Consensus-based framework for illuminant chromaticity estimation

Simone Bianco
Francesca Gasparini
Raimondo Schettini
University of Milan-Bicocca
Department of Informatics, Systems and Communication
viale Sarca 336
20126 Milan, Italy
E-mail: schettini@disco.unimib.it

Abstract. *Several algorithms were proposed in the literature to recover the illuminant chromaticity of the original scene. These algorithms work well only when prior assumptions are satisfied, and the best and the worst algorithms may be different for different scenes. We investigate the idea of not relying on a single method but instead consider a consensus decision that takes into account the responses of several algorithms and adaptively chooses the algorithms to be combined. We investigate different combining strategies of state-of-the-art algorithms to improve the results in the illuminant chromaticity estimation. Single algorithms and combined ones are evaluated for both synthetic and real image databases using the angular error between the RGB triplets of the measured illuminant and the estimated one. Being interested in comparing the performance of the methods over large data sets, experimental results are also evaluated using the Wilcoxon signed rank test. Our experiments confirm that the best and the worst algorithms do not exist at all among the state-of-the-art ones and show that simple combining strategies improve the illuminant estimation.* © 2008 SPIE and IS&T. [DOI: 10.1117/1.2921013]

1 Introduction

White balance is the process of removing unrealistic color casts from digital images, mostly due to the acquisition conditions. From a computational perspective, automatic white balance is a two-stage process: the illuminant is estimated, and the image colors are then corrected on the basis of this estimate. The correction generates a new image of the scene as if it were taken under a known, canonical illuminant. A generic image acquired by a digital camera is mainly characterized by three physical factors: the illuminant spectral power distribution $I(\lambda)$, the surface spectral reflectance $S(\lambda)$, and the spectral sensitivities $\mathbf{C}(\lambda)$ of the sensor. Using this notation, the sensor responses at the spatial point with coordinates (x,y) can be then described as

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

where ω is the wavelength range of the visible spectrum, $\boldsymbol{\rho}$ and $\mathbf{C}(\lambda)$ are three-component vectors. Because the three spectral sensitivities of the sensor $\mathbf{C}(\lambda)$ are usually respectively more sensitive to low, medium, and high wavelengths, the three-component vector of the sensor response $\boldsymbol{\rho}=(\rho_1,\rho_2,\rho_3)$ is also referred to as the sensor or camera $RGB=(R,G,B)$ triplet. Assuming that the color $\mathbf{I}$ of the illuminant in the scene observed by the camera only depends on the illuminant spectral power distribution $I(\lambda)$ and on the spectral sensitivities $\mathbf{C}(\lambda)$ of the sensor, automatic white balance is equivalent to the estimation of $\mathbf{I}$ by

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

given only the sensor responses $\boldsymbol{\rho}(x,y)$ across the image. This is an underdetermined problem and therefore cannot be solved without further assumptions and/or knowledge, such as some information about the camera being used, and/or assumptions about the statistical properties of the expected illuminants and surface reflectances.

The estimation of the color of the illuminant could be performed if an achromatic patch is present in the image. This is because the spectral reflectance $S(\lambda)$ of an achromatic surface is approximately constant over a wide range of wavelengths, and thus the sensor response $\boldsymbol{\rho}$ is proportional to $\int_{\omega} I(\lambda)\mathbf{C}(\lambda)\mathrm{d}\lambda$, that is, the RGB of the achromatic patch is proportional to that of the incident light.

To reduce the dimensionality of the problem, one common method is to not estimate the whole triplet of the illuminant color, but a two-dimensional (2D) projection of it in a chromaticity space. In fact, it is more important to estimate the chromatic components of the scene than its overall intensity.

The color correction is usually based on a diagonal model of illumination change derived from the von Kries hypothesis. This model assumes that two acquisitions of the same scene with the same imaging device but under different illuminants are related by an independent gain regulation of the three imaging channels.[1] A diagonal model is generally a good approximation of change in illumination, as shown by Finlayson *et al.*[2]

Paper 07153R received Jul. 26, 2007; revised manuscript received Nov. 30, 2007; accepted for publication Dec. 19, 2007; published online May 14, 2008.
1017-9909/2008/17(2)/023013/9/$25.00 © 2008 SPIE and IS&T.

The colors in a scene, acquired under an unknown illuminant U can be transformed as they were taken under the canonical illuminant C by

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} R^C/R^U & 0 & 0 \\ 0 & G^C/G^U & 0 \\ 0 & 0 & B^C/B^U \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (3)$$

where $RGB=(R,G,B)$ is a color in the image acquired under the unknown illuminant, $RGB'=(R',G',B')$ is the color in the corrected image, $RGB^U=(R^U,G^U,B^U)$ are the sensor responses of a camera to a reference white under the unknown illuminant, and $RGB^C=(R^C,G^C,B^C)$ are the corresponding responses under the canonical illuminant. Supposing RGB^C is known, to obtain the color correction matrix, we have to estimate the illuminant color RGB^U. To this aim several algorithms exist in literature, each with different assumptions.[2–11] To improve the illuminant estimation, Schaefer *et al.*[12] introduced a combined physical and statistical color constancy algorithm that integrates the statistics-based color by correlation method with a physics-based technique, based on the dichromatic reflectance model, using a weighted combination of their likelihoods for a given illumination set and taking the maximum likelihood entry. Cardei and Funt[13] obtained good illuminant estimation by combining the results of gray world, white patch, and neural net methods, considering both linear and nonlinear committee methods. Taking these results as points of departure, we investigate the idea of not relying on a single white estimation method but instead considering a consensus decision that takes into account the compendium of the responses of several algorithms. In this work, we consider not only fixed weight linear combinations of algorithms (e.g., Refs. 12 and 13), but also methods that adaptively choose the algorithms to combine. The white balance algorithms considered are briefly described in Sec. 2. In Sec. 3, the suggested combining methods are reported and exemplified. In Sec. 4, the combining methods proposed are experimentally evaluated on two databases of synthetic and *RGB* images. The results are evaluated using the Wilcoxon signed rank test[14] on the angular error in the illuminant white point estimations.[15] Finally, conclusions and future works are reported in Sec. 5.

2 Algorithms

We have implemented five white estimation algorithms. They are gray world,[16] white point,[6,17] edge-based,[7] color by correlation,[9,10] and color in perspective.[8] These are very different algorithms based on very different assumptions: the gray world and white point algorithms belong to a separate class of methods that rely on simple statistics, and the edge-based, the color by correlation, and the color in perspective algorithms rely on more advanced statistics. With respect to our previous works,[18,19] we have replaced the iterative white balance[20] with the edge-based automatic white balance[7] algorithms because the former was never the best estimate of the color of the illuminant in the preliminary experiments[18,19] and thus it was never selected in the combining methods. We have also excluded our self-tunable color balancing algorithm[5,21] in this study because it can be considered a high-level algorithm as it is based on a semantic analysis of the content of the image. We have also decided to exclude it from this study to make the results easily repeatable. We have chosen these five white estimation algorithms, but our combining methods can be applied without modifications to a different number and kind of algorithms.

The *gray world* algorithm assumes that given an image of sufficiently varied colors under a canonical illuminant, the average surface color in the image is gray.[16] The shift from gray of the measured averages on the three channels corresponds to the color of the illuminant, that is, $RGB^U = [\text{mean}(R), \text{mean}(G), \text{mean}(B)]$.

Assuming that there is always some white in the scene, the *white point* algorithm looks for it in the image; its color will then be the color of the illuminant. The white point algorithm determines this white as the maximum R, maximum G, and maximum B found in the image[6,17] that is, $RGB^U=[\max(R), \max(G), \max(B)]$.

The *edge-based* automatic white balancing with linear illuminant constraint has been introduced by Chen *et al.*[7] to reduce the effect of large regions of uniform colors on the illuminant estimate. Their algorithm is based on the gray world assumptions and employs both the edge detection and the illuminant constraint in the estimation. It consists of three steps. First, it looks for the edge points in the image. Second, it determines the average and standard deviation values of chromaticity of the edge points. Then, it uses such statistical values to find achromatic pixels in the image. Finally, the average chromaticity of the achromatic pixels is assumed to be the illuminant chromaticity.

The *color by correlation* algorithm has been introduced by Finlayson *et al.*[9,10] The basic idea is to precompute a correlation matrix that describes the extent to which the proposed illuminants are compatible with the occurrence of image chromaticities. Each row in the matrix corresponds to a different training illuminant. The matrix columns correspond to possible chromaticity ranges resulting from a discretization of the *rg*-chromaticity space (r,g), ordered in any convenient manner. The *rg*-chromaticity space is the 2D projection of the three-dimensional (3D) red, green, blue (RGB) space of the sensor responses $RGB=(R,G,B)$ defined as

$$\begin{cases} r = \dfrac{R}{(R+G+B)} \\ g = \dfrac{G}{(R+G+B)} \end{cases}. \quad (4)$$

In a further refinement, the correlation matrix has been set up to compute the probability that the observed chromaticities are due to each of the training illuminants. The best illuminant can then be chosen, using a maximum likelihood estimate, for example, or other methods described in the literature.[3,4] There are different versions of the color by correlation algorithm; in this paper, we have implemented the one described by Finlayson *et al.*[10]

The *color in perspective* algorithm, developed by Finlayson[8] is based on Forsyth's gamut-mapping approach.[11] The gamut-mapping algorithm considers the set of all possible (R,G,B) due to surfaces in the world under the known, canonical illuminant. This set is convex and is represented by its convex hull. The set of all possible (R,G,B) under the unknown illuminant is similarly represented by its convex hull. Under the diagonal assumption of illumination change, these two hulls are a unique diagonal mapping (a simple 3D stretch) of each other. Because the observed set is normally a proper subset, the mapping to the canonical illuminant is not unique, and Forsyth[11] provides a method for effectively computing the set of possible diagonal maps, which is a convex set in the space of mapping coefficients. Finlayson's color in perspective algorithm adds two additional ideas.[8] First, the gamut-mapping method can be used in the rg-chromaticity space (r,g). Second, the diagonal maps can be further constrained by restricting them to ones corresponding to expected illuminants.

There is a substantial difference between the last two algorithms and the previous ones: the last two cannot be applied to any image coming from any digital camera; this is because they respectively need a precomputation of the correlation matrix and of the color gamut under the canonical illuminant, and these are different for different capture devices. To build such a matrix, for the synthetic image data set, we have used a database of spectral reflectances[22] and computed their *RGB* values using the sensor spectral sensitivities of a camera available on the Web.[23] A different method has been used for the real image data set: because the data set was acquired by a camera whose intrinsic color characteristics are unknown. This data set, which is composed of more than 11 000 images, has been divided into illuminant sets based on the illuminant chromaticity measurements. Then for each illuminant set, 20% of the images (i.e., more than 2200 images) were used for training the color by correlation and the color in perspective algorithms and the remaining 80% for testing. The choice of using only 20% of the images for the training has been made to leave more data available (i.e., more than 8800 images) for testing the single algorithms and the combining methods.

3 Combining Methods

In this paper, we analyze different combining methods of the white estimation algorithms to improve the results in the illuminant chromaticity estimation. In this work, we propose different ways to achieve a consensus estimation. The underlying idea is that algorithms that give similar illuminant chromaticity estimations have to be trusted more than algorithms that give an estimate that is far from the others.

Let us suppose we have n different algorithms. For the sake of comparison, all the n estimations of the illuminant color given by the considered algorithms are projected into the rg-chromaticity space. Let then $\mathbf{rg}_1,\ldots,\mathbf{rg}_n$ be the rg-chromaticity estimations of the n algorithms considered. For a better understanding, it is worthwhile to emphasize that an rg-chromaticity estimation is a two-vector, that is, $rg=(r,g)$, and that all the operations in the following are done independently on the two components. The combining methods investigated are listed in the following with a brief description.

A. Mean: mean value of the results given by the n algorithms[13]; in formula

$$\mathbf{rg}_{Mean} = \frac{\sum_{i=1}^{n} \mathbf{rg}_i}{n}. \tag{5}$$

B. Nearest2: mean value of the two closest results of the n algorithms according to the Euclidean distance d in the chromaticity space; in formula

$$\mathbf{rg}_{Nearest2} = \left\{ \frac{\mathbf{rg}_m + \mathbf{rg}_n}{2} : d(\mathbf{rg}_m, \mathbf{rg}_n) = \min_{i,j;i\neq j} d(\mathbf{rg}_i, \mathbf{rg}_j) \right\}. \tag{6}$$

C. Nearest-N%: mean value of the results of the algorithms with relative distances below the $(100+N)$% of the distance of the two closest results of the n algorithms. N is an arbitrary parameter (in this work, we have chosen $N=10,30$). In formula

$$\mathbf{rg}_{Nearest\text{-}N\%} = \left\{ \frac{\sum_{i\in I} \mathbf{rg}_i}{\#I} : i \in I \Leftrightarrow \exists\, j : d(\mathbf{rg}_i, \mathbf{rg}_j) \leq \frac{100+N}{100} d_{Nearest2} \right\}, \tag{7}$$

where $\#I$ denotes the cardinality of the set I.

D. No-N-max: let $D_j=\sum_{i=1,\ldots,n;i\neq j} d(\mathbf{rg}_j,\mathbf{rg}_i)$, that is, D_j represents the sum of the distances of the estimate of the algorithm j from all the other estimates; we reorder the n algorithm responses $\mathbf{rg}_1,\ldots,\mathbf{rg}_n$ as $\mathbf{rg}_{p1},\ldots,\mathbf{rg}_{pn}$ where $D_{p1}<D_{p2}<\cdots<D_{pn}$, that is, we reorder the estimates form the one with the smallest distance from all the other estimates to the one with the greatest distance from all the others. This combining method takes the mean value of the results of the uncombined methods excluding the N estimates with the highest distance from the others (i.e., the last N estimates in the reordered sequence); in formula

$$\mathbf{rg}_{No\text{-}N\text{-}Max} = \frac{\sum_{i=D_{p_1},\ldots,D_{p_{n-N}}} \mathbf{rg}_i}{n-N}. \tag{8}$$

E. Median: extraction of the result with the lowest distance from all the others, that is, the first element of the reordered sequence introduced in the no-N-max estimation; in formula

$$\mathbf{rg}_{Median} = \left\{ \mathbf{rg}_j : D_j = \min_{i=1,\ldots,n} (D_i) \right\} = \{\mathbf{rg}_{D_{p1}}\}. \tag{9}$$

The combining methods apply equally well to a different number n of algorithms. The only thing to take into account is that no-N-max is equivalent to nearest2 if $N=n-2$.

These combining methods can be better understood with the aid of Fig. 1, where five points (numbered from 1 to 5) simulate the possible different estimations of the illuminant chromaticity obtained from the five independent algo-

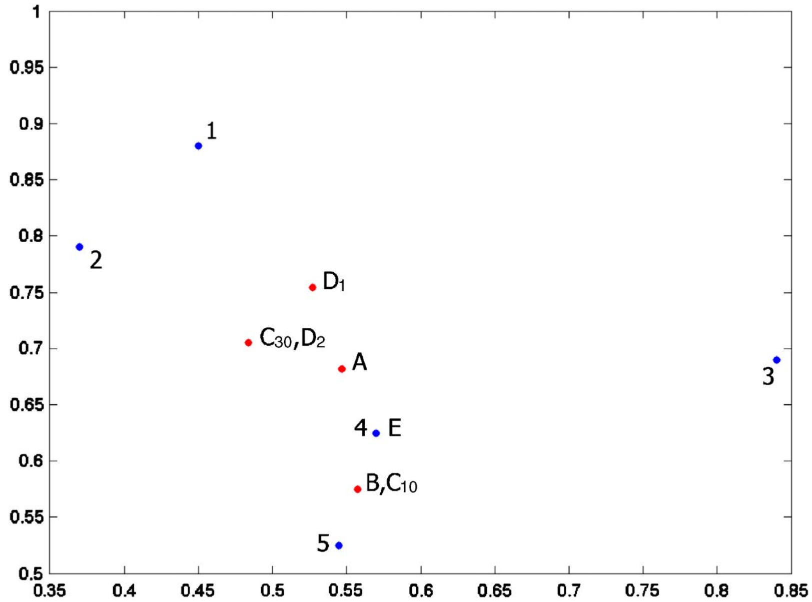

Fig. 1 A typical distribution of illuminant cromaticity estimations obtained using the uncombined algorithms considered (points from 1 to 5) and the combined methods (points A, B, C_{10}, C_{30}, D_1, D_2, E). The combined algorithms use, respectively, the chromaticity points listed in the following: A: mean (1-2-3-4-5); B: nearest2 (4-5); C_{10}: nearest-10% (4-5); C_{30}: nearest-30% (2-4-5); D_1: no-1-max (1-2-4-5); D_2: no-2-max (2-4-5); E: median (4).

rithms, and the investigated combining methods give the chromaticity estimations, named, respectively, A, B, C_{10}, C_{30}, D_1, D_2, E.

4 Experimental Results

4.1 *Error Measurement*

To evaluate the performances of the algorithms and combined methods considered, we need to specify an error measure. Because in estimating the scene illuminant it is more important to estimate its color than its overall intensity, we have to use an intensity-independent error measure. As suggested by Hordley and Finlayson,[15] for error measurement, we use the angle between the RGB triplets of the illuminant color ($\boldsymbol{\rho}_w$) and the algorithm's estimate of it ($\widehat{\boldsymbol{\rho}_w}$)

$$e_{Ang} = \arccos\left(\frac{\boldsymbol{\rho}_w^t \widehat{\boldsymbol{\rho}_w}}{\|\boldsymbol{\rho}_w\| \, \|\widehat{\boldsymbol{\rho}_w}\|}\right). \tag{10}$$

In Sec. 3, we explained that we have projected all the illuminant color estimates for homogeneity in the rg-chromaticity space to apply the combining methods. Next, we have to recover their RGB triplets given their rg-chromaticities (r,g). This can be easily done, up to a scale factor β, depending on its intensity, by Eq. (11)

$$\begin{cases} R = \dfrac{\beta r}{g} \\ G = \beta \\ B = -\dfrac{\beta(r+g-1)}{g} \end{cases}. \tag{11}$$

Because e_{Ang} is intensity-independent, it can be shown that it is blind to α; consequently, the angular error will be the same whatever the choice for α is. In this work, $\alpha = 1$ was chosen.

Hordley and Finlayson[15] showed that the underlying distribution of the angular error cannot be well modeled by a standard distribution and thus the mean error is not representative. Instead they suggest also using a test together with single summary statistics, which is able to compare the whole error distributions. An appropriate test for comparing angular error distributions is the Wilcoxon signed rank test, as it does not make any assumption about the underlying distribution.[14] Here is a brief description of the test already made by Hordley and Finlayson:[15] let X and Y be random variables representing the angular error of the illuminant chromaticity estimation of algorithms X and Y. The Wilcoxon test is used to test the hypothesis that the random variables X and Y are such that $p = P(X > Y) = 0.5$. That is, we hypothesize that algorithms X and Y have the same performance. To test the hypothesis $H_0 : p = 0.5$, we consider independent pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ of errors for N different images. We denote by W the number of images for which $X_i > Y_i$. When H_0 is true, W is binomially distributed $[b(N, 0.5)]$ and the Wilcoxon test is based on this statistic. We can define an alternative hypothesis $H_1 : p < 0.5$, which if true implies that errors for algorithm X are lower than those for algorithm Y. We accept or reject the null hypothesis at a given significance level α if the probability of observing the results we observe is less than or equal to α. The value of α we choose defines the error rate we accept when we reject the null hypothesis. For the experiments reported in this paper, we choose $\alpha = 0.01$, that is, we accept an error rate of 1% (we erroneously reject the null hypothesis in 1% of the cases).

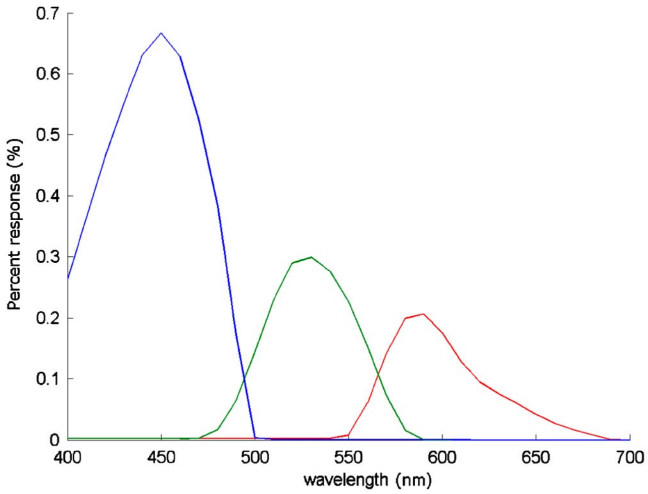

Fig. 2 Sony DXC-930 sensor spectral sensitivities. (Color online only.)

4.2 *Experiments*

We have analyzed the performance of the five uncombined algorithms and of seven combined method [mean (M), nearest2 (N2), nearest-10% (N10p), nearest-30% (N30p), no-1-max (N1M), no-2-max (N2M), and median (Med)] on two different sets of synthetic and real images. Although real image data sets are the most widely used, we have also included synthetic images in our experiments because they can be carefully controlled easily. Synthetic image data sets have already been used by Barnard *et al.*[3] and by Finlayson *et al.*[9,15]

The algorithms tested are gray world (GW), white point (WP), edge-based (EB), color by correlation (CbC), and color in perspective (CiP). For comparison, we have also added the do nothing (DN), which is not included to generate the responses of the combining methods.

4.2.1 *Synthetic image experiments*

We have computed the performance of the algorithms and of the combining methods on a test set of 6000 synthetic images, composed of six collections of 1000 scenes, each of them composed, respectively, of 2, 4, 8, 16, 32, and 64 surfaces.[15]

The reflectances of these surfaces have been randomly selected from a database of more than 40 000 measured reflectances representative of the world.[22] For each image that we have generated, the illuminant has also been randomly chosen from a set of 287 measured illuminants.[23] To generate synthetic sensor responses, we have adopted the spectral sensitivities of the Sony DXC-930 (Sony Corporation, New York, New York) digital video camera (depicted in Fig. 2), which are also available on the Web.[23]

The tristimulus values corresponding to each surface can thus be evaluated using Eqs. (12)

$$R=\sum_{\lambda=400\text{ nm}}^{700\text{ nm}} S_1(\lambda)I(\lambda)R(\lambda),$$

$$G=\sum_{\lambda=400\text{ nm}}^{700\text{ nm}} S_2(\lambda)I(\lambda)R(\lambda),$$

$$B=\sum_{\lambda=400\text{ nm}}^{700\text{ nm}} S_3(\lambda)I(\lambda)R(\lambda), \tag{12}$$

where $I(\lambda)$ is the spectral power distribution of the illuminant considered, $R(\lambda)$ the reflectance considered, and $S_1(\lambda)$, $S_2(\lambda)$, $S_3(\lambda)$ the spectral sensitivities of the sensor device. For the 6000 scenes, the angular error in the estimated illuminant chromaticity is evaluated for each of the 13 methods considered: 5 uncombined algorithms, 7 combined methods, and the do nothing method. The Wilcoxon signed rank test applied to the 13 angular error distributions gives a 13-by-13 matrix, whose entries are +, −, or = (see Tables 1 and 2). A plus sign in the i'th row and j'th column of the matrix means that algorithm i is statistically better than algorithm j when judged according to the Wilcoxon test. A minus implies that it is worse, and an equal sign implies that the two algorithms are statistically equivalent.

Counting the number of plus signs for each row of the matrix gives us a score; this is representative of the number of the algorithms with respect to which the algorithm considered results to be statistically better.

4.2.2 *Real image experiments*

We have analyzed the performance of the illuminant estimation algorithms and of the combining methods on a data set of carefully controlled real scene images acquired by Ciurea and Funt[24] with a Sony VX-2000 camera. The image data set consists of approximately 11 000 images in which the *RGB* color of the ambient illuminant in each scene is measured. The data set is built using a digital video camera with a neutral gray sphere attached to the camera so that the sphere always appeared in the field of view. Examples of images within this data set are reported in Fig. 3. To avoid the use of the gray ball as the reference to estimate the illuminant, we have properly cropped all the images of the database as shown in Fig. 4.

4.3 *Results*

In Tables 1 and 2, we, respectively, report the Wilcoxon signed rank test results on the synthetic and real image data sets. Wilcoxon scores and other single summary statistics for the algorithms on the test sets of synthetic and real images are reported, respectively, in Tables 3 and 4. The columns denoted with AVG, MAX, and STD report, respectively, the average, the maximum, and the standard deviation of the algorithm angular errors. The column %DN reports the percentual improvement on the angular error with respect to the do nothing method, and the %best reports the improvement with respect to the best uncombined algorithm. The last two columns, named best count and worst count, report the percention of images on which the algorithm has been considered the best (i.e., the algorithm whose illuminant chromaticity estimation is closest to the actual one) and the worst one (i.e., the algorithm whose illuminant chromaticity estimation is furthest from the actual one). A 0% in the last column does not mean that the corresponding algorithm has estimated the correct illumi-

Table 1 Wilcoxon signed rank test results on the synthetic images data set.

	DN	GW	WP	EB	CbC	CiP	M	N2	N10p	N30p	N1M	N2M	Med
DN		–	–	–	–	–	–	–	–	–	–	–	–
GW	+		–	+	–	+	–	–	–	–	–	–	–
WP	+	+		+	–	+	–	=	–	–	–	–	–
EB	+	–	–		–	+	–	–	–	–	–	–	–
CbC	+	+	+	+		+	+	+	+	+	+	–	–
CiP	+	–	–	–	–		–	–	–	–	–	–	–
M	+	+	+	+	–	+		+	+	+	=	–	–
N2	+	+	=	+	–	+	–		–	–	–	–	–
N10%	+	+	+	+	–	+	–	+		=	–	–	–
N30%	+	+	+	+	–	+	–	+	=		–	–	–
N1M	+	+	+	+	–	+	=	+	+	+		–	–
N2M	+	+	+	+	+	+	+	+	+	+	+		+
Med	+	+	+	+	+	+	+	+	+	+	+	–	

nant chromaticity, but rather that its illuminant chromaticity estimation was never the furthest from the correct one.

The last two columns of Tables 3 and 4 sum to a value greater than 100%. This is due mainly to two aspects. First, the median combining method does not generate a new chromaticity estimate but takes one already given by a white estimation algorithm; consequently, if this algorithm is considered the best one or the worst one for the image in

Table 2 Wilcoxon signed rank test results on the real images data set.

	DN	GW	WP	EB	CbC	CiP	M	N2	N10p	N30p	N1M	N2M	Med
DN		–	–	–	–	–	–	–	–	–	–	–	–
GW	+		+	–	+	+	–	+	+	+	–	+	–
WP	+	–		–	–	–	–	–	–	–	–	–	–
EB	+	+	+		+	+	+	+	+	+	–	+	+
CbC	+	–	+	–		–	–	–	–	–	–	–	–
CiP	+	–	+	–	+		–	–	–	–	–	–	–
M	+	+	+	–	+	+		+	+	+	–	+	=
N2	+	–	+	–	+	+	–		=	–	–	–	–
N10%	+	–	+	–	+	+	–	=		–	–	–	–
N30%	+	–	+	–	+	+	–	+	+		–	=	–
N1M	+	+	+	+	+	+	+	+	+	+		+	+
N2M	+	–	+	–	+	+	–	+	+	=	–		–
Med	+	+	+	–	+	+	=	+	+	+	–	+	

Fig. 3 Examples of images within the data set of real images.

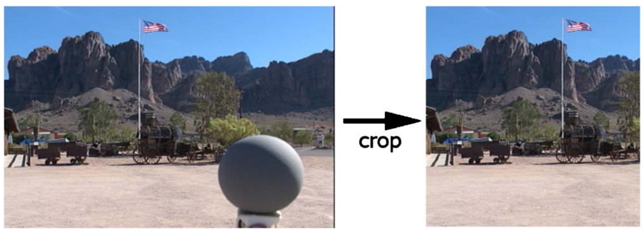

Fig. 4 Example of the cropping used to avoid the use of the gray ball in the illuminant estimation.

consideration, the median combining method will also be considered in the same way. Second, two different strategies may give the same chromaticity estimate.

The best strategy on the synthetic image data set is the no-2-max, which improves the average error of the best uncombined algorithm by 14.93%, and improvements in the maximum error and in the standard deviation are, respectively, 8.39% and 40.37%. The best strategy on the real image data set is the no-1-max, which improves the average error of the best uncombined algorithm by 1.28%, and improvements in the maximum error and in the standard deviation are, respectively, 33.81% and 11.70%. Looking at the last two columns of Tables 3 and 4, we may note that the best and the worst white estimation algorithms depend on the image content. This further supports our idea of not considering fixed weight combinations of the responses of

Table 3 Summary statistics on the synthetic images data set.

Method	WST score	AVG	MAX	STD	%DN	% best	best count	worst count
DN	0	15.87	39.86	9.81	*	*	3.46%	55.46
GW	3	5.30	39.43	3.83	66.60%	*	7.80%	1.22%
WP	4	4.86	48.97	4.84	69.39%	*	12.72%	1.20%
EB	2	8.25	51.90	7.82	48.06%	*	5.02%	14.54%
CbC	10	5.11	44.08	6.86	67.81%	*	28.62%	6.18%
CiP	1	10.06	25.03	5.04	36.65%	*	5.26%	21.40%
Mean	8	4.34	26.10	3.25	72.65%	10.65%	12.96%	0.00%
Nearest2	4	4.73	48.99	4.65	70.19%	2.60%	3.58%	0.00%
Nearest-10%	6	4.69	48.99	4.62	70.46%	3.48%	3.50%	0.00%
Nearest-30%	6	4.62	48.99	4.57	70.87%	4.81%	3.30%	0.00%
No-1-Max	8	4.51	29.42	3.87	71.61%	7.23%	9.74%	0.00%
No-2-Max	12	4.13	40.38	4.09	73.96%	14.93%	8.94%	0.00%
Median	11	4.24	39.43	3.84	73.28%	12.71%	9.20%	0.06%

Table 4 Summary statistics on the real images data set.

Method	WST score	AVG	MAX	STD	%DN	%best	best count	worst count
DN	0	11.10	27.41	6.72	*	*	15.91%	45.74%
GW	8	7.68	44.03	5.43	30.76%	*	30.99%	19.31%
WP	1	10.14	27.41	6.27	8.63%	*	13.20%	33.27%
EB	11	7.40	38.83	5.04	33.34%	*	17.82%	5.72%
CbC	2	9.34	24.85	5.27	15.87%	*	6.74%	24.87%
CiP	3	8.45	22.78	4.78	23.82%	*	7.84%	7.13%
Mean	9	7.47	26.09	4.40	32.70%	−0.96%	3.17%	0.00%
Nearest2	4	7.99	34.35	4.71	28.01%	−8.00%	1.76%	0.00%
Nearest-10%	4	7.97	34.35	4.68	28.22%	−7.69%	1.75%	0.00%
Nearest-30%	6	7.91	24.15	4.64	28.69%	−6.97%	1.80%	0.00%
No-1-Max	12	7.30	25.70	4.45	34.19%	1.28%	7.41%	0.00%
No-2-Max	6	7.93	24.35	4.86	28.51%	−7.24%	6.03%	0.00%
Median	9	7.46	26.92	4.47	32.81%	−0.79%	5.27%	0.40%

the algorithms but rather adaptive ones. Note that some of the combined methods instead are never considered the worst. This is due to the way they are obtained by combining the original algorithms.

5 Conclusions

Automatic estimation of illuminant chromaticity from digital images is an underdetermined problem and thus impossible to solve in the most general case.[17] Several algorithms have been proposed in the literature, each one based on different assumptions. We have considered here some well-known and widely used algorithms that are based on color image statistics, and we have shown that on both synthetic and real images, the best and the worst algorithms do not exist at all. We have thus studied different consensus-based strategies for illuminant chromaticity estimation. According to our approach, different algorithms are adaptively chosen and combined simply on the basis of their proximity in the chromaticity space. The proposed combining strategies can be applied without modifications to different numbers and kinds of algorithms.

Experimental results on large databases of both synthetic images and real RGB images have shown that combined methods are never considered the worst, and that they improve the accuracy of the illuminant estimation. Among the investigated combining strategies the most affordable one is the no-N-max, that takes the mean value of the results of the uncombined methods excluding the N estimates with the highest distance from the others.

Gijsenij and Gevers[25] have investigated how high-order image statistics can be used to select proper color constancy algorithms, and they have shown that for certain scene categories, one specific color constancy algorithm can be used. We have also observed in our experimental results that the best and the worst white estimation algorithms strongly depend on the image content. In future work, we thus plan to integrate combining strategies together with content-based image analysis and image classification.[21,26]

References

1. M. D. Fairchild, *Color Appearance Models*, Addison Wesley, Boston, Mass. (1997).
2. G. D. Finlayson, M. S. Drew, and B. Funt, "Diagonal transform suffice for color constancy," in *Proc. IEEE Int. Conf. Computer Vision*, pp. 164–171 (1993).
3. K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms; part one: methodology and experiments with synthetic images," in *IEEE Trans. Image Process.* **11**(9), 972–984 (2002).
4. K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms; part two: experiments with image data," in *IEEE Trans. Image Process.* **11**(9), 985–996 (2002).
5. A. Bruna, F. Gasparini, F. Naccari, and R. Schettini, "Multidomain pixel analysis for illuminant estimation and compensation," in *Proc. Digital Photography II*, N. Sampat, J. M. DiCarlo, and R. A. Martin, Eds., *Proc. SPIE* **6069**, 60690C (2006).
6. V. Cardei, B. Funt, and K. Barndard, "White point estimation for uncalibrated images," in *Proc. IS&T/SID 7th Color Imaging Conf.*, pp. 97–100 (1999).
7. H. H. Chen, C. Shen, and P. Tsai, "Edge-based automatic white balance with linear illuminant constraint," in *Visual Communications and Image Processing 2007*, C. W. Chen, D. Schonfeld, and J. Luo, Eds., *Proc. SPIE* **6508**, 65081D (2007).
8. G. D. Finlayson, "Color in perspective," *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 1034–1038 (1996).
9. G. D. Finlayson, P. Hubel, and S. Hordley, "Color by correlation," in *Proc. IS&T/SID 5th Color Imaging Conf.: Color Science, Systems and Applications*, pp. 6–11 (1997).
10. G. D. Finlayson, S. D. Hordley, and P. M. Hubel, "Color by correlation: a simple, unifying framework for color constancy," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 1209–1221 (2001).
11. D. Forsyth, "A novel algorithm for color constancy," *Int. J. Comput. Vis.* **5**, 5–36 (1990).

12. G. Schaefer, S. Hordley, and G. Finlayson, "A combined physical and statistical approach to colour constancy," in *Proc. Int. Conf. Comput. Vis. Pattern. Recogn.*, pp. 148–153 (2005).
13. V. C. Cardei and B. Funt, "Committee-based colour constancy," in *Proc. IS&T/SID 7th Color Imaging Conf.*, pp. 311–313 (1999).
14. R. V. Hogg and E. A. Tanis, *Probability and Statistical Inference*, Prentice Hall, Upper Saddle River, N.J. (2001).
15. S. D. Hordley and G. D. Finlayson, "Re-evaluating color constancy algorithms," in *Proc. 17th Int. Conf. Pattern Recogn.*, pp. 76–79 (2004).
16. G. Buchsbaum, "A spacial processor model for object color perceptions," *J. Franklin Inst.* **310**, 1–26 (1980).
17. B. Funt, K. Barnard, and L. Martin, "Is machine colour constancy good enough?," in *Proc. 5th European Conf. Comput. Vis.*, pp. 445–459 (1998).
18. S. Bianco, F. Gasparini, and R. Schettini, "Combining strategies for white balance," in *Proc. Digital Photography III*, R. A. Martin, J. M. DiCarlo, and N. Sampat, Eds., *Proc. SPIE* **6502**, 65020D (2007).
19. S. Bianco, F. Gasparini, and R. Schettini, "Combining strategies for automatic white estimation in real images, in *Proc. IEEE Computational Color Imaging Workshop*, pp. 175–178 (2007).
20. J. Huo, Y. Chang, J. Wang, and X. Wei, "Robust automatic white balance algorithm using gray color points in images," *IEEE Trans. Consum. Electron.* **52**, 541–546 (2006).
21. F. Gasparini and R. Schettini, "Color balancing of digital photos using simple image statistics," *Pattern Recogn.* **37**(6), 1201–1217 (2004).
22. "Graphic technology—standard object colour spectra database for colour reproduction evaluation (SOCS)," Technical Report ISO/TR 16066:2003(E), International Standards Organization, Geneva, Switzerland (2003).
23. Simon Fraser University, Computational Vision Lab Data, www.cs.sfu.ca/colour/data.
24. F. Ciurea and B. Funt, "A large image database for color constancy research," in *Proc. IS&T/SID 11th Color Imaging Conf.*, pp. 160–164 (2003).
25. A. Gijsenij and Th. Gevers, "Color constancy using natural image statistics," in *Proc. Int. Conf. Comput. Vis. Pattern Recogn.*, pp. 18–23 (2007).
26. R. Schettini, C. Brambilla, C. Cusano, and G. Ciocca, "Automatic classification of digital photographs based on decision forests," *Int. J. Pattern Recognit. Artif. Intell.* **18**(5), 819–845 (2004).

Simone Bianco obtained the BSc and the MSc degree in Mathematics from the University of Milan-Bicocca, Italy, respectively in 2003 and 2006. He is currently a PhD student at the DISCo, Departement of Informatic, Systems and Communication, University of Milan-Bicocca, Italy.

Francesca Gasparini took her degree (Laurea) in nuclear engineering at the Polytechnic of Milan in 1997 and her PhD in science and technology in nuclear power plants at the Polytechnic of Milan in 2000. Since 2001, she has been a fellow at the ITC Imaging and Vision Laboratory of the Italian National Research Council. Francesca is currently a research scientist at DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione), University of Milano Bicocca, where her research focuses on image and video enhancement, face detection, and image quality assessment.

Raimondo Schettini is an associate professor at the University of Milano Bicocca (Italy). He is Vice-Director of the Department of Informatics, Systems and Communication, and head of Imaging and Vision Lab (www.ivl.disco.unimib.it). He has been associated with Italian National Research Council (CNR) since 1987. He has been team leader in several research projects and published more than 180 refereed papers on image processing, analysis and reproduction, and on image content-based indexing and retrieval. He is an associated editor of the *Pattern Recognition Journal*. He was a co-guest editor of three special issues about Internet imaging (*Journal of Electronic Imaging*, 2002), color image processing and analysis (*Pattern Recognition Letters*, 2003), and color for image Indexing and Retrieval (*Computer Vision and Image Understanding*, 2004). He was member of the CIE TC 8/3 and Chairman of the 1st Workshop on Image and Video Content-based Retrieval (1998), of the First European Conference on Color in Graphics, Imaging and Vision (2002), of the EI Internet Imaging Conferences (2000-2006), and of the EI Multimedia Content Access: Algorithms and Systems 2007 Conference (2007, 2009).