. 29, no. 7, pp. 1199–1210, Jul 2012. [Online]. Available: <http://www.cat.uab.cat/Public/Publications/2012/FVS2012>
- [13] B. K. P. Horn, "Exact reproduction of colored images," *Computer Vision, Graphics, and Image Processing*, vol. 26, no. 2, pp. 135–167, 1984.
- [14] G. D. Finlayson and M. S. Drew, "The maximum ignorance assumption with positivity," in *IS&T/SID Fourth Color Imaging Conference: Color, Science, Systems and Applications*, 1996, pp. 202–205.
- [15] J. A. S. Viggiano, "Minimal-knowledge assumptions in digital still camera characterization i.: Uniform distribution, toeplitz correlation," in *IS&T/SID Ninth Color Imaging Conference: Color, Science, Systems and Applications*, 2001, pp. 332–336.
- [16] G. D. Finlayson and J. Paul, "Minimal knowledge versus the real world," in *IS&T/SID Tenth Color Imaging Conference: Color, Science, Systems and Applications*, 2002, pp. 133–238.
- [17] M. S. Drew and B. V. Funt, "Natural metamers," *Computer Vision, Graphics, and Image Processing: Image Understanding*, vol. 56, pp. 139–151, 1992.
- [18] L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *Journal of the Optical Society of America A*, vol. 3, pp. 1673–1683, 1986.
- [19] CIE, "Recommendations on uniform color spaces, color-difference equations, psychometric color terms," Bureau Central de la CIE, Publication 15, Supp.2, (E1.3.1)(TC1.3), 1978.
- [20] N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M. R. Luo, and T. Newman, "The ciec02 color appearance model," in *IST/SID 10 th Color Imaging Conference*, 2002, pp. 23–27.
- [21] M. Luo, G. Cui, and B. Rigg, "The development of the cie 2000 colour-difference formula: Ciede2000," *Color Research and Application*, vol. 26, pp. 340–350, 2001.
- [22] X. Zhang and B. A. Wandell, "A spatial extension of cielaab for digital color image reproduction," in *Soc. Inf. Display Symp. Tech. Dig. vol. 27*, 1996, pp. 731–734.
- [23] M. D. Fairchild and G. M. Johnson, "icam framework for image appearance, differences, and quality," *Journal of Electronic Imaging*, vol. 13, pp. 126–138, 2004.
- [24] I. Lissner, J. Preiss, P. Urban, M. Scheller, Lichtenauer, and P. Zolliker, "Image-difference prediction: from grayscale to color," *IEEE Transactions on Image Processing*, vol. 22, pp. 435–445, 2013.
- [25] J. Y. Hardeberg and F. Schmitt, *Recent Progress in Color Management and Communications*. IS&T, 1998, ch. Color Printer Characterization Using a Computational Geometry Approach, pp. 88–91.
- [26] G. D. Finlayson, M. Mackiewicz, and A. Hurlbert, "Root polynomial colour correction," in *IS&T/SID 19th Color Imaging Conference: Color, Science, Systems and Applications*, 2011, p. 115.
- [27] J. McElvain and W. Gish, "Camera color correction using two-dimensional transforms," in *IS&T/SID 21st Color Imaging Conference: Color, Science, Systems and Applications*, 2013, p. 250.
- [28] G. C. Holst, *CCD arrays, cameras, and displays*. JCD Pub. ; SPIE Optical Engineering Press Winter Park, FL : Bellingham, Wash, 1998.
- [29] G. D. Finlayson and S. Susstrunk, "Spherical Sampling and Color Transformations," in *Proc. IST/SID 9th Color Imaging Conference*, vol. 9, 2001, pp. 321–325.
- [30] J. Nakamura, *Image Sensors and Signal Processing for Digital Still Cameras*. Boca Raton, FL, USA: CRC Press, Inc., 2005.
- [31] J. Jiang, D. Liu, J. Gu, and S. Susstrunk, "What is the space of spectral sensitivity functions for digital color cameras?" in *Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV)*, ser. WACV '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 168–179. [Online]. Available: <http://dx.doi.org/10.1109/WACV.2013.6475015>
- [32] "<http://www.image-engineering.de/library-m/test-reports>."
- [33] K. Barnard, L. Martin, B. Funt, and A. Coath, "A data set for colour research," *Color Research and Application*, vol. 27, no. 3, pp. 147–151, 2002.
- [34] G. Wyszecki and W. Stiles, *Color science: concepts and methods, quantitative data and formulae*, 2nd ed. John Wiley & Sons, 1982.
- [35] G. Sharma, W. Wu, and E. N. Dalal, "The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations," *Color research and application*, vol. 30, no. 1, pp. 21–30, 2005. [Online]. Available: <http://www.ece.rochester.edu/~gsharma/ciede2000/>
- [36] X. Otazu, C. A. Parraga, and M. Vanrell, "Toward a unified chromatic induction model," *Journal of Vision*, vol. 10(12), no. 6, 2010. [Online]. Available: <http://www.cat.uab.cat/Public/Publications/2010/OPV2010>
- [37] D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," *J. Opt. Soc. Am. A*, vol. 23, no. 10, pp. 2359–2372, Oct 2006. [Online]. Available: <http://josaa.osa.org/abstract.cfm?URI=josaa-23-10-2359>
- [38] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.