Improving Color Constancy Using Indoor–Outdoor Image Classification

Simone Bianco, Gianluigi Ciocca, Claudio Cusano, and Raimondo Schettini

***Abstract*—In this work, we investigate how illuminant estimation techniques can be improved, taking into account automatically extracted information about the content of the images. We considered indoor/outdoor classification because the images of these classes present different content and are usually taken under different illumination conditions. We have designed different strategies for the selection and the tuning of the most appropriate algorithm (or combination of algorithms) for each class. We also considered the adoption of an uncertainty class which corresponds to the images where the indoor/outdoor classifier is not confident enough. The illuminant estimation algorithms considered here are derived from the framework recently proposed by Van de Weijer and Gevers. We present a procedure to automatically tune the algorithms' parameters. We have tested the proposed strategies on a suitable subset of the widely used Funt and Ciurea dataset. Experimental results clearly demonstrate that classification based strategies outperform general purpose algorithms.**

***Index Terms*—Color constancy, decision forests, indoor and outdoor image classification.**

I. Introduction

COMPUTATIONAL color constancy aims to estimate the actual color in an acquired scene disregarding its illuminant. The different approaches can be broadly classified into color invariant and illuminant estimation [1]. The former approaches derive from the image data invariant color descriptors without estimating explicitly the scene illuminant. The latter is actually a two stage procedure: the scene illuminant is estimated from the image data, and the image colors are then corrected on the basis of this estimate to generate a new image of the scene as if it were taken under a known, canonical illuminant. Many illuminant estimation solutions have been proposed in the last few years although it is known that the problem addressed is actually ill-posed as its solution lack uniqueness or stability. To cope with this problem, different solutions usually exploit some assumptions about the statistical properties of the expected illuminants and/or of the object reflectances in the scene. We have recently considered some well known and widely used color constancy algorithms that are based on color image statistics, and we have shown that on large datasets of both synthetic and real images the best and the worst algorithms do not exist at all [2].

Hordley in his review paper [1] gives an excellent review of illuminant estimation algorithms and highlighted two research areas that are important in the context of improving the performance of color constancy algorithms: making additional measurements at the time of image capture (i.e., use more color pixel information), and algorithm combining (i.e., using two or more estimations of the illuminants). In this paper we investigate a third hypothesis: the use of high level visual information to improve illuminant estimation.

The use of content-based image analysis for automatic color correction has been originally proposed by Schröoder and Moser [3]. They classify the images into several signal-oriented generic classes (e.g., scene with high color complexity) and, after the class-specific application of a set of color correction algorithms (White Patch and Gray World), they combine the results in such a way as to take into account the class-specific reliabilities of each algorithm. Their proposal is based on a hierarchical Bayesian image content analysis consisting of feature extraction and unsupervised clustering. They also suggest that semantic classes (e.g., indoor, outdoor, vegetation scene, mountain scene, etc.) and specific image degradation classes (e.g., underexposure, strong color cast, etc.) could be used in a similar way. Gasparini and Schettini [4] applied an adaptive mixture of the white balance and gray world procedures. In order to avoid the mistaken removal of an intrinsic color, regions identified as probably corresponding to skin, sky, sea or vegetation, are temporarily removed from the analyzed image. Van de Weijer *et al.* [5] proposed high-level visual information to improve illuminant estimation. They modeled the image as a mixture of semantic classes such as grass, skin, road, and building and exploited this information to select the best illuminant out of a set of possible ones. They applied several illuminant estimation approaches to compute a set of possible illuminants. For each of them an illuminant color corrected image is evaluated on the likelihood of its semantic content and the illuminant resulting as the most likely semantic composition of the image is selected as the illuminant color. They tested their method on a small subset of the Ciurea and Funt database [6] that is composed of a variety of both indoor and outdoor scenes and has shown that their top-down approach on outdoor images works better than any other tested algorithms.

Gijsenij and Gevers [7] used natural image statistics to identify the most important characteristics of color images and achieve selection and/or combining of color constancy algorithms. To this end, they used the Weibull parameterization to capture the image characteristics, applied a k-means algorithm to cluster the parameters in a predefined set of clusters, and then

Manuscript received April 02, 2008; revised July 30, 2008. Current version published November 12, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Luca Lucchese.

The authors are with the Dipartimento di Informatica, Sistemistica e Comunicazione (DISCo), Università degli Studi di Milano-Bicocca, 20126 Milano, Italy (e-mail: bianco@disco.unimib.it; ciocca@disco.unimib.it; cusano@disco.unimib.it; schettini@disco.unimib.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2008.2006661

associated the best-suited color constancy algorithm to each cluster. Unseen images are assigned to the computed clusters, and the best color constancy algorithm for that image is chosen. Gijsenij and Gevers [7] have also suggested that for certain scene categories, such as forest, open country and streets, a specific color constancy algorithm can be used.

The idea investigated here is that the effectiveness of automatic illuminant estimation techniques may be improved if information about the content of the images is taken into account. To this end, we designed an illuminant estimation approach which exploits the information provided by an image classifier. We considered, as suggested by Szummer and Picard [8], the indoor and outdoor classes, which correspond to categories of images with different content, usually taken under different illumination conditions. Therefore, these two classes of images may require different color processing procedures.

In this paper, we derive and experimentally compare four different strategies for illuminant estimation.

- Class-Independent (CI): the same algorithm is applied without taking into account the image class. Among the available algorithms with the parameters optimized on a training set, the best one is chosen on the basis of a robust statistical test.
- Class-Dependent Parameterization (CDP): two instantiations of the same algorithm are used. They differ in the value of the parameters which are optimized for the individual classes. The best algorithm is selected on the basis of the statistical test of the performance on the whole training set. Given an unseen image, it is first classified as indoor or outdoor, and then processed with the algorithm tuned for that class.
- Class-Dependent Algorithms (CDA): for each class a different algorithm is applied. The parameters of each algorithm are optimized for the corresponding class. The best algorithm for indoor and the best algorithm for outdoor are selected on the basis of the statistical test of the performance on the corresponding subsets of the training set. Given an unseen image, it is firstly classified, and then processed with the algorithm selected for the predicted class.
- Class-Dependent Algorithms with Uncertainty Class (CDAUC): a rejection class is introduced to identify images on which the classifier is not confident enough. Therefore, the images are classified as indoor, outdoor, or uncertain. Images classified as indoor or outdoor are processed according to the CDA strategy. Images classified as uncertain are processed according to the CI strategy.

The above strategies are independent from the illuminant estimation algorithms. In this paper, we consider eight algorithms: six derived from the framework recently proposed in [16], and two obtained by a linear and a nonlinear algorithm combinations.

In our experiments, we used the database presented by Ciurea and Funt [6]. Since the images in this database were frames extracted from video clips, and they show high correlation between each other, we applied a video-based analysis to automatically select the images to be included in our data set. Experimental results show that the accuracy of color constancy techniques can be significantly improved if specific algorithms or combinations of algorithms are chosen for each image class.

The paper is organized as follows. Section II introduces the problem of automatic color constancy and presents our approach. Section III describes the data we used for performance evaluation and the training processes. Finally, Section IV reports and comments the results obtained and Section V summarizes our work and our future plans about this topic.

II. Proposed Approach

An image acquired by a digital camera can be seen as a function $\boldsymbol{\rho}$ 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, the sensor responses at the pixel with coordinates (x,y) can be thus 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 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 $\mathbf{RGB} = (R,G,B)$ triplet. In the following, we adopt the convention that $\mathbf{RGB}$ triplets are represented by row vectors.

Assuming that the color $\mathbf{I}$ of the scene illuminant as seen by the camera only depends on the illuminant spectral power distribution $I(\lambda)$ and on the sensor spectral sensitivities $\mathbf{C}(\lambda)$, then computational color constancy 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. It has been demonstrated that this is an under-determined problem [9] and, thus, further assumptions and/or knowledge are needed to solve it. Typically, some information about the camera being used are exploited, and/or assumptions about the statistical properties of the expected illuminants and surface reflectances.

The basic idea of our illuminant estimation approach is the following: let us suppose to have several distinct color constancy algorithms; based on their performances on a training set that is manually labeled into indoor and outdoor classes, it is possible to identify the best algorithm for the two classes considered and the best algorithm for the whole training set. On the basis of this idea, we derived four strategies which use the output of an image classifier to select the appropriate illuminant estimation algorithms.

A. Image Classification

There have been several efforts to automate the classification of digital images to date. Szummer and Picard [8] have constructed algorithms for indoor/outdoor image classification. Vailaya *et al.* [10] have considered the hierarchical classification of vacation images: at the highest level the images are sorted into indoor/outdoor classes, outdoor images are then assigned to city/landscape classes, and finally landscape images are classified in sunset, forest, and mountain categories. In Schettini *et al.* [11] low-level visual features are related to semantic image categories, such as indoor, outdoor and close-up, using CART classifiers. Several classification strategies were designed and experimentally compared, producing a classifier that can provide a reasonably good performance and robustness. In this work, we use a similar approach to classify the images into indoor and outdoor classes.

Each image is described by a set of low-level features related to color, texture, and edge distribution. The extracted features, organized in a feature vector, are fed to a decision forest trained to distinguish between indoor and outdoor images. After a feature selection phase, which consisted in training several classifiers and evaluating them on an independent validation set, we selected four features.

Information about color distribution is captured by spatial color moments: we transform the image into the YCbCr color space, divide it into seven horizontal bands, and compute the mean and the standard deviation of each of the three color bands. Since the YCbCr color space decorrelates luminance and chrominance components it is commonly used in image classification tasks. The subdivision in horizontal bands adequately describes some characteristics which are very useful for indoor/outdoor classification (images with blue sky in the upper part, or green grass in the lower part...). Color moments are less useful when the bands contain heterogeneous color regions. Therefore, a global color histogram has been selected as a second color feature. The RGB color space has been subdivided in 27 bins by a uniform quantization of each component in three ranges.

To describe the most salient edges we used an 18 bin edge direction histogram (ten degrees for each bin): the gradient of the luminance image is computed using Gaussian derivative filters tuned to retain only the major edges. Only the points for which the magnitude of the gradient exceeds a set threshold contribute to the histogram.

Texture information is extracted computing a set of features based on a multiresolution analysis. A three level wavelet transform of the luminance image is computed, yielding to ten different sub-bands. For each band, we then compute the average absolute value of the coefficients and their standard deviation. Summing up, each image is described by a feature vector of $2 \times 3 \times 7 + 27 + 18 + 10 \times 2 = 107$ components.

For classification, we used decision trees built according to the CART methodology [12]. The CART approach to classification presents several advantages: first of all it is a nonparametric and nonmetric method so that no *a priori* knowledge about the distribution of the values of the features is needed and the issue of feature normalization may be ignored. The hierarchical structure of the trees is rather easy to analyze making it possible to understand which features play a major role in the classification process. Moreover, CART trees have been previously applied with satisfactory results to other image classification problems [11], [13], [14].

Briefly, decision trees are produced by recursively partitioning a training set of feature vectors labeled with the correct class. Each split consists of a comparison between the value of a single feature and a threshold. Once a tree has been constructed, a class is assigned to each of the terminal nodes, and when a new case is processed by the tree, its predicted class is the class associated with the terminal node into which the case finally moves on the basis of the values of the features. The construction process is based on a training set of cases of known class. A function of impurity of the nodes, $i(t)$, is introduced, and the decrease in its value produced by a split is taken as a measure of the goodness of the split itself. For each node, all the possible splits on all the features are considered and the split which minimizes the average impurity of the two sub-nodes is selected. The function of node impurity we have used is the Gini diversity index

$$i(t) = 1 - \sum_{c=1}^{C} p(c|t)^2 \tag{3}$$

where $p(c|t)$ is the resubstitution estimate of the conditional probability of class c (in this case $C = 2$) in node t, that is, the probability that a case found in node t is a case of class c. When the difference in impurity between a node and best subnodes is below a threshold, the node is considered as terminal. The class assigned to a terminal node t is the class c^* for which

$$p(c^*|t) = \max_{c \in \{1,\cdots,C\}} p(c|t). \tag{4}$$

In CART methodology, the size of a tree is treated as a tuning parameter, and the optimal size is adaptively chosen from the data. A very large tree is grown and then pruned, using a cost-complexity criterion which governs the tradeoff between size and accuracy. Although the pruning process prevents overfitting, pruned trees still present instability (a small change in data may result in a very different tree). Decision forests can be used to overcome this problem improving, at the same time, generalization accuracy [15]. The trees of a decision forest are generated by running the training process on bootstrap replicates of the training set. The classification results produced by the single trees are combined by majority vote. The number of concordant votes may also be used as a measure of confidence of the combined classifier on the classification result.

B. Color Constancy Algorithms

Several computational color constancy algorithms exist in the literature, each based on different assumptions. In this paper, we chose eight algorithms (six algorithms and two combining strategies), but a different set of algorithms could be used.

Recently Van de Weijer *et al.* [16] have proposed a framework which unifies a variety of algorithms. These algorithms correspond to instantiations of the following equation:

$$\left(\int\int |\nabla^n \boldsymbol{\rho}_\sigma(x,y)|^p \, \mathrm{d}x\mathrm{d}y \right)^{\frac{1}{p}} = k\mathbf{I} \tag{5}$$

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 so that the illuminant color I has unit length. In this paper, by varying the three variables (n, p, σ) we have generated six algorithm instantiations that correspond to well known and widely used color constancy algorithms.

1) Grey World (GW) algorithm [17], which is based on the assumption that the average reflectance in a scene is achromatic. It can be generated setting $(n, p, \sigma) = (0, 1, 0)$ into (5).
2) White Point (WP) algorithm [18], also known as Maximum RGB, which is based on the assumption that the maximum reflectance in a scene is achromatic. It can be generated setting $(n, p, \sigma) = (0, \infty, 0)$ into (5).
3) Shades of Gray (SG) algorithm [19], which is based on the assumption that the pth Minkowski norm of a scene is achromatic. It can be generated setting $(n, p, \sigma) = (0, p, 0)$ into (5).
4) General Grey World (gGW) algorthm [16], [20], which is based on the assumption that the pth Minkowski norm of a scene after local smoothing is achromatic. It can be generated setting $(n, p, \sigma) = (0, p, \sigma)$ into (5).
5) Gray Edge (GE1) algorithm [16], which is based on the assumption that the pth Minkowski norm of the first order derivative in a scene is achromatic. It can be generated setting $(n, p, \sigma) = (1, p, \sigma)$ into (5).
6) Second Order Gray Edge (GE2) algorithm [16], which is based on the assumption that the pth Minkowski norm of the second order derivative in a scene is achromatic. It can be generated setting $(n, p, \sigma) = (2, p, \sigma)$ into (5).

As can be noticed, the instantiations of GW and WP have all three parameters (n, p, σ) fixed; SG instead, has the parameter p that can be opportunely tuned for a particular image; while gGW, GE1 and GE2 have two parameters (p and σ) which must be tuned.

With the aim of improving the illuminant color estimation of the single algorithms, we have also implemented a linear and a nonlinear combining algorithm. The first one is the Least Mean Squares Committee (LMS) proposed by Cardei and Funt [21]. The response of this combining algorithm is given by multiplying the responses of all the six single algorithms considered for a fixed 18×3 weight matrix $\mathbf{W}$. Formally

$$\mathbf{RGB}_{LMS} = [\mathbf{RGB}_1 \cdots \mathbf{RGB}_6]\mathbf{W}. \tag{6}$$

The second one is an instantiation of the No-N-Max (NNM) algorithm, which has been recently proposed by the authors and showed to perform well on synthetic and real images [2]. The underlying idea is that algorithms that give similar illuminant color estimations have to be trusted more than algorithms that give estimates that are far from the others, and, thus, the latter ones have to be automatically discarded.

Let $\mathbf{rgb}_i = \mathbf{RGB}_i \|\mathbf{RGB}_i\|^{-1}$, $i = 1, \cdots, 6$ be the normalized versions of the illuminant color estimates given by the six single algorithms considered, and let $D_j = \sum_{i=1, i \neq j}^{6} d(\mathbf{rgb}_i, \mathbf{rgb}_j)$, $j = 1, \cdots, 6$ be the sum of the Euclidean distances of the illuminant color estimate of the algorithm from all the other estimates. We reorder the six algorithm estimates $\mathbf{rgb}_1, \cdots, \mathbf{rgb}_6$ as $\mathbf{rgb}_{p_1}, \cdots, \mathbf{rgb}_{p_6}$, where $D_{p_1} \leq D_{p_2} \leq \cdots \leq D_{p_6}$. In other words, we reorder the estimates from the one with the smallest distance from all the others to the one with the highest distance from all the others. This combining method takes the mean value of the estimates of the single algorithms automatically discarding the N estimates with the highest distance from all the others (i.e., the last N in the reordered sequence). Formally

$$\mathbf{RGB}_{\mathrm{NNM}} = \frac{\sum_{i=p_1, \cdots, p_{6-N}} \mathbf{rgb}_i}{6 - N} \tag{7}$$

which is finally normalized as defined before.

C. Algorithm Selection Approaches

We propose two algorithm selection approaches. In the first one, we used the output of the classifier to decide which color constancy algorithm (and/or which set of parameters) to apply to the specific image under consideration as shown in Fig. 1(a). This approach is used in the CDP and CDA strategies. In the second selection approach, we take into account the confidence measure provided by the classifier as shown in Fig. 1(b). The difference with respect to the first one is the introduction of an uncertainty class constituted by the images for which the classifier is not sure about their membership to indoor or outdoor class (i.e., CDAUC strategy).

Let P be the confidence measure, that is, the fraction of concordant votes in the trees in the forest w.r.t. the output class. Given an input image, we consider the image to be indoor if the predicted class is indoor and $P \geq T_{\mathrm{INDOOR}}$; the image is considered outdoor if the predicted class is outdoor and $P \geq T_{\mathrm{OUTDOOR}}$; we consider the image to be in the uncertainty class otherwise. We chose to use two different thresholds because the classifier adopted does not guarantee a uniform confidence measure for both classes. The two thresholds are tuned by analyzing the final performance of the illuminant estimation algorithms.

If the input image is classified as indoor or outdoor and satisfies the constraints above, the image is processed with the best color balancing algorithm for that class. Otherwise, since the membership of the image to either class cannot be reliably inferred, applying a specific algorithm tuned for indoor or outdoor images may significantly worsen the appearance of the image processed. In these cases, a more conservative, general-purpose algorithm is instead applied.

III. Experimental Setup

In our approach, the classifier needs to be trained, using a set of images manually labeled as indoor or outdoor, and the parameters of some color constancy algorithms need to be tuned. The last point requires a dataset of images labeled with ground truth illuminants, and the definition of a suitable error measure which can quantify the accuracy of the algorithms in the estimation of the illuminant. In the following sections we introduce the datasets and the setup procedures.

A. Illuminant Dataset Selection

In [6], Ciurea and Funt presented an image database to be used as a common data set in the evaluation of colour constancy algorithms. In this database, 15 digital video clips were recorded (at 15 frames per second) in different settings such as indoor, outdoor, desert, markets, cityscape, etc... for a total of two hours of videos. From each clip, a set of images was extracted resulting in a database of more than 11,000 images. In

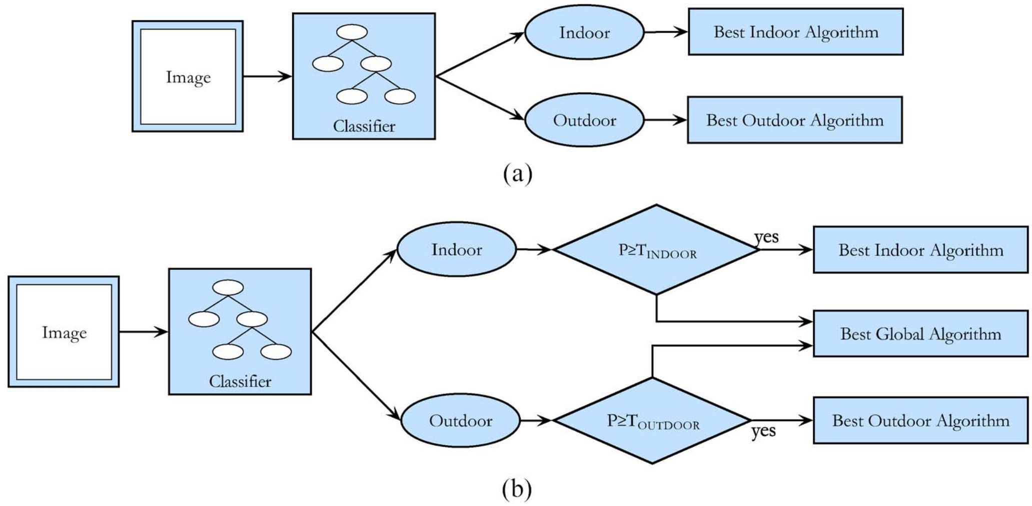

Fig. 1. Scheme of proposed approaches for the selection of the algorithms. In the first approach, the image is classified by the decision forest and the output class is used to select the color constancy algorithm. This approach is used in the CDP and CDA strategies. In the second approach, the appropriate color constancy algorithm is selected according to the output class and the confidence measure. This approach is used in the CDAUC strategy.

each image, a grey sphere appears in the bottom right corner of the images. This sphere was used to estimate the colour of the scene illuminant. The database is, thus, composed by images taken at different locations each one coupled with the measured illuminant.

Since the database sources were video clips, the images extracted show high correlation. To remove this correlation only a subset of images should be used from each set. Taking into account that the image sets came from video clips, we applied a video-based analysis to select the image to be included in the final illuminant dataset. The frames which show redundancy in terms of visual content are removed and only the most representative are retained (see [22] and [23]). Applying this procedure, we reduced the original dataset to 1,135 images.

Note that the aim of our procedure is to decorrelate the *pictorial content* of the images. Decorrelation of the illuminant is not guranteed. This fact, in our opinion, does not invalidate our evaluation strategy since we think that the performance of illuminant estimation algorithms depends more on the content of the image than on the illuminant itself.

More details about this procedure can be found in Appendix I.

B. Decision Forest Training

In order to obtain a classifier with good generalization capabilities, a rather large dataset of images is needed. For this purpose, we collected 6,785 images, downloaded from the web, or acquired by a scanner or digital cameras. All the material varied in size, resolution, and quality. The images were resized to 256 pixels on the largest dimension, and proportionally on the other dimension in such a way that the aspect ratio was maintained. The images have been manually annotated yielding to 2,105 indoor images and 4,680 outdoor images. No enhancement procedure (such as white balancing) has been applied to the images.

In order to select the features and to determine the size of the forest we partitioned the images into a training set of 2,000 images (1,000 indoor and 1,000 outdoor), and a validation set containing the remaining 4,785 images. As a result, we selected the four features described in Section II-A, and we set the size of the decision forest to 50 trees.

TABLE I
CONFUSION MATRIX OBTAINED ON THE TEST SET BY THE DECISION FOREST. THE NUMBER OF MISCLASSIFICATIONS WAS 169 (86 INDOOR AND 83 OUTDOOR IMAGES)

	Predicted indoor	Predicted outdoor
True indoor	82.1%	17.9%
True outdoor	12.7%	87.3%

Within this framework we obtained a classification accuracy of about 93.1% on the validation set. On the images of the illuminant dataset (Section III-A) we obtained the results summarized in Table I. Note that the test images have been manually annotated as indoor (481 images) or outdoor images (654 images).

The overall classification accuracy obtained on the test set is 85.1%. The difference with respect to the performance obtained on the validation set can be explained considering that the test set includes several images with little information about the context in which they were taken. For instance, the test set includes several close-ups of various objects, which cannot be classified as indoor or outdoor without exploiting high level knowledge and reasoning. Fig. 2 shows a sample of the misclassified images.

C. Error Measurement Definition

In order to evaluate the performance of the algorithms considered, we have used an intensity independent error measure. This choice is because it is more important in estimating the scene illuminant to estimate its color than its overall intensity. As suggested by Hordley and Finlayson [24], we use the angle between the RGB triplets of the illuminant color ($\boldsymbol{\rho}_w$) and the algorithm's estimate of it ($\hat{\boldsymbol{\rho}_w}$) as error measure

$$e_{\text{ANG}} = \arccos\left(\frac{\boldsymbol{\rho}_w^T \hat{\boldsymbol{\rho}_w}}{\|\boldsymbol{\rho}_w\| \|\hat{\boldsymbol{\rho}_w}\|}\right). \tag{8}$$

Fig. 2. Example of indoor images misclassified as outdoor images (a–l) and some outdoor images misclassified as indoor images (m–x).

It has been shown also [24] that the median error is a good descriptor of the angular error distribution.

D. Automatic Parameters Tuning

Four of the color constancy algorithms considered (SG, gGW, GE1, GE2), needed a training phase to opportunely tune the parameters (n, p, σ). We use the median error as the main index in the evaluation of the different algorithm performances. Being the median error a nonlinear statistic, we needed a multidimensional nonlinear optimization algorithm: our choice was to use a Pattern Search Method (PSM). PSMs are a class of direct search methods for nonlinear optimization [25], [26]. Their popularity is given by simplicity and by the fact that they work very well in practice on a variety of problems. Furthermore, global convergence can be established under certain regularity assumptions of the function to minimize [27]. PSMs are also extremely simple to implement and do not require any explicit estimate of derivatives.

We need a training set on which to perform the tuning of the parameters. To this end, 300 images (150 indoor and 150 outdoor) were randomly extracted from the 1,135 images of the illuminant dataset and used as training set. Since we expect different behavior of the algorithms on the different classes considered, the parameters were tuned independently for the indoor class, for the outdoor class and for the whole training set.

The form of a general pattern search algorithm can be described in the following way. At each step k, we have the current iterate $\mathbf{x}_k$, a set D_k of vectors which identify the search directions, and a step-length parameter Δ_k. Usually, the set D_k is the same for all iterations. For each direction $\mathbf{d}_k \in D_k$, we set $\mathbf{x}^+ = \mathbf{x}_k + \Delta_k \mathbf{d}_k$ (the "pattern") and we examine $f(\mathbf{x}^+)$ where f is the function to be minimized. If $\exists \mathbf{d}_k \in D_k : f(\mathbf{x}^+) < f(\mathbf{x}_k)$, we set $\mathbf{x}_{k+1} = \mathbf{x}^+$ and $\Delta_{k+1} = \alpha_k \Delta_k$ with $\alpha_k > 1$; otherwise, we set $\mathbf{x}_{k+1} = \mathbf{x}_k$ and $\Delta_{k+1} = \beta_k \Delta_k$ with $\beta_k < 1$. The algorithm stops when the step Δ_k is smaller than a fixed threshold or when the maximum number of iterations has been reached. In this paper, we have chosen to fix the maximum number of iterations $n = 50$, $\alpha_k = \alpha = 2$, $\beta_k = \beta = 0.5$, $D_k = \{NW, N, NE, E, SE, S, SW, W\}$, and $\Delta_0 = 0.1$. The same starting point has been chosen for the four algorithms that needed a training phase (SG, gGW, GE1 and GE2): $\mathbf{x}_0 = (p_0, \sigma_0) = (1, 0)$. If during the minimization process the parameter p exceeded a given threshold (in our case we set it at 200), we considered it to be infinity.

TABLE II
PARAMETERS FOUND BY THE PATTERN SEARCH ALGORITHM. ONLY THE VALUES REPORTED IN BOLD HAVE BEEN COMPUTED, THE OTHERS HAVE BEEN SET ACCORDING TO THE DEFINITION OF THE ALGORITHMS

	Indoor			Outdoor			General purpose		
	n	p	σ	n	p	σ	n	p	σ
GW	0	1	0	0	1	0	0	1	0
WP	0	∞	0	0	∞	0	0	∞	0
SG	0	**1.27**	0	0	$\boldsymbol{\infty}^*$	0	0	**1.06**	0
gGW	0	**1.32**	**1.00**	0	$\boldsymbol{\infty}^*$	**0.00**	0	**1.08**	**0.83**
GE1	1	**0.60**	**1.72**	1	**1.10**	**0.83**	1	**1.10**	**1.08**
GE2	2	**1.06**	**2.96**	2	**1.91**	**0.04**	2	**1.55**	**1.83**

* Values which diverge towards infinity.

The parameters found by the pattern search algorithm are reported in Table II. It can be seen that the optimal values found for the parameter p for the indoor class are lower than the ones for the outdoor class. The optimal values found for the parameter σ instead, are higher for the indoor class than the ones for the outdoor class. While for the indoor class the pattern search optimization found different values for the parameters of the different color constancy algorithms, we can see that in the choice of the parameters for the outdoor class of the gGW and SG algorithms tended to be asymptotically convergent to the WP's ones. Regarding the general-purpose class, we can see instead that the optimal choice for the parameters of the SG algorithm is very similar to the GW's ones.

Two combining algorithms are also considered, the No-2-Max (N2M) [2] and the Least Mean Squares Committee (LMS) [21]. The first does not require any training process, while the weight matrices for the LMS need to be computed on the training set using a least squares regression. Thus, for each of the classes considered, we have a different weight matrix for the LMS Committee which multiplies the responses of the six algorithms with the tuned parameters for that class. The LMS weight matrices are reported in Appendix II.

We underline that the fact of having three identical algorithms for the outdoor class (WP, SG, and gGW) does not pose any problems for the LMS: it is possible to obtain the same illuminant estimations if we sum the coefficients found for the three algorithms and use for example only the WP algorithm with the summed coefficient in the LMS committee. A different consideration has to be made for the N2M: as it discards image by image the two algorithms that have given an illuminant estimation that is farthest from the others, then for every image, the three algorithms that give the identical estimation (WP, SG, and gGW) are always retained together with the algorithm that gives the estimation that is closest to theirs. Furthermore, as the output of the N2M is the mean of the estimations of the algorithms retained, the estimation of the WP, SG, and gGW has more influence on the final estimation. In order to maintain the symmetry with the other class and to maintain a greater generality of the methods, we decided to still consider WP, SG, and gGW as distinct algorithms for both the instantiations of the N2M and LMS for the outdoor class.

IV. Experimental Results

The focus of our experimentation is to establish which of the following color constancy strategies is preferable (at least on the dataset considered).

- Class-Independent (CI): the parameters of the algorithms are tuned without considering the class of input images (see the "general purpose" column in Table II).
- Class-Dependent Parameterization (CDP): for a given algorithm the parameters to use are selected on the basis of the class predicted by the decision forest (see the "indoor" and "outdoor" columns in Table II).
- Class-Dependent Algorithms (CDA): for each class the best algorithm (and its corresponding parameters) is selected.
- Class-Dependent Algorithms with Uncertainty Class (CDAUC): the same as above, but with the introduction of the uncertainty class. Images falling in the uncertainty class are processed by the algorithm that has proved to be the best class-independent algorithm.

In order to perform such a comparison Hordley and Finlayson [24] showed that together with the summary statistics of the angular errors a test able to compare the whole error distribution between different algorithms is needed. Since standard probability models cannot well represent the underlying errors, we need a test that does not make any *a priori* assumptions about the underlying error distributions. In this work, to compare the performance of two color constancy algorithms we have used the Wilcoxon Sign Test (WST) [28]. Let X and Y be random variables representing the angular errors between the illuminant estimations of the two algorithms and the real illuminants; let μ_X and μ_Y be the medians of such random variables. The Wilcoxon signed-rank test can be used to test the null hypothesis $H_0 : \mu_X = \mu_Y$. To test H_0, we consider the difference of independent error pairs $(X_1 - Y_1), \cdots, (X_N - Y_N)$ for N different images. We rank these error pairs according to their absolute differences. If H_0 is correct, the sum of the ranks W will approximate zero. If W is much larger or smaller than zero, the alternative hypothesis $H_1 : \mu_X \neq \mu_Y$ is true. We can test H_0 against H_1 at a given significance level α. We reject and accept if the probability of observing the error differences we obtained is less than or equal to α. In this work, we have used the alternative hypothesis $H_1 : \mu_X < \mu_Y$ with a significance level $\alpha = 0.01$. Comparing every color constancy algorithm with all the others, we generated a score representative of the number of times that the null hypothesis H_0 has been rejected for the given algorithm, i.e., the number of times that the performance of the given algorithm has been considered to be better than the others.

TABLE III
MEDIAN ANGULAR ERROR OBTAINED BY THE COLOR CONSTANCY ALGORITHMS ON THE TRAINING SET. THE BEST RESULTS FOR EACH COLUMN ARE REPORTED IN BOLD

	Indoor Images†		Outdoor Images†		Whole Training Set	
	Median	WSTs	Median	WSTs	Median	WSTs
GW*	4.91	3	7.86	0	5.62	1
WP*	11.83	0	2.81	2	7.76	0
SG	**4.31**	**6**	2.81	2	5.56	1
gGW	4.32	**6**	2.81	2	5.57	1
GE1	5.40	1	3.72	1	5.45	1
GE2	5.57	1	**2.48**	**7**	5.47	1
N2M	5.13	3	2.83	2	5.02	6
LMS	4.58	5	2.71	2	**4.50**	**7**

* Algorithms with fixed, class independent, parameters.
† Test with algorithms tuned specifically for the class.

In order to select the best algorithms for each class, we evaluated the algorithms on the training set introduced in Section III. The results are reported in Table III. More in detail, the first column of the table reports the results obtained on the indoor images of the training set by the algorithms whose parameters have been tuned for that specific class (that is, the class-dependent parameterization strategy). Similarly, the second column summarizes the results obtained on outdoor images. The third column details the results of the general purpose version of the algorithms (that is, the class-independent strategy) when applied to the whole training set.

The results with the general purpose algorithms show that combinatorial algorithms, and in particular LMS, performed better than the others. The Wilcoxon test confirms that the difference in performance is significant (even the difference of 0.43 degrees between N2M and GE1). However, if we consider class-specific algorithms, SG and GE2 outperform the combining methods for indoor and outdoor images, respectively.

On the basis of the results obtained we can select the best algorithm for each class. For indoor images both SG and gGW may be chosen in fact, the WST index shows that they exhibit an indistinguishable behavior. We selected the SG because its angular error is slightly better, and is computationally less expensive than the gGW. For outdoor images the GE2 algorithm is clearly superior to the others.

We can see that the errors on the two classes are very different. In particular, with the exception of the GW algorithm, all the errors obtained on the outdoor class are significantly lower than

TABLE IV
MEDIAN ANGULAR ERROR OBTAINED BY THE COLOR CONSTANCY ALGORITHMS ON THE TEST SET USING THE CI AND CDP STRATEGIES. THE BEST RESULTS FOR EACH COLUMN ARE REPORTED IN BOLD

	CI Strategy		CDP Strategy	
	Median	WSTs	Median	WSTs
GW*	5.95	0	5.95	0
WP*	5.48	3	5.48	1
SG	5.80	0	4.08	**4**
gGW	5.80	0	5.39	1
GE1	4.47	5	4.32	3
GE2	4.65	5	**3.94**	**4**
N2M	4.79	4	4.01	**4**
LMS	**4.18**	**7**	4.05	**4**

* Algorithms with fixed, class independent, parameters.

TABLE V
SUMMARY OF THE RESULTS OBTAINED ON THE TEST SET BY THE FOUR STRATEGIES PROPOSED

Strategy	Underlying algorithms	Median	WSTs
CI	LMS, general purpose parameters	4.18	0
CDP	GE2, indoor and outdoor parameters	3.94	1
CDA	SG for indoor and GE2 for outdoor	3.78	1
CDAUC	SG ind., GE2 out., LMS gen. purpose	**3.54**	**3**

those obtained on the indoor class. This can be explained by analyzing the dataset. The majority of the outdoor images were shot under a near ideal illumination condition (clear sky without any color cast). Moreover, the outdoor illuminants do not present the same variability of the indoor illuminants. Looking at the parameters and results of the WP, SG, and gGW algorithms on the outdoor images, it can be seen that they behave identically as a WP algorithm. This can be explained considering that real outdoor images tend to exhibit color channel clipping in the high intensity range. These very bright pixels are taken as reference white by the three WP-like algorithms and this reference is often very close to the real scene illuminant.

Table IV summarizes the results on the 835 images (331 indoor and 504 outdoor) in the test set. The first column reports the results obtained by the algorithms using the class-independent strategy while the second reports those using the class-dependent parameterization strategy. For the sake of brevity, we omit the results obtained on the two classes. The errors of the GW and WP algorithms are the same for both strategies because their parameters are actually class independent. The results of the CI strategy are lower than those obtained on the training set. This is due to the fact that in the test set the outdoor class is more represented than the indoor class (504 versus 331 images). The introduction of a class-dependent parameterization, improved the results of all the algorithms (with the obvious exception of GW and WP). Four different algorithms (SG, GE2, N2M, and LMS) obtained a median error of about four degrees, which is better than the error obtained by the best class-independent algorithm.

In the last two experiments, we considered the application of different algorithms, on the basis of the classification outcome (namely the CDA and CDAUC strategies). According to the results obtained on the training set, the CDA strategy consists in applying the SG algorithm to the images classified as indoor, and the GE2 algorithm to those classified as outdoor. The CDAUC strategy consists in applying the same algorithms as the CDA but adding the uncertainty class for which we used the general purpose version of the LMS algorithm. The two thresholds of the CDAUC strategy were chosen using a five-fold cross validation approach on the 300 images of the training set. The final thresholds ($T_{\mathrm{INDOOR}} = 0.82$, $T_{\mathrm{OUTDOOR}} = 0.67$) are the average of the best thresholds found by each cross validation iteration. Using these values about 19.75% of the images in test set were classified as "uncertain." The results obtained on the test set are reported in Table V. They are compared with the performance of the best algorithms in the CI and CDP strategies (LMS and GE2, respectively). It can be noted that the median error decreases as more complex strategies are applied. However, according to the Wilcoxon Sign Test, it is not possible to claim that class dependent strategy is superior to class-dependent parameterization. Our results demonstrate that, at least on the dataset we considered, a classification based strategy which uses an uncertainty class outperforms a general purpose algorithm.

To determine the influence of the performance of the classifier on the final color contancy performance, we analyzed the results obtained with the CDA strategy considering correctly classified and misclassified images separately. On correctly classified indoor images the median angular error is 4.85; on indoor images misclassified as outdoor we obtained an error of 9.79. For outdoor images, the median errors on correctly classified and misclassified images are 2.31 and 5.07, respectively. So, image misclassification approximately doubles median angular errors. To avoid this decrease in performance an accurate classifier is crucial so that misclassifications occur rarely. To assess how much angular error may be improved using a better classifier, we compared our results with those of an "optimal" classifier (i.e., a classifier which correctly classifies all test images). By running the CDA strategy with the optimal classifier we obtained a median angular error of 3.48. We also considered a "random" classifier (i.e., a classifier which randomly misclassifies the images with a probability of 0.5) obtaining an error of 5.63. We can note that the results obtained using our classifier (3.78 of median angular error) are much closer to the results of the optimal classifier than to the random classifier ones. The results of this experiment give us the upper bound of 0.3 degrees of angular error to the improvement that could be achieved by adopting a more powerful classification methodology (Support Vector Machines [29], boosting [30], etc.).

We can summarize the results of our experimentation as follows.

- If no knowledge of the image content is exploited (CI strategy), combining methods perform better than the single ones.
- The algorithms that can be tuned on the basis of image contents benefit by the classification process.
- For the indoor class the SG (or equivalently the gGW) shows better results than the other methods. For outdoor class the best performance is obtained by the GE2 algorithm. For the specific classes, combining methods do not seem to be the best choice.
- When the same algorithm is used on both classes but with different parameters settings (Class-Dependent Parameterization strategy), there is not a single best algorithm. Four

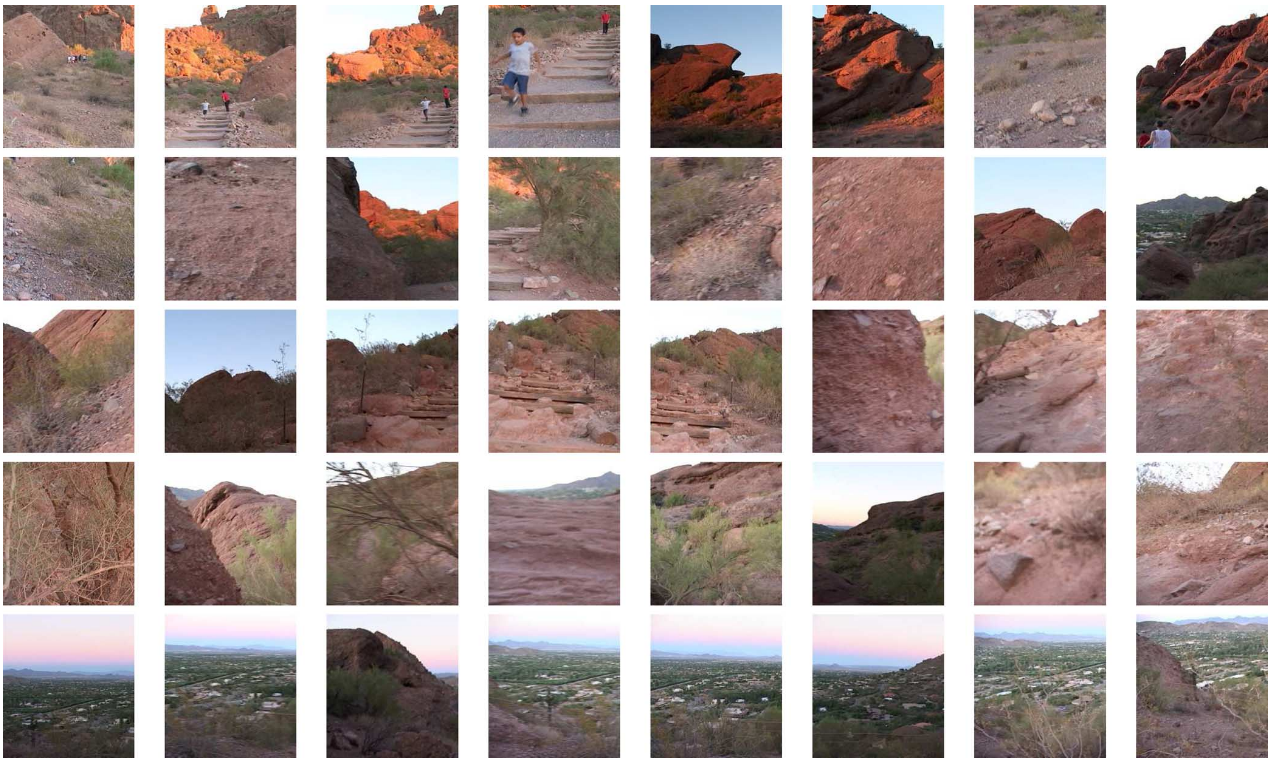

Fig. 3. The 40 key frames extracted from the "Camelback" video clip by the key frame extraction algorithm.

algorithms performed equally well obtaining the best result. Among these, being the less computational expensive, SG seems to be the best choice.

- Using different algorithms for indoor and outdoor images (Class-Dependent Algorithm strategy), improves the results with respect to the Class-Independent strategy. From our experiments, the best combination of algorithms consisted in the selection of the SG algorithm for the indoor images and GE2 for the outdoor images. We also observed a small improvement with respect to the Class-Dependent Parameterization strategy.
- The introduction of a third image class containing the images on which the classifier is not confident enough (CDAUC strategy), further improves the results. The algorithms selected in the CDAUC strategy are the SG for indoor images, GE2 for outdoor images and LMS (with the general purpose parameterization) for the other images. The improvement is statistically significant with respect to the other strategies considered.

V. Conclusion

We presented four strategies for automatic illuminant estimation using single algorithms or combination of algorithms. The strategies proposed can be used with any set of color constancy algorithms. In this work we used a set of algorithms derived from the framework recently proposed by Van de Weijer and Gevers for which a novel procedure is proposed to automatically tune the algorithms' parameters. We also used two combining frameworks proposed by Ciurea and Funt [6], and Bianco *et al.* [2].

In this paper, to improve illuminant estimation accuracy, an image classifier is trained to classify the images as indoor and outdoor, and different experimental framework are proposed to exploit this information in order to select the best performing algorithm on each class. The solutions investigated here included: an indoor/outdoor parameterization strategy according to which, given a color constancy algorithm, its parameters are set on the basis of the class predicted by the classifier; an indoor/outdoor algorithm selection strategy according to which the best algorithm (and its corresponding parameters) are set on the basis of the class predicted by the classifier; an indoor/uncertain/outdoor algorithm selection strategy which differs from the previous for the introduction of an uncertainty class. Images falling in the uncertainty class are processed by the algorithm that has proved to be the best class-independent algorithm.

We tested the strategies on a suitable subset of the widely used Funt and Ciurea dataset. To this end a method for extracting uncorrelated images from the dataset is used. Our results demonstrate that a classification based strategy which also uses an uncertainty class outperforms all the other strategies considered.

Future works will include the collection of a large dataset with high content and illuminant variability. Such a dataset would allow us to investigate if the introduction of additional classes may further improve the results. We also plan to experiment our strategies with other color constancy algorithms. For instance, having additional knowledge about the imaging device color

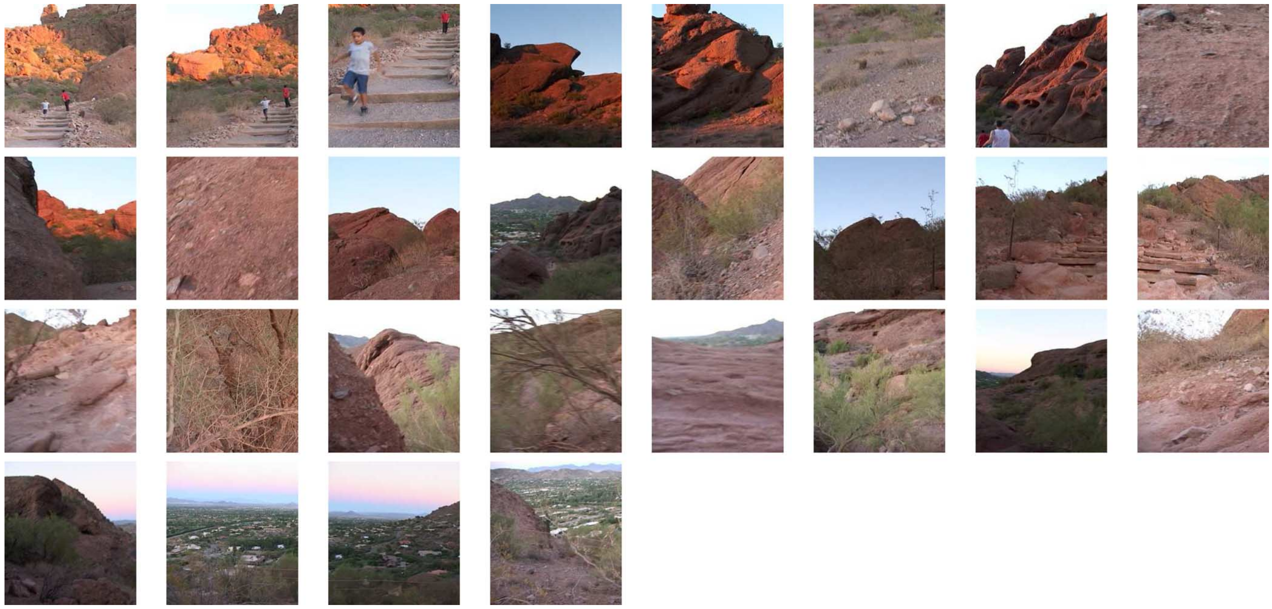

Fig. 4. "Camelback" images included in the final data set. Among the hierarchy of visual summaries generated by the visual summary post processing algorithm, the one containing exactly 28 key frames is selected.

TABLE VI
COMPOSITION OF THE IMAGES IN THE ORIGINAL 15 VIDEO CLIPS, NUMBER OF EXTRACTED KEY FRAMES AND NUMBER OF IMAGES WE INCLUDED IN OUR DATA SET

	Clip	# Frames	# Key frames	# Required frames
1	Apache	1,273	181	127
2	Burnabay	953	136	95
3	Camelback	276	40	28
4	CIC1	985	156	99
5	CIC2	406	75	41
6	CIC3	499	72	50
7	Deerlake	956	167	96
8	Fallcreek	708	114	71
9	Marine	513	82	51
10	Market1	555	86	56
11	Market2	1,098	144	110
12	Metrotown	1,313	206	131
13	Scottsdale	541	80	54
14	SFU	1,198	184	120
15	WhiteCliff	81	6	8 (6)
	TOTAL	11,355	1,729	1,135

gamut, it would be worth to investigate gamut-based color constancy [31], [32].

APPENDIX I
DETAILS ABOUT DATASET SELECTION

A video clip is reconstructed from each set of images removing the right part of the images containing the gray sphere. This video clip is then fed to a key frame extraction algorithm [22] which selects a set of candidate images. These images are dynamically selected within the video clip by analyzing the pictorial differences between consecutive images. The algorithm first identifies the shots present in the clip, i.e., uninterrupted sequences of images recorded with a single camera and usually taken in a single location. From each shot, a variable number of images (key frames) are extracted. The number of key frames depends on the visual contents of the shot. That is, shots showing high variability in their pictorial contents will have a high number of images extracted. Shots showing little or none variability will have only a single image extracted. The algorithm, by analyzing each shot independently, is able to remove most of the correlated images while preserving the overall structure and contents of the video clip. The set of key frames extracted corresponds to the visual summary of the clip. An example of key frames extracted from the "Camelback" clip is shown in Fig. 3.

The dataset to be used for testing the algorithms should be of significant size and without correlated images. As a trade-off between the number of images and the correlation problem, we decided to extract a number of images from each video clip corresponding to 10% of the size of the clip. To do this, we set the parameters of the key frames extraction algorithm so that the generated visual summary will contain at least that number of key frames. The same parameters were used for all the video clips. To select the exact number of required key frames for each clip, we processed the visual summary further with a visual summary postprocessing algorithm [23]. The algorithm is composed of three processing steps: key frame removal, key frame grouping and selection of the default visual summary. For this work, we exploited only the second processing step. All the key frames selected by the previous algorithm were processed and the final visual summary was selected according to the number of key frames required for that video clip. An example is shown in Fig. 4.

This algorithm is able to iteratively build a hierarchy of visual summaries. The set of initial key frames is processed using a hierarchical clustering algorithm that, at each step, merges consecutive key frames that are visually similar. Thus, the clustering algorithm further removes redundancies within the set of key frames. At each step a new visual summary is generated which contains one image less than the previous one until only a single key frame remains in the summary. The hierarchy can be used to

TABLE VII
LMS WEIGHT MATRIX FOR INDOOR IMAGES

Algorithm	Channel	weight R	weight G	weight B
GW	R	0.0141	0.0120	-0.0261
	G	-0.4729	0.1118	0.3611
	B	0.0000	0.0000	0.0000
WP	R	0.5074	-0.1092	-0.3857
	G	0.5066	-0.1094	-0.3848
	B	0.5138	-0.1102	-0.3911
SG	R	-1.3991	0.7291	0.6700
	G	0.0000	0.0000	0.0000
	B	-1.5044	0.7540	0.7504
gGW	R	-0.0075	0.0023	0.0051
	G	-0.0671	0.0053	0.0618
	B	0.0000	0.0000	0.0000
GE1	R	0.0000	0.0000	0.0000
	G	0.0336	0.0025	-0.0361
	B	-0.0055	0.0019	0.0037
GE2	R	0.8928	-0.6321	-0.2607
	G	0.0000	0.0000	0.0000
	B	1.0000	-0.6455	-0.3545

TABLE VIII
LMS WEIGHT MATRIX FOR OUTDOOR IMAGES

Algorithm	Channel	weight R	weight G	weight B
GW	R	1.0000	0.1234	-1.0763
	G	0.2328	0.0535	-0.2393
	B	0.9344	0.1038	-0.9912
WP	R	0.0000	0.0000	0.0000
	G	0.0652	0.0163	-0.0815
	B	-0.0495	-0.0083	0.0578
SG	R	0.0000	0.0000	0.0000
	G	-0.0268	-0.0101	0.0369
	B	0.0660	0.0044	-0.0704
gGW	R	0.0969	0.0180	-0.1149
	G	0.0000	0.0000	0.0000
	B	0.0644	0.0198	-0.0842
GE1	R	-0.4179	-0.0604	0.4784
	G	0.0000	0.0000	0.0000
	B	-0.3464	-0.0683	0.4146
GE2	R	-0.8044	-0.0813	0.8858
	G	0.0000	0.0000	0.0000
	B	-0.7718	-0.0640	0.8358

browse the clip contents at different levels of details, and we selected the one which contains the required number of key frames for the given sequence.

Table VI shows the statistics of each video clip: video name, number of images contained in each video clip, the number of key frames automatically extracted by the key frame extraction algorithm and the number of images required in our data set (corresponding to 10% of the images in each clip). It can be noted that the number of key frames extracted is very low compared to the number of images in the clip, even with the algorithm parameters set to deliberately extract a large number of key frames. This is an indication that the images within the video clip are truly highly correlated. For similar clips (e.g., CIC2 and CIC3) a similar number of key frames were extracted. Note also that for the "WhiteCliff" video, the key frames extracted are less than the required number of images. In this case, we kept all six images.

APPENDIX II
LMS WEIGHT MATRICES

We report the weight matrices corresponding to the indoor (Table VII), outdoor (Table VIII), and the general purpose (Table IX) instantiations of the LMS combining strategy, obtained as described in Section III-D.

Each row corresponds to the weights associated to one color channel of the illuminant estimated by a single input algorithm. Each column corresponds to one channel of the LMS output. The weights are scaled so that the maximum value is unitary for each matrix.

Note that, since the sum of each input and output illuminant estimation is unitary, the characteristic matrix of the linear regression system is singular. This fact explains the rows of zeroes in the three matrices.

TABLE IX
LMS WEIGHT MATRIX FOR THE GENERAL PURPOSE CASE

Algorithm	Channel	weight R	weight G	weight B
GW	R	0.8051	0.2636	-0.8313
	G	1.0000	0.5148	-1.2774
	B	0.1353	0.3456	-0.2435
WP	R	0.0000	0.0000	0.0000
	G	0.0054	0.0271	-0.0325
	B	0.0818	0.0035	-0.0853
SG	R	0.0000	0.0000	0.0000
	G	-0.5722	-0.3932	0.9655
	B	0.8567	-0.1032	-0.7535
gGW	R	-0.3006	-0.0307	0.3313
	G	-0.3124	0.0157	0.2967
	B	0.0000	0.0000	0.0000
GE1	R	0.0000	0.0000	0.0000
	G	-0.0433	-0.0418	0.0852
	B	-0.5214	-0.0696	0.5909
GE2	R	-0.3006	-0.1612	0.4618
	G	0.0000	0.0000	0.0000
	B	-0.6057	-0.1300	0.7358

ACKNOWLEDGMENT

The authors would like to thank Prof. B. Funt for making the data available for our experiments. Researchers interested in testing their methods on the same selection of the dataset used in this paper may contact the authors who will gladly provide the list of images used.

REFERENCES

[1] S. D. Hordely, "Scene illuminant estimation: Past, present, and future," *Color Res. Appl.*, vol. 31, no. 4, pp. 303–314, 2006.
[2] S. Bianco, F. Gasparini, and R. Schettini, "A consensus based framework for illuminant chromaticity estimation," *J. Electron. Imag.*, vol. 17, pp. 023013-1–023013-9, 2008.
[3] M. Schröder and S. Moser, "Automatic color correction based on generic content-based image analysis," in *Proc. Color Imaging Conf. 9*, 2001, pp. 41–45.
[4] F. Gasparini and R. Schettini, "Color balancing of digital photos using simple image statistics," *Pattern Recognit.*, vol. 37, no. 6, pp. 1201–1217, 2004.

[5] J. van de Weijer, C. Schmid, and J. Verbeek, "Using high-level visual information for color constancy," in *Proc. IEEE 14th Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[6] F. Ciurea and B. Funt, "A large image database for color constancy research," in *Proc. IS&T/SID 11th Color Imag. Conf.*, 2003, pp. 160–164.

[7] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics," in *Proc. Int. Conf. Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[8] M. Szummer and R. Picard, "Indoor-outdoor image classification," in *Proc. Int. Workshop Content-Based Access of Image and Video Databases*, 1998, pp. 42–51.

[9] B. Funt, K. Barnard, and L. Martin, "Is machine colour constancy good enough?," in *Proc. 5th Eur. Conf. Comput. Vis.*, 1998, pp. 445–459.

[10] A. Vailaya, M. Figueiredo, A. K. Jain, and H.-J. Zhang, "Image classification for content-based indexing," *IEEE Trans. Image Process.*, vol. 10, no. 1, pp. 117–130, Jan. 2001.

[11] R. Schettini, C. Brambilla, C. Cusano, and G. Ciocca, "Automatic classification of digital photographs based on decision forests," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 18, no. 5, pp. 819–845, 2004.

[12] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. New York: Wadsworth and Brooks/Cole, 1984.

[13] R. Schettini, C. Brambilla, C. Cusano, and G. Ciocca, "On the detection of pornographic digital images," in *Proc. Visual Communications and Image Processing*, 2003, vol. 5150, SPIE, pp. 2105–2113.

[14] R. Schettini, C. Brambilla, G. Ciocca, A. Valsasna, and M. De Ponti, "A hierarchical classification strategy for digital documents," *Pattern Recognit.*, vol. 35, no. 8, pp. 1759–1769, 2002.

[15] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, pp. 123–140, 1996.

[16] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based Color Constancy," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2207–2214, Sep. 2007.

[17] G. Buchsbaum, "A spacial processor model for object color perception," *J. Franklin Inst.*, vol. 310, pp. 1–26, 1980.

[18] V. Cardei, B. Funt, and K. Barnard, "White point estimation for uncalibrated images," in *Proc. IS&T/SID 7th Color Imaging Conf.*, 1999, pp. 97–100.

[19] G. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. IS&T/SID 12th Color Imaging Conf.*, 2004, pp. 37–41.

[20] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms; part two: Experiments with image data," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 985–996, Sep. 2002.

[21] V. C. Cardei and B. Funt, "Committee-based colour constancy," in *Proc. IS&T/SID 7th Color Imag. Conf.*, 1999, pp. 311–313.

[22] G. Ciocca and R. Schettini, "An innovative algorithm for key frame extraction in video summarization," *J. Real-Time Image Process.*, vol. 1, no. 1, pp. 69–88, 2006.

[23] G. Ciocca and R. Schettini, "Supervised and unsupervised classification post-processing for visual video summaries," *IEEE Trans. Consum. Electron.*, vol. 2, no. 2, pp. 630–638, May 2006.

[24] S. D. Hordley and G. D. Finlayson, "Re-evaluating color constancy algorithms," in *Proc. 17th Int. Conf. Pattern Recognit.*, 2004, pp. 76–79.

[25] R. M. Lewis and V. Torczon, "Pattern search algorithms for bound constrained minimization," *SIAM J. Optim.*, vol. 9, pp. 1082–1099, 1999.

[26] R. M. Lewis and V. Torczon, "Pattern search methods for linearly constrained minimization," *SIAM J. Optim.*, vol. 10, pp. 917–941, 2000.

[27] R. M. Lewis and V. Torczon, "On the convergence of pattern search algorithms," *SIAM J. Optim.*, vol. 7, pp. 1–25, 1997.

[28] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, pp. 80–83, 1945.

[29] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[30] R. E. Shapire, "A brief introduction to boosting," in *Proc. 16th Int. Joint Conf. Artificial Intelligence*, 1999, pp. 1401–1406.

[31] G. D. Finlayson, "Color in perspective," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 1034–1038, Oct. 1996.

[32] 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.*, vol. 23, no. 11, pp. 1209–1221, Nov. 2001.

Simone Bianco received the B.Sc. and M.Sc. degrees in mathematics from the University of Milan-Bicocca, Italy, in 2003 and 2006, respectively. He is currently pursuing the Ph.D. degree in computer science in the Dipartimento di Informatica, Sistemistica e Comunicazione (DISCo), University of Milano-Bicocca, working on image processing.

The main topics of his current research concern digital still cameras processing pipelines, color space conversions, optimization techniques, and characterization of imaging devices.

Gianluigi Ciocca received the degree (Laurea) in computer science from the University of Milan, Italy, in 1998.

He is a Fellow with the Institute of Multimedia Information Technologies, Italian National Research Council, where his research has focused on the development of systems for the management of image and video databases and the development of new methodologies and algorithms for automatic indexing. He is currently an Assistant Professor in computer science at the Dipartimento di Informatica, Sistemistica e Comunicazione (DISCo), University of Milano-Bicocca, working on video analysis and abstraction.

Claudio Cusano received the Ph.D. degree in computer science from the Department of Information Science, Systems Theory, and Communication (DISCo), University of Milano-Bicocca, Italy.

He is currently a postdoctorate student at DISCo. Since April 2001, he has been a Fellow of the the ITC Institute of the Italian National Research Council. The main topics of his current research concern 2-D and 3-D imaging, with a particular focus on image analysis and classification and face recognition.

Raimondo Schettini is an Associate Professor at the Department of Information Science, Systems Theory, and Communication (DISCo), University of Milano Bicocca, Italy, where he is in charge of the Imaging and Vision Lab. He has been associated with Italian National Research Council (CNR) since 1987. He has been team leader in several research projects and has published more than 160 refereed papers in image processing, analysis and reproduction, and image content-based indexing and retrieval.

Dr. Schettini is an Associate 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 General Co-Chairman of the First Workshop on Image and Video Content-Based Retrieval (1998), the First European Conference on Color in Graphics, Imaging and Vision (2002), and the EI Internet Imaging Conferences (2000–2006).