Color Constancy by Category Correlation

Javier Vazquez-Corral, Maria Vanrell, Ramon Baldrich, Francesc Tous

***Abstract*—Finding color representations which are stable to illuminant changes is still an open problem in computer vision. Until now most approaches have been based on physical constraints or statistical assumptions derived from the scene, while very little attention has been paid to the effects that selected illuminants have on the final color image representation.**

The novelty of this work is to propose perceptual constraints that are computed on the corrected images. We define the *category hypothesis*, which weights the set of feasible illuminants according to their ability to map the corrected image onto specific colors. Here we choose these colors as the universal color categories related to basic linguistic terms which have been psychophysically measured. These color categories encode natural color statistics, and their relevance across different cultures is indicated by the fact that they have received a common color name.

From this category hypothesis we propose a fast implementation that allows the sampling of a large set of illuminants. Experiments prove that our method rivals current state-of-art performance without the need for training algorithmic parameters. Additionally, the method can be used as a framework to insert top-down information from other sources, thus opening further research directions in solving for color constancy.

***Index Terms*—Color constancy, color naming, color categories, category correlation**

I. Introduction

Color is derived from three components: the reflectance of the object, the sensitivity of cones, and the illuminant spectra. Of these components, the illuminant spectrum is the least stable. Illumination changes depending on different aspects: time of the day (daybreak, midday, sunset), or indoor/outdoor situations, for example. Thus the problem for computer vision is that the color of an object depends on the light under which we are looking at it. The human visual system solves this problem thanks to the so-called color constancy property [1]. This property allows humans to identify the color of an object independently of the color of the light source.

Color constancy is important for human vision, since color is a visual cue that helps in solving different visions tasks such as tracking, object recognition or categorization. Therefore, several computational methods have tried to simulate human color constancy abilities to stabilize machine color representations. Two different kinds of approach have been used: normalization and constancy. Whilst color normalization creates a new representation of the image by cancelling illuminant effects [2], [3]; color constancy directly estimates the color of the illuminant in order to map the image colors to a canonical version. In this paper we focus on this second kind of approach.

Authors are with the Computer Vision Center, Department of Computer Sciences, Universitat Autònoma de Barcelona,08193, Bellaterra, Barcelona , Spain, e-mail: javier.vazquez@cvc.uab.cat.

This work has been partially supported by projects TIN2007-64577, TIN2010–21771-C02-1 and Consolider-Ingenio 2010 CSD2007-00018 of Spanish MEC (Ministery of Science).

Computational color constancy is an under-constrained problem and therefore it does not have a unique solution. Due to this ill-posed nature, a large number of methods have been proposed over a period spanning more than 20 years [4], and yet a widely accepted solution to the illuminant estimation problem [5] is still elusive. Existing solutions can be divided into two main families, *statistical* and *physical*.

Statistical methods can also be split in three types. The first type, based on *simple image statistics*, are the most common color constancy methods. In this group we have *Grey-World* [6], *White-Patch* [7], *Shades of Grey* [8], *Grey-Edge* [9], and *Bag-of-Pixels* [10]. A second type are *Gamut Mapping methods*. These methods are based on the seminal work of Forsyth [11], where he introduced the C-Rule algorithm. Improvements on C-Rule have been reported in [12] and [5]. However, these methods have a significant drawback for computer vision applications: they need calibrated conditions, that is, to know the camera sensitivities. The last type of statistical methods are *Probabilistic* or *Bayesian methods*, where prior information is used to correct the illuminant. Color-by-Correlation [13], Bayesian Color Constancy [14], [15] and Voting methods, such as [16], belong to this group.

Physical methods use a more general model of image formation than that used in statistical approaches. While statistical methods assume surfaces are Lambertian, physical methods assume Shafer's dichromatic model [17]. Some examples of methods using this approach are found in [18], [19], [20], [21].

All the above mentioned works try to solve the ill-posed nature of the constancy problem either by constraining the size of the feasible set of solutions (reducing either the number of illuminants or number of the reflectances that can be found in scenes) or by making physical or statistical assumptions about the scene and the image content.

None of the previous computational approaches have introduced perceptual constraints. Consequently, very little attention has been paid to how the selected illuminant affects the perception of the content of the corrected image. Evidence derived from experimental psychology on natural images gives support to the conclusion that several different perceptual mechanisms contribute to achieve constant images [1]. Different mechanisms based on different visual cues such as the local and global contrast [22], [23], highlights [24], mutual reflections [25], categorical or naming stability [26] and color memory of known objects [27], [28] are responsible for the almost perfect behaviour of the human constancy system. In this paper we focus on the definition of a color constancy method that considers the perceptual effects of categorization on the corrected image.

In this work we concentrate on the naming stability cue. We propose the *naming hypothesis* as a criterion to constrain the feasible illuminants. We propose to use the capability of

categorizing, or assigning basic color names, in the corrected image as the basis to weight all feasible illuminants. In this sense, preferred illuminants will produce a color categorized image with useful properties for further recognition tasks. Moreover, our process can be justified as it produces an image labelled with the color categories that encode natural color statistics which have evolved as relevant across different cultures by receiving a common color name. The existence of the basic color category terms was noted for the first time by Berlin and Kay [29], who recorded 11 basic terms. These basic terms were lately measured by Boynton and Olson [30] in psychophysical experiments.

Using the category hypothesis, we propose a computational approach that is a probabilistic method similar to illuminant voting [16] or color by correlation [13], but with two essential novelties that we list below.

Firstly, the method gives a compact framework that allows prior-knowledge from learnt-color categories to be easily introduced. Illuminant selection is done through the *category hypothesis*, which is defined as the preference of illuminants that assign color categories in the corrected images. In particular, we want to stress that this new algorithm can also be seen as a generalisation of simpler methods, such as, *WhitePatch* where we only consider the white category. This opens up a new way of generalizing simple methods to allow greater complexity (i.e. not only by increasing their statistical complexity).

Secondly, we present a fast algorithm that builds a weighted feasible set for a fine sampling of the feasible illuminants. This fast algorithm can also be seen as a fast implementation of the Color by Correlation approach [13] for the 3D case [31] in the particular case of a diagonal model of illuminant change. This fast algorithm requires the representation of the weighted feasible set in logarithm space. This in turn improves the illuminant selection step, since multiple solutions can be easily considered using a compact representation.

To evaluate the performance of the proposed approach, we compare our results with the existing state-of-the-art in terms of how well the illuminant is estimated. The results suggest that our approach achieves the performance of the other methods, whilst also incorporating the advantages mentioned above.

The paper has been organised as follows. In section II we explain the basic color term categories. Afterwards, in section III we introduce the category hypothesis, and we report the results compared to other current methods in sections IV and V. We conclude in section VI.

II. Basic Term Categories

Basic color term categories were first defined by Berlin and Kay [29], and they were deduced from a large anthropological study based on speakers of 20 different languages and specific documentation from a further 78 languages. They concluded that the universal basic color terms defined in most evolved languages are *white, black, red, green, yellow, blue, brown, purple, orange, pink* and *gray*. In subsequent works, psychophysical experiments have generated data that allow these basic categories to be specified accurately [30], [32], [33]. These datasets give 11 categories where colors have been labelled with a unique name. They are obtained from the averaged judgements given by all subjects in the experiment.

Basic color categories are derived from anthropological and psychophysical experiments that bring us to the conclusion that relevant colors are those that receive a common color name across different cultures. A similar conclusion about the relevance of these specific color categories has also been derived from a biological model of the human color sensors [34]. This work provides strong evidence that color coding in human vision favours these color categories. There are evidences that basic color terms are likely to be encoding fundamental natural color statistics [35]. That makes sense in an evolutionary theory as they would capture the most relevant information to survive.

In this work we make use of a mapping of these categories onto CIELab space provided by Benavente-*et al*- [33]. The first row in figure 1 shows the chromaticity of the convex-hull of these mapped colors at three different levels of intensity in the CIELab space. These polyedron contain the parts of the color space that are judged as pure colors (or focal colors); i.e. those colors named with a unique basic term. We will use these sets of colors as the anchor categories that will determine the corrected images. These sets are the focal points (F_i) of the corresponding color. We use the CIELab space for figure 1 for explanatory purposes but in the rest of the paper we refer to RGB space that is the space used in all the reported experiments on the standard datasets. To build the category matrix in RGB we use the reflectances corresponding to the named colors, the canonical light (white illuminant) and the RGB color matching functions.

In order to also encode common changes of these colors in real scenes, such as those in shadowed areas or textured surfaces, or even colors reproduced in man-made objects, we are going to experiment with some extensions of these basic categories, whilst not extending them beyond the convex-hull of the basic terms. Therefore, we define our categories depending on the distance to the focal points, whilst constraining them to remain inside the Convex Hull of the focal terms. Thus, a category C_i^β is defined as

$$C_i^\beta = \{p : d(p, F_i) < \beta, p \in CH(F)\} \tag{1}$$

where p is a point in RGB space, $F = \{F_i\}_{i=1:11}$ is the set of focal colors presented in [33], CH represents the convex hull of a set of points and d refers to the euclidean distance.

Then, from these equations, we are able to define a family of category sets by changing the β value. In Figure 1 we show some examples for these sets, where the first row represents the original basic categories ($\beta = 0$) as horizontal cross-sections in Lab space ($L = 25$, $L = 45$, and $L = 65$), and the second and third rows represent two different sets, $\beta = 10$ and $\beta = 20$ respectively. The grey background in all the different plots represents the global convex hull, which is the growing limit. To discretize category membership we will use a characteristic function defined as:

$$\mathcal{X}_{C_i^\beta}(p) = \begin{cases} 1 & \text{if} \quad p \in C_i^\beta, p \notin C_{j,j<i}^\beta \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $C^\beta = \{C_i^\beta\}_{i=1:11}$, i encodes each one of the eleven basic terms, namely {*white, black, red, green, yellow, blue, brown, purple, orange, pink* and *gray*} and p is a color representation vector. Condition $j < i$ is imposed to do not count twice those colors falling in the intersection of two categories. The order of the categories is not important for our results since different categories are equally weighted in our approach.

III. Category methods

We base our approach on the idea that color constancy aims to produce corrected images where important contents are stable. We refer to these important contents as basic color categories. These anchor categories constitute prior knowledge that is useful for general image understanding. Therefore we seek to correct images towards a new representation where these basic categories are anchors. This idea is formulated in the following hypothesis for color constancy:

Category Hypothesis: *Feasible illuminants can be weighted according to their ability to anchor the colors of an image to basic color categories.*

Thus, we will call Category Methods those that, applying this hypothesis, compute a weighted feasible illuminant set according to the set of anchor categories being used, and select one of them that allows us to obtain a corrected image whose colors falls into these categories.

In Figure 2 we show some examples of the results provided by the proposed hypothesis using the basic color terms categories. The original images are shown in the second column, while the first column presents the categorisation of these images. In the third column we give the corrected images and their corrected categorisation is given in the fourth column. Hence, from the first and the fourth column we can see how color categorization is changed, from the original to the corrected image, towards a more colorful image representation that in turn makes it more stable (e.g. sky is blue, the road is grey). Clearly, our proposal is simply a bottom-up approach that pursues a corrected, or more stable, image that needs further processing for full image understanding.

We will now explain our method in three parts: first, we will define the general mathematical formulation; secondly, we will explain the fast implementation of this mathematical formulation; and finally, we will explain the illuminant selection criteria.

A. Mathematical formulation

Let us define $P(e|I)$ as the probability of having illuminant e in image I. This is approximated as

$$P(e|I) \approx \frac{f(e)}{\sum_{\check{e} \in FS} f(\check{e})} = k_1 \cdot f(e) \tag{3}$$

where FS is the feasible set of illuminants (in the C-Rule sense, considering as canonical gamut the whole RGB cube) and the function $f(e)$ is defined in a voting procedure in the same manner as Sapiro in [16]. This voting function is defined as

$$f(e) = \sum_{p \in RGB_I} P(e|p) \tag{4}$$

where RGB_I represents the different colors appearing in the image, and $P(e|p)$ is the probability of having illuminant e given color p in the image. This probability is defined to follow the category hypothesis introduced earlier, thus

$$P(e|p) = P(e|p, C^\beta) = \frac{\sum_{C_i^\beta \in C^\beta} \mathcal{X}_{C_i^\beta}(p \cdot \mathrm{diag}(e)^{-1})}{\sum_{C_i^\beta \in C^\beta} \sum_{q \in RGB} (\mathcal{X}_{C_i^\beta}(q))} \tag{5}$$

quantifies the ability of illuminant e to categorize color p in the set of anchor categories denoted as C^β, and is normalized by the total amount of nameable colors. $\mathcal{X}_{C_i^\beta}(x)$, defined in equation 2, is responsible for counting the number of colors falling in each one of the categories for the specific illuminant.

To simplify the previous formulation, the denominator in equation 5 is substituted by a constant

$$k_2 = 1/ \sum_{C_i^\beta \in C^\beta} \sum_{q \in RGB} (\mathcal{X}_{C_i^\beta}(q)) \tag{6}$$

and we therefore rewrite $P(e|I)$ as

$$P(e|I) \approx k_1 \cdot k_2 \sum_{p \in RGB_I} \sum_{C_i^\beta \in C^\beta} \mathcal{X}_{C_i^\beta}(p \cdot \mathrm{diag}(e)^{-1}). \tag{7}$$

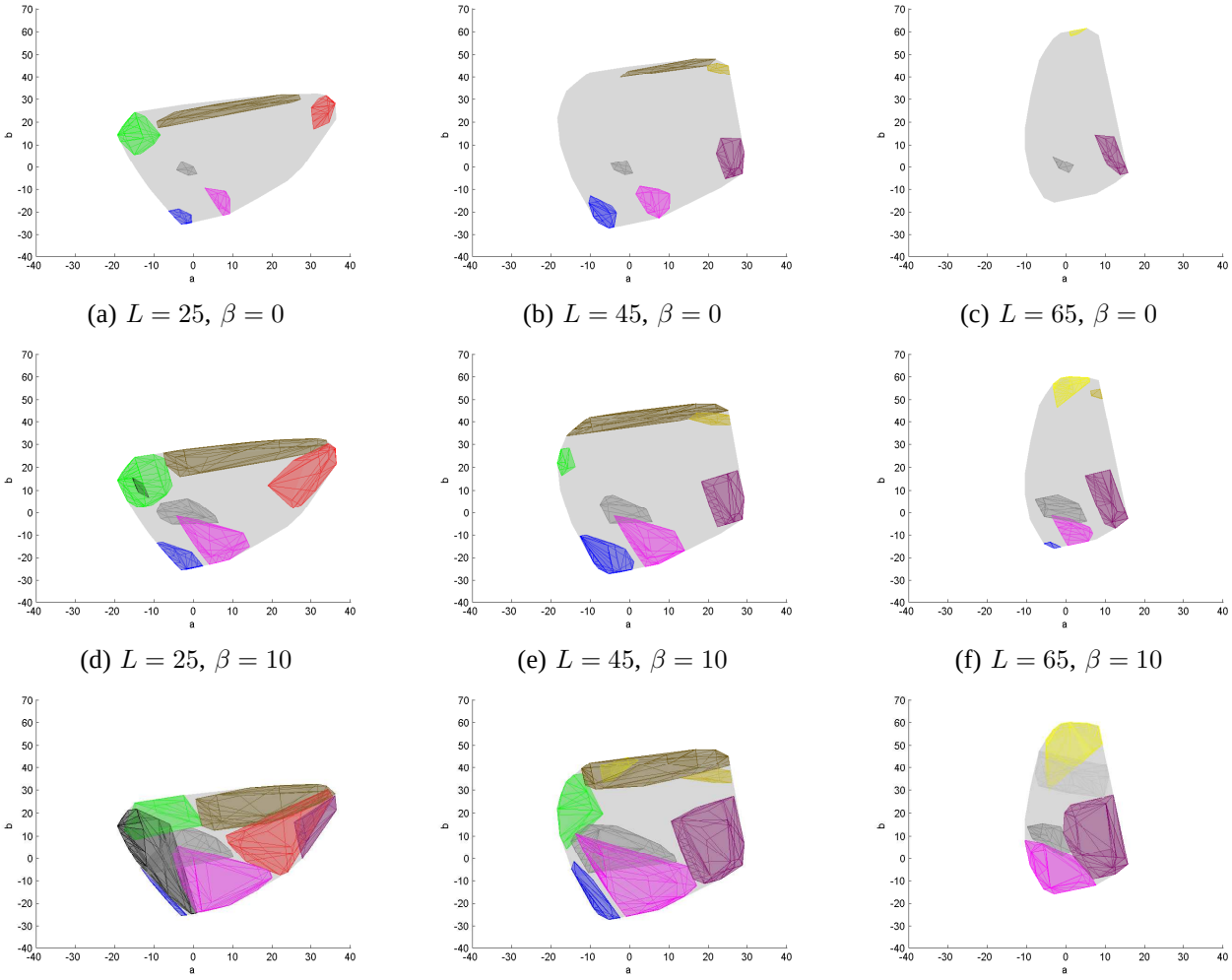
We want to highlight here, that this compact formulation could be used for a different set of categories than those used in this paper. Indeed, existing color constancy methods can be incorporated within this framework. For instance, using white as a unique category means that the method acts as a White-Patch algorithm, while taking all possible color values for a certain device as different categories behaves like the Color-by-Correlation [13] solution in the diagonal case for a 3D color space.

B. Fast implementation

The main problem of this formulation is its cpu time, which is large due to the double summation term. Therefore, in order to reach a fast implementation of the proposed voting approach, we reformulate equation (7) by reordering sums and obtaining

$$P(e|I) \approx k_1 \cdot k_2 \cdot \sum_{C_i^\beta \in C^\beta} \sum_{p \in RGB_I} \mathcal{X}_{C_i^\beta}(p \cdot \mathrm{diag}(e)^{-1}) \tag{8}$$

in this way, the inner summation is equivalent to a product of two functions histn and $\mathcal{X}_{C_i}^\beta$, where histn is the normalized histogram of the image I and $\mathcal{X}_{C_i^\beta}$ is the characteristic function of a category C_i^β. Both functions are defined over the complete RGB domain which allows the reformulation of the previous equation as



(a) $L = 25,\ \beta = 0$ (b) $L = 45,\ \beta = 0$ (c) $L = 65,\ \beta = 0$

(d) $L = 25,\ \beta = 10$ (e) $L = 45,\ \beta = 10$ (f) $L = 65,\ \beta = 10$

(g) $L = 25,\ \beta = 20$ (h) $L = 45,\ \beta = 20$ (i) $L = 65,\ \beta = 20$

Fig. 1. (a) Color name categories with luminance 25 in Lab space (b) Color name categories with luminance 45 in Lab space (c) Color name categories with luminance 65 in Lab space, (d), (e) and (f) first extension of the categories. (g),(h) and (i) second extension

$$P(e|I) \approx k_1 \cdot k_2 \cdot \sum_{C_i^\beta \in C^\beta} \sum_{r \in RGB} \text{histn}(r \cdot \text{diag}(e)^{-1}) \cdot \mathcal{X}_{C_i^\beta}(r). \tag{9}$$

Note that from now on, the inner summation is over the set of possible RGBs instead of over the values appearing in the image.

At this point we propose to estimate this probability by removing constants k_1 and k_2 and introducing a $\log$ monotonic function in the image domain. This implies that

$$\begin{aligned} P(e|I) &\approx k_1 \cdot k_2 \cdot \hat{P}(e|I) \\ &\propto \hat{P}(e|I) \\ = \sum_{C_i^\beta \in C^\beta} \sum_{r \in RGB} \widehat{\text{histn}}(\log(r \cdot \text{diag}(e)^{-1})) &\cdot \hat{\mathcal{X}}_{C_i^\beta}(\log(r)) \end{aligned} \tag{10}$$

where the membership function and the histogram function have been redefined in log space as $\hat{\mathcal{X}}_{C_i^\beta}(r) = \mathcal{X}_{C_i^\beta}(\exp(r))$ and $\widehat{\text{histn}}(x) = \text{histn}(\exp(x))$. Furthermore, considering that taking logarithms transforms products into additions, we can write

$$\begin{aligned} \hat{P}(e|I) = \\ \sum_{C_i^\beta \in C^\beta} \sum_{r \in RGB} \widehat{\text{histn}}(\log(r) + \text{diag}(\log(e))^{-1}) \cdot \hat{\mathcal{X}}_{C_i^\beta}(\log(r)) \end{aligned} \tag{11}$$

which brings us to compute a linear correlation of two functions

$$\hat{P}(e|I) = \sum_{C_i^\beta \in C^\beta} (\widehat{\text{histn}} * \hat{\mathcal{X}}_{C_i^\beta})(e) \tag{12}$$

that can be computed in the Fourier space as a simple product of functions. Using the *Fast Fourier Transform (FFT)* this can be done with a computational cost $O(n^3 \log(n))$.

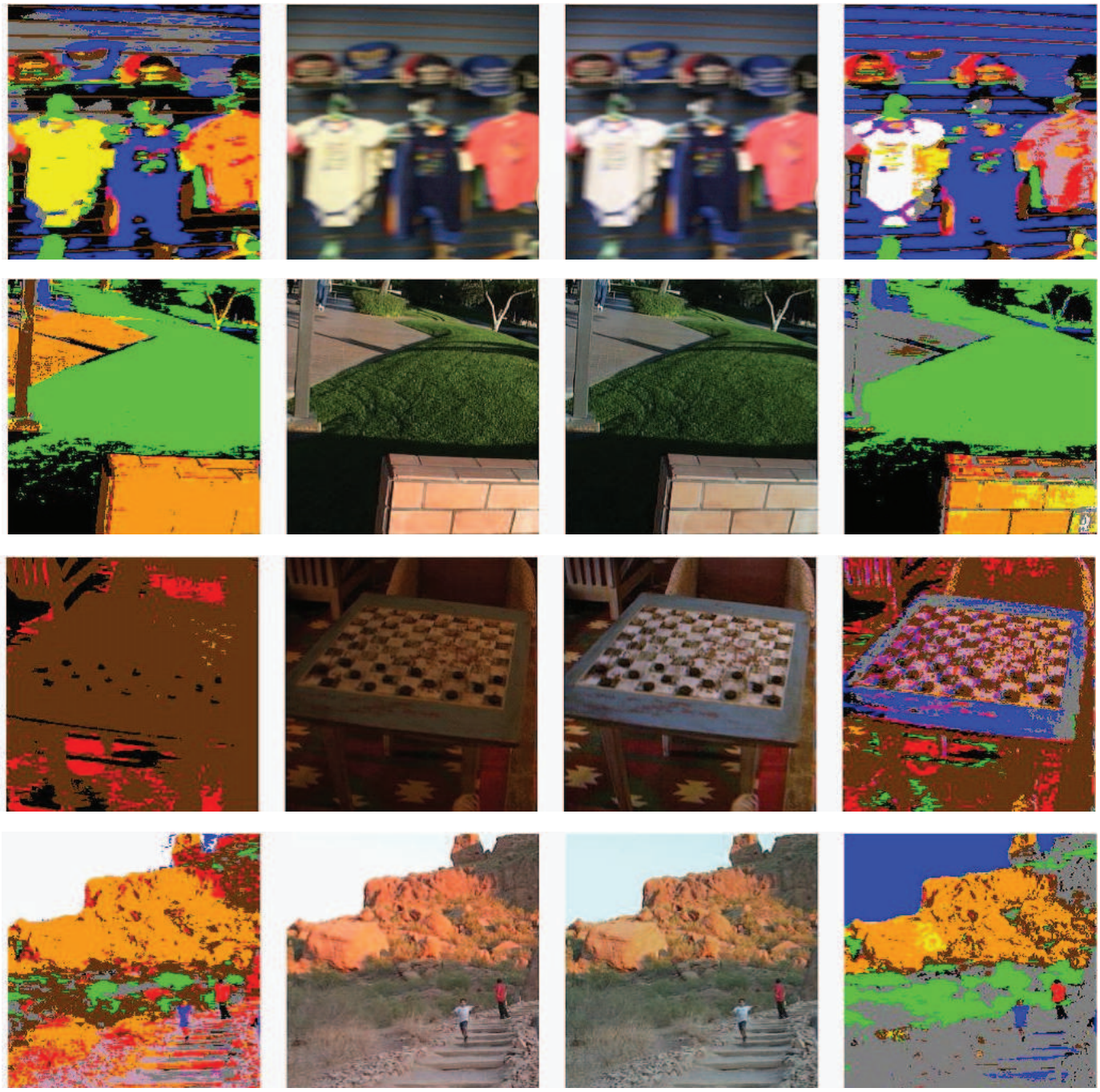


Fig. 2. Categorized original image (left), original image (center-left), corrected image (center-right), categorized corrected image(right)

C. Illuminant selection

In the foregoing sections we defined a computational framework that provides a weighted set of feasible solutions. The proposed algorithm assigns different probabilities to all plausible illuminants accordingly with the category hypothesis. The next step is to select the most relevant illuminant by using some specific criterion. To evaluate the performance of the hypothesis we set up experiments with two different criteria: i) selecting the illuminant with the maximum probability, which is the most common approach in probabilistic methods; and ii) selecting the illuminant by combining our feasible solutions with solutions provided by other methods which are based on a complementary hypothesis. In this way we can evaluate whether the category hypothesis can be improved by combining it with, for example, an edge-based hypothesis. This combination criterion can be seamlessly integrated within the proposed algorithm, which is another advantage of this framework. The use of a global convolution in the log-RGB space is the basis that allows the probabilities for a large sample of illuminants within the feasible set to be calculated, and allows us to work directly with these probabilities.

Using a maximum criterion we can formulate Category

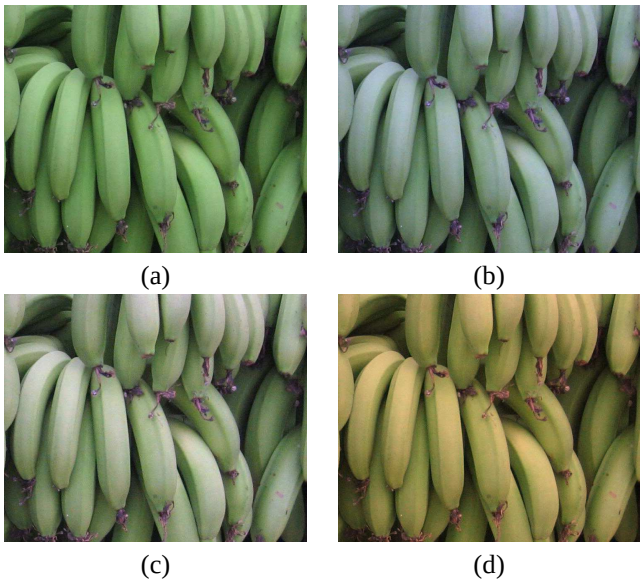


Fig. 3. Different feasible solutions for the same scene providing different explanations of that scene

Correlation methods (heretofore *CaC*) to deliver a unique solution, which is given by

$$\hat{e} = \arg\max_{e \in FS} P(e|I) \tag{13}$$

where $\hat{e}$ is the estimated illuminant for the scene based on equation (3).

Using a combination criterion we are assuming that our weighted feasible set is providing different plausible explanations of the corrected image. For instance, in some particular images such as the bananas shown in Figure 3, we can see that disambiguating the scene illuminant from the object reflectances is an unsolvable problem. In this case most of the solutions in the feasible set could be equally plausible since they could correspond to different ripeness of the fruit or different illuminants. The four images in Figure 3 have been obtained from a clustering with standard k-means with four classes onto the feasible set and extracting the illuminant with maximum probability as the representative of each cluster. In this case, the original image was close to the green bananas given in solution (a).

Accordingly with the previous observation we can state that working with multiple solutions can be an improvement to classical constancy approaches. One of the strengths of our method relies on the fact that a large sample of likely illuminants has already been computed. In this way we can extract multiple solutions by directly thresholding onto the weighted feasible set. Then, a multiple solution set for a given image I is given by

$$S^{\alpha} = \{e \in FS : P(e|I) > \alpha\}, \tag{14}$$

which denotes the set of illuminants having a probability higher than α. Providing multiple solutions allows us to delegate the final selection either to other visual processes with contextual information or to other top-down selective tasks. This approach has been used in [36] where an illuminant is selected to improve a scene recognition task from a variety of solutions from different constancy methods (and after a learning step). There are also other methods selecting a unique solution from a set of precomputed ones [37], [38]. These last methods use classifications techniques such as decision forest to this end.

Here in this work, we propose a criterion that estimates the best illuminant by selecting the solution from $S^{\alpha} = \{S_i\}_{i=1,\cdots,n}$ that is the most voted-for by solutions derived from other methods based on different hypotheses and which are denoted as $\{T_j\}_{j=1,\cdots,m}$. Formally, we select the most voted-for illuminant by computing

$$\hat{e} = S_{\arg\max_i \#\{v_j \in v : v_j = i\}} \tag{15}$$

where $v = \{v_j\}_{j=1,\cdots,m}$ encodes the solution of S^{α} that is closest to a solution in $\{T_j\}_{j=1,\cdots,m}$, and

$$v_j = \arg\min_i \mathrm{ang}(S_i, T_j) \tag{16}$$

where ang is the angular error distance between two given illuminants.

With this criterion we select an illuminant which has a high probability based on our own hypothesis and is reinforced by being close to the solutions provided by other hypotheses.

IV. Experiments

To evaluate our hypothesis we have run our method under different parameters, varying both the category sets and the selection criteria. We have used three different datasets and we have compared our results with the current state-of-the-art.

We denote our method as CaC^{β}_{sc} where sc denotes the selection criterion used and β refers to the category threshold defined earlier. The selection criterion will be m for selection based on maximum probability and c for a combined selection. For both selection criteria the value of β takes one out of four possible values: 0 (in order to use the basic categories), 10, 20 and 400. This last value has been defined in order to select the complete convex hull (grey polygon in Figure 1). In all the experiments our methods have worked with a $\log RGB$ cube of 50 bins, which implies a sampling of 50^3 different illuminants.

Specifically for the combined criterion, we have selected our solutions by setting $\alpha = 0.95 \cdot \max(P(e/I))$. We have combined these solutions with 24 solutions coming from different applications of the grey-edge hypothesis. We have used a wide range of statistical combinations of this hypothesis by fixing the following parameters $p = 1, 6, 11, 16$, $\sigma = 1, 3$ and $n = 0, 1, 2$ where p is the Minkowski Norm, σ the smoothness parameter and n the differentiation order

Here we compare our method with a range of previous approaches. These methods are divided in two groups: calibrated and uncalibrated. The first group includes C-Rule (maximum volume (GM-MV) and average (GM-AVE)) [11] and Gamut Constrained illuminant estimation (GCIE) [5]. This last method is constrained with a set of illuminants. We have

TABLE I
ANGULAR ERROR ON THE DIFFERENT DATASETS.

Method	Dataset 1		Dataset 2		Dataset 3	
	RMS	95%	RMS	95%	RMS	95%
Our approach						
$\mathbf{CaC_m^0}$	$\mathbf{14.57^\circ}$	$\mathbf{26.69^\circ}$	$\mathbf{9.38^\circ}$	$\mathbf{16.96^\circ}$	$\mathbf{8.82^\circ}$	$\mathbf{16.06^\circ}$
$\mathbf{CaC_c^0}$	$\mathbf{14.63^\circ}$	$\mathbf{27.58^\circ}$	$\mathbf{9.42^\circ}$	$\mathbf{18.30^\circ}$	$\mathbf{8.19^\circ}$	$\mathbf{16.11^\circ}$
CaC_m^{10}	$\mathbf{14.43^\circ}$	$\mathbf{26.69^\circ}$	9.89°	18.28°	8.29°	16.11°
CaC_c^{10}	14.55°	27.19°	9.60°	18.30°	7.66°	14.87°
CaC_m^{20}	14.72°	27.84°	$\mathbf{8.98^\circ}$	$\mathbf{16.96^\circ}$	7.34°	15.20°
CaC_c^{20}	14.74°	28.09°	9.43°	17.08°	7.23°	14.85°
CaC_m^{400}	14.76°	27.59°	8.99°	$\mathbf{16.96^\circ}$	7.23°	14.67°
CaC_c^{400}	14.79°	27.42°	9.32°	17.08°	$\mathbf{7.05^\circ}$	$\mathbf{14.34^\circ}$
Uncalibrated methods						
Grey-Edge	14.62°	27.17°	9.48°	21.42°	8.56°	18.96°
Shades-of-Grey	14.77°	27.57°	10.07°	22.32°	8.73°	20.50°
Max-RGB	15.89°	30.30°	9.58°	26.37°	11.76°	26.54°
Grey-World	15.97°	30.60°	13.02°	27.61°	13.56°	29.41°
no-correction	20.32°	37.67°	9.75°	26.37°	19.64°	34.95°
Color by Correlation	-	-	-	-	10.09°	-
Neural Networks	-	-	-	-	11.04°	-
Calibrated methods						
GCIE 87 lights	-	-	-	-	7.11°	-
GCIE 11 lights	-	-	-	-	6.88°	-
GM-MV	-	-	-	-	6.89°	-
GM-AVE	-	-	-	-	6.86°	-

used two different constraints: the set of 11 illuminants used in the image dataset and a set of 87 illuminants including the previous set. In the second group we include Grey-Edge [9], Shades of grey [8], Max-RGB [7], Grey-World [6], Color-by-Correlation [13] and Neural Networks [39].

We have run the Grey-Edge algorithm provided by the author [9], and have considered the following set of parameters: $0 \leq n \leq 2$, $0 \leq \sigma \leq 5$, $0 \leq p \leq 15$. For Shades-of-Gray the values are $0 \leq \sigma \leq 5$, $0 \leq p \leq 15$. For the training of these two methods, we used 33% of the images to set the parameters, and we applied these parameters to the rest of the images. In this way, independence between training and testing sets is preserved.

The same experiments have been performed using three different images datasets that we list below:

Dataset 1. *Real-World Images* This dataset, created by Ciurea and Funt [40], is composed of images captured with a grey sphere in the image field of view. This sphere allows the estimation of the scene illuminant. In our experiments the ball has been excluded in order to avoid any influence on the results. This image dataset is gamma corrected, therefore we have removed this correction using $\gamma = 2.2$, which is a typical value used in RGB devices. Furthermore, since this dataset was recorded with a video-camera, all the image scenes within each of the 15 scenarios have a high correlation of image content. To avoid the effects derived from this fact we have followed a similar procedure from previously reported experiments. In particular, we have used the frames extracted in [41], that constitute the biggest independent image dataset that can be extracted from the Ciurea-Funt dataset. The total amount of images is 1135, but with a different number of images for each scenario. Both for Grey-Edge and Shades-of-Gray we have used 5 scenarios for training and 10 scenarios for testing.

Dataset 2. *Barcelona Calibrated dataset* This dataset was firstly defined in [42] with 83 images, and is composed of images captured within the Barcelona area. This dataset is calibrated and was also acquired with a grey ball in the field of view. Again, the ball has been excluded. From this dataset, we have randomly selected two thirds of the images as a test set and the other one third as a training set for the Grey-Edge and Shades-of-Gray methods.

Dataset 3. *Controlled Indoor scenes* This dataset, created at Simon Fraser University [43], comprises 321 indoor images. It consists of 31 scenes captured under 11 different conditions, totalling 321 images. This dataset is formed by raw images, therefore no gamma correction is needed. In this experiment we trained both Grey-Edge and Shades-of-Gray by using 10 scenes for training and 21 to test.

In order to analyse whether the category hypothesis delivers meaningful solutions, we have used the root mean square (RMS) of the angular error between the solution and the known scene illuminant. Low RMS error rates imply that images are generally corrected towards the correct illuminant. We have also computed the 95% error to get an idea on how robust the different methods are.

V. RESULTS AND DISCUSSION

Results obtained from these experiments are summarized in Table I. Results are divided into three parts: our results, uncalibrated methods and calibrated methods in this order. The first rows of the table are related to our method. In particular, from the first two rows we can observe that our method achieves equivalent results to state-of-the-art methods by using a completely new hypothesis and, furthermore, without the

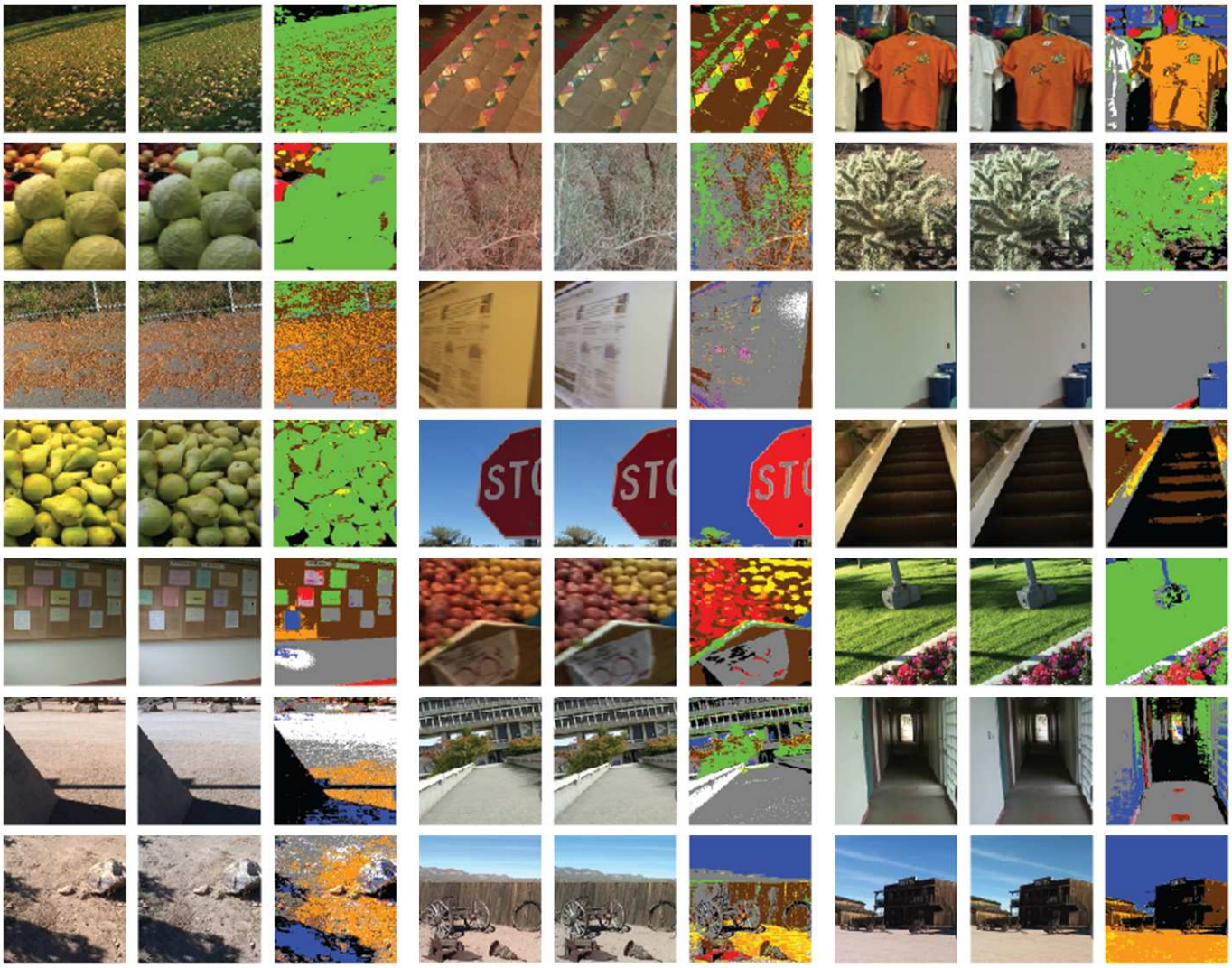


Fig. 4. Examples from a Real-World dataset. Original image (left), corrected image (center), categorized corrected image (right).

need for a training step that can tune parameters to the dataset content. In these first two rows we applied the basic method CaC^0 that simply uses the focal colors of the 11 basic color categories. Here, the combination criterion does not introduce critical changes to the performance. In subsequent rows we study the effect on the performance of our method when changing the basic categories and in order to compare our results.

In the second part of the table we report the performance of different uncalibrated methods. From those methods, we have reported the results on the three datasets for those methods where we could run the code; for the remaining methods (Neural Networks [39] and Color-by-Correlation [13]) we report the results provided in the literature that were just for dataset 3. For the case of calibrated methods we report the results for GM-MV and GM-AVE [11] computed by us, and we have transcribed from previous works the results for GCIE-11 and GCIE-87 [5]. A clear advantage is shown by calibrated methods which use the information derived from knowing camera sensitivities.

TABLE II
ANGULAR ERROR PERFORMANCE BOUND BY SELECTING THE BEST SOLUTION DURING THE COMBINATION ON THE DIFFERENT DATASETS.

Method	Dataset 1	Dataset 2	Dataset 3
$CaC^0_{p_b}$	11.91°	6.86°	7.12°
$CaC^{10}_{p_b}$	11.53°	6.83°	6.27°
$CaC^{20}_{p_b}$	11.81°	6.21°	5.70°
$CaC^{400}_{p_b}$	11.99°	6.16°	5.54°

Before analysing the results obtained when changing the size of the basic color categories, it is worth noting an important observation provided by experiments not reported here. We have found that increasing the size of categories beyond the convex-hull of the basic color categories results in a significant decrease in performance. This observation supports the idea that using the basic color terms as centered anchors is adequate to achieve good adaptation to the most common image content.

As we can see from the results, for the case of a big real-world dataset (dataset 1) the best results are obtained with

Fig. 5. Controlled indoor dataset. Original image (left), corrected image by proposed method (center-left), points weighting the selected illuminant (center-right), categorisation of corrected images (right)

the smallest categories CaC^0 and CaC^{10}. This result agrees with the general hypothesis of the method, which contends that basic color categories encode natural color statistics, since dataset 1 is mostly populated by natural images.

Dataset 2 contains a mix of man-made objects and natural images. The results for this dataset show that CaC^0 outperforms state-of-the-art methods. However, better results can also be achieved by increasing the size of categories. This result is most likely due to an increase in the percentage of man-made objects. In general, man-made objects may take any color (i.e. they are less likely to be basic colors) and may occur as big homogeneous surfaces (non-textured). The size of the basic color categories usually agrees with their texture appearance; for example a big green category correlates with highly textured green areas in natural vegetation, while yellow and red correspond with small category volumes correlating with their less frequent appearance in natural environments. Big homogeneous areas induced by man-made objects imply histograms with sharp peaks, in turn provoking an increase in the number of solutions that can achieve a high weight, which clearly implies a likely increase in the error measure.

Finally, for the indoor dataset (dataset 3), the best results are achieved when we use the biggest sizes of categories, that is, the full convex hull of the color categories. This fact can be explained by the high amount of non-natural and non-basic colors, such as turquoise or other intermediate colors which are not basic and appear in big areas of the images. Again, these images present histograms with sharp peaks due to the absence of natural textures. It is for this last reason that the combination criterion works very well in this dataset. Many different interpretations are plausible, therefore the use of different cues becomes more important. We can see how CaC_c^{400} reaches almost the level of calibrated methods when the categories are adapted to the dataset content.

Apart from the results shown, we want to outline a further advantage derived from the method. The estimated illuminant provides us with an annotated image that gives information about which parts of the images have been selected as anchors and with which color. In Figure 4 we show some results of CaC_m^0 using basic color categories and maximum selection, for images in dataset 1. From left to right, the first column shows the original image, the second column corresponds to the corrected image and the third column displays the categorized image. In Figure 5 we show a similar example for dataset 3 with the same basic method. In this case, the first and second columns show the original and corrected images respectively, while the third column shows the points that have been annotated with basic names in the selected solution. Finally, the fourth column presents the categorisation of the corrected images with basic terms.

Here we have also computed the performance bound we can obtain by improving the illuminant selection step in S^α. We want to emphasize again that all the images selected in this set were highly categorized with basic colors due to our selection of the value α. The results for these performance bound are shown in Table II. These results reinforce our hypothesis since they prove that a proper solution is included within the set of higher categorized images.

The proposed method opens the possibility for further research related to the introduction of top-down knowledge

TABLE III
TOP-DOWN APPROACH

Method	Dataset 1	Dataset 2	Dataset 3
CaC^{0}_{TD}	$11.51°$	$6.75°$	$7.70°$

from the image content that can further constrain the number of solutions and consequently allow even better performance. By top-down knowledge we refer to further processes on the image content that can provide clues to select which are the best color categories and even where they should be located in the image. For example, additional visual cues informing about the existence of, say, a tree in the image will direct the method to find green color in that location of the image. To evaluate the effects of this kind of top-down knowledge onto the performance of our method, we have done one further experiment, that is reported in table III.

In this experiment we have applied a pre-computation step that has provided the basic color categories appearing in the image under the canonical illuminant. In this way, this specific set of categories has been used to apply the basic algorithm to each image. In table III we show the results of estimating the illuminant by selecting the maximum probability from the feasible set built using specific categories for each image. We can see that by introducing information from other top-down visual processes the improvement in the performance is substantial.

VI. Conclusions and Further Work

The main novelty of this work is the definition of a new hypothesis for color constancy that relies on a set of reflectances, or color categories, that encode relevant color information in natural scenes. These categories are those that receive a name across different languages and cultures. These colors are distributed around the achromatic reflectances and we hypothesize that they can act as anchors for image correction.

We propose a color constancy method that estimates the best illuminant accordingly with its ability to label image points with these basic color categories. We use representatives for these categories obtained from psychophysical experiments. . Other categories sets could be tested in further work. Uncalibrated naming experiments have provided a bigger number of observers in [44] and, in [45], authors propose the use of different color name dictionaries depending on the background of the user.

The method we propose builds a set of feasible illuminants that are weighted accordingly with the hypothesis. A fast implementation is easily defined by working in log-space. The proposed algorithm allows to obtain a large sampling of the feasible solutions that is the basis for a useful framework. Having a set of multiple solutions allows the provision of different selection criteria and an open framework to introduce new cues from complementary visual processes.

We show that our methods achieve current state of art with some advantages. Our method is a purely bottom-up method providing a framework for further combination with complementary visual information. The method is based on general psychophysical data that can be modified depending on the application. Lastly, and most importantly, our results are achieved without the need for a training step, as is required in many other approaches.

The proposed method can be framed within the family of statistical methods that estimates the illuminant by voting. The method can be seen as a generalization of previous approaches such as *WhitePatch*, which results from using a single achromatic category in our method, or *Color-by-Correlation* (for the $3D$ case) where categories are represented by the full set of reflectances used.

Further research is now possible to exploit the advantages of using the weighted feasible set. Complementary visual cues, or constraints derived from specific visual tasks, can provide further information to decide on the final illuminant.

VII. Acknowledgments

The authors thank J. van de Weijer and D. Connah for their insightful comments.

References

[1] A. Hurlbert, "Colour vision: Is colour constancy real?" *Current Biology*, vol. 9, no. 15, pp. 558–561(4), 1999.
[2] G. Finlayson and M. Drew, "White-point preserving color correction," in *Proc. IST/SID 5th Color Imaging Conference, pp. 258-261, 1997.*, 1997. [Online]. Available: citeseer.ist.psu.edu/finlayson97whitepoint.html
[3] T. Gevers and A. W. M. Smeulders, "Color based object recognition," *Pattern Recognition*, vol. 32, pp. 453–464, 1999.
[4] S. D. Hordley, "Scene illuminant estimation: past, present, and future,," *Color Research and Application*, vol. 31, no. 4, pp. 303–314, 2006.
[5] G. D. Finlayson, S. D. Hordley, and I. Tastl, "Gamut constrained illuminant estimation," *Int. J. Comput. Vision*, vol. 67, no. 1, pp. 93–109, 2006.
[6] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst*, vol. 310, p. 126, 1980.
[7] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Am.*, vol. 61, no. 1, pp. 1–11, 1971. [Online]. Available: http://www.opticsinfobase.org/abstract.cfm?URI=josa-61-1-1
[8] G. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Color Imaging Conference*, 2004, pp. 37–41.
[9] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2207–2214, 2007. [Online]. Available: http://staff.science.uva.nl/ gijsenij/
[10] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy beyond bags of pixels," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008.
[11] D. A. Forsyth, "A novel algorithm for color constancy," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 5–35, 1990.
[12] G. D. Finalyson, "Color in perspective," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 1034–1038, 1996.
[13] G. Finlayson, S. Hordley, and P. Hubel, "Color by correlation: A simple, unifying framework for color constancy," *PAMI*, vol. 23, no. 11, pp. 1209–1221, November 2001.
[14] D. H. Brainard and W. T. Freeman, "Bayesian color constancy," *Journal of the Optical Society of America A*, vol. 14, pp. 1393–1411, 1997.
[15] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 06 2008, pp. 1–8. [Online]. Available: http://vision.eecs.ucf.edu/
[16] G. Sapiro, "Color and illuminant voting," *PAMI*, vol. 21, no. 11, pp. 1210–1215, November 1999.
[17] S. A. Shafer, "Using color to separate reflection components," *Color Research and Application*, vol. 10, no. 4, pp. 210–218, 1985.
[18] B. V. Funt, M. S. Drew, and J. Ho, "Color constancy from mutual reflection," *Int. J. Comput. Vision*, vol. 6, no. 1, pp. 5–24, 1991.
[19] G. Klinker, S. Shafer, and T. Kanade, "A physical approach to color image understanding," *International Journal of Computer Vision*, vol. 4, no. 1, pp. 7–38, January 1990.

[20] H. Lee, "Method for computing the scene-illuminant chromaticity from specular highlights," *Journal of the Optical Society of America A*, vol. 3, pp. 694–1699, 1986.
[21] S. Tominaga and B. A. Wandell, "Standard surface-reflectance model and illuminant estimation," *Journal of the Optical Society of America A*, pp. 70–78, 1992.
[22] E. Land, "The retinex," *Am Sci*, vol. 52, pp. 247–264.
[23] D. H. Foster, "Does colour constancy exist?" *Trends in Cognitive Science*, vol. 7, no. 10, 2003. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1364661303001980
[24] A. Hurlbert, "283-322," *Perceptual Constancy*, vol. Edited by Walsh V., Kulikowski J., pp. 558–561.
[25] J. M. Kraft and D. H. Brainard, "Mechanisms of color constancy under nearly natural viewing," *Proc. Nat. Acad. Sci. USA*, vol. 96, pp. 307–312, 1999.
[26] M. Olkkonen, T. Hansen, and K. R. Gegenfurtner, "Categorical color constancy for simulated surfaces," *J. Vis.*, vol. 9, no. 12, pp. 1–18, 11 2009. [Online]. Available: http://journalofvision.org/9/12/6/
[27] T. Hansen, M. Olkkonen, S. Walter, and K. R. Gegenfurtner, "Memory modulates color appearance," *Nature Neuroscience*, vol. 9, no. 11, pp. 1367–1368, October 2006. [Online]. Available: http://dx.doi.org/10.1038/nn1794
[28] T. Hansen, S. Walter, and K. R. Gegenfurtner, "Effects of spatial and temporal context on color categories and color constancy," *Journal of Vision*, vol. 7, no. 4, 2007. [Online]. Available: http://www.journalofvision.org/content/7/4/2.abstract
[29] B. Berlin and P. Kay, *Basic Color Terms: Their Universality and Evolution*. Berkeley, CA: University of California Press, 1969.
[30] R. M. Boynton and C. X. Olson, "Locating basic colors in the osa space," *Color Research and Application*, vol. 12, no. 2, pp. 94–105, 1987.
[31] K. Barnard, L. Martin, and B. Funt, "Colour by correlation in a three dimensional colour space," in *ECCV2000*, 2000, pp. 275–289.
[32] J. Sturges and T. W. A. Whitfield, "Locating basic colours in the munsell space," *Color Research and Application*, vol. 20, pp. 364–376, 1995.
[33] R. Benavente, M. Vanrell, and R. Baldrich, "A data set for fuzzy colour naming," *Color Research and Application*, vol. 31, no. 1, pp. 48–56, Feb 2006. [Online]. Available: http://www.cat.uab.cat/Publications/2006/BVB06
[34] D. Philipona and J. ORegan, "Color naming, unique hues and hue cancellation predicted from singularities in reflection properties," *Visual Neuroscience*, vol. 3-4, no. 23, pp. 331–339, 2006.
[35] S. N. Yendrikhovskij, "Computing color categories from statistics of natural images," *Journal of Imaging Science and Technology*, vol. 45, pp. 409–417, 2001.
[36] J. van de Weijer, C. Schmid, and J. Verbeek, "Using high-level visual information for color constancy," in *International Conference on Computer Vision*, oct 2007. [Online]. Available: http://lear.inrialpes.fr/pubs/2007/VSV07b
[37] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini, "Automatic color constancy algorithm selection and combination," *Pattern Recogn.*, vol. 43, pp. 695–705, March 2010. [Online]. Available: http://portal.acm.org/citation.cfm?id=1660180.1660643
[38] S. Bianco, F. Gasparini, and R. Schettini, "A consensus based framework for illuminant chromaticity estimation," *Journal of Electronic Imaging*, vol. 17, pp. 023 013–1–9, 2008.
[39] V. C. Cardei, B. Funt, and K. Barnard, "Estimating the scene illumination chromaticity by using a neural network," *Journal of the Optical Society of America A*, vol. 19, no. 12, pp. 2374–2386, 2002.
[40] F. Ciurea and B. Funt, "A large image database for color constancy research," in *Color Imaging Conference*, 2003, pp. 160–164.
[41] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini, "Improving color constancy using indoor-outdoor image classification," *IEEE Transactions on Image Processing*, vol. 17, no. 12, pp. 2381–2392, December 2008.
[42] J. Vazquez-Corral, C. Párraga, M. Vanrell, and R. Baldrich, "Color constancy algorithms: Psychophysical evaluation on a new dataset," *Journal of Imaging Science and Technology*, vol. 53, no. 3, May-June 2009.
[43] K. Barnard, L. Martin, B. Funt, and A. Coath, "A data set for colour research," *Color Research and Application*, vol. 27, pp. 147–151, 2002.
[44] N. Moroney, "Thousands of on-line observers is just the beginning," *Human Vision and Electronic Imaging*, p. 724005, 2009.
[45] G. Woolfe, "Natural language color editing," *ISCC Annual Meeting*, 2007.