

The Photometry of Intrinsic Images

Marc Serra¹, Olivier Penacchio^{1,3}, Robert Benavente¹, Maria Vanrell¹, Dimitris Samaras²

¹ Computer Vision Center / Computer Science Dept., Universitat Autònoma de Barcelona

² Image Analysis Lab, Computer Science Dept., Stony Brook University

³ School of Psychology & Neuroscience, University of St. Andrews

{mserra, penacchio, robert, maria}@cvc.uab.cat, samaras@cs.stonybrook.edu

Abstract

Intrinsic characterization of scenes is often the best way to overcome the illumination variability artifacts that complicate most computer vision problems, from 3D reconstruction to object or material recognition. This paper examines the deficiency of existing intrinsic image models to accurately account for the effects of illuminant color and sensor characteristics in the estimation of intrinsic images and presents a generic framework which incorporates insights from color constancy research to the intrinsic image decomposition problem. The proposed mathematical formulation includes information about the color of the illuminant and the effects of the camera sensors, both of which modify the observed color of the reflectance of the objects in the scene during the acquisition process. By modeling these effects, we get a "truly intrinsic" reflectance image, which is invariant to changes of illuminant or camera sensors. This model allows us to represent a wide range of intrinsic image decompositions depending on the specific assumptions on the geometric properties of the scene configuration and the spectral properties of the light source and the acquisition system, thus unifying previous models in a single general framework. We demonstrate that even partial information about sensors improves significantly the estimated reflectance images, thus making our method applicable for a wide range of sensors. We validate our general intrinsic image framework experimentally with both synthetic data and natural images.

1. Introduction

In an intrinsic image decomposition [6], intrinsic images of a given scene are the images depicting a single physical characteristic of the scene (such as reflectance, illumination, orientation, distance, transparency, specularity, luminosity, and so on). While intrinsic image decomposition methods initially focused on providing reflectance and shading image estimates of a scene, other subfields in computer vision estimate different intrinsic characteristics. Shape from shading methods [9] estimate the shape (*i.e.* orientation, depth...) of the objects given a shading image. Color constancy methods [16] estimate the illuminant of the scene. Highlight removal techniques [2] estimate image specularities. Intrinsic

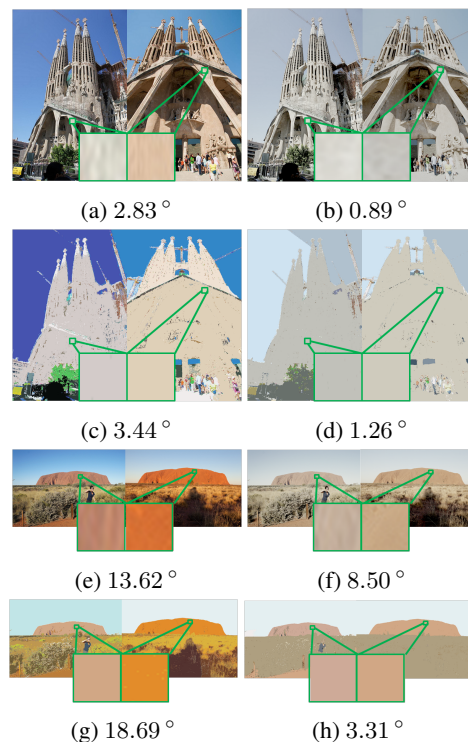


Figure 1: Our model applied to natural images. (a) and (e) are images of a given landmark, taken with different cameras under different illumination conditions. (b) and (f) are the images in (a) and (e) respectively, after removing the effects of the illuminant and the camera sensors. (c), (d), (g) and (h) are the estimated intrinsic images of (a), (b), (e) and (f), respectively. Chromatic angular error values are given in degrees.

characterization of scenes is fundamental in multiple computer vision applications. The complex ways in which light interacts with shapes and materials continue to confound solutions in areas that range from 3D shape reconstruction to object recognition to material identification. In this paper we argue that in order to increase accuracy in such applications, multiple intrinsic properties have to be studied in conjunction, so that the appropriate intrinsic image for each property can be estimated.

Most early models of intrinsic image estimation either worked with grayscale images [21, 8, 30] or assumed scenes with Lambertian surfaces (*i.e.* they reflect the same amount of light in all directions) and a single white light source [28, 14], thus simplifying the decomposition into intrinsic components to the product of its intrinsic images of shading and reflectance. This simplified formulation has been commonly used ever since [19, 26, 15, 24]. Recently, in [7], the Lambertian assumption was relaxed and a specular term was added to the model. Specularity detection is a hard problem in itself and several specularity removal techniques can be found in the literature (see [2] for a survey). Lately, in [5], the authors relaxed the white light assumption while keeping the Lambertian surfaces and single light source constraints. In their work, the shading image was modeled as a function of the shape of the objects in the scene, which can be described in many ways (depth maps, normal maps, etc.), and the color and geometry (*i.e.* direction) of the light source of the scene. All these formulations are consistent with the physically based dichromatic reflection model [25], which explains how light is reflected in a scene under some simplifying assumptions.

However, the color value at each image pixel is not only determined by the reflection model. During the image acquisition process, three factors influence the color value that we finally measure at each pixel: the reflectance of the objects in the scene, the illuminant of the scene, and the spectral response of the camera sensors. While reflectance is an intrinsic property of the objects, the illuminant and the camera sensors modify the way we see this reflectance in the images. That is why we must study and describe separately these three factors by isolating the effects of both illuminant and camera sensors. So far, however, existing methods on intrinsic image decomposition provide reflectance intrinsic images which contain mixed information about the color of the illuminant (which is assumed white in many models), object reflectance values and camera sensor effects. To the best of our knowledge camera sensors have not been studied before in the area of intrinsic image decomposition.

We draw insights upon the rich literature in color constancy (hitherto somewhat disconnected from intrinsic image research), which also aims to a stable representation of object color across different images. The aim of color constancy from the start [13] has been to estimate the color of the scene under a canonical light source, which is similar to the problem of reflectance estimation. Multiple works [16] have provided different methods to estimate the scene illuminant from a single image, which in turn allows to remove its effects and obtain a canonical color image. Constancy research has also devoted some attention to the effects of the imaging sensors for the final recovery of the canonical illuminants. The narrow band property of the spectral sensitivity of the sensors simplifies the illuminant change characterization [11] and therefore its subsequent estimation. In this paper we build upon the connection between these two fields by extending the intrinsic model to include the effects

of the light source color and the biases introduced by the sensors. Both factors have direct effects on the reflectance computation.

In Fig. 1 we can see how our model works on natural images. We can observe that the intrinsic reflectances of a given material (estimate using the method of Serra *et al.* [24]) are much closer after removing the effects of the illuminants and camera sensors, both qualitatively and quantitatively, than the intrinsic reflectances of the same material that we get from the original images.

In the formulation we propose, it is possible to study the effects of the camera sensors and the color of the illuminant and isolate them from the reflectance image. As a result, we obtain a new reflectance image which is invariant to changes of the camera sensors and the color of the illuminant. This reflectance image, which we call absolute reflectance, is a truly intrinsic image. Indeed in the second and forth rows of Fig. 1 we see that the difference between the intrinsic images has been minimized. The invariance of our absolute reflectance image provides many practical advantages in different scenarios. Our method is practically applicable because such sensor information is sometimes publicly available for many sensors and the color of the light can be estimated with any state of the art color constancy method. As we demonstrate in this paper even partial knowledge of sensor characteristics leads to significant improvement in the estimated intrinsic images. Furthermore, we believe that for widely used sensors such as the Kinect, such corrections could even be incorporated in the standard libraries for wide use by developers for any application that relies on accurate matching of surface appearances across different images.

2. Reflectance and color fundamentals

2.1. Color image formation: physics

When the light source is similar to a point source, the spectral radiance outgoing from an object *obj* at an infinitesimal patch at x along the direction v_r was modeled in [18] as

$$L_{obj}(x, v_r, \lambda) = f_r(x, v_r, v_i, \lambda) L_i(\lambda) \cos(\theta_i). \quad (1)$$

In this expression, $L_i(\lambda)$ is the incident radiance (*i.e.* irradiance), $\cos(\theta_i)$ captures the geometry of the scene by expressing the reduction in the amount of light impinging the surface at x due to the angle θ_i between the surface normal and the incident direction v_i , and $f_r(x, v_r, v_i, \lambda)$ is the bidirectional reflectance distribution function (brdf) of the object that specifies how much of the incident light coming from direction v_i is reflected into the viewing direction v_r per unit wavelength λ at x . The brdf captures how light and material interact to shape the appearance of an object. In particular, the spectral power of the light outgoing from an object depends on both the spectrum of the light source and the reflectance of the object.

Several reflection models have been proposed in the field of computer vision and computer graphics [22] which assume different simplifications of the physics of light. Here

we use Shafer’s Dichromatic Reflection Model (DRM) [25] because it offers a good trade-off between realism and practicability.

In Shafer’s DRM, light reflection is modeled as two independent reflection processes, each having a characteristic spectral power distribution whose magnitude varies with the direction of illumination and viewing. In the DRM, the amount of light reflected per unit wavelength in direction v_r by a small surface patch at a point x of an object obj is defined as

$$L_{obj}(x, v_r, \lambda) = m_b(x, v_r, v_i)c_b(\lambda) + m_s(x, v_r, v_i)c_s(\lambda), \quad (2)$$

where c_b and c_s are functions describing the spectral distributions of the body and the surface of the object, respectively, and m_b and m_s , called diffuse and specular magnitudes, are geometric factors that weight the amount of light coming from the body and the surface of the object, respectively, and only depend on the geometry of the scene. Eq.1 and the DRM in Eq.2 are equivalent (see supplementary material for the derivation).

To sum up, at the physical level the spectral composition of the light reflected from objects in a scene is already a mixture of the intrinsic reflectance of the object and of the spectrum of the light. This interdependency can be modeled mathematically, which will prove to be useful for the formulation of the intrinsic decomposition below.

2.2. Color image formation: sensors

As explained above the light reflected from an object depends on both its material reflectance and the color of the light. Additionally, in the acquisition process, the resulting image values are also affected by a third factor, namely the camera sensors.

Cameras use a finite set of sensor responses to describe the continuous light spectrum. These sample measurements are obtained by spectral integration [20]. Camera sensors vary widely with the characteristics of the camera, and different cameras usually produce different measurements. The values measured by the sensors of a camera are

$$p_k = \int L_{obj}(x, v_r, \lambda) S_k(\lambda) d\lambda, \quad (3)$$

where $S_k(\lambda)$, $k = 1, 2, 3$, are functions describing the absorption curves of the camera sensors and $L_{obj}(x, v_r, \lambda)$ is the radiance outgoing from objects in a scene.

For most computer vision problems, it is important to transform the measurements p_k made with a given camera to the measurements made with standard sensors $S_k^s(\lambda)$,¹

$$p_k^s = \int L_{obj}(x, v_r, \lambda) S_k^s(\lambda) d\lambda. \quad (4)$$

There exist several computational camera calibration models that estimate the sensitivity functions of the camera

[3, 23]. Given an estimate of the sensitivity functions of a camera, and standard sensitivity functions, a transformation matrix \mathbf{S}_{Sen} converts theoretical responses for standard sensors into those of the actual camera. Sensor transformations are often described by 3-by-3 matrices [17]. If \mathbf{p}^s denotes the vector whose standard coordinates are p_k^s , and \mathbf{p} denotes the response of the camera whose coordinates are p_k , we have $\mathbf{p} = \mathbf{S}_{Sen} \cdot \mathbf{p}^s$. In this work we use the standard RGB (sRGB) sensitivity curves [1] as the standard sensor.

Moreover, as we will see in the next section, a valuable property for a set of camera sensors is to be narrow band. We say that a set of camera sensors is narrow band when the overlap among these sensors is small, meaning that the color response for each of the sensors scarcely influences the responses of the other sensors. Narrow band responses are uncorrelated, which proves to be critical for most computer vision application among which color constancy.

2.3. Joint illumination/sensor modeling for intrinsic images

As we have seen in the previous section (Eq.3), both the spectral distribution of the light source and the sensor properties affect camera measurements. Defining color constancy algorithms able to provide image representations invariant to these dependencies has been a long-standing goal in the computer vision community. Originally, the term color constancy refers to the human ability to maintain a stable color representation of the world irrespective of the illuminant.

A classical approach in computational color constancy is to estimate the color of the illuminant of a scene using a single image and then subtract the illuminant to build a stable image under a canonical illuminant (see [16] for a survey). The canonical representation of the color of an image is given by

$$p_k^c = \int L_{obj}^c(x, v_r, \lambda) S_k(\lambda) d\lambda, \quad (5)$$

where $L_{obj}^c(x, v_r, \lambda)$ is the light reflected by the object under the canonical illuminant. In this work we use the CIE standard illuminant D65 as the canonical illuminant.

Let us now describe these transformations in a practical way. In the color constancy literature, 3-by-3 matrices have been commonly used to describe illuminant transformations [13]. We denote by \mathbf{L}_{CLig} the color conversion that transforms pixel values under the canonical illuminant to values under the actual illuminant. As for camera sensor transformations, we have $\mathbf{p} = \mathbf{L}_{CLig} \cdot \mathbf{p}^c$. Several methods have been proposed to find the appropriate transformation matrix \mathbf{L}_{CLig} [16] and most of them rely on the sharpness of the sensors and approximate the illuminant using a diagonal model [31, 11]. Spectral sharpening methods [12] are sensor transformations, \mathcal{T} , that convert a given set of sensor sensitivity functions into a new set of functions which are less overlapped,

$$\mathcal{T} \cdot \mathbf{p} \approx \mathbf{L}_{CLig} \cdot \mathcal{T} \cdot \mathbf{p}^c. \quad (6)$$

¹s superscript stands for “standard”

In general, spectral sharpening improves the performance of color constancy algorithms that are based on an independent adjustment of the sensor response channels.

Our objective is to describe the intrinsic components by discarding both the effects of the color of the light and the dependence on a particular set of camera sensors. The values

$$p_k^{s,c} = \int L_{obj}^c(x, v_r, \lambda) S_k^s(\lambda) d\lambda \quad (7)$$

represent the appearance of the objects under a canonical illuminant and standard camera sensors. The relationship between $p^{s,c}$ and \mathbf{p} is

$$\mathbf{p} = \mathbf{L}_{CLig} \cdot \mathbf{S}_{Sen} \cdot \mathbf{p}^{s,c}. \quad (8)$$

In this equation, \mathbf{L}_{CLig} and \mathbf{S}_{Sen} isolate lighting and sensor effects respectively.

3. Generalized color intrinsic image model

Thus far, intrinsic image decomposition works have obviated the photometric issues described above and have made some assumptions to simplify the problem. Under the assumptions of Lambertian surfaces and single white light, the intrinsic image decomposition problem is reduced to the estimation of shading and reflectance images [8, 30, 28, 14] according to

$$I(x, y) = I_{Shad}(x, y) \circ I_{Refl}(x, y)^2, \quad (9)$$

where I_{Shad} represents the amount of reflection arriving to the (x, y) point of the image from a specific point of the object surface considering the shape of the objects in the scene and the position of the light source, and I_{Refl} describes how the light is reflected by the corresponding point of the object considering the material reflectance properties. Since the model is assumed to be defined for any image point, the notation (x, y) is further omitted for the sake of simplicity.

In [7] a specular term was added to the model

$$I = I_{Shad} \circ I_{Refl} + I_{Spec}, \quad (10)$$

where I_{Spec} denotes the specular reflection of the objects in the scene.

Barron et al. [5] modeled the shading image as a function, M , of the shape of the scene, I_{Shape} , and the color and direction of the illuminant, $\mathbf{L} = [\mathbf{L}_{CLig}, \mathbf{L}_{GLig}]$, which led to the decomposition

$$I = I_{Shad} \circ I_{Refl} = M(I_{Shape}, \mathbf{L}) \circ I_{Refl}. \quad (11)$$

However, the effects of camera sensors and illumination have not been jointly considered. As explained above, image values are affected by both factors and, therefore, the reflectances recovered by previous methods depend on the illuminant and the sensor used to acquire the image. We refer to such recovered reflectances as relative reflectances.

²Here ' \circ ' denotes the Hadamard product

To overcome these dependence problems that relative reflectance images have, we propose a general framework for intrinsic image decomposition which takes into account both illuminant and camera sensor effects, thus allowing us to recover the reflectance images as if they were acquired with the standard sRGB sensor under the canonical illuminant. We define such images as absolute reflectance images, and denote them I_{Refl}^a .

As introduced in Eq.8 the physical properties of the scene at a pixel (reflection model, geometry, etc.) encoded in $\mathbf{p}^{s,c}$ can be isolated from the effects of the camera sensors and the illuminant of the scene, described respectively in \mathbf{S}_{Sen} and \mathbf{L}_{CLig} . According to the DRM, we can decompose these values, $\mathbf{p}^{s,c}$, into their diffuse and specular components as

$$\mathbf{p}^{s,c} = \mathbf{p}_{Shad}^{s,c} \circ \mathbf{p}_{Refl}^{s,c} + \mathbf{p}_{Spec}^{s,c}, \quad (12)$$

where $\mathbf{p}_{Shad}^{s,c}$ and $\mathbf{p}_{Refl}^{s,c}$ describe the magnitude and composition of the body (*i.e.* diffuse) reflection, and $\mathbf{p}_{Spec}^{s,c}$ represents the magnitude of the surface (*i.e.* specular) reflection. Hence, from Eqs.8 and 12 the general formulation at a pixel, \mathbf{p} , is

$$\mathbf{p} = \mathbf{L}_{CLig} \cdot \mathbf{S}_{Sen} \cdot (\mathbf{p}_{Shad}^{s,c} \circ \mathbf{p}_{Refl}^{s,c} + \mathbf{p}_{Spec}^{s,c}), \quad (13)$$

Eq. 13 can be extended from a pixel level to a whole image level, leading to our proposal for an intrinsic image model which deals with absolute reflectances:

$$I = \mathbf{L}_{CLig} \cdot \mathbf{S}_{Sen} \cdot (I_{Shad} \circ I_{Refl}^a + I_{Spec}). \quad (14)$$

3.1. Model particularities and relation to previous models

Depending on the knowledge we have about the scene and the acquisition conditions, or the assumptions we make on them, our model leads to different simplifications of the general formula proposed in Eq.14. Furthermore, existing formulations (Eqs.9-11) in the intrinsic image decomposition field can be derived from our model as specific cases. Let us describe some of these assumptions and discuss how they influence the general formulation of our model.

When surfaces in the scene are Lambertian, specularities can be discarded from the general formulation (*i.e.* $I_{Spec} = 0$) and Eq. 14 becomes:

$$I = \mathbf{L}_{CLig} \cdot \mathbf{S}_{Sen} \cdot (I_{Shad} \circ I_{Refl}^a). \quad (15)$$

When there is a single canonical light source, our illuminant transformation happens to be a diagonal matrix with a single scalar value for the three channels. Therefore, \mathbf{L}_{CLig} is substituted by the scalar value κ :

$$I = \kappa \cdot \mathbf{S}_{Sen} (I_{Shad} \circ I_{Refl}^a + I_{Spec}). \quad (16)$$

When camera sensors are narrow band, \mathbf{L}_{CLig} can be described by a diagonal matrix. When this happens, we can write $\mathbf{L}_{CLig} = I_{CLig}$, where I_{CLig} is a 3-channel image containing a single homogeneous color (*i.e.* the values of

the color are the values of the diagonal in \mathbf{L}_{CLig}), and use the Hadamard product to multiply this term with the shading and the absolute reflectance intrinsic images:

$$I = \mathbf{S}_{Sen} \cdot (I_{CLig} \circ I_{Shad} \circ I_{RefI}^a + I_{CLig} \circ I_{Spec}). \quad (17)$$

As explained in section 2.3, if camera sensors are not narrow band, spectral sharpening can be applied to make them more narrow band, which allows us to write $\mathbf{L}_{CLig} = I_{CLig}$ as in the previous case. Including the transform \mathcal{T} defined in Eq.6, the formulation above stands:

$$I = \mathcal{T}^{-1} \cdot [I_{CLig} \circ I_{Shad} \circ (\mathcal{T} \cdot \mathbf{S}_{Sen} \cdot I_{RefI}^a) + I_{CLig} \circ (\mathcal{T} \cdot \mathbf{S}_{Sen} \cdot I_{Spec})]. \quad (18)$$

When the camera sensors are the canonical sRGB sensors, \mathbf{S}_{Sen} is the identity matrix, resulting in

$$I = \mathbf{L}_{CLig} \cdot (I_{Shad} \circ I_{RefI}^a + I_{Spec}). \quad (19)$$

However, when the conditions described above are not fulfilled, as it is usually the case with natural images, we cannot estimate an accurate intrinsic absolute reflectance I_{RefI}^a . In such case, the absolute reflectance can only be approximated by a relative reflectance image, I_{RefI}^r . For instance, when we have no knowledge about the camera sensors and canonical sRGB sensors are assumed (i.e. estimating \mathbf{S}_{Sen} is not possible), the estimated reflectance is $I_{RefI}^r = \mathbf{S}_{Sen} \cdot I_{RefI}^a$.

Existing models for intrinsic image decomposition can also be derived from this generalized model by considering their specific assumptions. Several approaches [28, 27, 24] assume Lambertian objects ($I_{Spec} = 0$), white light ($\mathbf{L}_{CLig} = \kappa$) and unknown camera sensors ($\mathbf{S}_{Sen} = Id$), which yields

$$I = \kappa \cdot (I_{Shad} \circ I_{RefI}^r), \quad (20)$$

where $I_{RefI}^r = \mathbf{L}_{CLig} \cdot \mathbf{S}_{Sen} \cdot (I_{RefI}^a + I_{Spec})$. This is equivalent to Eq.9. Other approaches [7], deal with the specular image term while keeping the white light assumption ($\mathbf{L}_{CLig} = \kappa$) and having no knowledge of the camera sensors ($\mathbf{S}_{Sen} = Id$). This leads to

$$I = \kappa \cdot (I_{Shad} \circ I_{RefI}^r + I_{Spec}), \quad (21)$$

where $I_{RefI}^r = \mathbf{L}_{CLig} \cdot \mathbf{S}_{Sen} \cdot I_{RefI}^a$. This is, in fact, Eq.10. Finally, Barron et al.[5] (Eq.11) relaxed the white light assumption. If narrow band sensors are assumed ($\mathbf{L}_{CLig} = I_{CLig}$) our model becomes:

$$I = I_{Shad} \circ I_{CLig} \circ I_{RefI}^r, \quad (22)$$

where $I_{RefI}^r = \tilde{\mathbf{L}}_{CLig} \cdot \mathbf{S}_{Sen} \cdot (I_{RefI}^a + I_{Spec})$, and $\tilde{\mathbf{L}}_{CLig}$ represents the information of the light source not modeled by I_{CLig} when the sensors are not narrow-band.

Scenario	Camera Sensors	Scene Illuminant	Illuminant Model
SC1	Known	Known	Full
SC2	Known	Known	Diagonal
SC3	Unknown	Known	Diagonal
SC4	Unknown	Unknown	Diagonal

Table 1: Description of the scenarios used in the experiment.

4. Measuring light and sensor effects

In the previous sections we have proposed a new model for intrinsic image decomposition that revisits the simplistic image product formulation. Our approach includes constraints derived from considering the color of the light source and the bias introduced by the acquisition process. In this section we want to measure to what extent these considerations can improve intrinsic images decomposition in both, synthetic and real data.

4.1. Experiment 1: Synthetic data

We propose a set of experiments simulating different scenarios involving different assumptions on the sensors, the light source, or the illuminant model.

These scenarios differ by the degree of knowledge of the scene. The camera sensors may be either known (in which case we use the corresponding sensor responsivity functions), or unknown (standard sRGB sensors are assumed). We use the light source spectral power distribution when it is known or estimated, or a canonical D65 illuminant when information about the light source is not available. The model of illumination can either be a diagonal or a full 3-by-3 model. Taking these considerations into account, we define four different scenarios (summarized in Table 1):

Scenario 1 (SC1). It corresponds to the ideal case. We know both camera sensors and scene light source, and \mathbf{L}_{CLig} is approximated by a full illuminant model.

Scenario 2 (SC2). It is identical to SC1 except that a diagonal illumination model is assumed, just as most color constancy methods do.

Scenario 3 (SC3). Here the camera sensors are unknown (i.e. standard sensors are assumed) but the scene illuminant is known or can be estimated.

Scenario 4 (SC4). Here we have no knowledge of either the camera sensors (standard sensors are assumed) or of the lighting conditions (a canonical illuminant is assumed).

For this experiment we use a dataset of 1995 spectra compiled from several sources in Barnard *et al.* dataset [4]. These reflectances include the 24 Macbeth color-checker patches, 1269 Munsell chips, 120 Dupont paint chips, 170 natural objects, the 350 surfaces in Krinov dataset, and 57 additional reflectances. These data allow us to synthesize the scenes for the experiment. To simulate the light source of the scenes, we have used the spectral power distributions of 10 different illuminants, both Planckian and non-Planckian. These illuminants have been chosen from the same Barnard *et al.* dataset [4]. Their spectral power distributions are

shown in the supplementary material. To simulate the imaging process we have used 7 different sets of commercial cameras sensors (Canon EOS1D, Sigma Foveon D10, Kodak DCS420, Leica M8, Nikon D70, Olympus E400, Sony DXC930, and TVI MSC-1024RGB12).

We have performed the experiments on 10 different scenes, each one synthesized with 140 different reflectances randomly chosen from Barnard’s dataset. Each scene has been integrated with any of the 10 illuminants and subsequently with any of the 7 sets of camera sensors, providing the pixel values \mathbf{p} . Our ground-truth (*i.e.* absolute reflectances) are the same 10 scenes integrated with the canonical D65 illuminant spectral power distribution and standard sRGB sensitivity curves. The ground-truth values are denoted by $\mathbf{p}^{s,c}$.

We also acquire relative reflectance values by integrating each reflectance spectral distribution with the spectral power distribution of the canonical illuminant D65 and each of the 7 sets of camera sensors, resulting in a set of \mathbf{p}^c values. We approximate the camera sensor transformation matrix, \mathbf{S}_{Sen} , from $\mathbf{p}^{s,c}$ and \mathbf{p}^c using a 3-by-3 matrix estimated by least squares minimization. The illuminant transformation matrix \mathbf{L}_{CLig} can be estimated from \mathbf{p} and \mathbf{p}^c values following the same procedure.

Figure 2 illustrates the results of this experiment. To measure the error we use the chromatic angular error between ground truth values and the reflectance recovered in any of the scenarios. This error metric has been previously used in the literature to test the accuracy of color constancy methods [16]. We observe that the error behaves as expected: the less information we have about the camera sensors and the scene illuminant, the worse our reflectance estimate is. However, even when we have partial information we are still able to estimate the illuminant transformation matrix and the camera sensor behavior, which makes our method widely applicable.

4.2. Experiment 2: Natural images

In our second experiment, we use a set of images from Flickr (<http://www.flickr.com/>). We choose images from six well-known scenarios (Figure 3), taken with four popular DSLR camera models, namely Canon EOS 5D Mark II, Nikon D50, Nikon D7000 and Sony NEX-5.

The images selected for this experiment fulfill the basic assumptions of our model: scenes with a single light source and known camera sensors. The amount of pictures that we choose for each camera model and image category depends on the number of available pictures which satisfy these requirements.

When the set of images for this experiment has been selected, we first remove its gamma correction. Then, we manually choose two or three regions for each picture. These regions describe a single material under similar illumination conditions (*i.e.* overshadowed and saturated areas of the picture are avoided). The mean value of these regions is used as a color descriptor for this picture.

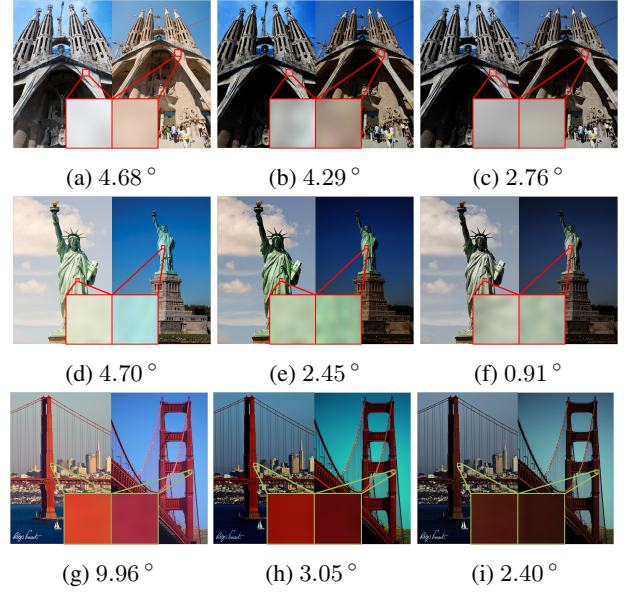


Figure 4: Results on 3 pairs of natural images. At each row, in the first column we see a pair of original images. In the second column a color constancy algorithm has been applied (Grey-Edge [29]). In the last column, we see the images of the second column after their camera sensor effects have been removed. The chromatic angular errors are expressed in degrees.

For each of the pictures we apply different color constancy methods (Shades of Grey [10] and Grey-Edge [29]). We finally remove the effects of the sensors for each of them (*i.e.* express the image values under sRGB sensors) by using the transformation matrices we obtain at DxOMark website (<http://www.dxomark.com/>).

Figure 4 shows some qualitative results where the improvement after applying a color constancy algorithm and removing the effect of the camera sensors is clear. In Figure 5 we present quantitative results using a color constancy algorithm to estimate the illuminant of the scenes. For each picture, we use the color descriptors described above together with the chromatic angular error to compare each possible pair of images. Using this measure we compute the amount of pairs of pictures which are closer, from a chromatic point of view, after applying our model. When the effects of both the illuminant and the camera sensors are removed, there is an average decrease in the error of 22.59%. These results are coherent with those we obtain in our previous experiment with synthetic data. Figure 1 shows how this model can be applied in practice to the problem of intrinsic image decomposition. More examples on synthetic and lab-acquired objects can be found in the supplementary material.

5. Discussion and conclusion

In this paper we have shown how the application of insights from the color constancy literature has led to an improved general framework that explicitly models the effects

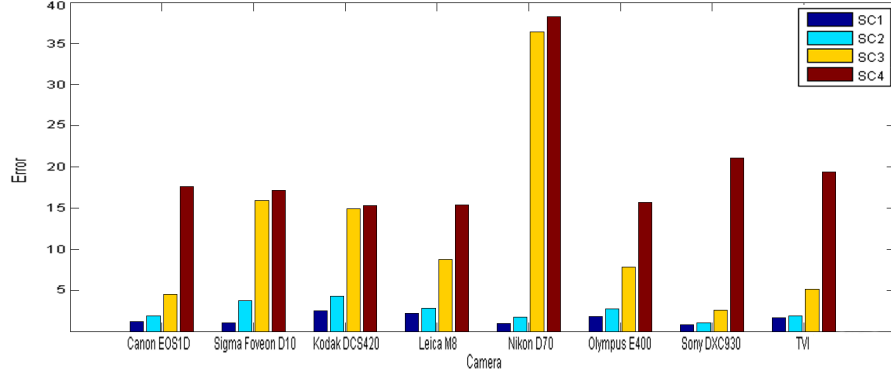


Figure 2: Mean angular error for different commercial cameras in multiple scenarios. The most information we have about our scene, the best our reflectance estimate will be. Scenario 4 always assumes canonical illumination and standard sensors, like most existing intrinsic image models. Scenarios 1 and 2 represent our model when there is some knowledge about scene illumination and camera sensors.



Figure 3: Image categories: From left to right, the Golden Gate bridge (San Francisco), the Statue of Liberty (New York City), the Kinkakuji temple (Kyoto), the church of Sagrada Família (Barcelona), the Uluru Rock (Australian desert) and St. Basil's cathedral (Moscow).

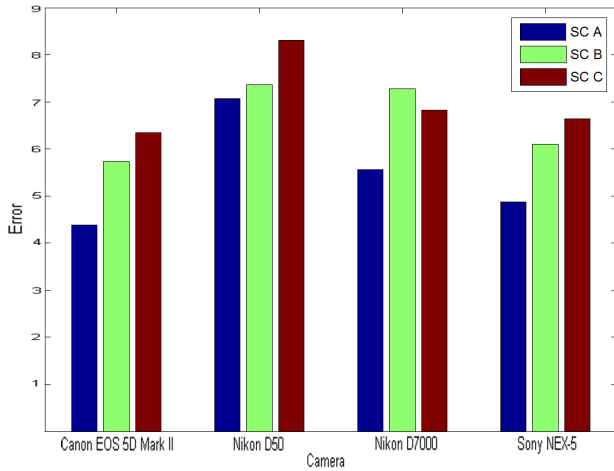


Figure 5: Mean angular error among pairs of images from our set of natural images. In scenario C we always assume canonical illumination and standard sensors. In scenario B standard sRGB sensors are assumed but we estimate the illuminant using a color constancy method (Shades of Grey [10]). In scenario A we estimate the illuminant and also remove the effects of the camera sensors. The error decreases accordingly to the available information about the illuminant and the camera sensors. Similar results are obtained when using Grey-Edge [29].

of illuminant and sensor responses. This enables the estimation of absolute reflectance images, an elaboration on the intrinsic image concept, which exhibit increased invariance to illuminant or sensor changes. As direct consequences of our framework we can model camera effects. When the camera sensors are sufficiently narrow band, the illuminant transformation model is described by means of a diagonal matrix. When sensor responses are not narrow band but known, spectral sharpening [12] can be applied. Otherwise, a full (*i.e.* non-diagonal) linear model is used. Furthermore, we can model scene illumination. This is an important step towards the normalization of color through different acquisitions with the same device.

Absolute reflectances could prove useful to many applications. Imagine a set of cameras placed outdoors with known camera sensors (*e.g.* transit cameras placed along a highway). Illumination effects will be very diverse depending on the location of each camera and the time of the day and the year when the images were taken. However, if we estimate the color of the illuminant using color constancy, then we could infer an absolute reflectance image which would be very useful for further computer vision tasks. Imagine we have multiple camera devices in a lab. We do not know anything about their sensors, but we have control over the scene illumination and know its spectral distribution. If we place an object with some known reflectance values in the scene, we can estimate a sensor transformation for each camera. This way, we can estimate

absolute reflectance images for any camera. Finally, imagine we have a camera with unknown sensors and we have no knowledge about the illuminant either. However, we know or can approximate some reflectance values which appear in the scene. This could lead to data-driven approaches for the inference of illuminant and sensor type. Furthermore, one could envision absolute reflectance descriptors, which would provide a characterization of object materials invariant to specific device characteristics.

Acknowledgements: This work has been supported by projects TIN2010-21771-C02-1 of Spanish Ministry of Science and Innovation, and 2009-SGR-669 of Generalitat de Catalunya. Dimitris Samaras acknowledges support from projects NSF IIS-1111047 and IIS-1161876 and the SUB-SAMPLE Project from the DIGITEO Institute, France.

References

- [1] IEC 61966-2-1 / FDIS: Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB. [3](#)
- [2] A. Artusi, F. Banterle, and D. Chetverikov. A survey of specular removal methods. *Computer Graphics Forum*, 30(8):2208–2230, 2011. [1](#), [2](#)
- [3] K. Barnard and B. Funt. Camera characterization for color research. *Color Research and Application*, 27(3):152–163, 2002. [3](#)
- [4] K. Barnard, L. Martin, B. Funt, and A. Coath. A data set for color research. *Color Research and Application*, 27(3):148–152, 2002. [5](#)
- [5] J. T. Barron and J. Malik. Color constancy, intrinsic images, and shape estimation. In *European Conference on Computer Vision*, pages 57–70, 2012. [2](#), [4](#), [5](#)
- [6] H. G. B. Barrow and J. M. Tenenbaum. Recovering intrinsic scene characteristics from images. In *Computer Vision Systems*, pages 3–26, 1978. [1](#)
- [7] S. Beigpour and J. van de Weijer. Object recoloring based on intrinsic image estimation. In *International Conference on Computer Vision*, pages 327–334, 2011. [2](#), [4](#), [5](#)
- [8] M. Bell and W. T. Freeman. Learning local evidence for shading and reflection. In *International Conference on Computer Vision*, pages 670–677, 2001. [2](#), [4](#)
- [9] J.-D. Durou, M. Falcone, and M. Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding*, 109(1):22–43, 2008. [1](#)
- [10] G. D. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *SID Color Imaging Conference*, pages 37–41, 2004. [6](#), [7](#)
- [11] G. Finlayson, M. Drew, and B. Funt. Color constancy: Generalized diagonal transforms suffice. *Journal of the Optical Society of America A*, 11(11):3011–3019, 1994. [2](#), [3](#)
- [12] G. D. Finlayson, M. S. Drew, and B. V. Funt. Spectral sharpening: Sensor transformations for improved color constancy. *Journal of the Optical Society of America A*, 11(5):1553–1563, 1994. [3](#), [7](#)
- [13] D. Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–35, 1990. [2](#), [3](#)
- [14] B. Funt, M. Drew, and M. Brockington. Recovering shading from color images. In *European Conference on Computer Vision*, pages 124–132, 1992. [2](#), [4](#)
- [15] P. V. Gehler, C. Rother, M. Kiefel, L. Zhang, and B. Schölkopf. Recovering intrinsic images with a global sparsity prior on reflectance. In *Neural Information Processing Systems*, pages 765–773, 2011. [2](#)
- [16] A. Gijsenij, T. Gevers, and J. van de Weijer. Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing*, 20(9):2475–2489, 2011. [1](#), [2](#), [3](#), [6](#)
- [17] P. M. Hubel, J. Holm, G. D. Finlayson, and M. S. Drew. Matrix calculations for digital photography. In *Color Imaging Conference: Color Science, Systems, and Applications*, pages 105–111, 1997. [3](#)
- [18] D. S. Immel, M. F. Cohen, and D. P. Greenberg. A radiosity method for non-diffuse environments. *ACM SIGGRAPH*, 20(4):133–142, 1986. [2](#)
- [19] X. Jiang, A. J. Schofield, and J. L. Wyatt. Correlation-based intrinsic image extraction from a single image. In *European Conference on Computer Vision*, pages 58–71, 2010. [2](#)
- [20] G. J. Klinker. *A Physical Approach to Color Image Understanding*. A. K. Peters, Ltd., 1993. [3](#)
- [21] E. Land and J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971. [2](#)
- [22] A. Ngan, F. Durand, and W. Matusik. Experimental analysis of brdf models. In *Eurographics Conference on Rendering Techniques*, pages 117–126, 2005. [2](#)
- [23] A. Parraga, R. Baldrich, and M. Vanrell. Accurate mapping of natural scenes radiance to cone activation space: a new image dataset. In *European Conference on Colour in Graphics, Imaging, and Vision*, pages 50–57, 2010. [3](#)
- [24] M. Serra, O. Penacchio, R. Benavente, and M. Vanrell. Names and shades of color for intrinsic image estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 278–285, 2012. [2](#), [5](#)
- [25] S. A. Shafer. Using color to separate reflection components. *Color Research and Applications*, 10(4):210–218, 1985. [2](#), [3](#)
- [26] L. Shen and C. Yeo. Intrinsic images decomposition using a local and global sparse representation of reflectance. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 697–704, 2011. [2](#)
- [27] M. F. Tappen, E. H. Adelson, and W. T. Freeman. Estimating intrinsic component images using non-linear regression. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1992–1999, 2006. [5](#)
- [28] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(9):1459–1472, 2005. [2](#), [4](#), [5](#)
- [29] J. van de Weijer and A. Gijsenij. Edge-based color constancy. *Transactions on Image Processing*, 16(9):2207–2214, 2007. [6](#), [7](#)
- [30] Y. Weiss. Deriving intrinsic images from image sequences. In *International Conference on Computer Vision*, pages 68–75, 2001. [2](#), [4](#)
- [31] G. West and M. H. Brill. Necessary and sufficient conditions for von kries chromatic adaptation to give color constancy. *Journal of Mathematical Biology*, 15(2):249–258, 1982. [3](#)