Multi-Illuminant Estimation With Conditional Random Fields

Shida Beigpour, Christian Riess, Joost van de Weijer, and Elli Angelopoulou

***Abstract*— Most existing color constancy algorithms assume uniform illumination. However, in real-world scenes, this is not often the case. Thus, we propose a novel framework for estimating the colors of multiple illuminants and their spatial distribution in the scene. We formulate this problem as an energy minimization task within a conditional random field over a set of local illuminant estimates. In order to quantitatively evaluate the proposed method, we created a novel data set of two-dominant-illuminant images comprised of laboratory, indoor, and outdoor scenes. Unlike prior work, our database includes accurate pixel-wise ground truth illuminant information. The performance of our method is evaluated on multiple data sets. Experimental results show that our framework clearly outperforms single illuminant estimators as well as a recently proposed multi-illuminant estimation approach.**

***Index Terms*— Color constancy, CRF, multi-illuminant.**

I. Introduction

THE vast majority of existing color constancy algorithms are based on the assumption that there exists a single illuminant in the scene. Many images, however, exhibit a mixture of illuminants with distinct chromaticities. Consider, for example, indoor scenes which are lit by both indoor light sources and outdoor light coming through the windows. Or think of an outdoor scene, where parts of the image are in direct sunlight, while others are in shadow which is illuminated by blue skylight. Another example, where single illuminant white balancing is known to give unsatisfactory results, is pictures taken with a camera-flash. Illuminant estimation methods that assume uniform illumination cannot accurately recover the illuminant chromaticity and its variations across such scenes. Examples of multi-illuminant pictures, and the color-coded pixelwise influence of each illuminant, can be seen in Fig. 1.

Extending existing color constancy methods to successfully compute multi-illuminant estimates is a challenging problem. Consider two of the most popular branches of existing color constancy approaches: statistics-based methods and physics-based ones. The success of statistics-based techniques [1]–[4] depends on the size of the statistical sample. Applying these methods to small image regions introduces inaccuracies [5] and is unlikely to yield stable results. Physics-based methods either assume purely diffuse scenes, e.g., [6], [7], which is not often applicable in real scenes, or exploit the presence of specularities in an image, e.g., [8], which typically occur very sparsely. As a result, a direct extension of global (image-wide) color constancy methods to region-based ones is likely insufficient. Spatial constraints between the estimates will be required to obtain acceptable results.

Fig. 1. Sample images from our datasets. The bottom row shows the relative influence of the two incident illuminants color-coded in blue and red.

We propose a multiple illuminant estimation method which first extracts local estimates. We overcome the inherent instability of local measurements by globally solving the illuminant labelling problem by means of a Conditional Random Field (CRF). We prove that several existing approaches, namely statistics- and physics-based methods, can be written in the form of a CRF. The CRF formulation provides a natural way to: 1) combine various approaches into a single multi-illuminant estimate and 2) incorporate spatial information about the illuminant distribution. We show that representing these methods by such a model allows us to robustly extend them to multi-illuminant estimation. Furthermore, we created two new datasets for multi-illuminant color constancy with highly accurate, computationally extracted (instead of manually annotated) pixelwise ground truth. The first database contains laboratory images, designed for evaluation under close-to-ideal conditions. The second dataset is composed of real-world multi-illuminant scenes (indoor and outdoor), which more closely approximate real-world scenarios.

Manuscript received November 24, 2012; revised July 27, 2013; accepted September 24, 2013. Date of publication October 18, 2013; date of current version November 7, 2013. The work of J. van de Weijer was supported by the Spanish under Project TIN2009-14173. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. David S. Taubman.

S. Beigpour and J. van de Weijer are with the Computer Vision Center, Universidad Autonoma de Barcelona, Barcelona 08193, Spain (e-mail: shida@cvc.uab.es; joost@cvc.uab.es).

C. Riess and E. Angelopoulou are with the Pattern Recognition Laboratory, University of Erlangen-Nuremberg, Erlangen 91058, Germany (e-mail: christian.riess@fau.de; elli@i5.cs.fau.de).

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.2013.2286327

In summary, the main contributions of this paper are:

- The formulation of multi-illuminant estimation as a CRF model.
- The expression of existing bottom-up approaches to color constancy as an energy minimization problem.
- The creation of two new datasets for multi-illuminant estimation.
- An extensive experimental evaluation which shows that the proposed method addresses the intrinsic challenges in multi-illuminant scenes, i.e. the estimation of the illuminant colors and their spatial distribution, with superior accuracy compared to prior work.

This paper is organized as follows. In Section II, we present related work in color constancy, and in particular methods that have been used as a foundation for the proposed approach. An overview and the theoretical foundation of the proposed framework is introduced in Section III. Unary and pairwise potentials for the CRF are derived in Section IV and Section V, respectively. In Section VI, we provide implementation details of our algorithm. Section VII contains a description of the new multi-illuminant datasets, as well as the computation of the ground truth. Experimental results are presented and discussed in Section VIII. We summarize the findings of this paper in Section IX. Additional mathematical details are presented in the appendix.

II. Related Work

Statistics-based color constancy methods derive the estimate of the illuminant color from assumptions on the statistics of reflectances in the world. The gray-world algorithm [1] is the most well-known method of this family. It computes the illuminant of a scene by assuming that the average scene reflectance is gray. Another popular method is the MAX-RGB algorithm which computes the illuminant in a scene from the maximum responses in the RGB channels [9]. It was noted by Gershon *et al.* [10] that it is often beneficial to assume that the average of the intensity values of a scene is equal to the average reflectance of a database. Finlayson and Trezzi [3] showed that both the gray-world and the MAX-RGB algorithms are instantiations of the more general shades-of-gray method which estimates the illuminant of images by computing the Minkowski norm of the pixels in a scene. Van de Weijer *et al.* [2] further extended this theory to also include image derivatives. Finally, Gijsenij *et al.* [4] showed that weighting edges according to their physical cause (shadow, specularity, or material transition) can further improve results.

In comparison, physics-based methods exploit the interaction between light and material to infer the illuminant color in an image. Some methods, e.g. [6], [7], assume the scene is entirely composed of diffuse surfaces, while others, e.g. [8], [11], exploit the presence of specular highlights. These latter methods are based on dichromatic reflectance [12] which models the reflected light as a combination of diffuse and specular reflection. Based on the assumption of neutral interface reflection, the color of specular highlights is the same as the illuminant color and therefore an important cue for color constancy (see, [8], [12]).

Gamut based methods exploit the fact that only a limited set of RGB values can be observed under a known canonical illuminant. This set of RGB values can be represented by a canonical convex hull in RGB space [13]. Thus, feasible illuminants can be estimated by computing all possible mappings from a single image's convex hull to the canonical convex hull. The scene illuminant is heuristically selected from the feasible illuminants. This method was further extended by Finlayson *et al.* [14] by constraining the possible illuminants to be on the Planckian locus. Gijsenij *et al.* [15] expanded this theory to higher-order derivative structures of the images. For a more complete overview of color constancy, see the recent overview articles [16]–[18].

There are illuminant estimation methods explicitly designed to handle varying illumination. In 1997, Barnard *et al.* [19] were the first ones to develop a methodology that automatically detects non-uniform illumination. They then proceeded with removing the illumination variation, at which point they could apply any gamut-based color constancy method. Though this method was pioneering at that time, its smooth illumination assumption restricts its applicability on real-world images. Ebner [20] followed a different approach of applying a diffusion-based technique on pixel intensities. However, he too assumes a smoothly varying illumination, which together with his underlying theory of regional gray-world can result in inaccuracies, especially in colorful scenes [21]. More recently, Kawakami *et al.* [22] proposed a physics-based method specifically designed to handle illumination variations between shadowed and non-shadowed regions in outdoor scenes. Due to its explicit assumption of hard shadows and sky-light/sunlight combination (or even more general Planckian illuminants), this method does not generalize well on arbitrary images. Gijsenij *et al.* [23] recently proposed an algorithm for scenes with two light sources. The reported experimental results are promising. However, it is not clear how to extend this methodology for non-local illuminant cues. When the chromaticity of the two incident illuminants is known, Hsu *et al.* [24] proposed an algorithm for high quality white-balanced images. However, their assumption of two known illuminants limits the applicability of the method to close-to laboratory conditions. Thus, by construction, none of the existing multi-illuminant estimation methods can handle arbitrary images and as such, none of them has been extensively tested on a large variety of real-world images.

The usage of a CRF for estimating the intrinsic properties of a scene, has been applied in several other papers. Lu and Drew [25] apply a Markov Random Field (MRF) for shadow segmentation. Their method is based on a Planckian illuminant assumption, which allows them to formulate a pairwise potential which enforces label consistency when chromaticity variations could be explained by an illuminant change. However, their method does not explicitly compute the illuminants of a scene. Lalonde *et al.* [26] identify shadow-regions based on local features. They subsequently use a CRF to group the shadow edges into coherent shadow contours. Serra *et al.* [27] use a CRF approach to estimate both reflectance and shading intrinsic images of scenes. However, their method assumes a single white illuminant.

Algorithm 1 MIRF Overview

1: Divide an image into N patches $p_1, p_2, ..., p_N$.
2: Extract local illuminant-color estimates $\mathbf{i}_j$ per patch p_j.
3: Determine the dominant illuminants via K-means clustering on the $\mathbf{i}_j$'s.
4: Find the optimal labelling by minimizing the CRF energy function. The energy function encourages both the local illuminant labels $\mathbf{x}_j$ to be close to the local illuminant estimates $\mathbf{i}_j$, and neighboring patches to have the same label, when applicable.

III. Methodology

A. Overview

When multiple illuminants are present in a scene, the spatial distribution of the illumination conditions becomes very important. We propose a new algorithm for multiple illuminant estimation based on a Conditional Random Field (CRF) framework. Our approach, called Multi-Illuminant Random Field (MIRF), estimates illuminants locally and uses the random field to resolve topological inconsistencies in the extracted illuminant colors. More specifically, MIRF is composed of the steps outlined in MIRF Overview.

A key component of MIRF is the setup of the conditional random field. The nodes in the graph represent image patches $p_1, p_2, \ldots, p_N$ with local observations $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_N$. The labels correspond to local illuminant colors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, while the edges connect neighboring patches. In such a representation local illuminant estimation becomes equivalent to finding the maximum a posteriori (MAP) labelling of the CRF. Such a framework facilitates both the local computation of illuminant color, as well as the incorporation of spatial information about the distribution of illuminants.

B. Conditional Random Field Setup

We will first introduce our conditional random field setup. Details on how we adapt it to the task of multiple illuminant estimation will be presented in the following sections. A conditional random field can be viewed as an undirected graph model, globally conditioned on observations. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the set of nodes representing the N patches and $\mathcal{E}$ is the set of edges connecting neighboring patches. We define a discrete random field $\mathcal{X}$ over the graph $\mathcal{G}$. Each node $i \in \mathcal{V}$ is associated with a random variable $\mathbf{X}_i \in \mathcal{X}$, which can take on a value $\mathbf{x}_i$ from the illuminant-color label set $\mathcal{L} = \{\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_k\}$. At each node $i \in \mathcal{V}$ we also have a local observation $\mathcal{F}_i$, which is the set of (R, G, B) values of all the pixels belonging to the corresponding patch p_i together with their spatial distribution. The probability $P(\mathcal{X} = \check{\mathbf{x}}|\mathcal{F})$ of a particular labelling $\check{\mathbf{x}} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ conditioned on the observations $\mathcal{F}$ of the entire image will be denoted as $P(\check{\mathbf{x}}|\mathcal{F})$. Then according to the Hammersley-Clifford theorem,

$$P(\check{\mathbf{x}}|\mathcal{F}) \propto \exp\left(-\sum_{c \in C} \xi_c(\check{\mathbf{x}}_c|\mathcal{F})\right), \tag{1}$$

where $\xi_c(\check{\mathbf{x}}_c|\mathcal{F})$ are potential functions defined over the observations $\mathcal{F}$ and the variables $\check{\mathbf{x}}_c = \{\mathbf{x}_i, i \in c\}$ belonging to clique c. A clique c is a subgraph in which every node is connected to all other nodes in the subgraph. C is the set of all cliques in $\mathcal{G}$. Finding the labelling $\check{\mathbf{x}}^*$ with the maximum a posteriori (MAP) probability $\check{\mathbf{x}}^*$ is then equal to

$$\check{\mathbf{x}}^* = \arg\max_{\check{\mathbf{x}} \in \mathbf{L}} P(\check{\mathbf{x}}|\mathcal{F}) = \arg\min_{\check{\mathbf{x}} \in \mathbf{L}} E(\check{\mathbf{x}}|\mathcal{F}), \tag{2}$$

where $\mathbf{L}$ is the set of all possible labellings on $\mathcal{X}$ and $E(\check{\mathbf{x}}|\mathcal{F})$ is the corresponding Gibbs energy defined as

$$E(\check{\mathbf{x}}|\mathcal{F}) = \sum_{c \in C} \xi_c(\check{\mathbf{x}}_c|\mathcal{F}). \tag{3}$$

Hence, computing the MAP labelling is equal to finding the labelling which minimizes the energy $E(\check{\mathbf{x}}|\mathcal{F})$. In our case, this means that obtaining the MAP assignment of illuminants to patches can be accomplished by finding that assignment which minimizes the corresponding Gibbs energy. Considering only up to pairwise clique potentials, the energy function becomes

$$E(\check{\mathbf{x}}|\mathcal{F}) = \sum_{i \in \mathcal{V}} \phi(\mathbf{x}_i|\mathcal{F}_i) + \theta_p \sum_{(i,j) \in \mathcal{E}} \psi((\mathbf{x}_i, \mathbf{x}_j)|(\mathcal{F}_i, \mathcal{F}_j)), \tag{4}$$

where ϕ denotes the unary potential and ψ the pairwise potential. The unary potentials ϕ penalize the discrepancy between the observations, i.e., the colors of the pixels in a patch $\mathcal{F}_i$, and the solution, i.e., the illuminant-color label assigned to the patch. The pairwise potentials ψ provide a definition of smoothness. Our CRF uses the image data to decide when to allow differences in labels of adjacent patches. Consider, for example, an object in the foreground that is illuminated by a camera flash. Such an object may have neighboring patches that belong to the background which is illuminated by ambient light. In this particular case, dissimilarities in the labels of the adjacent patches should be allowed. The constant $\theta_p > 0$ controls the balance between smoothness and data fit. When we use a pairwise potential function that enforces a single label for all patches, we obtain similar result as traditional single illuminant estimation methods. In the next section we propose various unary potentials which allow us to represent several well-known illumination estimation algorithms as CRFs.

IV. Unary Potentials

Unary potentials measure how well the local evidence supports the assigned label. In the context of illuminant color estimation we need to compare the locally obtained illuminant estimate with the color of the dominant illuminant assigned to a node. Different illuminant estimation methods can be used. Thus, MIRF can express several existing color constancy methods as an error minimization problem.

A. Statistics-Based Color Constancy

There exists a family of color constancy methods which are based on the statistics of reflectances in the world. Examples of this group of methods are the gray-world, the gray-edge and the MAX-RGB algorithms [1]–[3]. We show that several of these algorithms, when used in MIRF, can be formulated within an error minimization problem.

Let $\mathbf{f}^j = (f_R^j, f_G^j, f_B^j)^{\mathrm{T}}$ be the observed color at the j-th pixel in an image. We assume that an image is segmented into a number of patches $P = \{p_1, p_2, \ldots, p_N\}$ where p_i contains the indices to the pixels in patch i. From the set of observations $\mathcal{F}_i$ in a patch we can obtain an estimate of the local illuminant color $\mathbf{i}(\mathcal{F}_i)$, which, for conciseness, we will denote as $\mathbf{i}_i$.

If the estimate is computed with the gray-world algorithm, then the local illuminant color is determined by the average color in the patch, as defined by

$$\mathbf{i}_i = \frac{\sum\limits_{j \in p_i} \mathbf{f}^j}{\left\| \sum\limits_{j \in p_i} \mathbf{f}^j \right\|_2}, \tag{5}$$

where $\|.\|_2$ is the L_2 norm. It is applied to ensure that $\mathbf{i}_i$ has unit length.

Once the local illuminant is computed, its distance from the assigned label should be measured. A typical metric for comparing two normalized illuminant colors $\mathbf{i}_i$ and $\mathbf{i}_j$ is their angular distance given by

$$\alpha(\mathbf{i}_i, \mathbf{i}_j) = \arccos\left((\mathbf{i}_i)^{\mathrm{T}} \mathbf{i}_j\right). \tag{6}$$

In MRFs and CRFs it is preferred to use normalized distances in the [0, 1] range when comparing vertices. Thus, we can use

$$\varphi(\mathbf{i}_i, \mathbf{i}_j) = \left(1 - (\mathbf{i}_i)^{\mathrm{T}} \mathbf{i}_j\right). \tag{7}$$

We now define the statistics-based unary potential ϕ^s, which specifies the cost for patch p_i to take on illuminant $\mathbf{x}_i$ as

$$\phi^s(\mathbf{x}_i | \mathcal{F}_i) = w_i \rho(\mathbf{i}_i, \mathbf{x}_i) \tag{8}$$

where w_i is a scalar weight per patch, and ρ is the error norm. The influence of outliers on the unary potential can be reduced by choosing a robust error norm. One can directly use Eq. 7 as the error norm. We discuss several choices of error norms and weights at the end of this subsection. When using the gray-world algorithm an appropriate weight value is the sum of the intensities in a patch, $w_i = \|\sum_{j \in p_i} \mathbf{f}^j\|_1$. We, thus obtain the following unary potential

$$\phi^s(\mathbf{x}_i | \mathcal{F}_i) = \left\| \sum_{j \in p_i} \mathbf{f}^j \right\|_1 \varphi(\mathbf{i}_i, \mathbf{x}_i). \tag{9}$$

When the illuminant given by the label $\mathbf{x}_i$ and the illuminant derived directly from the observations are equal, this unary potential is zero. When they are maximally different, this unary potential is equal to the summed intensity of the patch. In Appendix IX-A we prove that this particular unary potential leads to the standard gray-world solution when we enforce a single illuminant label for all patches in Eq. 4.

Alternatively, one can also use the more general class of statistics-based illuminant estimation [2], given by

$$\mathbf{i}_i^{n,m} \approx \sqrt[m]{\sum_{j \in p_i} \left| \frac{\partial^n \mathbf{f}_{\sigma_{\mathrm{GW}}}^j}{\partial \mathbf{x}^n} \right|^m}, \tag{10}$$

where n is the order of the partial differentiation, m is the Minkowski norm, and $\mathbf{f}_{\sigma_{\mathrm{GW}}} = \mathbf{f} \otimes \mathbf{G}_{\sigma_{\mathrm{GW}}}$ is the image smoothed with a Gaussian of standard deviation σ_{GW}. Depending on the choice of parameters m and n, the illuminant estimation becomes the gray-world, shades of gray, or gray-edge algorithm. We propose the following unary potential for this general case of statistics-based estimators

$$\phi^s(\mathbf{x}_i | \mathcal{F}_i) = \left\| \sqrt[m]{\sum_{j \in p_i} \left| \frac{\partial^n \mathbf{f}_{\sigma_{\mathrm{GW}}}^j}{\partial \mathbf{x}^n} \right|^m} \right\| \varphi(\mathbf{x}_i, \mathbf{i}_i). \tag{11}$$

For $n = 1$ and $m = 1$, minimizing Eq. 4 with this unary potential results in the standard gray-edge algorithm [2].

In general, a large θ_p in Eq. 4, forces the whole image to have the same label resulting in a single illuminant estimate for the entire image. In that case, MIRF can encode a number of well-known color constancy algorithms by simply varying the error norm ρ and the patch weight w_i. If we look at the other extreme, where $\theta_p = 0$, every patch would take on the label of the illuminant which is closest (in an angular error sense) to its local estimate. However, the local estimates of the statistical color constancy algorithms are very noisy and this can lead to unsatisfactory results. This can be countered by choosing an intermediate θ_p (by means of cross validation), that enforces multiple neighboring patches to take on the same label, and thereby reducing the noise of the statistical estimate. Additionally, one can apply the following two adaptations to the unary potential in order to further improve robustness with respect to noisy statistical measurements.

1) Robust Error Norm: In order to reduce the influence of outliers on the energy functional, MRFs and CRFs often use an exponential error norm, e.g., [28], [29]. Thus, throughout this paper we use the robust error norm

$$\rho_{\sigma_r}(\mathbf{i}_i, \mathbf{x}_i) = 1 - \exp\left(-\frac{\alpha(\mathbf{i}_i, \mathbf{i}_j)^2}{2\sigma_r^2}\right), \tag{12}$$

where σ_r is a scaling parameter that controls the decay of the exponential function.

2) Uneven Color Balance: Statistical methods are known to be biased towards large segments of uniform color [30]. To counter this we propose the following adaptation:

$$\phi^s(\mathbf{x}_i | \mathcal{F}_i) = (w_i)^q \, \rho_{\sigma_r}\left((\mathbf{i}_i, \mathbf{x}_i)\right). \tag{13}$$

The parameter q allows the dampening of the effects of uneven color balance in an image. In our experiments we used $q \in \left\{0, \frac{1}{2}, 1\right\}$.

B. Physics-Based Color Constancy

Another family of color constancy methods is based on the physics of light-surface interactions [6]–[8], [11]. In this paper we focus on the approach by Tan *et al.* [8] because it is very competitive performance-wise and is applicable to a wider family of surfaces that exhibit a mixture of diffuse and specular reflectance. More specifically, we follow the extension by Riess *et al.* [31] which can be locally applied, even in patches that are just moderately specular.

Specularity-based approaches assume neutral interface reflection which states that the color of pure specularities is the color of the illuminant. These methods are often comprised of two steps: a specularity detection step, which identifies regions

which may contain specularities; and an illuminant estimation step, which exploits the color of the detected specular regions.

Tan *et al.* [8] introduce the inverse-intensity chromaticity (IIC) space. Per color channel, IIC is a 2D space where the horizontal axis represents the inverse intensity $1/\|\mathbf{f}^j\|_1$ and the vertical axis is a pixel's chromaticity. Thus, a pixel $\mathbf{f}^j = (f_R^j, f_G^j, f_B^j)^T$ is mapped to

$$f_C^j \rightarrow \left(\frac{1}{\|\mathbf{f}^j\|_1}, \frac{f_C^j}{\|\mathbf{f}^j\|_1} \right), \tag{14}$$

where $\|\cdot\|_1$ is the L_1 norm and $C \in \{R, G, B\}$.

The advantage of IIC is that the relationship between image chromaticity and illumination chromaticity becomes linear. According to [8], one generates per color channel a scatterplot of the pixels in IIC space. Pixels of the same material and albedo that are purely diffuse form a horizontal cluster. If, however, they also have a specular component, they form a triangular-shaped cluster. The base of the triangle intersects the diffuse horizontal cluster. The tip of the triangle intersects the vertical axis. This vertical-axis intercept is the corresponding illuminant chromaticity component i_C^j.

Riess *et al.* [31] also exploit this cluster geometry but they obtain multiple estimates per patch. More specifically, following [31], we subdivide each patch p_i into a rectangular grid of cells. The pixels of each cell k are projected onto IIC space (one IIC space per color channel). If the scatter plot of a cell k satisfies two straightforward shape criteria (i.e., it forms an elongated non-horizontal cluster) see [31], then an illuminant estimate $\mathbf{i}^k$ is obtained by analyzing the covariance ellipse of the scatter plot. The intersection of its major axis with the vertical axis of the IIC space becomes the illuminant color estimate i_k^c for color channel C in cell k.

The illuminant color estimates from all the cells in a patch p_i are collected in single color histograms $\mathcal{H}_{i,R}$, $\mathcal{H}_{i,G}$, $\mathcal{H}_{i,B}$, where $\mathcal{H}_{i,R}$ is the histogram of the red chromaticities of the illuminant estimates of the various cells in patch p_i. The other two histograms are similarly defined for the remaining two color channels. The illuminant estimate for a patch p_i is then determined by

$$\mathbf{i}_i = \underset{i_C^k}{\operatorname{argmax}}\, \mathcal{H}_{i,C}(i_C^k) \quad \forall C \in \{R, G, B\}, \tag{15}$$

where $\mathcal{H}_{i,C}(i_C^k)$ denotes the number of occurences of i_C in all the cells $k \in p_i$. For further details, please refer to [31].

Our physics-based unary potential is then defined as

$$\phi^p\left(\mathbf{x}_i|\mathcal{F}_i\right) = w_i \rho_{\sigma_r}\left(\mathbf{i}_i, \mathbf{x}_i\right), \tag{16}$$

where ρ_{σ_r} is the robust error norm of Eq. 12 and the weight w_i is proportional to the specular component of patch p_i. More specifically, we compute a measure of specularity as follows. Let S_i be the set of specular pixels in patch p_i. Similarly to Tan *et al.* [8], we consider a pixel to be specular if its brightness and saturation values exceed two corresponding thresholds $t_b = 0.2$ and $t_s = 0.8$. We define the following measure of a patch's specularity, $s_{\text{sp}} = \frac{1}{3}\sum_{j \in S_i}(f_R^j + f_G^j + f_B^j)$. One could set $w_i = s_{\text{sp}}$. However, the identification of specular pixels, especially when using simple thresholding, is error-prone. Therefore, in our implementation we use a binary weight, where $w_i = 1$ only if $s_{\text{sp}} \geq t_{\text{sp}}$, where t_{sp} is a patch-specularity threshold. Hence, the specularity-based illuminant estimate is only employed if a sufficiently large percentage of pixels within a patch are detected as specular.

C. Combining Statistics- and Physics-Based Illuminant Estimation

Both statistics- and physics-based illuminant estimation methods can be incorporated in a CRF framework using different unary potentials. An advantage of expressing each method as an energy minimization problem is that there is a natural way of combining them into a single color constancy technique by defining the local potential as

$$\phi\left(\mathbf{x}_i|\mathcal{F}_i\right) = (1 - \lambda_p)\phi^s\left(\mathbf{x}_i|\mathcal{F}_i\right) + \lambda_p \phi^p\left(\mathbf{x}_i|\mathcal{F}_i\right), \tag{17}$$

where λ_p is weighting the importance of the physics-based unary potential versus the statistics-based one. Thus, we can combine information from statistical cues, as well as specularities, into the final local illuminant estimate.

D. Constrained Illuminant Estimation

Constrained illuminant estimation methods have been popular because they allow the incorporation of prior knowledge about the illuminants. For example, several existing methods confine the illuminant set to lie on the Planckian locus [14]. Such constraints can be incorporated in the MIRF framework by placing restrictions on the illuminant label set $\mathcal{L}$. In this paper, we use a simple constraint where we exclude very saturated illuminants, i.e., $\{\forall m | \alpha\left(\mathbf{l}_m, \mathbf{i}_w\right) < \phi_d\}$, where $\mathbf{i}_w = \frac{1}{\sqrt{3}}(1, 1, 1)^{\mathrm{T}}$ is the white illuminant, $\mathbf{l}_m \in \mathcal{L}$ is an illuminant label and ϕ_d is an illuminant saturation threshold.

As a second constraint on the illuminants, we use the fact that in the majority of the multi-illuminant scenes only two dominant illuminants are typically present. Given a pair of labels $\mathbf{l}_m$ and $\mathbf{l}_n$, the optimal labeling $\check{\mathbf{x}}^*(m, n)$ for an observation $\mathcal{F}$ is determined by

$$\check{\mathbf{x}}^*(m, n) = \arg \min_{\check{\mathbf{x}} \in \mathbf{L}^{mn}} E\left(\check{\mathbf{x}}|\mathcal{F}\right), \tag{18}$$

where $\mathbf{L}^{mn}$ is the set of all possible labellings on $\mathcal{X}$ when using only illuminants $\mathbf{l}_m$ and $\mathbf{l}_n$. The two-illuminant constraint is enforced by finding the two labels which minimize the energy function. Thus, the selected illuminants are computed by

$$\hat{\mathbf{L}} = \underset{(\mathbf{l}_m, \mathbf{l}_n) \in \mathcal{L}^2}{\arg \min} \left(E(\check{\mathbf{x}}^*(m, n)|\mathcal{F}) \right). \tag{19}$$

Note that, this formulation also allows for single illuminant estimation when $m = n$.

V. Pairwise Potential

The purpose of the pairwise potential function, $\psi\left((\mathbf{x}_i, \mathbf{x}_j)|(\mathcal{F}_i, \mathcal{F}_j)\right)$ is to ensure, when appropriate, the smooth transition of labels in neighboring vertices. Similar to Boykov *et al.* [32] we consider pairwise potentials that resemble a well. In MRFs, especially as described in [32],

$\psi(\mathbf{x}_i, \mathbf{x}_j) = u(1 - \delta_{ij})$, where u is the well "depth" and the function $(1 - \delta_{ij})$ controls the shape of the well. In [32], u is defined as a constant and δ_{ij} is the unit impulse function.

In a CRF (see also [29]) the "depth" depends on the observations $u = h(\mathcal{F}_i, \mathcal{F}_j)$. Thus, our pairwise potential function has the form

$$\psi\left((\mathbf{x}_i, \mathbf{x}_j)|(\mathcal{F}_i, \mathcal{F}_j)\right) = h(\mathcal{F}_i, \mathcal{F}_j)(1 - \delta_{ij}), \qquad (20)$$

where we use $\delta_{ij} = \delta(\mathbf{x_i} - \mathbf{x_j})$. If two neighboring labels are distinct, then there are two possibilities. It can be that the two patches, though spatially close, are illuminated by distinct illuminants. In such cases, we should allow for a transition in labels and not significantly penalize the difference in their values. It may, however, be the case that an erroneous label was assigned and the two patches are illuminated by the same illuminant. The goal of the depth function $h(\mathcal{F}_i, \mathcal{F}_j)$ is to distinguish between erroneous versus justified differences in adjacent labels.

This is a challenging task. We tried a number of different depth functions. We experimented with a function that is based on the photometric classification of edges using the photometric quasi-invariants [33]. We also investigated the employment of the Retinex [9] heuristic which states that illumination is expected to vary slowly, thus large changes in surface reflectance are due to differences in material. Within our system we obtained the best results when we applied the following function:

$$h(\mathcal{F}_i, \mathcal{F}_j) = \text{boundary\_length}\left(\mathcal{F}_i, \mathcal{F}_j\right), \qquad (21)$$

where boundary_length is the number of pixels in patch p_i which have at least one neighboring pixel (assuming 4-connectedness) belonging to p_j. This depth function ensures that the pairwise potential between two patches p_i and p_j is proportional to the shared border between the two patches.

VI. MIRF Algorithm

The main concept of MIRF is to segment an image in patches, obtain local illuminant estimates per patch and then consolidate the local estimates using a CRF. Such a general framework allows for the incorporation of many existing methodologies. This, however, also implies that several design decisions should be made during implementation.

A. Implementation Details

The very first step of MIRF involves the formation of patches. There are several ways for dividing an image into subregions. Superpixels generated by a segmentation algorithm tend to follow object boundaries and significant illumination changes. A uniform grid, on the other hand, provides more diverse content per patch, and thus more information for statistics-based estimators. In MIRF we experimented with both types of tessellations. Overall, our methodology is not very sensitive to the image-subdivision method. Image segmentations based on [34] and [35] and uniform grids produced comparable results (see Section VIII). Hence, we decided to use a uniform-grid tessellation for ease of computation. All the reported results in this paper were produced using a grid.

Algorithm 2 MIRF Detailed Algorithm

1: Divide an image into a set of patches $P = \{p_1, p_2, ..., p_N\}$
2: Extract at each patch the local illuminant colors $\mathbf{I} = \{\mathbf{i}_1, \mathbf{i}_2, ..., \mathbf{i}_N\}$ (using Eq. 10 and Eq. 15).
3: Cluster the illuminants (using K-means) and get the K centers. Add an avg image illuminant estimate $\mathbf{I}_0$.
4: Decrease the number of labels by grouping labels whose angular distance is less than 0.5°. $\mathcal{L}$ now denotes the reduced set of illuminant labels.
5: **for all** $\mathbf{l}_m$ and $\mathbf{l}_n \in \mathcal{L}$ **do**
6: Calculate $\breve{\mathbf{x}}^*(m, n)$ (Eq. 18)
7: **end for**
8: Select the pair of illuminants $\hat{\mathbf{L}}$ (Eq. 19) which generates the MAP labelling $\breve{\mathbf{x}}^*$.
9: Backproject $\breve{\mathbf{x}}^*$ to form an illumination map M.
10: Apply Gaussian smoothing on M to soften possible artificial edges induced by the image tessellation.

Once an image is split into patches, we obtain per patch a statistics-based illuminant estimate using Eq. 10 and, if the patch is sufficiently specular (see end of Section IV-B), an additional physics-based estimate according to Eq. 15. We then create our set of illuminant labels, by separately grouping the statistics-based estimates into K^s clusters and the physics-based ones into K^p clusters, for a total of $K = K^s + K^p$ clusters. In both cases, we use K-means clustering based on the angular distance (see Eq. 6) of the illuminant chromaticity values $i_C/\|\mathbf{i}\|_1$, where $C \in \{R, G, B\}$. The mean values of the K clusters become the set of illuminant labels $\mathcal{L}$. We also add a single "ambient" illuminant estimate $\mathbf{I}_0$ to the illuminant labels, $\mathcal{L} = \mathcal{L} \cup \{\mathbf{I}_0\}$, where $\mathbf{I}_0$ is obtained by applying Eq. 10 on the entire image. In order to reduce the computational cost, we further decrease the cardinality of $\mathcal{L}$ by merging labels which differ by less 0.5°.

We then evaluate Eq. 18, i.e. for every possible pair of labels $\mathbf{l}_m, \mathbf{l}_n \in \mathcal{L}$. We use Bagon's [36] graph-cut Matlab implementation to minimize the CRF's energy function shown in Eq. 4. Unary potentials are computed using Eq. 13 and Eq. 16. The pair of labels and their corresponding labelling that minimizes the CRF energy function is then chosen (see Eq. 19) as the solution to the multi-illuminant labelling problem. Finally, the derived label colors are assigned to the corresponding patches and the estimated illumination map M is generated. In the last step of the algorithm, a Gaussian smoothing filter with standard deviation σ_p is applied to M as a post-processing step in order to reduce artifacts created by patch boundaries. The methodology is compactly presented below.

B. Parameters

Unless explicitly stated otherwise, the following parameter settings were used in generating the results reported in this paper. First of all, we always used a 20×20 uniform rectangular grid to tessellate the images into patches. The only exception was the outdoor images of Gijsenij *et al.* [23], where, due to their smaller size, we employed a 10×10 grid. In all cases

the resulting patch-size was approximately 15×20 pixels. The number of cluster centers for the K-means algorithm was set to $K^p \leq K^s = \sqrt{N}$, where N is the number of patches.

For the physics-based illuminant estimates, we set the brightness and saturation thresholds to $t_b = 0.2$ and $t_s = 0.8$ respectively, and the overall specularity threshold to $t_{\text{sp}} = 10$ (pixel values are in $[0, 1]$). The cell size within a single patch was 20×20 pixels with a step size of 10 pixels,[1] as in [31]. The parameters of the CRF itself were as follows. The saturation threshold ϕ_d for illuminant labels (see Section IV-D) was set to 15°. The parameter σ_r in Eq. 12 for robust thresholding on the unary potentials was 2.5°. Finally, we used a standard deviation $\sigma_p = 10$ for the Gaussian smoothing of the reprojected illuminant labels.

Three parameter values were determined via two-fold cross validation on each of the datasets on which MIRF was evaluated. These were the weighting between unary and pairwise potentials θ_p (see Eq. 4), the power q (see Eq. 13) for computing the unary potentials, and finally, λ_p (see Eq. 17) for the relative influence of the statistics- and physics-based estimators.

VII. Multi-Illuminant Multi-Object Dataset

Several datasets are available for single illuminant estimation. The first datasets were taken under laboratory settings with controlled illumination [30], [37]. Later datasets – often much larger – consisted of images of real-world scenes where the ground truth was computed from a reflectance target in the scene, which was either a gray ball [38] or a Macbeth color checker [39]. Recently, Gijsenij *et al.* [23] introduced a multi-illuminant dataset, where the ground truth consisted of manually annotating the areas where each light source was dominant. Such manual annotations, however, are difficult to do on complex scenes and error-prone.

To address these limitations, we created two new datasets[2] for multi-illuminant estimation: one contains images taken under a controlled laboratory setting, and one consists of real-world indoor and outdoor scenes. Each of the sets includes complex scenes with multiple reflectances and specularities. A variety of lighting conditions and illuminant colors are present in the datasets. Instead of manually annotating the ground truth like [23], we computed a pixel-wise ground truth for our images. This way we avoid the subjective task of manually segmenting the image and obtain high-resolution ground truth. In addition, this further allowed us to estimate the contribution of each illuminant per pixel, rather than making a binary decision on the dominant illuminant at each pixel. As can be seen in Fig. 2 large regions in the image are lit by both illuminants.

A. Data Acquisition

We used a Sigma SD10 single-lens reflex (SLR) digital camera which uses a Foveon X3 sensor. We specifically chose this photographic equipment for its Bayer-pattern-free image acquisition and for its lossless raw 12 bit per color high quality output. We captured the images in linear RAW format, i.e. without additional gamma or JPEG compression. The original image size is 2304×1531 pixels, i.e. roughly 3.5 megapixels. However, in order to reduce the computational load of our experimental evaluation, we rescaled the images to 20% of their original size.

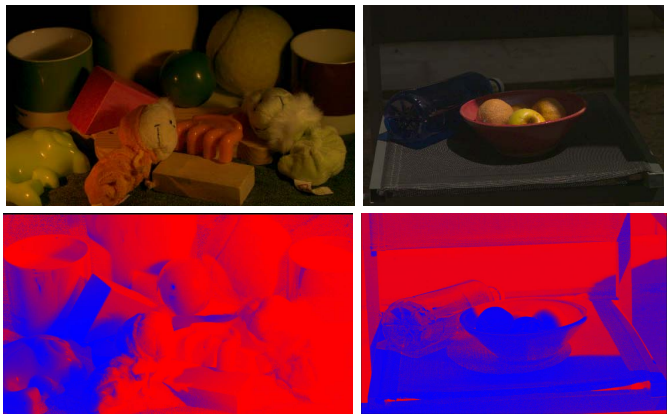

Fig. 2. Example images from our laboratory (top-left) and real-world (top-right) datasets. In the bottom row, the relative influence of the two illuminants is shown, color-coded blue and red.

1) Laboratory Dataset: The first dataset is taken in a controlled laboratory setting. The scenes vary from simple single-object setups to more challenging multi-object (cluttered) settings. The scene content also varies between diffuse objects, specular objects, and mixtures of diffuse and specular objects. In total, the laboratory database consists of 10 scenes, each taken under 6 distinct illumination conditions. We used three differently colored lights, referred to as "blue", "white" and "red", with a chromatic difference of 5.9° between blue and white, 6.1° between white and red, and 11.4° between blue and red. Each scene was lit by distinct pairs of illuminants from two different angles (referred to as "left" and "right" illumination). For computing the ground truth, each scene was also captured under only a single illuminant from each position. In a couple of images the ground truth and the multi-illuminant images were misaligned. As a result, after removal of these images, our laboratory dataset is composed of a total of 58 benchmark images. To reduce the influence of ambient illumination, the data acquisition for this set was performed in a box with black diffuse walls.

The top-left images in Figs. 1 and 2 show two example scenes illuminated by our red light source from the left side and our white illuminant from the right. The lion in Fig. 1 is an example of a single object scene and the toys in Fig. 2 are an example of a mixture of specular and diffuse objects. The bottom row shows the influence of the two illuminants. A stronger blue component denotes stronger influence of the left illuminant, while red represent the illuminant on the right.

2) Real-World Dataset: In order to evaluate our framework on more challenging real world images, we captured 20 additional indoor and outdoor scenes. Here the data is converted to sRGB to mimic a more typical user setting. We also

[1]Note that this may lead to obtaining a single estimate per patch, i.e. the voting aspect is essentially eliminated. However, when MIRF is applied on larger images (or patches, respectively), the histogram voting is employed.

[2]http://www5.cs.fau.de/research/data/two-illuminant-dataset-with-computed-ground-truth/

generated perceptually enhanced versions of the real-world images (using the publicly available code of Parraga *et al.* [40]) in order to increase the overall spread of colors in an image. The scenes contain two dominant illuminants, namely an ambient light source and a direct light. In the outdoor images, shadow regions correspond to ambient light. In the indoor images, the overall room illumination is treated as ambient light. A direct light source is added either via a projector, the sun, or an additional light bulb. The top-right images in Figs. 1 and 2 depict two such examples. One scene shows a two-illuminant indoor scene, while the second shows a strong color shadow in an outdoor scene. The main difference between the two datasets is that in the real-world scenes, the ambient illuminant is present on almost the entire image area, while the direct illuminant covers only a part of each scene.

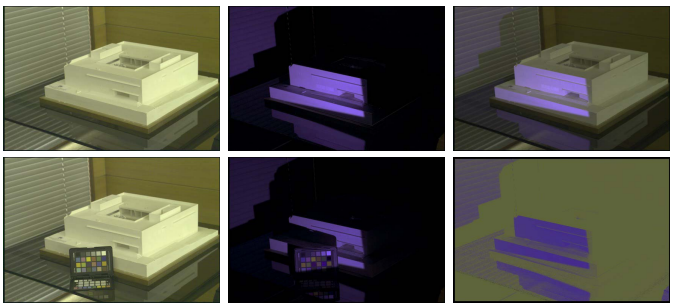

Fig. 3. Example images required fot the ground truth computation. In the top row, from left to right: a scene under illuminant $\mathbf{i}_a$, $\mathbf{i}_b$ and $\mathbf{i}_a + \mathbf{i}_b$. In the bottom row: the same scene together with a Macbeth color checker under illuminants $\mathbf{i}_a$ only and $\mathbf{i}_b$ only. On the bottom right, the ground truth is shown, i.e. the two illuminant colors and their spatial distribution. Gamma is added to the images for visualization.

B. Ground Truth Computation

In order to compute the contribution of each illuminant per pixel, we need to capture additional images. Consider, for example, a scene that is simultaneously illuminated by two light sources $\mathbf{i}_a$ and $\mathbf{i}_b$. Without loss of generality let $\mathbf{i}_a$ be the left illuminant and $\mathbf{i}_b$ be the right illuminant. We take two additional images of exactly the same scene: one under only the left illuminant $\mathbf{i}_a$ and one under only the right one $\mathbf{i}_b$. We also capture two further single illuminant images of the same scene again under only $\mathbf{i}_a$ and under only $\mathbf{i}_b$ but this time we also include either a Macbeth color checker or a gray reflectance target. From these last two pictures we can estimate the illuminant color. Fig. 3 shows a multi-illuminant scene (top-left) together with the additional images required for the ground truth computation and the extracted ground truth image. For the laboratory dataset we ony capture three images per multi-illuminant scene. We do not capture the pictures with the color calibration target. Instead, it suffices to take the single illuminant images of the empty capture box (which acts as a backdrop for all laboratory acquisitions) with a gray reflectance target that covers most of the scene.

To compute the ground truth we exploit the linearity of light: we use the fact that the scene taken under both illuminants is equal to the sum of the two scenes taken under a single illuminant. We verified that this assumption approximately holds for our Sigma SD10 SLR camera. A pixel $\mathbf{f}_{ab}$ from a two-illuminant scene is equal to the sum of the pixels from the two images of the same scene under a single illuminant, i.e. $\mathbf{f}_{ab} = \mathbf{f}_a + \mathbf{f}_b$. As a first step, we use the single illuminant images to derive per pixel the contribution of each illuminant in the multi-illuminant scene. The chromaticity of each illuminant, $\mathbf{i}_a$ and $\mathbf{i}_b$, is extracted from the color calibration target. Under the von Kries assumption, when we divide each single illuminant image by its respective illuminant chromaticity, $\hat{f}_{a,C} = f_{a,C}/i_{a,C}$ and $\hat{f}_{b,C} = f_{b,C}/i_{b,C}$ where $C \in \{R, G, B\}$, we obtain the scenes under white illumination. The observed intensity differences in $\hat{\mathbf{f}}_a$ and $\hat{\mathbf{f}}_b$ are due to the individual contribution of each illuminant.

The per-pixel relative contribution of illuminant $\mathbf{i}_a$ in the multi-illuminant image $\mathbf{f}_{ab}$ for the green channel is then obtained by

$$r = \frac{\hat{f}_{a,G}}{\hat{f}_{a,G} + \hat{f}_{b,G}}. \tag{22}$$

The details of this derivation can be found in Appendix IX-B. In principle, any color channel could be used. However, we found that the green channel yielded the most stable (noise-resilient) results.

The ground truth illuminant color $\mathbf{i}_{ab}$ at a pixel f_{ab} of the mixed-illuminant image is then a pixelwise linear interpolation of $\mathbf{i}_a$ and $\mathbf{i}_b$,

$$\mathbf{i}_{ab} = r \cdot \mathbf{i}_a + (1 - r)\mathbf{i}_b. \tag{23}$$

Note that, since the sum of the two single illuminants is equivalent to the two-illuminant scene, two of the three scenes suffice for the ground truth computation. We use this fact for the scenes where it was not possible to obtain the two single-illuminant images. Consider, for example, an indoor scene with indoor illuminant and outdoor light coming through the window (which we were unable to block). In such a case we can take two images - one with both illuminants and one with only the outdoor lighting after switching off the indoor light source - and still compute the ground truth.

VIII. Experiment

In this section we compare the performance of the proposed MIRF method to several other approaches. Our error metric per image is the mean pixelwise angular distance (Eq. 6) between the estimated illuminant color and the corresponding ground truth. Very dark pixels (i.e., for our 12-bit images, pixels with intensities below 50) were excluded from the evaluation due to their relatively low signal-to-noise ratio. We used the per-image errors, to compute the median and mean errors per dataset. The evaluation was conducted on three datasets: our laboratory dataset, our real-world dataset, and the outdoor dataset that has been used by Gijsenij *et al.* [23].

There is little prior work on methodologies explicitly designed for non-uniform illumination. As a baseline, we evaluated the performance on multi-illuminant scenes of a number of well-established color constancy algorithms that assume uniform illumination. We also implemented the recent

TABLE I
COMPARATIVE PERFORMANCE ON OUR LABORATORY DATASET

	Single-illuminant		Gijsenij *et al.*		MIRF	
	Mean	Median	Mean	Median	Mean	Median
DN	10.6°	10.5°	-	-	-	-
GW	3.2°	2.9°	6.4°	5.9°	3.1° (-3%)	2.8° (-3%)
WP	7.8°	7.6°	5.1°	4.2°	3.0°(-41%)	2.8°(-33%)
GE1	3.1°	2.8°	4.8°	4.2°	2.7°(-13%)	**2.6°** (-7%)
GE2	3.2°	2.9°	5.9°	5.7°	**2.6°**(-19%)	**2.6°**(-10%)
IEbV	8.5°	8.3°	-	-	4.5°(-47%)	3.0°(-64%)

TABLE II
COMPARATIVE PERFORMANCE ON OUR REAL-WORLD DATASET

	Single-illuminant		Gijsenij *et al.*		MIRF	
	Mean	Median	Mean	Median	Mean	Median
DN	8.8°	8.9°	-	-	-	-
GW	5.2°	4.2°	4.4°	4.3°	3.7°(-16%)	3.4°(-19%)
WP	6.8°	5.6°	4.2°	3.8°	**4.1°** (-2%)	**3.3°**(-13%)
GE1	5.3°	3.9°	9.1°	9.2°	4.0°(-25%)	3.4°(-13%)
GE2	6.0°	4.7°	12.4°	12.4°	4.9°(-18%)	4.5° (-4%)
IEbV	6.0°	4.9°	-	-	5.6° (-7%)	4.3°(-12%)

TABLE III
COMPARATIVE PERFORMANCE ON THE GAMMA-CORRECTED VERSION OF THE OUTDOOR DATASET BY GIJSENIJ *et al.* [23]

	Single-illuminant		Gijsenij *et al.*		MIRF	
	Mean	Median	Mean	Median	Mean	Median
DN	4.4°	3.6°	-	-	-	-
GW	15.0°	13.8°	12.2°	13.8°	10.0°(-18%)	10.1°(-27%)
WP	10.3°	11.3°	10.0°	8.4°	7.7°(-23%)	6.4°(-24%)
GE1	10.1°	10.1°	8.5°	7.6°	**7.1°**(-16%)	**4.7°**(-38%)
GE2	8.7°	8.5°	8.1°	7.4°	7.2°(-11%)	5.0°(-32%)
IEbV	10.0°	7.3°	-	-	9.3° (-7%)	7.3° (-0%)

multi-illuminant method by Gijsenij *et al.* [23], as it showed very competitive performance in a number of experiments.

Both the method by Gijsenij *et al.* [23] and MIRF use as input illuminant estimates with small spatial support. Such illuminant estimates can be obtained from different estimators. We chose to use gray world ("GW"), which can be obtained from Eq. 10 by using the parameters $n = 0$, $m = 1$, $\sigma_{\text{GW}} = 0$, white patch ("WP", with $n = 0$, $m = \infty$, $\sigma_{\text{GW}} = 0$), first order gray-edge ("GE1", with $n = 1$, $m = 1$, $\sigma_{\text{GW}} = 1$) and second-order gray edge ("GE2", with $n = 2$, $m = 1$, $\sigma_{\text{GW}} = 1$). Additionally, we use the physics-based estimator, as presented in Eq. 15, denoted as "IEbV" (derived from "Illuminant Estimation by Voting"). We used these base estimators for comparing the performance of the three families of methods as described above. Additionally, we provide a "do nothing" ("DN") error, where the image is assumed to be already perfectly white balanced.

A. Comparing Single- and Multi-Illuminant Methods

In Table I, we present the mean and median errors on our laboratory dataset. The "single-illuminant" column, shows the results of a single global illuminant estimate. The columns "Gijsenij *et al.*" and "MIRF" report results for the multi-illuminant methods by Gijsenij *et al.* [23] and our proposed algorithm "Multi-Illuminant Random Field" accordingly.

Some single-illuminant estimators, namely GW, GE1 and GE2, already perform relatively well on our laboratory dataset. This is mainly because the two colors of the two light sources illuminating the scene are not very distinct (see Section VII). Thus, the overall error can be small, even if only one of the two illuminants (or a color in between the two) is reported as global estimate. However, in all cases, MIRF improves over these estimates. The physics-based estimates for IEbV yield a considerably weaker performance in the mean error, which might be due to the fact that the individual patches are relatively small and thus the voting becomes ineffective. The method by Gijsenij *et al.* [23] performed surprisingly unsatisfactorily, even worse than the single-illuminant estimators. We investigated this case more closely. It turned out that relatively often, weak candidate estimates were selected by this method, which in turn penalized the overall algorithm. MIRF avoids this particular problem, as the remaining energy from the energy minimization is used as a criterion for the quality of a solution. In Section VIII-B, we excluded this source of error by directly comparing the performance of only determining the distribution of illuminants.

Table II shows the performance of the same algorithms on our real-world dataset. One can observe similar tendencies in the results. Note that, the overall errors are higher, mainly because these images were perceptually enhanced in order to increase the overall spread of colors in an image. The largest performance gain is obtained by the localized estimates of the physics-based method, IEbV. This improvement is mostly due to the robust error metric, which suppresses gross outliers in the physics-based estimates.

We also evaluated the same set of algorithms on the outdoor dataset by Gijsenij *et al.* [23], see Table III. Note that, the reported results for the method by Gijsenij *et al.* deviate from what the authors reported in their paper. Upon further investigation, we realized that the evaluation in [23] was conducted on the non-gamma-corrected images. Without gamma correction, we obtain the same numbers as reported in [23]. In our implementation, we performed gamma correction on the input images, as it was also originally intended by [23]. Overall, the errors are higher than in the previous two experiments. This can be attributed to the nature of the images in this dataset: they are relatively small, consisting mostly of two relatively homogeneous regions of sunlight and shadow. Thus, the underlying localized illuminant color estimators are extracting information from relatively uninformative input. Note that, we did not evaluate on the laboratory data by Gijsenij *et al.*, as we found upon manual inspection that the ground truth for these images is not very reliable.

Lastly, we evaluated the performance of MIRF in single-illuminant settings. We used the SFU single-illuminant laboratory dataset [41], which contains 30 scenes taken under eleven known illuminants. After removal of several images, which were considered unusable, the dataset consisted of 321 images. The pixelwise ground truth was computed by assigning the single illuminant to all the pixels in the image. We applied two-fold cross-validation to train the parameters of our algorithm. Table IV shows the results of our method in comparison to Gijsenij *et al.* [23] and the single illuminant methods. Our method outperforms the multi-illuminant

TABLE IV
COMPARATIVE PERFORMANCE ON THE SFU SINGLE-ILLUMINANT LABORATORY DATASET

	Single-illuminant		Gijsenij *et al.*		MIRF	
	Mean	Median	Mean	Median	Mean	Median
DN	17.3°	15.6°	-	-	-	-
GW	9.8°	7.0°	13.3°	11.6°	9.7°(-1%)	**5.2°**(-26%)
WP	9.1°	6.5°	11.6°	9.6°	8.8° (-3%)	5.4°(-17%)
GE1	6.9°	5.8°	9.9°	9.0°	**6.8°**(-1%)	**5.2°**(-10%)
GE2	8.3°	7.8°	10.5°	9.5°	8.7°(+5%)	8.1° (+4%)

TABLE V
PERFORMANCE ANALYSIS OF THE SPATIAL DISTRIBUTION. METHODS EVALUATED ON OUR LABORATORY DATASET. THE TRUE ILLUMINANT COLORS ACT AS DOMINANT ILLUMINANTS TO THE METHODS

	Gijsenij *et al.*		MIRF	
	Mean	Median	Mean	Median
GW	2.4°	2.3°	2.3°	2.3°
WP	2.2°	2.1°	2.0°	1.9°
GE1	2.1°	2.0°	**1.8°**	**1.7°**
GE2	2.2°	2.1°	1.9°	1.8°

method of Gijsenij *et al.* The cross validation actually learns that the smoothing parameter should be maximal and forces the algorithm to a single illuminant solution for all patches in an image. MIRF's deviations from the standard single illuminant methods are mainly caused by our use of the robust error norm. In most cases (except GE2) this leads to improved results even in this single illuminant dataset.

B. Benchmarking Separate Components of the Algorithm

Estimating multiple illuminants can be considered as two interleaved tasks, namely estimating the illuminant colors and their spatial distribution. The recovery of the spatial distribution was not required for single-illuminant estimators. Hence, we empirically investigated the capability of finding the proper spatial distribution. We isolated the spatial distribution aspect of the two multi-illuminant methods, i.e. MIRF and [23] by providing the ground truth illuminant colors as dominant illuminants.

The results on our laboratory dataset are shown in Table V. In the left two columns, it can be seen that the performance of the method by Gijsenij *et al.* greatly improved, compared to Table I. This implies that the selection of the correct dominant illuminant color is one of the major challenges in the method of Gijsenij *et al.* In the right columns, we show the performance of our proposed method. The best performing local estimator in both methods is the first order gray edge, with a median error at MIRF of 1.7°. This shows that the spatial distribution of the illuminants is well approximated by our proposed framework.

In another experiment, we investigated the impact of the various improvements we have introduced to MIRF (see Table VI). We focus on only one local illuminant estimator, gray world ("GW"), since it was consistently among the best local performers. If we remove the constraint of two illuminants and allow an arbitrary number of illuminants, the error increases significantly on our two datasets. Similarly, the robust error norm (see Eq. 12) yields an important performance

TABLE VI
GRAY-WORLD RESULTS FOR DIFFERENT CONFIGURATIONS OF THE PROPOSED FRAMEWORK FOR EACH DATASET

	Laboratory data		Real-world data		Gijsenij *et al.*	
	Mean	Median	Mean	Median	Mean	Median
original MIRF	3.1°	2.8°	3.7°	3.4°	10.0°	10.1°
any # of lights	4.6°	4.0°	4.2°	4.0°	10.0°	10.2°
w/o Eq. 12	3.9°	3.7°	4.3°	4.0°	10.1°	10.1°
$q = 1$	3.0°	2.8°	3.6°	3.3°	10.7°	10.3°
w/o ϕ_d	3.6°	3.3°	4.6°	3.2°	11.2°	10.1°
superpixels	3.1°	2.8°	3.5°	3.6°	-	-
w/o smoothing	3.2°	2.9°	3.8°	3.4°	10.0°	10.1°

TABLE VII
IMPACT OF CLUSTERING ON MIRF. GRAY-WORLD IS USED AS THE LOCAL ILLUMINANT ESTIMATOR

	Laboratory data		Real-world data	
	Mean	Median	Mean	Median
$K^s = \sqrt{N}$	3.1°	2.8°	3.7°	3.4°
$K^s = 50$	2.9°	2.7°	3.8°	3.7°
$K^s = 100$	3.0°	2.7°	3.6°	3.3°
no clustering	4.6°	4.0°	4.2°	4.0°

TABLE VIII
COMBINATION OF PHYSICS-BASED AND STATISTICS-BASED METHODS ON OUR LABORATORY DATASET

	Laboratory data		Real-world data	
Combination variant	Mean	Median	Mean	Median
IEbV-GW	3.0°	2.8°	4.2°	4.3°
IEbV-WP	2.7°	2.5°	4.0°	3.4°
IEbV-GE1	2.6°	2.4°	4.5°	4.2°
IEbV-GE2	2.8°	2.7°	4.7°	3.9°

gain on both our datasets. Removing the parameter q from Eq. 13 which counters uneven color balances only affects results on the Gijsenij dataset. Finally, removing the saturation constraint ϕ_d on the illuminants (see Section IV-D) results in a performance drop on all datasets. We also evaluated the impact of using superpixels [34] instead of a regular grid. The results were identical for our laboratory dataset. In our real-world dataset superpixels result in a lower mean error but a larger number of outliers. This is mainly caused by the type of scenes in our databases. Our laboratory images, unlike our real-world ones, include mostly textureless objects. A color segmentation method in textured scenes will generate small uniform patches, which often handicap statistics-based methods. Due to the small size of the images in the Gijsenij dataset, we were unable to obtain meaningful segmentation results. Another factor we examined was the influence of the Gaussian smoothing of the derived illumination map M (Step 10 of the MIRF Detailed Algorithm). As can be seen in Table VI (last row), this post-processing step only affects few values by an improvement of 0.1°.

We also examined the effect of the number of clusters K on the accuracy of MIRF. We once again focus on only one of the best-performing local illuminant estimator, i.e. gray world ("GW"). We analysed the impact of clustering on our two datasets. We considered four different values of K^s (recall $K = K^s + K^p$ and K^s acts as an upper bound of K^p):

Fig. 4. Examples of automatic white balance (WB). From top to bottom the rows present: original images, the corresponding WB images using the ground truth, global gray world, Gijsenij *et al.* [23], and MIRF accordingly. Note that the images have been converted to sRGB for better visualization. The captions of each image denote the mean angular error in the WB image.

$K^s = \sqrt{N}$ which is the default value for all our MIRF evaluations, $K^s = 50$, $K^s = 100$, and no clustering. This last option allows every local estimate to be in the label set without any additional constraints on the dominant illuminants. As can be seen in Table VII, clustering helps not only with the computational speed but also with accuracy since it diminishes the effect of outlier local-estimates. The number of clusters itself does not significantly influence the outcome of MIRF, resulting in approximately 0.1° variation in performance.

One last aspect we analyzed was the data set dependence of the optimal parameters learned through cross-validation. We, therefore, tested MIRF on our laboratory dataset using the average value of the parameters we obtained via cross-validation on the real-world data and vice-versa. In both cases we observed an approximately 0.1° improvement in accuracy. When we used the parameters of the real-world data on the laboratory dataset, the mean and median angular errors dropped to 3.0° and 2.7°, respectively. Similarly, the mean and median errors dropped to 3.5° and 3.3° respectively, when we used the laboratory-data parameters on the real-world dataset. We believe that this improvement is caused by our using the parameters of one fold in evaluating another fold, which can sometimes result in over-training in some folds.

C. Combination of Statistics- and Physics-Based Estimates

Table VIII demonstrates another benefit of our framework. By defining the unary potentials as a weighted sum of statistics-based and physics-based potentials, we are able to combine cues from multiple methods in a natural way. To determine appropriate values of the unifying parameters θ_p, q and λ_p (see Section VI-B) we performed a full cross-validation over them. It turns out, that such a combination can indeed further improve the results (compare Table VIII (left) and Table I), in particular for the white patch and first order gray edge estimates. On the other hand, the MIRF performance slightly dropped for the combination of IEbV with GE2. Thus, there is no guarantee that a combination of the unary potentials brings a performance gain.

The two right columns of Table VIII show the performance on our proposed real-world dataset. It is interesting to note that the impact of the combined unary potentials on the overall performance is quite different from the experiments on the laboratory data. Here, the majority of the results are slightly worse than those reported in Table II. This behavior, however, is not consistent. For instance, the mean error of IEbV-WP is only slightly smaller than the reported error in Table II. A similar behavior is observed for the median error for IEbV-GE2. From these results, we conclude that the framework is general enough to allow the straightforward integration of multiple cues. However, whether such a combination indeed brings the desired performance gain has to be investigated on a case-by-case basis.

D. Automatic White Balance

Example results for automatic white balancing using the derived illuminant colors and distributions are shown in Fig. 4. All images are contrast-enhanced for improved visualization. In the top row, from left to right, the input scenes "toys", "lion", "camera", and "detergents" are presented. The second row shows perfectly white balanced output using the computed ground truth. At the bottom right corner of the remaining images we display the average angular error between the perfectly white-balanced and the derived white-balanced image. The third row shows white balancing results for a single global gray-world estimator. The resulting images suffer from a color cast, as both illuminant colors in the scene are corrected with only one estimate. Using the same local estimator within the framework by Gijsenij *et al.* [23] (fourth row) clearly improves over the global estimator. However, the images look more greyish and with faded colors as the local estimations were not able to fully separate the effects of illumination from the object color. Also the "lion" appears more reddish on the right side. Finally, in the last row, the output of the proposed MIRF method is shown. One can notice here the absence of global color cast, due to the improved selection of the dominant illuminant color. Nonetheless, some inaccuracies in the estimation of the spatial distribution of the illuminants may lead to local color casts (e.g., several bluish "blobs" overlay considerable regions of the "camera" image). However, the overall performance of MIRF is, in general, consistently strong, as demonstrated in the "toys" and "detergents" images.

IX. Conclusion

We developed a new algorithm, "Multi-Illuminant Random Field" (MIRF), for color constancy under non-uniform illumination. In scenes that are illuminated by multiple light sources, one should estimate not only the illuminant colors but also their spatial distribution. In our approach, these two tasks are jointly solved within an energy minimization framework. Our methodology is general enough to allow: 1) the natural combination of different illuminant estimators, like statistics- and physics-based techniques; and 2) the incorporation of additional cues when available, like, for instance, estimates for illuminant edges.

For quantitative evaluation, we created two highly accurate ground truth datasets for scenes under two illuminants. One database consists of 58 laboratory images and the other of 20 real-world pictures. In contrast to prior work, the spatial distribution of the illuminant colors is computed from multiple, spatially aligned input images. Performance evaluation on these images and on the real-world dataset by Gijsenij *et al.* [23] are promising. MIRF outperforms single-illuminant estimators. Furthermore, we showed that MIRF's joint estimation of the illuminant color and its spatial distribution consistently outperforms the recently proposed method by Gijsenij *et al.* [23], which solves these two steps separately. In an experiment with ground-truth illuminant colors, we also showed that the individual tasks of color estimation and localization perform superiorly. A combination of statistics- and physics-based estimates yields competitive results.

In this paper, influenced by our own laboratory setup, we mainly focused on two-dominant illuminant images. Determining the number of dominant illuminants in a scene is an interesting problem that calls for further investigation. We believe that an analysis of the distribution of local illuminant estimates in chromaticity space should provide an insight in this topic. We are planning to further explore this aspect as part of our future work. Another extension to this paper that is worth investigating is the incorporation of top-down semantic cues into the framework [42], [43]. Recognition of common materials in the scene such as grass, stone, and faces could further improve multi-illuminant estimation. An important aspect of using a CRF in multi-illuminant setups is the identification of image regions with a relatively abrupt transition in illumination, so that the pairwise potential can be accordingly adjusted. This, however, is a non-trivial task. We are already investigating more sophisticated solutions than the ones proposed in this paper.

Appendix

A. Equivalence With Gray World Algorithm

In this appendix we prove that using Eq. 9 as the unary potential and for a large θ_p in Eq. 4 MIRF yields the same estimate as the standard gray-world algorithm. If the weight of the pairwise potential θ_p is sufficiently large, no label changes will be allowed between neighboring patches. As a consequence, all patches will have the same illuminant estimate, which will essentially be determined by the unary potential. MIRF will then select the illuminant which yields, summed over all patches, the lowest energy. Let $\mathbf{x}$ the illuminant choice for all patches (we drop the subscript on $\mathbf{x}$ since the estimate is equal for all patches), MIRF's energy function can then be written as

$$E(\mathbf{x}|\mathcal{F}) = \sum_{i\in\mathcal{V}} \phi(\mathbf{x}|\mathcal{F}_i) = \sum_i \left\| \sum_{j\in p_i} \mathbf{f}^j \right\| \big(\varphi\,(\mathbf{i}_i, \mathbf{x})\big), \qquad (24)$$

where we used $\varphi\,(\mathbf{i}_i, \mathbf{x}) = (1 - \mathbf{i}_i^{\mathrm{T}}\mathbf{x})$. Replacing $\mathbf{i}_i$ with Eq.5 we obtain

$$E(\mathbf{x}|\mathcal{F}) = \sum_i \left\| \sum_{j\in p_i} \mathbf{f}^j \right\| - \sum_i \Big(\sum_{j\in p_i} \mathbf{f}^j \Big)^{\mathrm{T}} \mathbf{x}, \qquad (25)$$

where only the second term depends on $\mathbf{x}$. Since we want to compute $\operatorname{argmin}_{\mathbf{x}} E(\mathbf{x}|\mathcal{F})$, this is equal to maximizing the second term of the equation

$$\mathbf{x}^* = \operatorname*{argmax}_{\mathbf{x}} \sum_i \left(\sum_{j \in p_i} \mathbf{f}^j \right)^{\mathrm{T}} \mathbf{x} = \operatorname*{argmax}_{\mathbf{x}} \left(\sum_i \sum_{j \in p_i} \mathbf{f}^j \right)^{\mathrm{T}} \mathbf{x}, \tag{26}$$

since the inner product is distributive over vector addition. From this it follows that

$$\mathbf{x}^* \propto \sum_i \sum_{j \in p_i} \mathbf{f}^j, \tag{27}$$

which is he gray-world solution for the entire image. In conclusion, we have seen that by choosing Eq. 9 as the unary potential, the standard gray-world method can be expressed as an energy minimization problem. Hence, when θ_p is chosen sufficiently large, minimizing Eq. 4 leads to the same result as the gray-world algorithm. It should be noted, that this is only true when the solution of the gray-world algorithm is in the illuminant label set $\mathcal{L}$. In practice this can easily obtained by choosing the solution of the gray-world as one of the labels.

A similar derivation can be used to prove that using Eq. 11 as the unary potential and minimizing Eq. 4 yields the gray-edge algorithm. Enforcing exactly one label leads to the same answer as the single illuminant for $p = 1$ in Eq. 10.

B. Estimation of the Two-Illuminant Ground Truth

This appendix provides additional details on the derivation of Eq. 22. As in Section VII-B let $f_{a,G}$ and $f_{b,G}$ denote aligned pixels from the green channel of two images, which are exposed to illuminant $\mathbf{i}_a$ and illuminant $\mathbf{i}_b$, respectively. We seek to derive the influence of $\mathbf{i}_a$ and $\mathbf{i}_b$ in $f_{ab,G}$ where both illuminants are additive, i.e., $f_{ab,G} = f_{a,G} + f_{b,G}$. Intuitively, if a pixel is brighter in $f_{a,G}$ than in $f_{b,G}$, then the influence of $\mathbf{i}_a$ is stronger in $f_{ab,G}$. This brightness difference is caused by: 1) the intensity of the illuminant and 2) the different angles between the light source vector and the surface normal (for instance, in the laboratory dataset the lights are located at the left and right sides of the scene). Thus, we want to compute per pixel a weighting factor r, such that

$$\mathbf{i}_{ab} = r \cdot \mathbf{i}_a + (1 - r) \cdot \mathbf{i}_b, \tag{28}$$

i.e., the illuminant color $\mathbf{i}_{ab}$ at a pixel is a weighted sum of the colors of the two illuminants.

To obtain r, we first compute illumination-normalized versions of the original images $\hat{f}_{a,C} = f_{a,C}/i_{a,C}$ and $\hat{f}_{b,C} = f_{b,C}/i_{b,C}$, where $C \in \{R, G, B\}$, based on the von Kries assumption. r is then obtained by computing the relative contribution of $\hat{f}_{a,G}$ to respect to the two-illuminant image,

$$r = \frac{\hat{f}_{a,G}}{\hat{f}_{a,G} + \hat{f}_{b,G}}. \tag{29}$$

Assuming Lambertian reflectance and sharpened sensors, $\hat{f}_{a,G} = \rho_G e_a w_a$, where e_a and w_a denote scaling factors due to the intensity and the geometry of the light source, respectively, and ρ_G is the pixel's albedo. Note that the illuminant color is omitted, as it has been neutralized. Expanding Eq. 29, the ratio of a pixel under both illuminant corresponds to the ratio of their scaling factors e_a and w_a,

$$r = \frac{\rho_G e_a w_a}{\rho_G (e_a w_a + e_b w_b)} = \frac{e_a w_a}{e_a w_a + e_b w_b}, \tag{30}$$

as albedo and neutral illuminant are identical in $\hat{f}_{a,G}$ and $\hat{f}_{b,G}$. This leads directly to the formulation in Eq. 22.

In practice, we clip the weight r if one of the illuminants is t_{B} times brighter than the other, i.e.,

$$r = \begin{cases} 1 & \text{if } \hat{f}_{a,G}/\hat{f}_{b,G} > t_{\mathrm{B}} \\ 0 & \text{if } \hat{f}_{b,G}/\hat{f}_{a,G} > t_{\mathrm{B}}, \\ \tilde{r}\,\hat{f}_{a,G}/\hat{f}_{b,G} & \text{otherwise} \end{cases} \tag{31}$$

where $\tilde{r}$ normalizes the range of values between 0 and 1. For our dataset, we empirically determined that $t_{\mathrm{B}} = 40$ is a reasonable threshold.

In real-world images, the assumption of sharpened sensors and Lambertian reflectance are typically violated. We alleviate these issues with two "engineering decisions". First of all, we use only the green channel, as an approximation to a sharp sensor. Pixels with a specular component are indirectly handled via clipping. In these cases, the intensity of the specularity often exceeds the clipping range, thus the respective pixel is assigned to the specular illuminant (which agrees with the neutral interface assumption [12]). Finally, note that interreflections are in general not well modeled by this approach. Despite these shortcomings, we manually investigated all scenes, and concluded that the cases that violate our assumptions are rare or do not considerably influence the result. Thus, the proposed approach is an economic, feasible way to obtain pixelwise multi-illuminant ground truth on real-world scenes.

ACKNOWLEDGMENT

S. Beigpour and C. Riess contributed equally to this work.

REFERENCES

[1] G. Buchsbaum, "A spatial processor model for color perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, Jul. 1980.
[2] 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.
[3] G. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. IS&T/SID Color Imag. Conf.*, 2004, pp. 37–41.
[4] A. Gijsenij, T. Gevers, and J. van de Weijer, "Improving color constancy by photometric edge weighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 918–929, May 2012.
[5] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Troeger, *et al.*, "Color constancy and non-uniform illumination: Can existing algorithms work?" in *Proc. IEEE Workshop Color Photometry Comput. Vis.*, Nov. 2011, pp. 774–781.
[6] D. H. Brainard and W. T. Freeman, "Bayesian color constancy," *J. Opt. Soc. Amer. A*, vol. 14, no. 7, pp. 1393–1411, 1997.
[7] J.-M. Geusebroek, R. Boomgaard, A. Smeulders, and T. Gevers, "Color constancy from physical principles," *Pattern Recognit. Lett.*, vol. 24, no. 11, pp. 1653–1662, Jul. 2003.
[8] R. T. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse-intensity chromaticity space," *J. Opt. Soc. Amer. A*, vol. 21, no. 3, pp. 321–334, Mar. 2004.
[9] E. H. Land, "Lightness and the retinex theory," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, Dec. 1977.
[10] R. Gershon, A. D. Jepson, and J. K. Tsotsos, "From [R,G,B] to surface reflectance: Computing color constant descriptors in images," in *Proc. 10th Int. Joint Conf. Artif. Intell.*, vol. 2. Aug. 1987, pp. 755–758.

[11] H.-C. Lee, "Method for computing the scene-illuminant chromaticity from specular highlights," *J. Opt. Soc. Amer. A*, vol. 3, no. 10, pp. 1694–1699, 1986.
[12] S. A. Shafer, "Using color to separate reflection components," *J. Color Res. Appl.*, vol. 10, no. 4, pp. 210–218, 1985.
[13] D. Forsyth, "A novel algorithm for color constancy," *Int. J. Comput. Vis.*, vol. 5, no. 1, pp. 5–36, 1990.
[14] G. D. Finlayson, S. D. Hordley, and I. Tastl, "Gamut constrained illuminant estimation," *Int. J. Comput. Vis.*, vol. 67, no. 1, pp. 93–109, 2006.
[15] A. Gijsenij, T. Gevers, and J. van de Weijer, "Generalized gamut mapping using image derivative structures for color constancy," *Int. J. Comput. Vis.*, vol. 86, nos. 2–3, pp. 127–139, Jan. 2010.
[16] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2475–2489, Sep. 2011.
[17] B. Li, "Evaluating combinational color constancy methods on real-world images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1–8.
[18] D. Foster, "Color constancy," *Vis. Res.*, vol. 51, no. 7, pp. 674–700, Apr. 2011.
[19] K. Barnard, G. Finlayson, and B. Funt, "Color constancy for scenes with varying illumination," *Comput. Vis. Image Understand.*, vol. 65, no. 2, pp. 311–321, Feb. 1997.
[20] M. Ebner, "Color constancy using local color shifts," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 276–287.
[21] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand, "Light mixture estimation for spatially varying white balance," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 70:1–70:7, Aug. 2008.
[22] R. Kawakami, K. Ikeuchi, and R. T. Tan, "Consistent surface color for texturing large objects in outdoor scenes," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2005, pp. 1200–1207.
[23] A. Gijsenij, R. Lu, and T. Gevers, "Color constancy for multiple light sources," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 697–707, Feb. 2011.
[24] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand, "Light mixture estimation for spatially varying white balance," in *Proc. ACM SIGGRAPH*, 2008, pp. 70:1–70:7.
[25] C. Lu and M. S. Drew, "Shadow segmentation and shadow-free chromaticity via Markov random fields," in *Proc. IS&T/SID Color Imag. Conf.*, 2005, pp. 125–129.
[26] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan, "Detecting ground shadows in outdoor consumer photographs," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 322–335.
[27] M. Serra, O. Penacchio, R. Benavente, and M. Vanrell, "Names and shades of color for intrinsic image estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 278–285.
[28] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jul. 2001, pp. 105–112.
[29] P. Kohli, L. Ladický, and P. Torr, "Robust higher order potentials for enforcing label consistency," *Int. J. Comput. Vis.*, vol. 82, pp. 302–324, Jan. 2009.
[30] K. Barnard, L. Martin, A. Coath, and B. Funt, "A comparison of computational color constancy algorithms—Part II: Experiments with image data," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 985–996, Sep. 2002.
[31] C. Riess, E. Eibenberger, and E. Angelopoulou, "Illuminant color estimation for real-world mixed-illuminant scenes," in *Proc. IEEE Workshop Color Photometry Comput. Vis.*, Nov. 2011, pp. 782–789.
[32] Y. Boykov, O. Veksler, and R. Zabih, "Markov random fields with efficient approximations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1998, pp. 648–654.
[33] J. van de Weijer, T. Gevers, and J.-M. Geusebroek, "Edge and corner detection by photometric quasi-invariants," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 625–630, Apr. 2005.
[34] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, 2004.
[35] O. Veksler, Y. Boykov, and P. Mehrani, "Superpixels and supervoxels in an energy optimization framework," in *Proc. Eur. Conf. Comput. Vis.* 2010, pp. 211–224.
[36] S. Bagon. (2006, Dec.). *Matlab Wrapper for Graph Cut* [Online]. Available: http://www.wisdom.weizmann.ac.il/~bagon
[37] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms—Part I: Methodology and experiments with synthesized data," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 972–983, Sep. 2002.
[38] F. Ciurea and B. Funt, "A large image database for color constancy research," in *Proc. IT&S/SID Color Imag. Conf.*, 2003, pp. 160–164.
[39] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[40] C. A. Parraga, R. Baldrich, and M. Vanrell, "Accurate mapping of natural scenes radiance to cone activation space: A new image dataset," in *Proc. Eur. Conf. Colour Graph., Imag., Vis.*, 2010, pp. 50–57.
[41] K. Barnard, L. Martin, B. Funt, and A. Coath, "A data set for color research," *Color Res. Appl.*, vol. 27, no. 3, pp. 147–151, 2002.
[42] J. van de Weijer, C. Schmid, and J. Verbeek, "Using high-level visual information for color constancy," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[43] S. Bianco and R. Schettini, "Color constancy using faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 65–72.

Shida Beigpour is an Associate Professor with the Faculty of Computer Science and Media Technology, Gjovik University College (Høgskolen i Gjøvik) and a Researcher at the Norwegian Colour and Visual Computing Laboratory. She received the M.Sc. degree in artificial intelligence and computer vision in 2009 and the Ph.D. in informatics (computer vision) in 2013 from Universitat Autonoma de Barcelona, Barcelona, Spain. Her main research interests are color vision, illumination and reflectance analysis, color constancy, and color in human vision. She has published articles in high-impact journals and conferences.

Christian Riess received the Diploma degree in computer science in 2007 and the Doctoral degree in 2013 from the University of Erlangen-Nuremberg, Erlangen, Germany. From 2007 to 2010, he was working on an industry project with Giesecke+Devrient on optical inspection. He is currently pursuing the post-doctoral research at the Radiological Sciences Laboratory, Stanford University, Stanford, CA, USA. His research interests include all aspects of image processing, in particular with applications in image forensics, medical imaging, optical inspection, and computer vision.

Joost van de Weijer is a Senior Scientist at the Computer Vision Center, Barcelona. He received the M.Sc. degree in applied physics from the Delft University of Technology in 1998. In 2005, he received the Ph.D. degree from the University of Amsterdam. From 2005 to 2007, he was a Marie Curie Intra-European Fellow in the LEAR Team, INRIA Rhone-Alpes, France. From 2008 to 2012, he was a Ramon y Cajal Fellow at the Universidad Autonoma de Barcelona. His main research is on the usage of color information in computer vision application. He has published in the fields of color constancy, color feature extraction and detection, color image filtering, color edge detection, and color naming.

Elli Angelopoulou received the Ph.D. degree in computer science from Johns Hopkins University in 1997. She was a post-doctoral at the General Robotics, Automation, Sensing and Perception Laborator, University of Pennsylvania. She was an Assistant Professor at the Stevens Institute of Technology. She is currently an Associate Research Professor at the University of Erlangen-Nuremberg. Her research focuses on computational colour analysis, multispectral imaging, capsule endoscopy, computer assisted surgery, image forensics, reflectance analysis of natural scenes, and 3-D reconstruction. She has published over 70 refereed publications, multiple patents and has received numerous grants, including the NSF CAREER Award. She has served on the program committees of ICCV, CVPR, and ECCV and is an Associate Editor of *Machine Vision and Applications* and the *Journal of Intelligent Service Robotics*.