Color Constancy Using Faces

Simone Bianco Raimondo Schettini
University of Milano-Bicocca
DISCo - Department of Informatics, Systems and Comunication
{bianco,schettini}@disco.unimib.it

Abstract

In this work, we investigate how illuminant estimation can be performed exploiting the color statistics extracted from the faces automatically detected in the image. The proposed method is based on two observations: first, skin colors tend to form a cluster in the color space, making it a cue to estimate the illuminant in the scene; second, many photographic images are portraits or contain people. The proposed method has been tested on a public dataset of images in RAW format, using both a manual and a real face detector. Experimental results demonstrate the effectiveness of our approach. The proposed method can be directly used in many digital still camera processing pipelines with an embedded face detector working on gray level images.

1. Introduction

Computational color constancy aims to estimate the actual color in an acquired scene disregarding its illuminant. Since it is an ill-posed problem, as its solution lacks uniqueness and stability [11], many different solutions exist in the literature, each based on different assumptions [21]. It is known that when these assumptions are not met the algorithm's estimate of the actual illuminant can be very poor. Recently, it has been shown that the universal best and the universal worst algorithms do not exist [5]: the algorithm that performs best for a specific image depends on the image content. For this reason a recent research area which has shown promising results aims to improve illuminant estimation by using visual information automatically extracted from the images. The existing algorithms exploit both low-level [4, 14], intermediate-level [3] and high-level [30] visual information. Hansen et al. [19] showed that in the human visual system, memory colors could be used as hints to give an estimate of the illuminant in the scene. Moreno et al. [24] obtained memory colors for three different objects (grass, snow and sky) using psychophysical experiments. They then used a supervised image segmentation method to detect memory color objects and exploit them to color correct the image using a weighted Von Kries formula. In [17] a color constancy approach is used to compensate for skin color variations to achieve accurate skin color segmentation, while the use of contextual information in the form of detected faces to improve skin pixel detection for tracking has been proposed by Soriano et al. [28]. Skin tones have been also used to automatically tune the parameters of different enhancement algorithms to correspond to human preferences regarding the appearance of people in an image [25]. However, they assume that the image has already been white balanced and mapped in a standard color space.

In this work we hypothesize that skin colors provide enough and reliable information to estimate the scene illuminant and suggest the use of faces to detect skin areas. Since it is known that skin colors tend to form a cluster in various color spaces [12, 23] we guess that the diversity between the gamut of skin pixels of the detected faces and the skin canonical gamut can be affordably used to estimate the scene illuminant. This approach could be generalized to other memory colors such as sky and grass; however their color distributions depict a much larger variation and there are not easily available and affordable detectors [26].

The proposed method is compared with state-of-the-art algorithms on a standard dataset of RAW camera images having a known color target [13]. Experimental results show that the proposed method, using both a manual and a real face detector, is able to improve illuminant estimation accuracy with respect to the state-of-the-art algorithms considered.

2. Problem formulation and existing methods

The image values for a Lambertian surface located at the pixel with coordinates (x, y) can be seen as a function $\boldsymbol{\rho}(x, y)$, mainly dependent on three physical factors: the illuminant spectral power distribution $I(\lambda)$, the surface spectral reflectance $S(\lambda)$ and the sensor spectral sensitivities $\mathbf{C}(\lambda)$. Using this notation $\boldsymbol{\rho}(x, y)$ can be expressed as

$$\boldsymbol{\rho}(x, y) = \int_{\omega} I(\lambda) S(x, y, \lambda) \mathbf{C}(\lambda) \mathrm{d}\lambda, \qquad (1)$$

978-1-4673-1228-8/12/$31.00 ©2012 IEEE

where ω is the wavelength range of the visible light spectrum, $\boldsymbol{\rho}$ and $\mathbf{C}(\lambda)$ are three-component vectors. Since the three sensor spectral sensitivities are usually respectively more sensitive to the low, medium and high wavelengths, the three-component vector of sensor responses $\boldsymbol{\rho} = (\rho_1, \rho_2, \rho_3)$ is also referred to as the sensor or camera $\mathrm{RGB} = (R, G, B)$ triplet.

The goal of color constancy is to estimate the color $\mathbf{I}$ of the scene illuminant, i.e. the projection of $I(\lambda)$ on the sensor spectral sensitivities $\mathbf{C}(\lambda)$:

$$\mathbf{I} = \int_{\omega} I(\lambda)\mathbf{C}(\lambda)\mathrm{d}\lambda. \tag{2}$$

As it is more important to estimate the chromaticity of the scene illuminant than its overall intensity, usually the illuminant color is estimated up to a scale factor. Since the only information available are the sensor responses $\boldsymbol{\rho}$ across the image, color constancy is an under-determined problem [11]; and thus further assumptions and/or knowledge are needed to solve it.

Several computational color constancy algorithms exist in the literature, each based on different assumptions. Recently Van de Weijer et al. [29] have unified a variety of algorithms. These algorithms estimate the illuminant color $\mathbf{I}$ by implementing instantiations of the following equation:

$$\mathbf{I}(n, p, \sigma) = \frac{1}{k}\left(\iint |\nabla^n \boldsymbol{\rho}_\sigma(x, y)|^p \,\mathrm{d}x\,\mathrm{d}y\right)^{\frac{1}{p}}, \tag{3}$$

where n is the order of the derivative, p is the Minkowski norm, $\boldsymbol{\rho}_\sigma(x, y) = \boldsymbol{\rho}(x, y) \otimes G_\sigma(x, y)$ is the convolution of the image with a Gaussian filter $G_\sigma(x, y)$ with scale parameter σ, and k is a constant to be chosen such that the illuminant color $\mathbf{I}$ has unit length (using the $2-$norm). The integration is performed over all pixel coordinates. Different (n, p, σ) combinations correspond to different illuminant estimation algorithms, each based on a different assumption. For example, the Gray World algorithm [6] – generated setting $(n, p, \sigma) = (0, 1, 0)$ – is based on the assumption that the average color in the image is gray and that the illuminant color can be estimated as the shift from gray of the averages in the image color channels; the White Point algorithm [7] – generated setting $(n, p, \sigma) = (0, \infty, 0)$ – is based on the assumption that there is always a white patch in the scene and that the maximum values in each color channel are caused by the reflection of the illuminant on the white patch, and they can be thus used as the illuminant estimation; the Gray Edge algorithm [29] – generated setting for example $(n, p, \sigma) = (1, 0, 0)$ – is based on the assumption that the average color of the edges is gray and that the illuminant color can be estimated as the shift from gray of the averages of the edges in the image color channels.

The Gamut Mapping assumes that for a given illuminant, one observes only a limited gamut of colors [10]. It has a training phase in which a canonical illuminant is chosen and the canonical gamut is computed observing as many surfaces under the canonical illuminant as possible. Given an input image with an unknown illuminant, its gamut is computed and the illuminant is estimated as the mapping that can be applied to the gamut of the input image, resulting in a gamut that lies completely within the canonical gamut and produces the most colorful scene. If the spectral sensitivity functions of the camera are known, the Color by Correlation approach could be also used [9]. Bayesian approaches [13] model the variability of reflectance and of illuminant as random variables, and then estimate illuminant from the posterior distribution conditioned on image intensity data.

Given a set computational color constancy algorithms, in [3] an image classifier is trained to classify the images as indoor and outdoor, and different experimental frameworks are proposed to exploit this information in order to select the best performing algorithm on each class. In [4] it has been shown how intrinsic, low level properties of the images can be used to drive the selection of the best algorithm (or the best combination of algorithms) for a given image. The algorithm selection and combination is made by a decision forest composed of several trees on the basis of the values of a set of heterogeneous features. In [14] the Weibull parameterization has been used to train a maximum likelihood classifier based on mixture of Gaussians to select the best performing color constancy method for a certain image.

In [8] a statistical model for the spatial distribution of colors in white balanced images is developed, and then used to infer illumination parameters as those being most likely under their model. High level visual information has been used to select the best illuminant out of a set of possible illuminants [31]. This is achieved by restating the problem in terms of semantic interpretability of the image. Several color constancy methods are applied to generate a set of illuminant hypotheses. For each illuminant hypothesis, they correct the image, evaluate the likelihood of the semantic content of the corrected image, and select the most likely illuminant color.

3. The proposed approach

In this work we propose a new method to estimate the scene illuminant using faces: more exactly, we detect faces in the image and extract the skin color information using a rough skin detector. Our method is based on the assumption that skin colors form a sufficiently compact cluster in the color space in order to represent a valid clue for illuminant estimation [19]. In order to confirm the feasibility of our assumption we have computed the RAW RGB values of a dataset of 697 measured skin reflectances [22] under 287 different illuminants [2] with the sensor spectral sensitivities of six commercially available cameras, using Equation 1. The cameras considered are the Nikon D1, Nikon

D70, Nikon D100, Nikon D5000, Sony DXC930 and Canon 500D. The convex hulls of the skin clusters imaged by the different cameras are reported in the HS plane of the HSV color space in Figure 1.

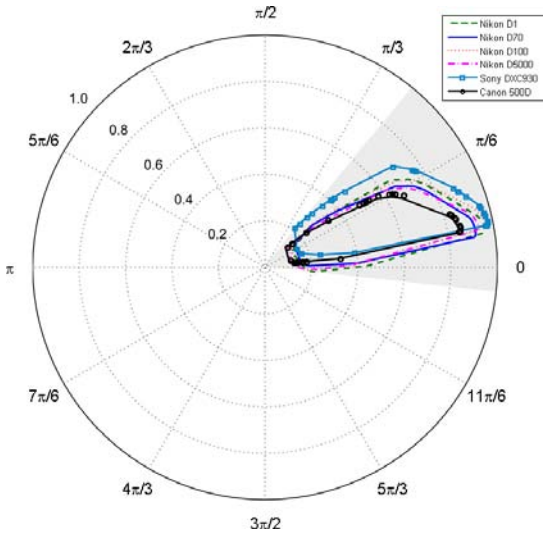

Figure 1. The convex hulls, in the HS plane of the HSV color space, of the skin reflectance database acquired under 287 illuminants by the six different cameras considered: Nikon D1, Nikon D70, Nikon D100, Nikon D5000, Sony DXC930 and Canon 500D.

Our approach applies gamut mapping to skin pixels only: the illuminant is estimated as the mapping that can be applied to the gamut of the skin colors in the input image, resulting in a gamut that lies completely within the skin canonical gamut.

The first step is the computation of the skin canonical gamut S_C: this is the convex hull of the skin colors of different people acquired under the chosen canonical illuminant.

Given an input image for which we want to estimate the illuminant, in which F faces are present, the masks $u_f(x,y)$, $f = 1,\ldots,F$, are obtained:

$$u_f(x,y) = \begin{cases} 1 & \text{if } (x,y) \in \text{face no. } f \wedge \text{is a skin pixel} \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

i.e. $u_f(x,y)$ assumes the value 1 only on the pixels inside the $f-$th detected face area, being classified as skin pixels. The extracted skin colors $\boldsymbol{skin}_f$ are then computed as

$$\boldsymbol{skin}_f = \{\boldsymbol{\rho}(x,y) : u_f(x,y) = 1\}, \tag{5}$$

Once we have extracted the skin colors for each detected face, these are converted into the YCbCr color space and luminance normalized, such that the average luminance $\overline{Y} = 0.5$. The skin gamut S_I is then computed as the convex hull of the converted values of all detected faces. The set of feasible mappings $\mathcal{M}$ is then determined: it consists of all mappings that can be applied to the skin gamut S_I of the input image and that result in a gamut that lies completely within the skin canonical gamut S_C:

$$\mathcal{M} = \{\mathcal{M}_i : \mathcal{M}_i S_I \in S_C\} \tag{6}$$

In this work each transformation $\mathcal{M}_i$ is be modeled by a diagonal mapping or von Kries Model [33]. The illuminant color $\mathbf{I}$ is then estimated as the inverse of the average of the feasible set [1].

Once we have estimated the illuminant color $\mathbf{I} = [I_R, I_G, I_B]$, given the choice of diagonal mappings, each pixel in the image is color corrected using the von Kries model [33], i.e.:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix}_{out} = \begin{bmatrix} 1/I_R & & \\ & 1/I_G & \\ & & 1/I_B \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{in}. \tag{7}$$

The operation flowchart of the proposed method is reported in Figure 2. The face detector module is run on the input image to detect any faces. If no faces are detected, the input image may be processed with any other state-of-the-art illuminant estimation algorithm. If one or more faces are detected, a skin detection module is run on the detected faces to filter out any non-skin and unreliable pixels.

Looping on all the faces detected, the first step of the skin detection is the conversion of the detected face pixels in the HSV color space. Then a technique based on scale-space histogram filtering [34] is used to identify the highest peak location and width of the histogram of the hue component, within the hue interval corresponding to feasible skin colors (see for example Fig.1). This involves generating a multi-scale description of the histogram $h(x)$ by convolving it with a series of Gaussians of increasing scale:

$$H(x,\sigma) = h(x) \otimes G_\sigma(x) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2\sigma^2}} \mathrm{d}t\,. \tag{8}$$

$H(x,\sigma)$ forms a two-dimensional surface called the scale-space image. The locations of valleys and peaks are easily located in terms of zero-crossing in the scale-space image, since at any value of σ, the extrema in the nth derivative of the smoothed histogram are given by the zero-crossing in the $(n+1)$th derivative, computed using the relation

$$\frac{\partial^n H(x,\sigma)}{\partial x^n} = h(x) \otimes \frac{\partial^n G_\sigma(x)}{\partial x^n}, \tag{9}$$

where the derivatives of the Gaussian are readily obtained. Although this technique applies to zeros in any derivative we will focus our attention on those in the second derivative [34].

The highest peak location P_h and width $2w$ are identified within the hue interval corresponding to feasible skin

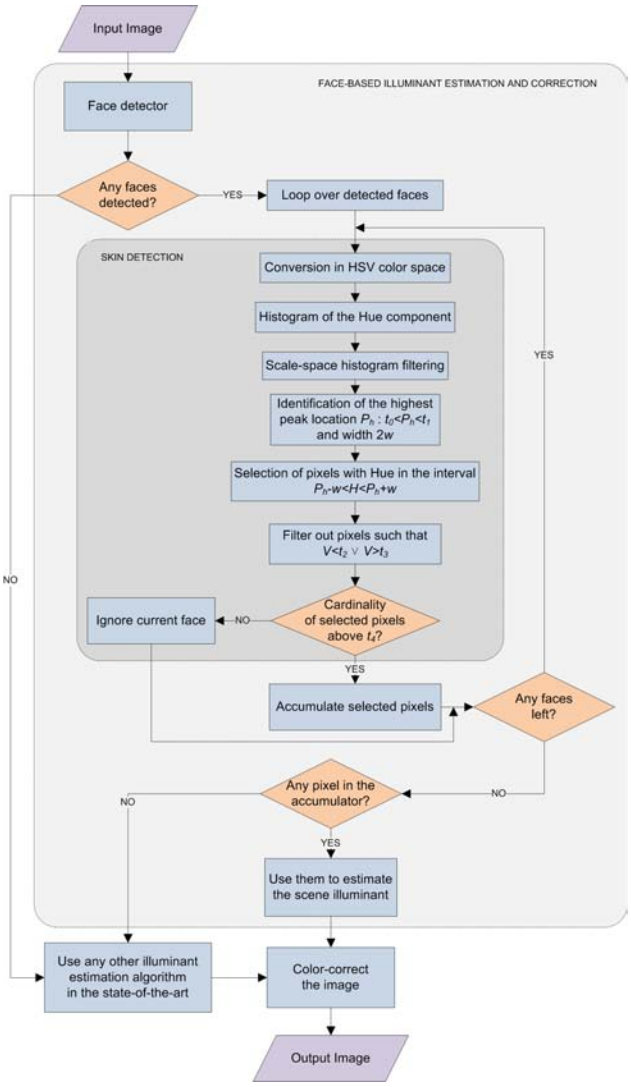

Figure 2. The operation flowchart of the proposed method

colors, i.e. $P_h \in [t_0\ t_1]$. Then only those pixels satisfying the condition $P_h - w < H < P_h + w$ are selected. This condition permits to implement a sort of adaptive skin detector, since the selected hue interval depends on the current peak location P_h. Any unreliable skin pixel as being too dark or too bright, and thus potentially clipped, is filtered out if it satisfies the condition $V < t_2 \vee V > t_3$. If the cardinality of the remaining skin pixels, normalized for the total number of pixels in the detected face, is above the threshold t_4, they are converted into the YCbCr color space where they are luminance normalized as previously described, and then they are added in an accumulator.

When the looping on all the faces detected is finished, the accumulator is analyzed: if it is non-empty, the color coordinates of the skin pixels in it are used to estimate the scene illuminant by first computing the skin gamut S_I and then using the skin-based gamut mapping; otherwise the scene illuminant can be estimated using any other state-of-the-art illuminant estimation algorithm. Given the illuminant estimation, the input image is then color corrected to give the final output image.

The main assumption of the proposed approach is that skin colors form a sufficiently compact cluster in the color space in order to represent a valid clue for illuminant estimation. Most of the skin detectors in the literature exploit this compactness property [23], and cluster boundaries have been given in various color spaces [12]. Unfortunately these boundaries cannot be used in our approach, since they are computed for standard color space, while here we are working with images encoded in a device-dependent color space. We have therefore decided to implement the adaptive skin detector previously described, as it does not require a training set to define the skin cluster boundaries in the color space used. It should be noted that the proposed method in the present form makes it possible to discard faces with an unnatural mixing illuminant condition (such as when the face is under a colored umbrella) due to the feasibility check on the hues.

An example of the combined effect of the implemented face and skin detector is reported in Figure 3.

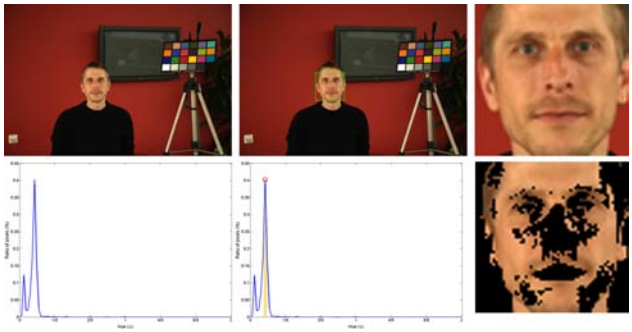

Figure 3. An example of the combined effect of the implemented face and skin detector. The original image (top left); the output of the face detector (top center); the face detected (top right); the histogram of the hue component of the detected face (bottom left); highest peak location and width after scale-space histogram filtering (bottom center); the output of the skin detector (bottom right).

4. Experimental Setup

The aim of this section is to investigate if the proposed algorithm can outperform state-of-the-art algorithms in the illuminant estimation on portraits and on images containing faces. To have an idea of the applicability of the proposed method, we have generated a generic query on Flickr (`http://www.flickr.com/`) – i.e. we searched the first 10,000 images with the tag "Canon" – and found that almost half of the returned images (more exactly 46% of them) were portraits or included faces. This result is also consistent with a qualitative analysis of the camera phone

photospace distribution reported by the International Imaging Industry Association (I3A) [18].

4.1. Image Dataset and Evaluation Procedure

To test the performance of the proposed algorithm, a standard dataset of RAW camera images having a known color target is used [13]. This dataset is captured using a high-quality digital SLR camera in RAW format, and is therefore free of any color correction. This dataset was originally available in sRGB-format, but Shi and Funt [27] reprocessed the raw data to obtain linear images with a higher dynamic range (14 bits as opposed to standard 8 bits). The dataset consists of a total of 568 images, 95 of which are portraits or include faces and constitute what we have called the portrait dataset. It is important to notice that the portrait dataset contains examples of people of different nationalities. The Macbeth ColorChecker (MCC) chart is included in every scene acquired, and this allows to accurately estimate the actual illuminant of each acquired image. During experiments the MCC has been masked to avoid biasing the algorithms.

Examples of images within the portrait dataset are reported in Figure 4.

Figure 4. Examples of images within the portrait dataset

4.2. Benchmark algorithms

In this work we have considered different benchmarking algorithms for color constancy. Six of them were generated varying the three variables (n, p, σ) in Equation 3, and correspond to well known and widely used color constancy algorithms. The values chosen for (n, p, σ) are reported in Table 1 and set as in [16].The algorithms are used in the original authors' implementation which is freely available online (`http://lear.inrialpes.fr/people/vandeweijer/code/ColorConstancy.zip`). The seventh algorithm is the pixel-based Gamut Mapping [15]. The value chosen for σ is also reported in Table 1. The other algorithms considered are the Bayesian (BAY [13]); the Natural Image Statistics (NIS [14]); the High Level Visual Information [31]: bottom-up (HLVI BU), top-down (HLVI TD), and their combination (HLVI BU&TD); the Spatio-Spectral statistics [8]: with Maximum Likelihood estimation (SS ML), and with General Priors (SS GP); the Automatic color constancy Algorithm Selection (AAS) [4] and the Automatic Algorithm Combination (AAC) [4].

Table 1. Values chosen for (n, p, σ) for the state-of-the-art algorithms which are instantiations of Eq.3.

Algorithm	n	p	σ
Gray World (GW)	0	1	0
White Point (WP)	0	∞	0
Shades of Gray (SoG)	0	4	0
general Gray World (gGW)	0	9	9
1st-order Gray Edge (GE1)	1	1	6
2nd-order Gray Edge (GE2)	2	1	1
Gamut Mapping (GM)	0	0	4

The last algorithm considered is the Do Nothing (DN) algorithm which gives the same estimation for the color of the illuminant ($\mathbf{I} = [1\ 1\ 1]$) for every image, i.e. it assumes that the image is already correctly balanced.

5. Results and Discussion

The state-of-the-art algorithms considered are run on the portrait dataset. The error metric considered, as suggested by Hordley and Finlayson [20], is the angle between the RGB triplet of estimated illuminant ($\boldsymbol{\rho}_w$) and the RGB triplet of the measured ground truth illuminant ($\hat{\boldsymbol{\rho}_w}$):

$$e_{ANG} = \arccos\left(\frac{\boldsymbol{\rho}_w^T \hat{\boldsymbol{\rho}_w}}{\|\boldsymbol{\rho}_w\| \|\hat{\boldsymbol{\rho}_w}\|}\right). \tag{10}$$

In Table 2 the minimum, the 10^{th}-percentile, the median, the average, the 90^{th}-percentile, and the maximum of the angular errors obtained by the considered state-of-the-art algorithms on the portrait dataset are reported.

In order to apply the proposed algorithm, we have to first calculate the convex hull of the skin colors under the canonical illuminant. Since we are considering RAW images, free of color correction, it is impossible to use information about skin clusters available in the state of the art [12]. The skin pixels in the train images are extracted and all the thresholds introduced in Sec.3 computed. The extracted skin colors are then white balanced using the ground truth illuminants, obtaining the skin colors under the canonical illuminant. The skin colors are then intensity normalized and their convex hull is calculated. The results are obtained using a leave-one-out cross validation. This means that the skin canonical gamut S_C is recomputed for each image. When calculating the S_C for each image, all the images containing the same person were left out too.

Table 2. Angular error statistics obtained by the state-of-the-art algorithms considered on the test set of the portrait dataset.

Algorithm	Min	10^{th}prc	Med	Avg	90^{th}prc	Max
DN	6.43	9.93	13.20	13.10	15.77	25.36
GW	0.25	1.68	5.38	5.75	10.02	14.84
WP	0.19	1.42	6.22	7.62	15.63	40.54
SoG	0.42	1.28	3.80	4.61	9.12	13.90
gGW	0.19	1.04	3.44	4.51	9.29	14.87
GE1	0.49	1.74	4.44	4.90	8.34	**11.50**
GE2	0.68	1.53	3.90	4.34	7.27	12.87
GM	**0.05**	**0.40**	2.61	4.24	11.33	20.00
BAY	0.10	1.22	4.05	4.79	9.02	17.33
NIS	0.61	1.40	3.34	4.05	7.33	12.88
HLVI BU	0.33	0.89	2.95	3.45	6.71	14.96
HLVI TD	0.37	0.84	2.88	3.67	6.87	15.88
HLVI BU&TD	0.37	0.80	**2.59**	**3.43**	**6.44**	15.88
SS ML	0.34	0.92	3.45	3.78	7.55	12.30
SS GP	0.43	0.85	3.28	3.67	7.14	12.09
AAS	0.19	1.17	3.57	4.70	10.18	18.08
AAC	0.12	0.81	2.87	3.79	7.83	13.31

In Table 3 the minimum, the 10^{th}-percentile, the median, the average, the 90^{th}-percentile, and the maximum of the angular errors obtained by the proposed algorithm on the portrait dataset are reported. The table contains two rows: the first one refers to the ideal results that would be obtained using a perfect face detector (i.e. the faces are manually detected), the second one refers to the results obtained using a real face detector which is an implementation of the widely used Viola-Jones [32], trained on an independent dataset to detect frontal faces. The statistics on the second row are referred to only those images in which the face detector was able to detect at least one face (i.e. 82.11% of the images of the portrait dataset).

Table 3. Angular error statistics obtained by the proposed algorithm on the portrait dataset.

Algorithm	Min	10^{th}prc	Med	Avg	90^{th}prc	Max
FACES$_{ideal}$	0.13	0.68	1.81	2.34	4.87	8.53
FACES$_{real}$	0.13	0.77	1.93	2.53	5.05	8.53

It can be noticed from the comparison of Table 3 with Table 2 that the proposed algorithm – in the instantiation with the manual face detector – is able to estimate the illuminant in the scene with a higher accuracy, reducing the median angular error by 30.65% with respect to the GM (which had the lowest median error). Furthermore the maximum angular error is reduced by 25.83% with respect to the GE1 (which had the lowest maximum error).

In Figure 5 the four images of the portrait dataset on which the proposed algorithm makes the highest four angular errors are reported. The images have been contrast stretched and the sRGB gamma correction has been applied for better visualization.

In Table 4 the minimum, the 10^{th}-percentile, the median, the average, the 90^{th}-percentile, and the maximum of the angular errors obtained on the whole dataset by the seven state-of-the-art algorithms considered and the proposed algorithm are reported. The table is divided into two parts: in the first one the errors obtained by the state-of-the-art algorithms considered are reported; in the second one we report the errors obtained by the proposed algorithm coupled with the different state-of-the-art algorithms considered when no faces are detected.

Table 4. Angular error statistics obtained by the state-of-the-art algorithms considered on the whole dataset.

Algorithm	Min	10^{th}prc	Med	Avg	90^{th}prc	Max
DN	3.72	10.38	13.55	13.62	16.45	27.37
GW	0.18	1.88	6.30	6.27	10.12	24.84
WP	0.08	1.38	5.61	7.46	15.68	40.59
SoG	0.18	1.04	4.04	4.85	9.71	19.93
gGW	**0.03**	0.82	3.45	4.60	9.68	22.21
GE1	0.16	1.82	4.55	5.21	9.78	19.69
GE2	0.26	2.06	4.43	5.01	8.93	16.87
GM	0.05	**0.40**	**2.28**	4.10	11.08	23.18
BAY	0.10	1.17	3.44	4.70	10.21	24.47
NIS	0.08	0.93	3.13	4.09	8.57	26.20
HLVI BU	0.06	0.75	2.54	**3.30**	**6.59**	17.51
HLVI TD	0.11	0.85	2.63	3.65	7.53	25.24
HLVI BU&TD	0.13	0.77	2.47	3.38	6.97	25.24
SS ML	0.06	0.85	2.93	3.55	7.23	15.25
SS GP	0.07	0.82	2.90	3.47	7.00	**14.80**
AAS	**0.03**	0.77	3.16	4.18	9.15	22.21
AAC	0.05	0.90	2.90	3.74	7.93	14.98
FACE$_{real}$+GW	0.13	1.32	5.66	5.70	9.98	24.84
FACE$_{real}$+WP	0.08	1.04	4.49	6.58	14.87	31.51
FACE$_{real}$+SoG	0.13	0.81	3.52	4.47	9.28	19.93
FACE$_{real}$+gGW	**0.03**	0.78	3.20	4.23	9.30	22.21
FACE$_{real}$+GE1	0.13	1.24	4.11	4.79	9.52	19.69
FACE$_{real}$+GE2	0.13	1.52	4.12	4.68	8.52	16.87
FACE$_{real}$+GM	0.05	**0.40**	**2.01**	3.67	9.50	23.18
FACE$_{real}$+BAY	0.10	1.05	2.99	4.29	9.58	24.47
FACE$_{real}$+NIS	0.08	0.80	2.85	3.78	8.39	26.20
FACE$_{real}$+HLVI BU	0.06	0.68	2.23	**3.05**	**6.33**	17.51
FACE$_{real}$+HLVI TD	0.11	0.81	2.38	3.37	6.97	25.24
FACE$_{real}$+HLVI BU&TD	0.13	0.71	2.22	3.12	6.38	25.24
FACE$_{real}$+SS ML	0.07	0.80	2.60	3.24	6.85	15.25
FACE$_{real}$+SS GP	0.08	0.75	2.57	3.18	6.67	**14.80**
FACE$_{real}$+AAS	**0.03**	0.67	2.71	3.74	8.31	22.21
FACE$_{real}$+AAC	0.05	0.81	2.60	3.42	7.54	14.98

From the analysis of Table 4 it is possible to notice that the use of the proposed algorithm, combined with the state-of-the-art algorithms in the case that no faces were detected, was able to reduce almost all the error statistics with respect to the use of the respective state-of-the-art algorithms alone (i.e. the the 10^{th}-percentile, the median, the average, and the 90^{th}-percentile). The only unchanged statistic is the maximum error (except for the WP), which is due to the chosen state-of-the-art algorithms used, and is relative to an image not belonging to the portrait dataset. It is important to notice that the improvement in performance is obtained despite the dataset used does not reflect the photospace distribution we have found in Section 4. In fact only the 17% of images included in the dataset used are portraits or contain people, while we have found this ratio to be 46% in the photospace.

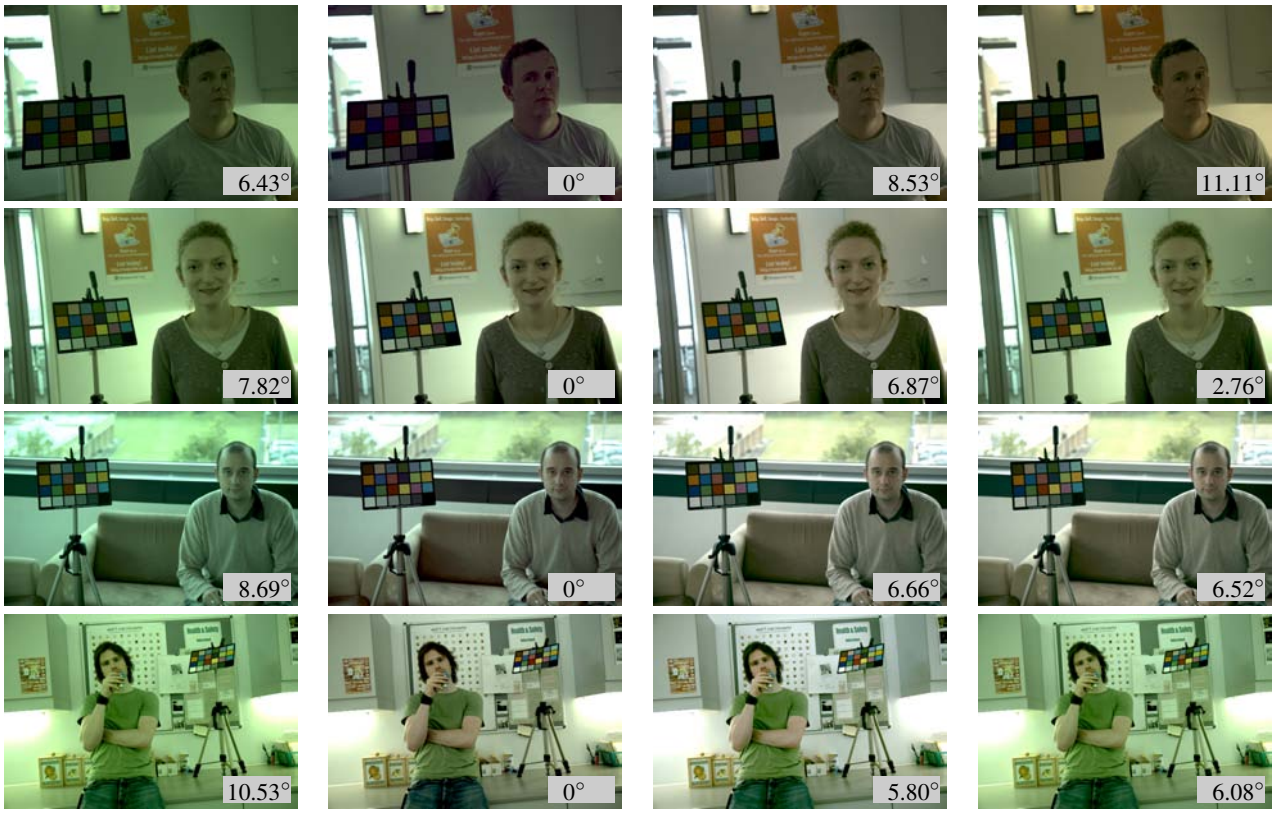

Figure 5. Example of the image in the portrait dataset on which the proposed algorithm makes the four highest angular error. Left to right: original image; ideal correction based on the MCC; correction with the proposed algorithm; correction with the best algorithm, on the whole dataset, in the state-of-the-art among the ones considered, i.e. GM.

6. Conclusions

We have shown here that it is possible to use skin tones to estimate the illuminant color. We have used a face detector to find faces in the scene, and the corresponding skin colors to estimate the chromaticity of the illuminant. The method was based on two observations: first, skin colors tend to form a cluster in the color space, making it a cue to estimate the illuminant in the scene; second, many photographic images are portraits or contain people. The method has been compared with a large number of algorithms in the state of the art on the only public dataset of images in RAW format.Its implementation in actual digital still camera processing pipelines is straightforward. The improvement in color constancy that our method is able to obtain could translate into an improved color-based face and skin detections.

We would like to improve the adaptive skin detector working in a different color space, and the reliability of the gray level face detector. We have assumed a single light source (as most of the illuminant estimation algorithms in the state of the art [16]). An extension of the proposed method to achieve color constancy for multiple light sources is under investigation, and a proper dataset is under construction. The method could be also extended to use "memory objects" in the scene for color constancy. If we are able to automatically recognize objects and logos that have intrinsic colors, we can use them for color constancy (see for example Fig. 6).

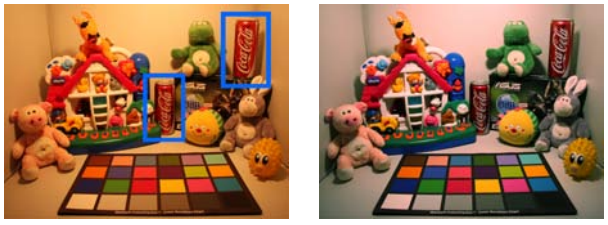

Figure 6. Examples of using objects for color constancy: images acquired and bounding boxes of the objects recognized (left column); images corrected compensating for the illuminant estimated on the objects recognized (right column)

References

[1] K. Barnard. Improvements to gamut mapping colour constancy algorithms. *European Conference on Computer Vision*, pages 390–403, 2000. 3

[2] K. Barnard, L. Martin, B. Funt, and A. Coath. A data set for colour research. *Color Research and Application*, 27(3):147–151, 2002. 2

[3] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini. Improving color constancy using indoor-outdoor image classification. *IEEE Transactions on Image Processing*, 17(12):2381–2392, 2008. 1, 2

[4] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini. Automatic color constancy algorithm selection and combination. *Pattern recognition*, 43(3):695–705, 2010. 1, 2, 5

[5] S. Bianco, F. Gasparini, and R. Schettini. Consensus based framework for illuminant chromaticity estimation. *Journal of Electronic Imaging*, 17(2):023013, 2007. 1

[6] G. Buchsbaum. A spacial processor model for object color perception. *J. Franklin Inst.*, 310:1–26, 1980. 2

[7] V. Cardei, B. Funt, , and K. Barnard. White point estimation for uncalibrated images. *Proceedings of the IS&T/SID 7th Color Imaging Conference*, pages 97–100, 1999. 2

[8] A. Chakrabarti, K. Hirakawa, and T. Zickler. Color constancy with spatio-spectral statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012. (DOI:10.1109/TPAMI.2011.252). 2, 5

[9] G. Finlayson, S. Hordley, and P. Hubel. Color by correlation: a simple, unifying framework for color constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, 2001. 2

[10] D. A. Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–36, 1990. 2

[11] B. Funt, K. Barnard, and L. Martin. Is machine colour constancy good enough? *Proceedings of the 5th European Conference on Computer Vision*, pages 445–459, 1998. 1, 2

[12] F. Gasparini, S. Corchs, and R. Schettini. Recall or precision oriented strategies for binary classification of skin pixels. *Journal of Electronic Imaging*, 17:023017, 2008. 1, 4, 5

[13] P. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, pages 1–8, 2008. 1, 2, 5

[14] A. Gijsenij and T. Gevers. Color constancy using natural image statistics and scene semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(33):687–698, 2011. 1, 2, 5

[15] A. Gijsenij, T. Gevers, and J. van de Weijer. Generalized gamut mapping using image derivative structures for color constancy. *International Journal of Computer Vision*, 86(2-3):127–139, 2010. 5

[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. 5, 7

[17] R. Gottumukkal and V. Asari. Skin color constancy for illumination invariant skin segmentation. *Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 5685*, pages 969–976, 2005. 1

[18] C. I. Group. Camera phone image quality phase 1: fundamentals and review of considered test methods. *I3A CPIQ Phase 1 document repository*, pages 1–82, 2007. 5

[19] T. Hansen, M. Olkkonen, S. Walter, and K. R. Gegenfurtner. Memory modulates color appearance. *Nature Neuroscience*, 9:1367–1368, 2006. 1, 2

[20] S. Hordley and G. Finlayson. Re-evaluating color constancy algorithms. *Proc. of the 17th International Conference on Pattern Recognition*, pages 76–79, 2004. 5

[21] S. D. Hordley. Scene illuminant estimation: past, present, and future. *Color Research & Application*, 31(4):303–314, 2006. 1

[22] ISO. Graphic technology - standard object colour spectra database for colour reproduction evaluation (socs). *Technical Report ISO/TR 16066:2003(E)*, 2003. 2

[23] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40:1106–1122, 2007. 1, 4

[24] A. Moreno, B. Fernando, B. Kani, S. Saha, and S. Karaoglu. Color correction: A novel weighted von kries model based on memory colors. *Computational Color Imaging Workshop (CCIW11), LNCS*, 6626:165–175, 2011. 1

[25] H. Nachlieli, R. Bergman, D. Greig, C. Staelin, B. Oicherman, G. Ruckenstein, and D. Shaked. Skin-sensitive automatic color correction. *HP Technical Report: HPL-2009-13*, pages 1–13, 2009. 1

[26] B. T. Ryu, J. Y. Yeom, C.-W. Kim, J.-Y. Ahn, D.-W. Kang, and H.-H. Shin. Extraction of memory colors for preferred color correction in digital TVs. *Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7241*, page 72410H, 2009. 1

[27] L. Shi and B. V. Funt. Re-processed version of the gehler color constancy database of 568 images. [online]. available: http://www.cs.sfu.ca/ colour/data/. last access: Nov. 1, 2011. 5

[28] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen. Using the skin locus to cope with changing illumination conditions in color-based face tracking. *Proc. IEEE Nordic Signal Processing Symposium (NORSIG 2000)*, pages 383–386, 2000. 1

[29] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Transactions on Image Processing*, 16(9):2207–2214, 2007. 2

[30] J. van de Weijer, C. Schmid, and J. Verbeek. Using high-level visual information for color constancy. *Proc. IEEE 14th Int. Conf. on Computer Vision*, pages 1–8, 2007. 1

[31] J. van de Weijer, C. Schmid, and J. Verbeek. Using high-level visual information for color constancy. *IEEE International Conference on Computer Vision (ICCV 2007)*, pages 1 –8, 2007. 2, 5

[32] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, pages 511–518, 2001. 6

[33] J. von Kries. Chromatic adaptation. *Festchrift der Albrecht-Ludwigs-Universitat, 1902 [Translation: D.L. MacAdam. Colorimetry-Fundamentals. SPIE Milestone Series, Vol. MS 77, 1993]*. 3

[34] A. Witkin. Scale-space filtering: a new approach to multi-scale description. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 9:150–153, 1984. 3