Color Constancy using Natural Image Statistics and Scene Semantics

Arjan Gijsenij, *Member, IEEE,* Theo Gevers, *Member, IEEE,*

Abstract—Existing color constancy methods are all based on specific assumptions such as the spatial and spectral characteristics of images. As a consequence, no algorithm can be considered as universal. However, with the large variety of available methods, the question is how to select the method that performs best for a specific image.
To achieve selection and combining of color constancy algorithms, in this paper, natural image statistics are used to identify the most important characteristics of color images. Then, based on these image characteristics, the proper color constancy algorithm (or best combination of algorithms) is selected for a specific image. To capture the image characteristics, the Weibull parameterization (*e.g.* grain size and contrast) is used. It is shown that the Weibull parameterization is related to the image attributes to which the used color constancy methods are sensitive to. A MoG-classifier is used to learn the correlation and weighting between the Weibull-parameters and the image attributes (number of edges, amount of texture and SNR). The output of the classifier is the selection of the best performing color constancy method for a certain image.
Experimental results show a large improvement over state-of-the-art single algorithms. On a data set consisting of more than 11,000 images, an increase in color constancy performance up to 20% (median angular error) can be obtained compared to the best-performing single algorithm. Further, it is shown that for certain scene categories, one specific color constancy algorithm can be used instead of the classifier considering several algorithms.

Index Terms—Color Constancy, Illuminant Estimation, Natural Image Statistics, Scene Semantics, Computer Vision.

1 Introduction

Differences in illumination cause measurements of object colors to be biased towards the color of the light source. Fortunately, humans have, to some extent, the ability of color constancy: they perceive the same color of an object despite large differences in illumination [1, 2]. Various computer vision related topics like human-computer interaction [3], color feature extraction [4, 5] and color appearance models [6] would benefit from a similar ability.

Many computational color constancy algorithms have been proposed, see [7, 8] for recent overviews. In general, color constancy algorithms can be divided into two groups. The first group consists of algorithms based on low-level image features that can directly be applied to images. One of the first color constancy methods is based on the Retinex-theory by Land and McCann [9]. Examples of algorithms which are derived from this theory include the White-Patch algorithm [10], the Grey-World algorithm [11], and more recently the Shades-of-Grey algorithm [12] and the Grey-Edge algorithm [13].

The second group consists of algorithms that use information acquired in a learning phase to obtain knowledge about the images, like possible light sources and the distribution of possible reflectance colors to be present in natural scenes. This information is then used to estimate the illuminant. One of the first algorithms of this type is the gamut mapping algorithm by Forsyth [14]. This algorithm is based on the assumption that in real-world images, for a given illuminant, only a limited number of colors can be observed. Using this assumption, the illuminant can be estimated by comparing the distribution of colors in the current image to a pre-learned distribution of colors (called the canonical gamut). Many algorithms have been derived from the original algorithm including Color-by-Correlation [15], the Gamut constrained illuminant estimation [16] and derivative-based gamut mapping [17]. Other approaches that use a learning phase include probabilistic methods [18, 19] and methods based on genetic algorithms [20].

Because the color constancy problem is an under-constrained problem (see for example [14, 15, 8]), all of the above color constancy methods are based on specific assumptions. These assumptions include constrained gamuts (limited number of image colors which can be observed under a specific illuminant) and the distribution of colors that are present in an image (*e.g.* white patch, grey-world, grey-edge etc.). As a consequence of the use of different assumptions, no color constancy algorithm can be considered as universal. With the large variety of available methods, the problem is how to select the method that is most suitable for different imaging setting and scenes. Furthermore, the subsequent question is how to combine the different algorithms in a proper way.

Little research has been published on the selection and fusion of color constancy methods. In [21], fusion is performed by a weighted average of several methods. More recently, a statistics-based method is combined

- *Both authors are with the Intelligent Systems Lab Amsterdam (ISLA), University of Amsterdam, The Netherlands.*
  E-mail: {a.gijsenij, th.gevers}@uva.nl
- *Source-code available at http://www.science.uva.nl/∼gijsenij*

with a physics-based method [22]. Both methods are based on weighting the output of the used color constancy algorithms, where the weights are optimized for a specific data set. However, the combination of the used algorithms still depends on the type of images being processed.

Therefore, in this paper, the most appropriate color constancy algorithm for an image is selected based on the statistical contents of this image. For instance, even though edge-based color constancy algorithms, *on average*, are preferred over pixel-based color constancy algorithms [13, 17], this does not imply that edge-based color constancy algorithms are the best methods for *all* images. If only few edges are present in an image, i.e. edge-based methods only have little information to work with, then it is likely that pixel-based methods will perform better. To this end, the Weibull parameterization is used to express these image characteristics in terms of grain size (texture) and contrast. Then, based on these image characteristics, the proper color constancy algorithm (or best combination of algorithms) is selected for a specific image. As Weibull distributions are derived from higher-order image statistics (*i.e.* image derivatives), the choice of a proper set of different color constancy methods should support this. To this end, the color constancy framework proposed by [13] is used. This framework incorporates higher-order statistics (*e.g.* first and second order derivatives). Further, it allows us to generate different color constancy algorithms in a systematic way.

This paper is organized as follows. First, in section 2, color constancy based on low-level image features is discussed. In section 3, the concepts of natural image statistics and scene semantics are provided. In section 4, several approaches are given to combine the different color constancy algorithms, and in section 5 the methods are evaluated on a large data set containing over 11000 images. Finally, in section 6 the results and contributions of the paper are discussed in more detail.

2 Color Constancy

In this paper, color constancy is defined as correcting an image that was taken under an unknown light source, so that it appears to be taken under a canonical (often white) light source. This is computed by first estimating the color of the light source, followed by a transformation of the original image (*e.g.* [8]). To model this process, in section 2.1, the image formation model is considered first. Then, in section 2.2, several algorithms to estimate the color of the light source are presented. Finally, in section 2.3, the model to transform images from one light source to another is discussed.

2.1 Reflection model

The image values $\mathbf{f} = (f_R, f_G, f_B)^T$ for a Lambertian surface depend on the color of the light source $I(\lambda)$, the surface reflectance $S(\mathbf{x}, \lambda)$ and the camera sensitivity function $\boldsymbol{\rho}(\lambda) = (\rho_R(\lambda), \rho_G(\lambda), \rho_B(\lambda))^T$, where λ is the wavelength of the light and $\mathbf{x}$ is the spatial coordinate (*e.g.* [8, 14, 15]):

$$f_c(\mathbf{x}) = m(\mathbf{x}) \int_\omega I(\lambda)\rho_c(\lambda)S(\mathbf{x}, \lambda)d\lambda, \tag{1}$$

where ω is the visible spectrum, $m(\mathbf{x})$ is Lambertian shading and $c = \{R, G, B\}$. It is assumed that the scene is illuminated by one light source and that the observed color of the light source $\mathbf{e}$ depends on the color of the light source $I(\lambda)$ as well as the camera sensitivity function $\boldsymbol{\rho}(\lambda)$:

$$\mathbf{e} = \begin{pmatrix} e_R \\ e_G \\ e_B \end{pmatrix} = \int_\omega I(\lambda)\boldsymbol{\rho}(\lambda)d\lambda. \tag{2}$$

Color constancy can be achieved by estimating the color of the light source $\mathbf{e}$, given the image values of $\mathbf{f}$, followed by a transformation of the original image values using this illuminant estimate. This transformation will leave the intensity of every pixel unaltered, as the proposed method will only correct for the chromaticity of the light source. More information on this transformation can be found in section 2.3. Since both $I(\lambda)$ and $\boldsymbol{\rho}(\lambda)$ are, in general, unknown, the estimation of $\mathbf{e}$ is an under-constrained problem that cannot be solved without further assumptions. Therefore, in practice, color constancy algorithms are based on simplifying assumptions such as restricted gamuts (limited number of image colors which can be observed under a specific illuminant), the distribution of colors that are present in an image (*e.g.* white patch, grey-world etc.) and the set of possible light sources. In this paper, the focus is on the distribution of colors that are present in an image as the major assumption. In the next section, a framework is discussed generating different color constancy methods where each method is based on a specific assumption about the presence of colors and color edges in images.

2.2 Illuminant estimation

Two well-established color constancy algorithms, using pixel values, are based on the Retinex Theory proposed by Land and McCann [9]. The White-Patch algorithm [10] is based on the White-Patch assumption, *i.e.* the assumption that *the maximum response in the RGB-channels is caused by a white patch*. The Grey-World algorithm [11] is based on the Grey-World assumption, *i.e. the average reflectance in a scene under a neutral light source is achromatic*. In [12], these two algorithms are proven to be important instantiations of the Minkowski-norm:

$$\mathcal{L}_c(p) = \left(\int f_c^p(\mathbf{x})d\mathbf{x} \right)^{\frac{1}{p}} = ke_c, \tag{3}$$

where $c = \{R, G, B\}$ and k is a multiplicative constant chosen such that the illuminant color, $\mathbf{e} = (e_R, e_G, e_B)^T$, has unit length. When $p = 1$ is substituted, equation (3) is equivalent to computing the average of $\mathbf{f}(\mathbf{x})$, *i.e.*

$\boldsymbol{\mathcal{L}}(1) = (\mathcal{L}_R(1), \mathcal{L}_G(1), \mathcal{L}_B(1))^T$ equals the Grey-World algorithm. When $p = \infty$, equation (3) results in computing the maximum of $\mathbf{f}(\mathbf{x})$, *i.e.* $\boldsymbol{\mathcal{L}}(\infty)$ equals the White-Patch algorithm. In general, to arrive at a proper value, p is tuned for the data set at hand. Hence, the value of this parameter may vary for different data sets.

The assumptions of the above color constancy methods are based on the distribution of colors (*i.e.* pixel values) that are present in an image. The incorporation of higher-order image statistics (in the form of image derivatives) is proposed in [13], where a framework is presented that incorporates the well-known methods like equation (3), as well as methods based on first- and second-order statistics:

$$\left(\int \left| \frac{\partial^n f_{c,\sigma}(\mathbf{x})}{\partial \mathbf{x}^n} \right|^p d\mathbf{x} \right)^{\frac{1}{p}} = k e_c^{n,p,\sigma}, \tag{4}$$

where $|\cdot|$ indicates the Frobenius norm, $c = \{R, G, B\}$, n is the order of the derivative and p is the Minkowski-norm. Further, the derivative is defined as the convolution of the image with the derivative of a Gaussian filter with scale parameter σ [23]:

$$\frac{\partial^{s+t} f_{c,\sigma}}{\partial x^s \partial y^t} = f_c * \frac{\partial^{s+t} G_\sigma}{\partial x^s \partial y^t} \tag{5}$$

where $*$ denotes the convolution and $s + t = n$.

Using equation (4), many different color constancy algorithms can be generated. For instance, algorithms based on zeroth-order statistics like Grey-World, White-Patch and general Grey-World can be generated by substituting $n = 0$:

1) $\mathbf{e}^{0,1,0}$ ($\equiv \boldsymbol{\mathcal{L}}(1)$) is equivalent to the Grey-World algorithm;
2) $\mathbf{e}^{0,\infty,0}$ ($\equiv \boldsymbol{\mathcal{L}}(\infty)$) is equivalent to the White-Patch algorithm;
3) $\mathbf{e}^{0,p,\sigma}$ is called the general Grey-World algorithm, where the values for p and σ are dependent on the type of images that are in the data set. In [13], it is shown that for real-world images, $p = 13$ and $\sigma = 2$ are found to produce proper results for a real-world data set, *i.e.* $\mathbf{e}^{0,13,2}$.

Equation (4) extends these instantiations to higher-order statistics. For instance, when taking the first-order or second-order derivative, the values for the Minkowski-norm and the smoothing parameter are used to produce different algorithms:

4. $\mathbf{e}^{1,p,\sigma}$ is the first-order Grey-Edge. The basic assumption here is that the average edge (*i.e.* derivative) is grey. The Minkowski-norm p and the smoothing parameter σ are dependent on the images in the data set. In [13], $p = 1$ and $\sigma = 6$ provide good results on real-world data, *i.e.* $\mathbf{e}^{1,1,6}$.
5. $\mathbf{e}^{2,p,\sigma}$ is the second-order Grey-Edge. The basic assumption here is that the average second-order derivative is grey. Again, the Minkowski-norm p and smoothing parameter σ depend on the data set used, where for a real-world data set $p = 1$ and $\sigma = 5$ provide good results [13], *i.e.* $\mathbf{e}^{2,1,5}$.

In conclusion, a wide variety of color constancy algorithms are obtained, corresponding to different instantiations of equation (4), where each color constancy method has its own basic assumption about the distribution of color values and edges in the image.

2.3 Diagonal model

The focus of this paper is on estimating the color of the light source. However, in many cases the color of the light source is of less importance than the appearance of the input image under a reference light (called canonical light source). Therefore, the aim of most of the color constancy methods is to transform all colors of the input image, taken under an unknown light source, to colors as they appear under this canonical light source. This transformation can be considered to be an instantiation of chromatic adaptation, *e.g.* [6]. Chromatic adaptation is often modeled using a linear transformation, which in turn can be simplified to a diagonal transformation when certain conditions are met [24, 25, 26]. Other possible chromatic adaptation methods include linearized Bradford [27] and CIECAM02 [28].

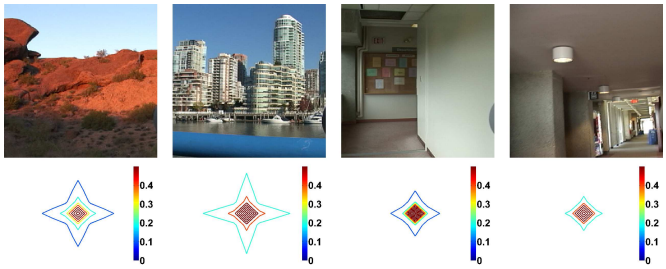
In this paper, the diagonal transform or *von Kries Model* [29] is used, without changing the color basis [30] or applying spectral sharpening [25]. These latter techniques are shown to be able to improve the quality of the output image with respect to the diagonal model, *i.e.* if the color of the light source is known, then these modified algorithms result in more realistic images than the diagonal model. For simplicity, these steps are omitted in this paper, so the diagonal model that is used is given by:

$$\mathbf{f}_t = \mathcal{D}_{u,t} \mathbf{f}_u, \tag{6}$$

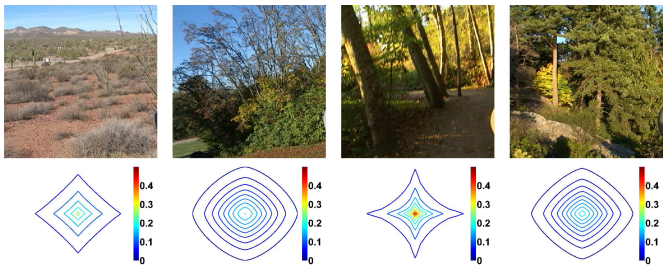
where $\mathbf{f}_u$ is the image taken under an unknown light source, $\mathbf{f}_t$ is the same image transformed, so it appears if it was taken under the canonical illuminant, and $\mathcal{D}_{u,t}$ is a diagonal matrix which maps colors that are taken under an unknown light source u to their corresponding colors under the canonical illuminant c:

$$\begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix} = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} R_u \\ G_u \\ B_u \end{pmatrix} \tag{7}$$

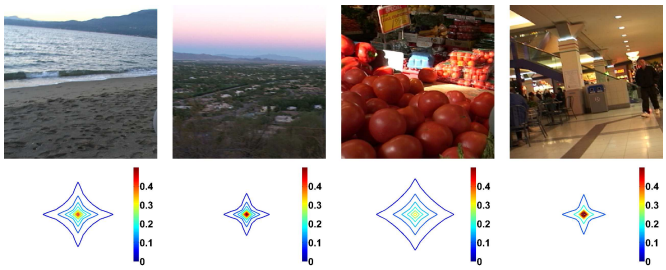
Even though this model is merely an approximation of illuminant change and might not accurately be able to model photometric changes due to disturbing effects like highlights and interreflections, it is widely accepted as color correction model [24, 25, 26] and it underpins many color constancy algorithms (*e.g.* the gamut mapping [14] and the used framework of methods [13]). The diagonal mapping is used throughout this paper to create output-images after correction by a color constancy algorithm, where a perfect white light, *i.e.* $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$, is used as canonical illuminant.



(a) 0^{th}-order method



(b) First-order method



(c) Second-order method

Fig. 1. Examples of images that can be considered to be characteristic for the corresponding color constancy algorithms, *i.e.* the corresponding color constancy algorithm will perform best on these type of images. Below each image, the distribution is plotted of edges in the intensity channel. The images come from the data set published in [31].

3 Natural Image Statistics and Scene Semantics

All methods that are comprised in the used color constancy framework [13] are based on assumptions on the distribution of colors (edges) that are present in an image. For instance, the Grey-World algorithm assumes that the average color in a scene taken under a neutral light source is achromatic, while the Grey-Edge algorithm assumes that the average *edge* is achromatic. It has also been shown that the incorporation of spatial dependencies between colors (*e.g.* edges) produce more constrained gamuts improving the accuracy of color constancy in general [17]. This means that the set of possible adjacent color values (*i.e.* color edges) in real-world images is more restricted than the set of possible pixel values. Hence, the use of local spatial information will provide more stable gamuts than pixel values to compute color constancy. Furthermore, a higher accuracy is obtained when there is a large variety of edges in a scene [17]. The same observation is valid for the Grey-World algorithm in terms of the number of different, *e.g.* [32, 33].

Hence, color constancy methods are largely dependent on the distribution of colors and color edges in an image. In the next section, image statistics are used to describe these distributions.

3.1 Spatial image structures

Image structures are valuable identification cues in determining which type of scene the image is taken from. In [34], the authors show that the power spectrum (distribution of edge responses) of an image is characteristic for the type of scene. Further, in [35], it is shown that this distribution of edge responses can be modeled by a Weibull distribution. In the context of scene classification, features derived from the power spectrum and Weibull distributions have successfully been applied [36, 35, 37]. In this article, we focus on modeling natural image statistics using the *two parameter integrated Weibull distribution* [35]:

$$w(x) = C \exp\left(-\frac{1}{\gamma}\left|\frac{x}{\beta}\right|^{\gamma}\right), \tag{8}$$

where x is the edge responses in a single color channel to the Gaussian derivative filter, C is a normalization constant, $\beta > 0$ is the scale parameter of the distribution and $\gamma > 0$ is the shape parameter. The parameters of this distribution are indicative for the edge statistics of an (natural) image. In fact, the *contrast* of the image is indicated by β (*i.e.* the width of the distribution), and the *grain size* by γ (*i.e.* the peakedness of the distribution). Hence, a higher value for β indicates more contrast, while a higher value for γ indicates a smaller grain size (more fine textures).

To fit the Weibull distribution, edge responses are computed by a Gaussian derivative filter. There exists a high correlation between the Weibull parameters that are fitted through the distribution of edges for the first derivative, second derivative and third derivative. Hence, a single filter type, although measured in different orientations, is sufficient to assess the spatial statistics of images [35].

In figure 1, examples are shown of images with their corresponding edge distributions which are approximated by a Weibull-fit. The intensity channel is chosen for the ease of illustration because a 6-dimensional edge distribution (*i.e.* β and γ for each R, G and B channel) is hard to visualize. The edge distributions and corresponding Weibull-fits computed for separate color channels show similar plots. The images are examples on which the different color constancy algorithms using the corresponding type of information (*i.e.* pixel-values, edges or second-order transitions) performs best (based on the angular error that is explained in section 5).

The relationship between the images in figure 1 and their corresponding color constancy algorithm becomes clear from the edge distributions that are shown together with the images in figure 1. Pixel-based algorithms (*i.e.* 0^{th}-order) perform better than higher-order methods (*i.e.* 1^{st}- and 2^{nd}-order) on images with only little texture. This reflects in an edge distribution that densely sampled around the origin, *i.e.* many edges with little or zero energy. Higher-order methods require more edge information for an accurate illuminant estimate, which is reflected in an edge distribution that is less sharply peaked. Important to note at this stage is the semantic similarity between images that correspond to the same type of color constancy algorithms. For instance, forest-like scenes show a similar edge distribution in figure 1(b) and are all best solved by a 1^{st}-order color constancy algorithm. Hence, scene semantics can steer the process of color constancy. Natural image statistics and scene semantics will therefore be used in the next sections to achieve a proper selection of color constancy algorithms.

4 Combination of Illuminant Estimation Methods

In this section, a novel strategy is proposed based on natural image statistics to select the color constancy method which performs best for a specific image. To combine and compare different fusion strategies, in section 4.1, a basic approach is discussed based on using the output of multiple algorithms. Then, in section 4.2, natural image statistics are used to identify the most important characteristics of color images. Based on these image characteristics, the proper color constancy algorithm is selected for a specific image. Finally, scene semantics are used to find a category-specific combination of color constancy algorithms.

4.1 Color Constancy using Standard Fusion

When using the output of multiple algorithms to generate a new estimate of the illuminant, the simplest method is to take the average of the estimates over all algorithms. A straightforward extension is to take the weighted average of the estimated illuminants. If n algorithms are combined, then the weighted average is defined as:

$$\overline{\mathbf{e}} = \sum_{i=1}^{n} w_i \mathbf{e}_i, \tag{9}$$

where $\sum_{i=1}^{n} w_i = 1$. The average is just a special instance of the weighted average: $w_1 = w_2 = ... = w_n$. The estimates can also be combined using a non-linear committee. However, in [21], it is shown that a non-linear neural network did not produce better results than the weighted average. In fact, the weighted average outperformed a multi-layer Perceptron neural network.

In [22], two algorithms (one statistics-based and one physics-based algorithm) were combined using a similar approach. However, the output of the two used algorithms are somewhat different than the output of a general color constancy algorithm. Both methods produce a vector of probabilities, where each element represents the probability that the corresponding illuminant is the illuminant that was used to create the current image. In the combination-phase, the weighted average of these two vectors is determined, after which the illuminant with the highest probability is selected to be correct. Since this method requires the output of the color constancy algorithms to comply to a specific (irregular) form, this approach is not further evaluated here.

4.2 Color Constancy using Natural Image Statistics

In section 3, the Weibull distribution is considered as the parameterization of the edge distribution of images. Several characteristics, like the number of edges and the amount of texture and contrast are captured by this parameterization *i.e.* β and γ. In this section, it is proposed to select different color constancy methods based on these statistics. In previous work [39], it is shown that applying the k-means clustering on the Weibull-features, combined with a Gaussian weighting function provides proper color constancy. In this paper, the k-means approach is generalized to a probabilistic approach, corresponding to a maximum likelihood classifier based on mixture of Gaussians (MoG). This provides a more principled and probabilistic basis than k-means to relate natural image statistics with color constancy.

This novel algorithm aims at combining the estimates of several color constancy algorithms into a single more accurate estimate. To be precise, let M be the set of algorithms that are to be combined, where M_i denotes algorithm i. Further, the accuracy of the estimate of algorithm i on image j (*i.e.* the performance of algorithm i on image j) is denoted by $\epsilon_i(j)$. The algorithm consists of the following steps.

- First, the image statistics $\omega \in \mathbb{R}^{p \times q}$ for all images are computed, where p is the number of features that are computed and q is the number of images, *i.e.* ω_{ij} is the i^{th} feature of the j^{th} image. For simplicity, the subscript i is omitted, so ω_j denotes the feature vector representing the image statistics of the j^{th} image.
- Then, all images that are in the training set are labelled. The label y_j of an image j is derived using the performance of the algorithms on image j:

$$y_j = \arg\min_i \{\epsilon_i(j)\} \tag{10}$$

- Apply the MoG-classifier [40] on the training data. The likelihood of the observed image statistics ω_j for image j given color constancy algorithm y_j is computed as a weighted sum of k Gaussian distributions:

$$p(\omega_j|y_j) = \sum_{m=1}^{k} \alpha_m G(\omega_j, \mu_m, \Sigma_m) \tag{11}$$

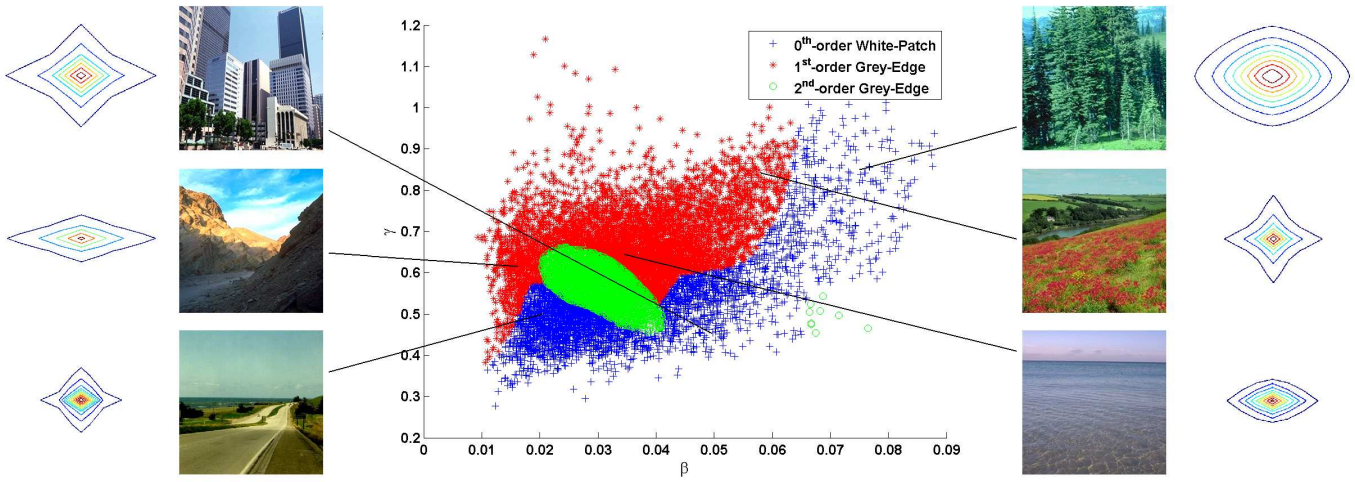


Fig. 2. A scatterplot of the β and γ of the gradient in the O_3-channel. Each point represents the Weibull-parameters of one image. The parameters of more than $11,000$ images (the real-world set by Ciurea and Funt [31]) are plotted . The differently colored parts in the graphs represent clusters of images that are best solvable by a single color constancy algorithm. The images are merely used as illustration, and are taken from [38].

Here, α_m are positive weights of the Gaussian components (with mean and variance defined as μ_m and Σ_m, respectively) such that $\sum_{m=1}^{k} \alpha_m = 1$. The parameters of the model are learned through training using the Expectation-Maximization algorithm.

- Apply the learned MoG-classifier on the test data, and assign to the current image j the algorithm that maximized the posterior probability.

Weibull-parameters can be computed for each R, G and B-channel separately. However, these color channels are highly correlated [41]. Therefore, the image is first transformed to a decorrelated color space before computing these parameters. To this end, the opponent color space is used:

$$O_1 = \frac{R-G}{\sqrt{2}} \quad (12)$$

$$O_2 = \frac{R+G-2B}{\sqrt{6}} \quad (13)$$

$$O_3 = \frac{R+G+B}{\sqrt{3}} \quad (14)$$

The selection of the most appropriate color constancy algorithm for the current image is done by computing the maximum posterior probability of the classifier. The corresponding color constancy algorithm is selected for the current image. The other algorithms are ignored.

In figure 2, an example of the output of the algorithm is shown. For this example, the Weibull-features representing the amount of texture and contrast (*i.e.* γ and β, respectively) that are computed using the gradient of the intensity-channel are used, along with three different color constancy algorithms: a zeroth-order algorithm ($\mathbf{e}^{0,\infty,0}$), a first-order algorithm ($\mathbf{e}^{1,1,1}$) and a second-order algorithm ($\mathbf{e}^{2,2,1}$). Again intensity (*i.e.* O_3) is used for the ease of illustration (instead of the 6-dimensional space *i.e.* β and γ for O_1, O_2 and O_3). Note that the final classifier is computed based on the complete 6-dimensional feature space. In this figure, the different color constancy methods that are assigned to the images are visualized by different colors. It can be observed that images on which a certain algorithm *generally* performs well are grouped together.

4.3 Color Constancy using Scene Semantics

Natural image statistics are known to provide identification cues for the classification of different types of scenes like forest, coast and street [34]. Further, in [42] it is shown that using high-level visual information in the form of class-specific information improves the accuracy of the estimated illuminant. Van de Weijer et al. [42] assume that an image can be modeled as a mixture of semantic classes. The information on the different classes that are present in an image is used to estimate the color of the light source. In this section, we aim at using scene semantics to find a category-specific combination of color constancy algorithms that optimizes the performance of the illuminant estimation.

Before discussing the category-specific combination method, the selection of the most appropriate color constancy algorithm based on scene semantics is examined first. In [38], a data set is provided consisting of eight urban and natural scene categories (*e.g.* Coast & Beach, Open country, Forest, Mountain, Highway, Street, City center and Tall building). The corresponding Weibull-parameters of the images of a selection of these categories is shown in figure 3, along with the Weibull-parameters of the images that are derived from the real-world data set [31]. It can be observed that images from the same category have similar edge distributions, resulting in similar Weibull-parameters.

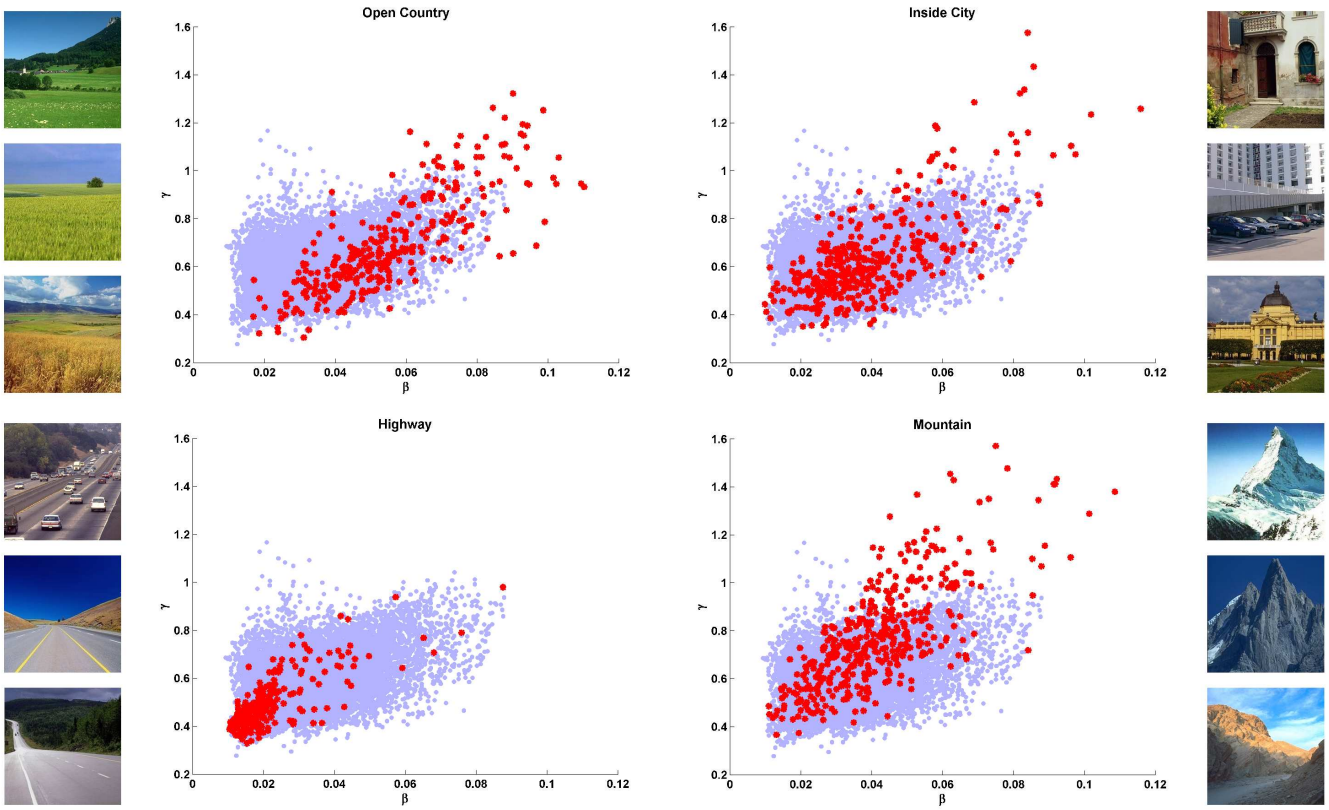


Fig. 3. Scatter plots of the Weibull-parameters based on O_3 derived from images coming from several categories (defined in [38]), overlayed on the Weibull-parameters of the images that are in the real-world set [31]. From this plot, it can be seen that images from the same scene category have similar Weibull-parameters.

Some categories have a larger variance in edge distribution than others. For instance, most of the images of the category *Highway* have a low value for β and a low value for γ, indicating a low contrast and few edges. Images of the category *Mountain*, on the other hand, generally have a large variance. However, even for this category, it can be observed that most images have higher values for β and γ, indicating higher contrast and more edges.

From these observations, a supervised selection of a color constancy algorithm for images from all scene categories can be achieved. By classifying an input image as one of these image categories (either supervised by user-intervention or unsupervised by a scene-recognition system like [36, 37, 43]), the corresponding color constancy algorithm can be applied to the image to obtain a performance that is similar to the proposed automatic selection-algorithm.

5 Experiments

In this section, the proposed method is evaluated and compared to state-of-the-art methods on large-scale data sets. First, the data sets that are used for training and testing are introduced in section 5.1 and the performance measure is discussed in section 5.2. In section 5.3, the performance of several state-of-the-art algorithms on a large-scale data set is shown, and in sections 5.4 and 5.5 the performance of the standard and the proposed fusion algorithms are discussed and analyzed. Finally, in section 5.6, experiments are performed using scene semantics.

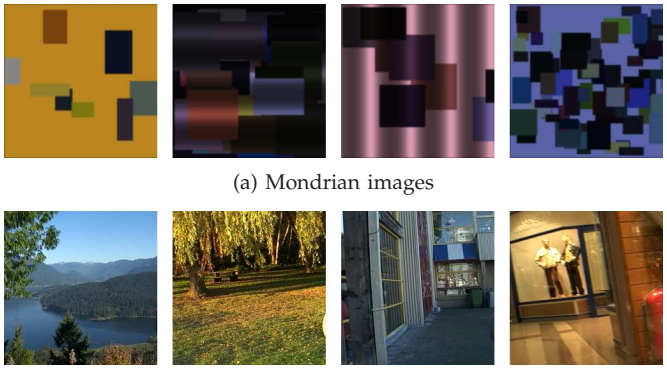
5.1 Data set

A training data set is created based on spectral reflectance data presented in [44]. This data set originally comprises only surface and illuminant spectra, which are first combined into (R, G, B)-values. Then, using these generated pixel colors, several Mondrian-like images are created, which all have different properties in the number of edges, the amount of texture and contrast. Since only material surfaces are present in the original data set, shadow gradients are added to several images to enlarge their photometrical variety. Note that the resulting images contain up to tens of different surfaces, and hence many different transitions, simulating the statistics of real-world images as close as possible. A few examples are shown in figure 4(a). This data set will be called the Mondrian data set in the remainder of the paper.

As a second set, the color constancy data set of Ciurea and Funt [31] is used. In this data set, over $11,000$ images are present, extracted from 2 hours of video recorded

under a large variety of imaging conditions (including indoor, outdoor, desert, cityscape, and other settings). In total, the images are divided into 15 different clips taken at different locations, see figure 4(b) for some example images that are in the data set. The data set provides the ground truth of the color of the illuminant. This ground truth is acquired by making use of the small grey sphere in the bottom right corner of the images. Note that this grey sphere is omitted while estimating the illuminant in the experiment. Further, since some of the images in this set are rather correlated due to the nature of the set, we make sure that, if applicable, this correlation is taken into account when separating the data set, *e.g.* for training purposes. Hence, in this way, images, or similar images, that are used for testing a method do not appear in the training set.

One of the basic assumptions in machine learning, is that the distribution of the test data should be similar to the distribution of the training data. Hence, the variety of images used for training the algorithm, should be similar to the variety of the images that are used to test it. To this end, we propose two different experiments. The first experiment consists of a training phase using the Mondrian-like images shown in figure 4(a) and a test phase using the real-world data shown in figure 4(b). In this way, the date sets for training and testing are completely different. This scenario reflects the case when little information on the data set (used for testing the method) is known. The second experiment is based on cross-validation: the data will be divided into 15 parts, where we make sure that correlated images are grouped in the same part. This means that images that are taken from the same video shot are *not* separated. Next, the method is trained on 14 parts of the data, and tested on the remaining part. This procedure is repeated 15 times, so every images is in the test set exactly once and all images from one shot will either be in the training set or in the test set.



(a) Mondrian images

(b) Real-world images

Fig. 4. Examples of images that are in the two data sets that are used in this paper. The first data set consists of images that are generated using surface reflectance spectra combined with illuminant spectra. The second data set consists of real-world images [31].

TABLE 1
Median angular errors for several (single) algorithms on the complete data set of $11,346$ images.

Method	Median
Do nothing (*i.e.* without color constancy)	6.7°
Grey-World ($\mathbf{e}^{0,1,0}$)	7.0°
White-Patch ($\mathbf{e}^{0,\infty,0}$)	5.3°
General Grey-World ($\mathbf{e}^{0,13,2}$)	5.5°
1^{st}-order Grey-Edge ($\mathbf{e}^{1,1,6}$)	5.2°
2^{nd}-order Grey-Edge ($\mathbf{e}^{2,1,5}$)	5.2°
Best 0^{th}-order algorithm: $\mathbf{e}^{0,9,0}$	5.3°
Best 1^{st}-order algorithm: $\mathbf{e}^{1,1,1}$	4.6°
Best 2^{nd}-order algorithm: $\mathbf{e}^{2,1,2}$	4.9°
Gamut mapping	
- trained on: *Mondrian data*	5.5°
- trained on: *real-world data*	4.8°

5.2 Performance measure

For all images in the data set, the correct color of the light source $\mathbf{e}_l$ is known a priori. To measure how close the estimated illuminant resembles the true color of the light source, the angular error ϵ is used:

$$\epsilon = \cos^{-1}\left(\frac{\mathbf{e}_l \cdot \mathbf{e}_e}{||\mathbf{e}_l|| \cdot ||\mathbf{e}_e||}\right), \tag{15}$$

where $\mathbf{e}_l \cdot \mathbf{e}_e$ is the dot product of the two vectors representing the true color of the light source $\mathbf{e}_l$ and the estimated color of the light source $\mathbf{e}_e$, and $||\cdot||$ indicates the Euclidean norm. To measure the performance of an algorithm on a data set, the median angular error is reported as summarizing statistic, as this is considered to be the most appropriate measure [33, 45].

5.3 Color Constancy algorithms

In table 1, the results for several (state-of-the-art) algorithms are shown for the real-world data set of Ciurea and Funt [31]. Results of the Grey-World algorithm are comparable to not applying any method at all (*i.e.* the estimate of the illuminant is set to white), while the White-Patch performs slightly better. Instantiations of the framework that have been previously reported as best-performing on real-world data [13] are also tested on the same data set that. The results for $\mathbf{e}^{0,13,2}$, $\mathbf{e}^{1,1,6}$ and $\mathbf{e}^{2,1,5}$ are shown in table 1 and the three instantiations have similar performance on this data set, outperforming the Grey-World and the White-Patch algorithms. However, choosing a more appropriate set of parameters can significantly improve the results (denoted by "Best ..." in table 1). Further, the performance of these more carefully selected parameters confirms the findings of van de Weijer et al. [13], where it is concluded that the 1^{st}-order Grey-Edge algorithm outperforms other algorithms on the real-world data. Note, however, the difference in performance between the two parameter settings, while both settings are found to be optimal for a certain data

set. The fact that for this data set a different optimal parameter setting exists than for the data sets used in [13] confirms the need for an algorithm to dynamically tune the parameters for the different instantiations.

For the experiments based on gamut mapping, the same implementation as in [32, 46] is used [1]. The software provides an option to apply segmentation as a preprocessing step, and several different parameter settings are tested (the reported results are computed by using the best parameter settings obtained for data set used). Further, experiments using two different canonical gamuts are performed. The first experiment uses a canonical gamut that is constructed using a large Mondrian data set [44], the second experiment uses a canonical gamut that is constructed on real-world data.

To avoid overfitting by learning on the test set, for the second experiment, the same setup is used that was described earlier as cross-validation. The data set is divided into 15 parts, keeping all images that are extracted from the same clip together. The canonical gamut is constructed on 14 parts, and the remaining part of the data is used for testing. This procedure is repeated 15 times, excluding each scene exactly once from the training data to use it as test data.

As expected, the best results are obtained when using the real-world data to construct the canonical gamut. When using the Mondrian data to calibrate the canonical gamut, the median angular error is $5.5°$. This is close to the performance of the White-Patch algorithm. However, by learning the canonical gamut from the real-world data, the error drops to a respectable $4.8°$. However, the best-performing algorithm remains the 1^{st}-order Grey-Edge. Note that the performance of the gamut mapping too is obtained by optimizing the segmentation parameters that can be set using the gamut-software provided by [32, 46] and hence the performance is considered to be best-case.

In figure 5, the angular errors of all the methods are shown, together with error bars indicating a confidence interval of 95%. These error bars are obtained using resampling: let ϵ^A be the set of error measurements for a method A on all images of the data set. By taking n bootstrap samples (*i.e.* samples drawn at random with replacement) from this set of measurements, n sets of error measurements are obtained. For every set of error measurements ϵ_i^A, the mean and median can be computed, resulting in n values for the mean and n values for the median (*i.e.* distributions for the mean and median). Finally, a $p\%$ confidence interval is obtained from the $p/2$ and $1-p/2$ quantiles of these distributions [33].

5.4 Color constancy using standard fusion

To evaluate the different combination strategies, the algorithms of [13] are used as instantiations of combination methods as described in section 4. Therefore, the best-performing instantiation is used as baseline to compare the algorithms with.

A simple average of the output results does not seem to improve the results as compared to the baseline algorithm (see table 2): the median angular error is slightly worse. Using a weighted average instead of a simple average, levels the median angular error with the best-performing single algorithm. The main advantage of this fusion method, is that there are less outliers. This results in a slightly better mean angular error ($5.5°$ over $5.7°$).

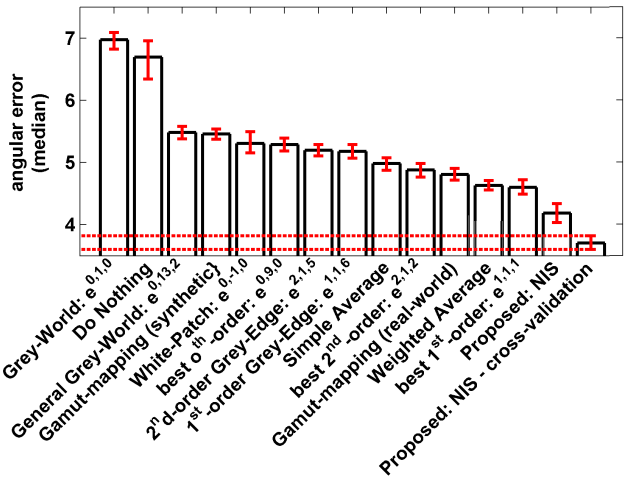


Fig. 5. Median angular errors for all the methods plotted with a 95% confidence interval.

5.5 Color constancy using natural image statistics

The proposed fusion-algorithm, learned on the Mondrian data set and tested on the real-world data set (denoted by NIS in table 2) performs significantly better than the baseline and the weighted average: an improvement of the median angular error of nearly 10% is obtained. In fact, in this experiment, test-data is unknown, and corresponds to the most generic approach of the proposed method. No learning step is required for a new data set, since the resulting classifier is independent of the test data.

Results of the second experiment, denoted by NIS - cross-validation in table 2, corresponds to the situation where the circumstances under which the system is used are known a priori. Training is done using part of the real-world data, and testing is done using an independent part, where images that are similar are either in the training set or in the test set. Hence, no similarities exist between the training data and test data. Results show a significant improvement over the first case (with no a priori knowledge): the median angular error drops to $3.7°$, an improvement of nearly 20% over the baseline method. In this experiment, test-data is known making

1. http://kobus.ca/research/programs/colour_constancy

TABLE 2
The performance of several different methods to combine the output of single color constancy methods. The color constancy methods used include one zeroth-order method (General Grey-World $\mathbf{e}^{0,5,1}$, one first-order method (Grey-Edge $\mathbf{e}^{1,1,2}$) and one second-order method (2^{nd}-order Grey-Edge $\mathbf{e}^{2,1,1}$). The entry NIS-Mondrian denotes the proposed algorithm trained on the *independent* Mondrian data set and tested on the real-world data set, the entry NIS-realworld denotes the proposed algorithm evaluated using cross-validation on the real-world data set.

Method	Median	
Baseline ($\mathbf{e}^{1,1,1}$)	4.6°	
Simple average	5.0°	$(+8.7\%)$
Weighted average	4.6°	$(\pm 0\%)$
Proposed: NIS	$\mathbf{4.2}^\circ$	$(\mathbf{-8.7}\%)$
Proposed: NIS - cross-validation	$\mathbf{3.7}^\circ$	$(\mathbf{-19.6}\%)$

the proposed method less generic. Hence, a learning step is required for each new data set.

The proposed algorithm requires data, with enough variety, to train the method. However, from the experiments, it can be concluded that the proposed method can be trained using a completely independent training set, and still perform significantly better than the baseline algorithm. Note also, that this baseline algorithm is the best-case performance of the single methods, since it was manually selected among the single algorithms as color constancy method with the highest performance on the current data set. Still, the worst-case performance of the proposed method (NIS-Mondrian in table 2) is significantly higher.

SNR performance analysis. In order to analyze the signal-to-noise sensitivities of pixel-based, edge-based and higher-order statistics-based methods, the following experiment is conducted: a Mondrian-like image is created using 8 patches with different colors, keeping the difference between two consecutive colors fixed. After that, an illuminant is selected at random from the set of light sources proposed by Barnard et al. [44]. Based on the diagonal model, this illuminant is used to transform the image so that it appears to be taken under a different light source. Finally, the three color constancy algorithms are used to estimate the illuminant and the performance of these algorithms is evaluated with respect to the signal-to-noise ratio (SNR). The signal-to-noise ratio is defined by the average difference between two consecutive patches a_i and a_j:

$$\text{SNR} = \frac{a_i - a_j}{\sigma_{\text{noise}}}, \tag{16}$$

where σ_{noise} is kept fixed.

In figure 6, the sensitivities of the 0^{th}-order, the 1^{st}-order and the 2^{nd}-order algorithm (*i.e.* White-Patch, 1^{st}-order Grey-Edge and 2^{nd}-order Grey-Edge, respectively)

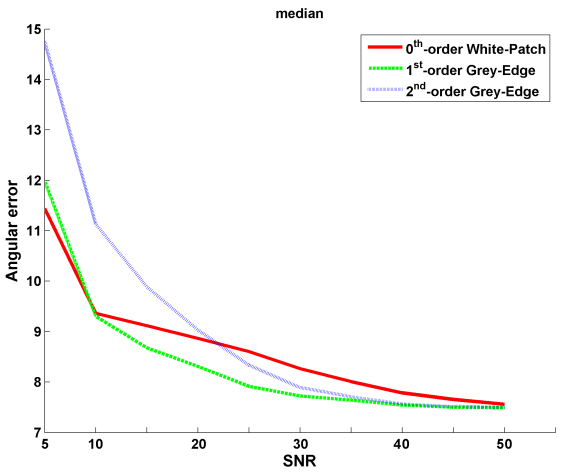


Fig. 6. Experiment monitoring the sensitivity of the different color constancy algorithms to the signal-to-noise ratio. It can be derived that all methods perform better as the SNR increases. Further, 1^{st}-order and 2^{nd}-order Grey-Edge have a higher error than the 0^{th}-order method for low signal-to-noise ratios, but a lower error for higher signal-to-noise ratios.

are shown with respect to the SNR. As expected, it can be derived that, for a higher signal-to-noise ratio, on average, the accuracy of all algorithms is improved. However, it is also shown that the performance of the pixel-based method degrades less than edge-based and higher-order statistics-based methods when the SNR decreases. For higher values of the SNR, the 1^{st} and 2^{nd}-order Grey-Edge algorithms outperform the pixel-based method; the performance of the three algorithms converge when the SNR reaches very high values.

To conclude, the Grey-Edge algorithms (1^{st} and 2^{nd}-order) provide better performance than the pixel-based one, but they need a medium to high SNR. Pixel-based methods obtain better performance for a low SNR. In other words, edge-based is preferred when the image contains a reasonable number of (sufficiently) contrasted edges. Otherwise, pixel-based color constancy is preferred. This principle is reflected by the Weibull-distribution parameters contrast β and grain size γ and forms the basis of the proposed algorithm. The MoG classifier learns the correlation between these parameters.

5.6 Color constancy using scene semantics

Color constancy using scene semantics is evaluated on the real-world data set as follows. Images from the real-world data set are annotated, to obtain images that can be classified as coming from the same scene category. In total, 75 images from 5 clips of the complete data set (15 images per category) are annotated as *open country* and 70 images from 7 clips (10 per category) as *street*, *indoor* and *forest*. During annotation, the selected images

(a) Original images

(b) Ideal correction

(c) Correction using Natural Image Statistics

(d) Correction using Grey-Edge (baseline)

(e) Correction using White-Patch

(f) Correction using Grey-World

Fig. 7. Examples of images that are in the data set used for evaluation in section 5 and results of some color constancy algorithms. The corresponding angular errors are shown in the lower right corner of the images.

are different enough to avoid the overfitting problem. The performance of the proposed method using scene semantics is compared to the combination-method using natural image statistics.

In table 3, the results are shown. In this table, *scene semantics* refers to applying one color constancy algorithm to all images of a specific scene category. This one color constancy algorithm is selected in a supervised manner, based on the Weibull-parameters of images from that scene category. It can be observed that for some categories the performance is similar to the method using natural image statistics, even though only one color constancy method is applied to all images of the same scene category.

6 Discussion

Natural image statistics are indicative for the performance of the different color constancy algorithms. For instance, the Grey-Edge algorithm performs better on

TABLE 3
Median angular error for several different methods on images from four categories.

Open Country	Median
Entire image (NIS-Mondrian)	5.5°
Proposed: Scene Semantics	6.7°
Street	Median
Entire image (NIS-C)	2.7°
Proposed: Scene Semantics	3.4°
Indoor	Median
Entire image (NIS-C)	5.1°
Proposed: Scene Semantics	5.5°
Forest	Median
Entire image (NIS-C)	5.6°
Proposed: Scene Semantics	5.5°

images with a reasonable number of edges than on images with only a few edges. Moreover, if the contrast of these edges is rather low, then the accuracy of the estimate, in general, will decrease. The effect of the SNR is analyzed in section 5.5. From this analysis, it is derived that if the SNR is high then it is beneficial to use methods based on higher-order statistics. However, the performance of such methods decreases as the SNR decreases, resulting in a preference for methods based on 1^{st}-order statistics for images with a medium SNR. For a very low SNR (*i.e.* no edge information), pixel-based methods are preferred.

This principle is also reflected in the resulting classifier that is learned on the Mondrian data set. From the analysis of this classifier, it is observed that the method based on 2^{nd}-order statistics is preferred, provided that there is enough information to give an accurate estimate. This means that the distribution of the 2^{nd}-order statistics is balanced and not easily biased towards noisy measurements. Since the available information of 2^{nd}-order statistics decreases rapidly, depending on the content of an image, the performance of color constancy methods based on 2^{nd}-order statistics decreases also. Hence, color constancy methods using edge-information are preferred for images with a medium amount of information, and pixel-based methods are preferred when the amount of information becomes low. The amount of information is dependent on the content of an image, and natural image statistics parameterize the amount of available information. For instance, a high contrast of colors (*i.e.* the β of the edge distributions of the O_1 and O_2-channels) combined with only few edges (*i.e.* a high value for γ) indicates that there are only a few edges with enough energy and that the estimate that is based on this information is likely to be biased towards the color of the high contrast. Consequently, in such situations it is better to select pixel-based color constancy methods. On the other hand, when the value of γ is low, this means that there are relatively many edges with a strong response, making the estimates of methods based on 1^{st} and 2^{nd}-order statistics more reliable.

In summary, the three basic indicators for the performance of color constancy methods based on pixel and edge information, are:

- Image features. The set of possible adjacent color values (*i.e.* color derivatives) in real-world images is more restricted than the set of possible pixel values. Hence, the use of derivatives provides more stable gamuts than pixel values alone to compute color constancy. The order of best performance in terms of image features is then 2^{nd}-order statistics first, followed by 1^{st}-order and 0^{th}-order statistics.
- Number of features. A higher color constancy accuracy is obtained when there is a large variety of different edges in an image [17]. The same argument holds for the Grey-World algorithm for the number of different surfaces in a scene, *e.g.* [32, 33]. From [17], it is derived that if the image contains a limited number of edges, pixel-based color constancy is preferred. In case of sufficient edges (*e.g.* more than 8 different edges), edge-based color constancy is preferred. The order of best performance in terms of the number of edges (from low to high) is then 0^{th}-order statistics first, followed by 1^{st}-order and 2^{nd}-order statistics.
- SNR. Pixel-based methods obtain a better performance for a low SNR. The order of best performance in terms of the amount of SNR (from low to high) is then 0^{th}-order statistics first, followed by 1^{st}-order and 2^{nd}-order statistics.

The above principles are reflected by the Weibull-distribution parameters contrast β and grain size γ. The amount of texture, the number of edges and contrast are captured by this parameterization. The proposed algorithm exploits this information to select and combine the different order color constancy methods as follows:

- Image features. The value γ corresponds to grain size and is related to the amount of texture.
- Number of features. The value γ corresponds grain size and is related to the number of edges.
- SNR. The value β corresponds to contrast and is related to SNR.

The MoG classifier computes the correlation between the values of the image features, number of features and SNR and the values of the Weibull-distribution parameters contrast β and grain size γ.

7 Conclusion

In this paper, the Weibull parameterization (texture and contrast) has been used to identify the most important characteristics of color images. It is shown that the Weibull parameterization (grain size and contrast) is related to the number of edges, amount of texture and signal-to-noise ratio to which the used color constancy methods are sensitive to. A MoG-classifier has been used to learn the correlation and weighting between the

Weibull-parameters (grain size and contrast) and these image attributes (number of edges, amount of texture and SNR). The output of the classifier is the selection of the best performing color constancy method for a certain image.

Experimental results show a large improvement over state-of-the-art single algorithms. On a data set consisting of more than 11,000 images, the best-performing single algorithm is found to be the 1^{st}-order Grey-Edge. Comparing the median angular error of this algorithm with our proposed algorithm, an increase of nearly 20% can be obtained when the circumstances under which the algorithm will be used are known a priori. When this is not the case, *i.e.* when no information about the test images is known, a pre-learned system can be used, that is trained on the Mondrian set. This system only needs to be trained once, and is completely independent of the test data. Using this classifier, the median angular error still decreases significantly (nearly 10%) compared to the best-case performance of the single algorithms. Finally, we showed that for certain scene categories, one specific color constancy algorithm can be used instead of the classifier considering several algorithms.

REFERENCES

[1] P. B. Delahunt and D. H. Brainard, "Does human color constancy incorporate the statistical regularity of natural daylight?" *Journal of Vision*, vol. 4, no. 2, pp. 57–81, 2004. 1

[2] D. H. Foster, K. Amano, and S. M. C. Nascimento, "Color constancy in natural scenes explained by global image statistics," *Visual Neuroscience*, vol. 23, no. 3–4, pp. 341–349, 2006. 1

[3] J. Yang, R. Stiefelhagen, U. Meier, and A. Waibel, "Visual tracking for multimodal human computer interaction," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1998, pp. 140–147. 1

[4] T. Gevers and A. Smeulders, "Color-based object recognition." *Pattern Recognition*, vol. 32, no. 3, pp. 453–464, 1999. 1

[5] T. Gevers and A. W. M. Smeulders, "Pictoseek: combining color and shape invariant features for image retrieval," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 102–119, 2000. 1

[6] M. D. Fairchild, *Color Appearance Models, 2nd Ed.*, ser. Wiley-IS&T Series in Imaging Science and Technology. Chichester, UK: John Wiley & sons, 2005. 1, 3

[7] M. Ebner, *Color constancy*, ser. Wiley-IS&T Series in Imaging Science and Technology. John Wiley & Sons, 2007. 1

[8] S. D. Hordley, "Scene illuminant estimation: past, present, and future," *Color Research & Application*, vol. 31, no. 4, pp. 303–314, 2006. 1, 2

[9] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Journal of the Optical Society of America A*, vol. 61, pp. 1–11, 1971. 1, 2

[10] E. H. Land, "The retinex theory of color vision," *Scientific American*, vol. 237, no. 6, pp. 108–128, December 1977. 1, 2

[11] G. Buchsbaum, "A spatial processor model for object colour perception," *Journal of the Franklin Institute*, vol. 310, no. 1, pp. 1–26, July 1980. 1, 2

[12] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. IS&T/SID Color Imaging Conference*, 2004, pp. 37–41. 1, 2

[13] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2207–2214, 2007. 1, 2, 3, 4, 8, 9

[14] D. A. Forsyth, "A novel algorithm for color constancy," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 5–36, 1990. 1, 2, 3

[15] G. D. Finlayson, S. D. Hordley, and P. M. Hubel, "Color by correlation: a simple, unifying framework for color constancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1209–1221, 2001. 1, 2

[16] G. D. Finlayson, S. D. Hordley, and I. Tastl, "Gamut constrained illuminant estimation," *International Journal of Computer Vision*, vol. 67, no. 1, pp. 93–109, 2006. 1

[17] A. Gijsenij, T. Gevers, and J. van de Weijer, "Generalized gamut mapping using image derivative structures for color constancy," *International Journal of Computer Vision*, vol. 86, no. 2–3, pp. 127–139, 2010. 1, 2, 4, 12

[18] D. H. Brainard and W. T. Freeman, "Bayesian color constancy," *Journal of the Optical Society of America A*, vol. 14, pp. 1393–1411, 1997. 1

[19] M. D'Zmura, G. Iverson, and B. Singer, "Probabilistic color constancy," in *Geometric Representations of Perceptual Phenomena*. Lawrence Erlbaum Associates, 1995, pp. 187–202. 1

[20] M. Ebner, "Evolving color constancy," *Pattern Recognition Letters*, vol. 27, no. 11, pp. 1220–1229, 2006. 1

[21] V. C. Cardei and B. V. Funt, "Committee-based color constancy," in *Proc. IS&T/SID Color Imaging Conference*, 1999, pp. 311–313. 1, 5

[22] G. Schaefer, S. D. Hordley, and G. D. Finlayson, "A combined physical and statistical approach to colour constancy," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2005, pp. 148–153. 2, 5

[23] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991. 3

[24] G. West and M. H. Brill, "Necessary and sufficient conditions for von kries chromatic adaptation to give color constancy," *Journal of Mathematical Biology*, vol. 15, no. 2, pp. 249–258, 1982. 3

[25] G. D. Finlayson, M.S. Drew, and B.V. Funt, "Spectral sharpening: sensor transformations for improved color constancy," *Journal of the Optical Society of*

America A, vol. 11, no. 5, pp. 1553–1563, 1994. 3

[26] B. V. Funt and B. C. Lewis, "Diagonal versus affine transformations for color correction," *Journal of the Optical Society of America A*, vol. 17, no. 11, pp. 2108–2112, 2000. 3

[27] K. M. Lam, "Metamerism and colour constancy," Ph.D. dissertation, University of Bradford, 1985. 3

[28] N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M. R Luo, and T. Newman, "The ciecam02 color appearance model," in *Proc. IS&T/SID Color Imaging Conference*, 2002, pp. 23–27. 3

[29] J. von Kries, "Influence of adaptation on the effects produced by luminous stimuli," in *Sources of Color Vision*, D. MacAdam, Ed. MIT Press, 1970, pp. 109–119. 3

[30] H. Y. Chong, S. J. Gortler, and T. Zickler, "The von kries hypothesis and a basis for color constancy," in *Proc. of the International Conference on Computer Vision*, 2007, pp. 1–8. 3

[31] F. Ciurea and B. V. Funt, "A large image database for color constancy research," in *Proc. IS&T/SID Color Imaging Conference*, 2003, pp. 160–164. 4, 6, 7, 8

[32] K. Barnard, V. C. Cardei, and B. V. Funt, "A comparison of computational color constancy algorithms; part i: Methodology and experiments with synthesized data," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 972–984, 2002. 4, 9, 12

[33] S. D. Hordley and G. D. Finlayson, "Reevaluation of color constancy algorithm performance," *Journal of the Optical Society of America A*, vol. 23, no. 5, pp. 1008–1020, May 2006. 4, 8, 9, 12

[34] A. Torralba and A. Oliva, "Statistics of natural image categories," *Network: computation in neural systems*, vol. 14, no. 3, pp. 391–412, 2003. 4, 6

[35] J.-M. Geusebroek and A. W. M. Smeulders, "A six-stimulus theory for stochastic texture," *International Journal of Computer Vision*, vol. 62, no. 1-2, pp. 7–16, 2005. 4

[36] A. Torralba, "Contextual priming for object detection," *International Journal of Computer Vision*, vol. 53, no. 2, pp. 169–191, 2003. 4, 7

[37] J. C. van Gemert, J. M. Geusebroek, C. J. Veenman, C. G. M. Snoek, and A. W. M. Smeulders, "Robust scene categorization by learning image statistics in context," in *CVPR Workshop on Semantic Learning Applications in Multimedia (SLAM)*, New York, USA, June 2006. 4, 7

[38] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001. 6, 7

[39] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, USA, June 2007, pp. 1–8. 5

[40] C. M. Bishop, *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press, 1996. 5

[41] D. L Ruderman, T. W. Cronin, and C. C. Chiao, "Statistics of cone responses to natural images: implications for visual coding," *Journal of the Optical Society of America A*, vol. 15, no. 8, pp. 2036–2045, 1998. 6

[42] J. van de Weijer, C. Schmid, and J. J. Verbeek, "Using high-level visual information for color constancy," in *Proc. of the International Conference on Computer Vision*, Rio de Janeiro, Brasil, 2007. 6

[43] J.-M. Geusebroek, "Compact object descriptors from local colour invariant histograms," in *British Machine Vision Conference*, vol. 3, 2006, pp. 1029–1038. 7

[44] K. Barnard, L. Martin, B. V. Funt, and A. Coath, "A data set for color research," *Color Research & Application*, vol. 27, no. 3, pp. 147–151, 2002. 7, 9, 10

[45] A. Gijsenij, T. Gevers, and M. Lucassen, "A perceptual analysis of distance measures for color constancy algorithms," *Journal of the Optical Society of America A*, vol. 26, no. 10, pp. 2243–2256, 2009. 8

[46] K. Barnard, L. Martin, A. Coath, and B. V. Funt, "A comparison of computational color constancy algorithms; part ii: Experiments with image data," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 985–996, 2002. 9

Arjan Gijsenij Arjan Gijsenij is currently Post-doctoral researcher at the Intelligent Systems Lab of the University of Amsterdam. He received his M.Sc. degree in Computer Science at the University of Groningen in 2005, and obtained his Ph.D. degree at the University of Amsterdam in 2010. His main research interests include color in computer vision and psychophysics, (color) image processing, visual categorization and statistical pattern recognition. He has served on the program committee of several major conferences and workshops. He is a member of the IEEE.

Theo Gevers Theo Gevers is an Associate Professor of Computer Science at the University of Amsterdam, The Netherlands, where he is also teaching director of the M.Sc. of Artificial Intelligence. He currently holds a VICI-award (for excellent researchers) from the Dutch Organisation for Scientific Research. His main research interests are in the fundamentals of content-based image retrieval, color image processing and computer vision specifically in the theoretical foundation of geometric and photometric invariants. He is chair of various conferences and he is associate editor for the IEEE Transactions on Image Processing. Further, he is program committee member of a number of conferences, and an invited speaker at major conferences. He is a lecturer of post-doctoral courses given at various major conferences (CVPR, ICPR, SPIE, CGIV). He is member of the IEEE.