Color Constancy and Non-Uniform Illumination: Can Existing Algorithms Work?

Michael Bleier[1], Christian Riess[1], Shida Beigpour[2], Eva Eibenberger[1], Elli Angelopoulou[1], Tobias Tröger[3], André Kaup[3]

[1]Pattern Recognition Lab,
University of Erlangen-Nuremberg
http://www5.cs.fau.de

[2] Computer Vision Center,
Universidad Autónoma de Barcelona
http://cat.uab.cat

[3]Multimedia Communications and Signal Processing,
University of Erlangen-Nuremberg
http://lnt.de

Abstract

The color and distribution of illuminants can significantly alter the appearance of a scene. The goal of color constancy (CC) is to remove the color bias introduced by the illuminants. Most existing CC algorithms assume a uniformly illuminated scene. However, more often than not, this assumption is an insufficient approximation of real-world illumination conditions (multiple light sources, shadows, interreflections, etc.). Thus, illumination should be locally determined, taking under consideration that multiple illuminants may be present. In this paper we investigate the suitability of adapting 5 state-of-the-art color constancy methods so that they can be used for local illuminant estimation. Given an arbitrary image, we segment it into superpixels of approximately similar color. Each of the methods is applied independently on every superpixel. For improved accuracy, these independent estimates are combined into a single illuminant-color value per superpixel. We evaluated different fusion methodologies. Our experiments indicate that the best performance is obtained by fusion strategies that combine the outputs of the estimators using regression.

1. Introduction

Accurate color measurements are important in a variety of computer vision applications ranging from object recognition and tracking, to stereo and image retrieval. The appearance of color, however, is heavily dependent on the illuminant. Computational color constancy (CC) algorithms attempt to alleviate this dependency. One common framework of such methods is to first explicitly estimate the illuminant color and then color correct the image accordingly.

Unfortunately, the recovery of the illumination color from a single image is an underconstrained problem. Every observed image pixel represents an unknown combination of surface reflectance and illumination. Many CC algorithms try to make this problem tractable by imposing different assumptions on the observed scene. For instance, Gray Edge algorithms assume that a derivative of the pixels sums up to 0 under canonical illumination [23]. Gamut Mapping assumes that the convex hull of the pixels in a suitably chosen color space encompasses most illumination changes [13]. Furthermore, most illuminant color estimators typically assume globally uniform illumination. This prerequisite is essential for collecting a sufficiently large number of samples from the whole image and thus increasing the accuracy and robustness of the methodology.

In practice, assuming uniform illumination is well justifiable for images taken under laboratory conditions. However, many real-world scenes consist of more than one illuminant. For instance, indoor photographs often exhibit a mixture of indoor illumination and sunlight coming from the windows. Flash photographs typically have people near the camera illuminated by the flashlight, whereas the background is illuminated by other light sources. In such situations it would be more appropriate to estimate the illuminant color locally. To our knowledge there have been very few methods that explicitly focus on such local illuminant estimation. Ebner [7], for example, applied a diffusion-based methodology on the pixel intensities. However, he locally assumed a gray-world which can result in inaccuracies, especially in colorful scenes [16]. More recently, Kawakami *et al*. [17] proposed a method specifically de-

2011 IEEE International Conference on Computer Vision Workshops
978-1-4673-0063-6/11/$26.00

signed to handle illumination variations between shadowed and non-shadowed regions in outdoor scenes. Their technique, thus, does not generalize well to arbitrary images.

In this paper we investigate whether existing CC methods, originally developed assuming uniform illumination, can be adapted to local illuminant color estimation. In order to obtain such localized estimates, we examine how the uniform-illuminant assumption of state-of-the-art CC methods can be relaxed. We use image sub-regions to compute the illuminant color locally. We then compensated for the loss of accuracy, by combining multiple independently obtained local estimates. Most of the existing fusion strategies try to solve the combination problem by extracting additional features from the image, e.g. color and texture statistics. Based on these features the best algorithm is selected or a weighted average of the estimates is computed. If the best algorithm could always be selected, we could gain in theory significant performance. Recent evaluations show, that algorithm combination based on image features does not perform substantially better than the best performing algorithm [14]. Computing regression on the estimates has been shown to be more robust than selecting a single estimate [2]. Thus, we expand this idea to make it applicable to our local estimates. An important part of such a comparative analysis is quantitative evaluation. To our knowledge, currently available databases do not provide sufficient information for evaluating multiple illuminant algorithms. Thus, we captured own multi-illuminant ground truth data. We then evaluated on this database different CC as well as fusion algorithms. We concluded that machine-learning based regression consistently outperformed all other combination strategies, as well as individual estimates.

2. Current CC Algorithms

There is a considerable body of work on Color constancy. In this section, we briefly review variants of the more well-known Gray World, Gamut Mapping and Bayesian Color Constancy.

2.1. Generalized Gray World Algorithm

The Gray World and Gray Edge algorithms can be combined into a common framework [12, 23]. The illuminant color is estimated as

$$\boldsymbol{l} = k \left(\int \left| \frac{\partial^n \boldsymbol{f}^\sigma(\boldsymbol{x})}{\partial \boldsymbol{x}^n} \right|^p \mathrm{d}\,\boldsymbol{x} \right)^{\frac{1}{p}} , \tag{1}$$

where $\boldsymbol{f}^\sigma(\boldsymbol{x})$ is the Gaussian smoothed image at location $\boldsymbol{x}$, n is the order of the derivative, p the parameter of the Minkowski norm, σ the standard deviation of the Gaussian smoothing kernel and k a scaling factor. This very compact formulation covers several well-known variants, *e.g.* the Retinex algorithm [19], the classical Gray Wold [6], the Shades of Gray [12] and the Gray Edge hypothesis [23].

The **White Patch Retinex** or **max-RGB** algorithm is based on the Retinex theory developed by Land *et al*. [19]. It assumes that somewhere in the image is a white patch, which reflects maximally and achromatically. Thus, the illuminant color can be directly recovered from the brightest pixel. In practice we take the maximum response of each color channel separately, potentially from different pixels. This algorithm is obtained by setting $p \rightarrow \infty$ in (1). It is sensitive to noise because a single bright pixel can lead to a bad estimate. The algorithm can be improved by computing histograms for each color channel (see *e.g*. [8, pp.118-119]). Another possibility is to select a value, where only a small percentage, e.g. 1%, of the pixels has a higher intensity.

The **Gray World** hypothesis was first formalized by Buchsbaum [6]. It assumes Lambertian reflection and an on-average achromatic scene. Instead of gray it is also possible to use the average of a reflectance database [1, p. 70]. A general formulation to exploit this assumption was provided by Finlayson and Trezzi [12],

$$\boldsymbol{l} = k \left(\frac{\int (\boldsymbol{f}(\boldsymbol{x}))^p \mathrm{d}\,\boldsymbol{x}}{\int \mathrm{d}\,\boldsymbol{x}} \right)^{\frac{1}{p}} , \tag{2}$$

called Shades of Gray. Note that the exponentiation of $\boldsymbol{f}(\boldsymbol{x})$ is a component-wise operation. It incorporates the classical formulation for $p = 1$.

Van de Weijer *et al*. proposed the **Gray Edge** method [23]. It assumes that the average of the reflectance differences is achromatic. If the assumption holds, the illuminant color is

$$\boldsymbol{l} = k \left(\frac{\int (\boldsymbol{f}^\sigma_{\boldsymbol{x}}(\boldsymbol{x}))^p \mathrm{d}\,\boldsymbol{x}}{\int \mathrm{d}\,\boldsymbol{x}} \right)^{\frac{1}{p}} , \tag{3}$$

where $\boldsymbol{f}^\sigma_{\boldsymbol{x}}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}) * \frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{G}^\sigma(\boldsymbol{x})$ and $\boldsymbol{G}^\sigma(\boldsymbol{x})$ is a Gaussian smoothing operation with standard deviation σ.

2.2. Gamut Constrained Methods

A key component of the gamut constrained methods is the definition of a canonical gamut $\boldsymbol{\Gamma}(\boldsymbol{C})$, which denotes the convex set of sensor responses $\boldsymbol{C} = \{\boldsymbol{q}_1, ..., \boldsymbol{q}_N\}$ to N surface reflectances under a canonical illuminant:

$$\boldsymbol{\Gamma}(\boldsymbol{C}) = \left\{ \sum_i \alpha_i \boldsymbol{q}_i \;\middle|\; \boldsymbol{q}_i \in \boldsymbol{C},\ \alpha_i \geq 0,\ \sum_i \alpha_i = 1 \right\}. \tag{4}$$

A mapping between the gamut of an unknown illuminant and the canonical gamut is reveals then the illuminant color. Based on Forsyth's original algorithm [13], a number of variations have been developed [10, 9, 11, 15].

The most well-known of these methods is **Gamut Mapping**, originally proposed by Finlayson. When transforming an image gamut $\boldsymbol{\Gamma}(\boldsymbol{I})$ to the canonical gamut $\boldsymbol{\Gamma}(\boldsymbol{C})$,

gamut mapping uses a diagonal matrix transform $\boldsymbol{D}$. Because only a limited number of surfaces is observed within a single image, the unknown gamut can only be approximated by the observed image gamut. The set of all possible mappings from $\boldsymbol{\Gamma}(\boldsymbol{I})$ to $\boldsymbol{\Gamma}(\boldsymbol{C})$ is calculated and the best mapping (w.r.t. a selection criterion) is selected.

In more detail, let $\boldsymbol{D}_{\boldsymbol{p},\boldsymbol{q}}$ be a diagonal matrix that maps a point $\boldsymbol{p} = (p_R, p_G, p_B)^T$ in the image gamut to a point $\boldsymbol{q} = (q_R, q_G, q_B)^T$ in the canonical gamut,

$$\forall \boldsymbol{p} \in \boldsymbol{\Gamma}(\boldsymbol{I}), \quad \boldsymbol{p}\boldsymbol{D} \in \boldsymbol{\Gamma}(\boldsymbol{C}) \quad , \tag{5}$$

where $\boldsymbol{D}_{\boldsymbol{p},\boldsymbol{q}}\boldsymbol{p} = \boldsymbol{q}$. The possible mappings for one point $\boldsymbol{p}$ in the image gamut can be calculated as

$$\boldsymbol{M}(\boldsymbol{p}) = \left\{ \boldsymbol{d}_{\boldsymbol{p},\boldsymbol{q}} \;\middle|\; \boldsymbol{d}_{\boldsymbol{p},\boldsymbol{q}} = \left(\frac{q_R}{p_R}, \frac{q_G}{p_G}, \frac{q_B}{p_B}\right)^T, \quad \boldsymbol{q} \in \boldsymbol{\Gamma}(\boldsymbol{C}) \right\} . \tag{6}$$

These sets are likewise convex. The feasible set $\tilde{\boldsymbol{M}}$ can be calculate by intersecting all elements of $\boldsymbol{M}(\boldsymbol{p})$ for each point $\boldsymbol{p} \in \boldsymbol{\Gamma}(\boldsymbol{I})$ in the image gamut, *i.e.*,

$$\tilde{\boldsymbol{M}} = \bigcap_{\boldsymbol{p} \in \boldsymbol{\Gamma}(\boldsymbol{I})} \boldsymbol{M}(\boldsymbol{p}) \quad . \tag{7}$$

Their intersection is also a convex set. Each map $\boldsymbol{m} \in \boldsymbol{M}$ corresponds to a possible illuminant. For the final decision, Forsyth proposed to choose the diagonal matrix transform with the **maximum** trace from the feasible set, which results in the most colorful gamut.

Finlayson and Hordley showed that Gamut Mapping can be also performed in a 2D chromaticity space, as only intensity information is lost [10, 9]. We refer to this method as **2D-Gamut Mapping**.

The 3D-vector $(R, G, B)^T$ is projected onto the plane at $B = 1$, yielding 2D chromaticities $r = R/B$, $g = G/B$, $b = 1$. For 2D-Gamut Mapping, the diagonal transform

$$\boldsymbol{D} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{8}$$

has only two parameters. The feasible set is constrained by the set of possible illuminants. Using Monte Carlo estimation, the explicit computation of the intersection between illuminants and the feasible set can be avoided. Rather, random points in 3D-sensor space are generated. The illuminant is the **mean** or **median** of a set of randomly chosen points lying within the feasible set.

2.3. Bayesian Color Constancy

Bayesian color constancy generates a probabilistic model for surface reflectances and illuminants. Assuming statistical independence of illuminants and surfaces, Bayes' rule is used to decide for the illuminant $\boldsymbol{l}$ according to a loss function $\mathcal{L}(\boldsymbol{l}, \boldsymbol{l}')$ [5, 21].

(a) (b) (c)

Figure 1. Examples for multi-illuminant situations. (a) Two dominant light sources, spatially separated (b) Two dominant light sources mix smoothly on the floor (c) Complex illumination situation.

Let the probability of coocurrence of illuminants and surface reflectances be known. The illuminant $\boldsymbol{l}^\star$ that minimizes the average loss is

$$\boldsymbol{l}^\star = \operatorname*{argmin}_{\boldsymbol{l}} \sum_{\boldsymbol{l}'} \mathcal{L}(\boldsymbol{l}, \boldsymbol{l}')\, p(\boldsymbol{l}'|\boldsymbol{I}) \quad , \tag{9}$$

where $\boldsymbol{I}$ is the observed image, and the loss function is

$$\mathcal{L}(\boldsymbol{l}, \boldsymbol{l}') = \sqrt{(r - r')^2 + (g - g')^2} \quad , \tag{10}$$

with $r = l_R/(l_R + l_G + l_B)$ and $g = l_G/(l_R + l_G + l_B)$. Using Bayes' rule, set

$$p(\boldsymbol{l}|\boldsymbol{I}) = \frac{p(\boldsymbol{I}|\boldsymbol{l})\, p(\boldsymbol{l})}{p(\boldsymbol{I})} = k\, p(\boldsymbol{I}|\boldsymbol{l})\, p(\boldsymbol{l}) \quad , \tag{11}$$

where $p(\boldsymbol{I})$ has a uniform prior density and k is a constant over the variables of interest.

Rosenberg *et al.* [21] proposed to model the likelihood $p(\boldsymbol{I}|\boldsymbol{l})$ using reflectances. The illuminant prior $p(\boldsymbol{l})$ can be estimated from training data, or assumed to be equally distributed.

3. From Uniform to Non-Uniform Illumination

The presented algorithms were designed for recovering a single, dominant illuminant. In real-world images, this assumption is often violated. Fig. 1 shows some examples of multi-illuminant situations. On the left image, the game console acts as a local second dominant light source on the face of the boy. In the middle, two dominant light sources mix smoothly on the floor of the church. On the right, several local light sources create a complex multi-illuminant scene.

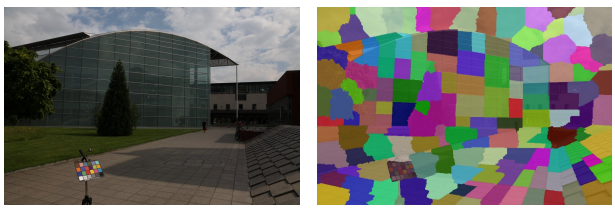
It is not straightforward to incorporate these different effects in a single multi-illuminant CC algorithm. When the influence of an illuminant is spatially limited to a distinct object (like the face in the left image), object segmentation and subsequent single-illuminant estimation may lead to a

satisfactory recovery of the illumination colors. In a different scenario, the church floor in the middle image is also a single "object" from a segmentation viewpoint. However, its surface is illuminated by distinct light sources at different locations. Thus, an object-based illumination extraction is inappropriate in this case. Pixel diffusion-based approaches like the one by Ebner [7] can be expected to solve the church floor example, but are expected to fail on object boundaries as in the left image.

We decided to compromise between these two extremes. We segment the image in a set of superpixels, *i.e.*, small image sub-regions such that all pixels in a single superpixel satisfy the same property, in our case color value. A collection of CC algorithms is then applied on each superpixel independently. The per-superpixel output of the algorithms is fused afterwards. For the superpixel segmentation, we used the algorithm by Veksler *et al.* [20, 4, 18, 3], though other segmentation methods can also be employed. It segments the image into non-overlapping, compact superpixels based on their RGB values (see for example Fig. 2.

The underlying assumption is, that the illumination is approximately locally constant on a single superpixel. Though this assumption may sometimes be violated, our experimental evaluation shows that this does not significantly impact the performance of our method. We then apply state-of-the-art CC algorithms on a per-superpixel basis. The superpixel segmentation can address, without any fine-tuning, a large range of multiple scenarios. Since the superpixels follow object boundaries, our method can handle object-specific illuminants as in Fig. 1(a). Because superpixels are small, a large object that is illuminated differently at distinct locations (like the church floor in Fig. 1(b)) is subdivided and its subregions are separately processed.

However, when applying established methods locally, a trade-off between spatial resolution and color constancy performance has to be made. For instance, gamut mapping and Bayesian color constancy draw their accuracy from extensive statistics over the range of colors in the image. A superpixel offers only a limited selection of the observable colors. Hence, a performance drop for statistics-based



(a) (b)

Figure 2. Example superpixel segmentation on an image from the Gehler database with the method by Veksler *et al.* Left: original image, right: the segmentation typically preserves object boundaries.

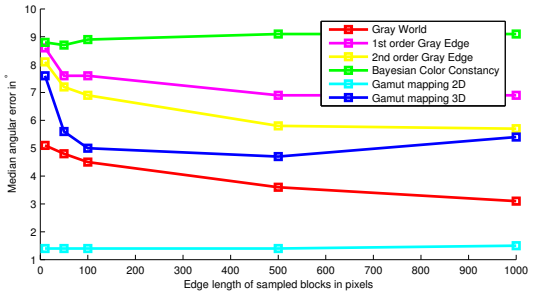


Figure 3. Error rates on rectangular subregions of different sizes.

methods is expected. Equivalently, 0-order gray world is clearly affected from the fact that superpixels typically contain pixels of similar colors. In order to partially alleviate these problems, we constrained the possible illuminants to the convex set of illuminants in our data.

As part of our analysis, we evaluated how the size of local subregions can affect the performance of CC algorithms. We selected 80 images from the re-processed Gehler database [14, 22] with approximately uniform illumination. We excluded the parts of the image containing the Macbeth chart. We selected rectangular regions of different sizes and applied Bayesian color constancy and several instances of the Generalized Gray World and Gamut Mapping on it. We use the median angular error as a performance measure. As shown in Fig. 3, blocks of size less than 100×100 pixels typically exhibit a performance drop. One notable exception is the 2-dimensional Gamut Mapping. Here, the fine-tuned set of possible illuminants dominated the overall estimation result, leading to a very low error rate.

3.1. Fusion of Multiple Illuminants

In order to improve the illumination estimation on small regions, we fused the outputs of different algorithms based on their error statistics. We evaluated different fusion approaches. As a straightforward baseline method, we computed the average of all combined estimates. Besides this, we used two approaches based on machine learning regression. First, we used Gradient tree boosting as a classical machine learning algorithm to combine multiple weak predictors to a single strong one. As an alternative, we used Random forest regression. It consists of a set of tree predictors, which are trained on randomly chosen, different training sets. The output is computed as the average response over all trees in the forest.

The implementations of both machine learning algorithms were taken from the `openCV` library. For Gradient tree boosting, we used a squared loss function, 200 learning iterations and a maximum depth of 20. The Random forest regression was trained with a maximum depth of 50 with

at most 100 trees. Note that, in difference to prior work, we did not use additional features to guide the fusion process. Instead, only the estimates (plus for the training set the ground truth, of course) were available for computing the regression. Thus, this approach can be seen as a "brute force" approach to localized illuminant estimation. Its outcome should serve as a cue whether we can use variants of these established algorithms also for illumination estimation on non-uniformly illuminated scenes.

4. Multi-illuminant Ground Truth Data Set

In order to obtain ground truth for multiple illuminants we generated our own data under laboratory conditions. Four scenes (see Fig. 4) were taken under 17 different illumination conditions, 9 of which were truly multi-illuminant, for a total of 36 multi-illuminant images. The different lighting setups were created by two Reuter lamps with LEE color filters. One Reuter lamp was positioned on the left side of the scene and was combined with the LEE filters 201, 202 and 281. The other Reuter lamp was positioned on the right side and was used with the LEE filters 204, 205 and 285. We also took images with only one filtered light on at a time, both Reuter lamps on without any filters and one under ambient illumination. As ground truth we spray-painted each scene gray and took a series of images under the exact same 17 illumination conditions. We used RAL 7035 and RAL 7047 spray paints which were verified with a Macbeth color checker. Note that this method of ground-truth generation eliminates interreflections. Our scene is mostly composed of diffuse materials, since all the evaluated CC methods assume Lambertian reflectance. The tin object in Fig. 4(c), the knife and some of the fruits in Fig. 4(e) are the only exceptions.

The data was captured with a Canon EOS 550D camera and a Sigma 17-70 lens. The aperture and ISO settings were the same for all the images. The RAW data was converted using dcraw with gamma set to 1.0 and without any white balancing. Different fixed shutter speeds were used across the 17 different illumination conditions in order to avoid under- and over-exposure. Note that the collected data, as well as the code for this work can be downloaded from the web[1].

5. Evaluation

In our evaluation, we used as a performance measure the angular error e_{angular} between the recovered and the true illuminant color. The results over multiple estimates are most often aggregated by computing the median of the errors. Additionally, we computed the mean, root mean square error and the maximum error for every estimator.

[1] http://www5.cs.fau.de

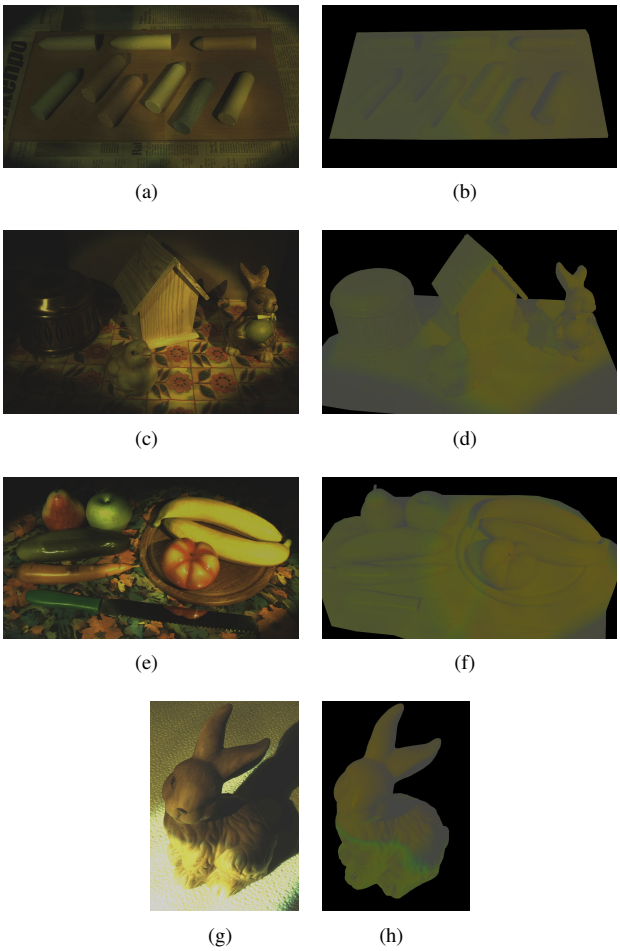


Figure 4. Captured scenes. Left: original scenes, right: same scene painted gray, for computing the ground truth (see text for further details).

For the evaluation we used leave-one-out cross validation. The base estimators for the fusion schemes have been trained on the reprocessed database by Gehler *et al*. Accordingly, the fusion algorithms have also been trained on the database by Gehler *et al*. Thus, our own captured testing data was for the algorithms completely unknown.

5.1. Evaluation on Uniform Illumination

First, we evaluated our implementations on the reprocessed database by Gehler *et al*. [14, 22]. We used this data set, as it is to our knowledge the largest real-world dataset, which is available as raw data. The reflection target has been masked out. These results show the overall performance of our implementations of individual algorithms — as there are always implementational ambiguities. They also serve as a basis for comparison to the subsequent evaluation on images containing non-uniform illumination. Even in the Gehler database, there are images taken under multi-

Algorithm	Angular error in °			
	Mean	Median	RMS	Max
Do nothing	13.5	13.6	13.6	21.2
Average illuminant	2.3	1.9	3.0	18.7
White patch Retinex	8.4	6.6	10.5	37.3
Gray World	6.4	5.3	7.6	25.2
1st order Gray Edge	4.7	3.8	5.6	24.3
2nd order Gray Edge	4.2	3.2	5.3	24.3
Best Gray World / Edge	4.2	3.2	5.3	24.3
Gamut Mapping (max)	6.6	5.2	8.4	31.6
Gamut Mapping (mean)	4.6	4.5	5.0	17.4
Bayesian Color Constancy	3.2	2.5	4.1	19.5
Average estimate	4.9	3.6	6.1	21.1
Gradient tree boosting	2.6	1.9	3.5	18.6
Random forest regression	2.7	2.2	3.6	18.7

Table 1. Mean, median, root mean square, and maximum errors for outdoor images from the reprocessed Gehler *et al*. database.

ple illuminants, even though a single illuminant is provided as ground truth. This is a source of error which is more prominent in the indoor scenes. Thus, we subdivided the images in outdoor and indoor images (see Tab. 1 and Tab. 2).

"Do nothing" assumes a white illuminant, and "Average illuminant" estimates always the average of all the training illuminants. For the generalized Gray World we used the following settings: We varied the Minkowski norm with $1 \leq p \leq 10$, $0 \leq \sigma \leq 4$ in steps of 1. We explicitly report only the results for "White patch Retinex" ($n = 0$, $p \to \infty$, $\sigma = 0$), "Gray World" ($n = 0$, $p = 1$, $\sigma = 0$), "1st order Gray Edge" ($n = 1$, $p = 1$, $\sigma = 1$), "2nd order Gray Edge" ($n = 2$, $p = 1$, $\sigma = 1$), and the best performing Generalized Gray World algorithm based on the median angular error. For Gamut Mapping, "Gamut Mapping (max)" denotes 3D Gamut Mapping that chooses the illuminant based on the maximum trace. "Gamut Mapping (mean)" denotes 2D Gamut Mapping with the mean selection strategy. Only the 2D Gamut Mapping uses illuminant constraints. "Bayesian color constancy" denotes Bayesian illumination estimation using the Euclidean distance as loss function. For the fusion of the estimates, "Average estimate" denotes the mean of the outputs of the fused estimators, "Gradient tree boosting" and "Random forest regression" denote the two machine learning-based regression approaches.

From Tab. 1, we observe that the error of choosing the mean illuminant is very small for outdoor images. Thus, the variability of illuminants is small for the outdoor images. The 2D Gamut Mapping variants benefit the most from that fact. The best performing generalized Gray World algorithm for outdoor images was $n = 2$, $p = 1$, $\sigma = 1$. Using regression and a collection of estimators, we were able to obtain results that are better than the best performing single algorithm. For the training and testing itself, we used k-fold cross-validation. Interestingly, the average illuminant from the ground truth performs still slightly better with respect to the mean and RMS error, which suggests that the variability of illuminants is not very high in this dataset. For indoor images (see Tab. 2), the best performing Gray World method was Gray Edge with $n = 1$, $p = 1$, $\sigma = 1$. Here, the variation of illuminants is significantly higher, as can be seen from the higher error of the "Average illuminant". Note that among the fusion schemes, Random forest regression performs best, and improves the final error for approximately $0.9°$, compared to the single estimators.

Algorithm	Angular error in °			
	Mean	Median	RMS	Max
Do nothing	13.8	13.4	14.4	27.4
Average illuminant	7.0	6.5	8.1	22.8
White patch Retinex	10.9	10.7	12.6	48.1
Gray World	6.6	6.3	7.7	24.8
1st order Gray Edge	4.7	4.3	5.3	15.0
2nd order Gray Edge	5.6	4.9	6.4	17.3
Best Gray World / Edge	4.7	4.3	5.3	15.0
Gamut Mapping (max)	9.5	8.9	10.9	27.4
Gamut Mapping (mean)	7.1	6.7	8.2	20.6
Bayesian Color Constancy	5.9	5.4	7.0	20.0
Average estimate	7.0	6.5	8.2	29.4
Gradient tree boosting	4.5	3.9	5.3	15.7
Random forest regression	3.9	3.4	4.7	16.2

Table 2. Mean, median, root mean square, and maximum errors for indoor images from the reprocessed Gehler *et al*. database.

5.2. Evaluation on Non-Uniform Illumination

The 36 multi-illuminant images have been segmented in superpixels. The segmentation was transferred to the ground truth images, and the per-segment ground truth was determined by averaging the ground truth over the superpixels. This averaging introduces inaccuracies on an illumination boundary. However, we considered it reasonable, because the algorithms produce illuminant estimates per-superpixel, and as such this is the level of detail in which we require ground truth.

An example segmentation and estimation is shown in Fig. 5. The segmentation parameters were chosen, so that an image was subdivided into approximately 30 superpixels, with the individual superpixel size varying between approximately $10,000$ and $50,000$ pixels.

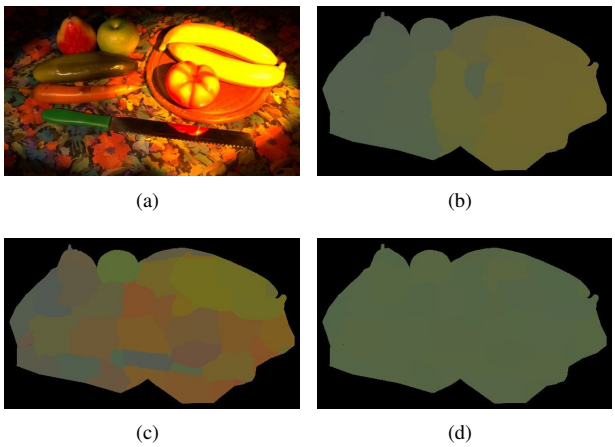
We trained the illuminant estimators on the indoor images from the reprocessed Gehler database. The results are shown in Tab. 3. The best gray world configuration was $n = 0$, $p = -1$, $\sigma = 1$. Note that the individual errors on each superpixel are largely increased, and range from a me-

Algorithm	Angular error in °			
	Mean	Median	RMS	Max
Do nothing	10.4	10.0	11.7	21.6
Average illuminant	7.7	7.8	8.3	15.4
White patch Retinex	5.8	5.1	7.0	21.6
Gray World	5.1	5.0	5.7	14.4
1st order Gray Edge	14.2	13.7	14.9	29.0
2nd order Gray Edge	12.8	12.6	13.4	27.3
Best Gray World / Edge	5.4	4.6	6.3	21.6
Gamut Mapping (max)	5.9	5.2	7.2	21.3
Gamut Mapping (mean)	6.4	6.3	7.1	14.6
Bayesian Color Constancy	6.5	4.8	8.1	20.7
Average estimate	7.8	7.7	8.4	16.2
Gradient tree boosting	6.2	6.0	6.8	17.5
Random forest regression	4.4	4.1	5.0	12.9

Table 3. Mean, median, root mean square, and maximum errors non-uniform illumination estimation on superpixels.

dian of about 4.6 degrees to 13.7 degrees. This has been expected. The superpixel segmentation aims to provide areas of approximately the same color, which is theoretically poor input for almost all applied estimators. For instance, Gamut Mapping expects colorful images, while gray world expects balanced colors. Picking a superpixel with mainly only one color undermines such algorithms. However, this effect is apparently limited. For instance, in the present case, the 1st and 2nd order gray world algorithms perform considerably worse than the statistical methods.

In this scenario, we performed all necessary training steps on the Gehler indoor database. Thus, for Gamut Mapping and Bayesian Color Constancy, as well as the fusion



(a) (b) (c) (d)

Figure 5. Example segmentation, ground truth and illuminant estimation (colors are scaled for the purpose of printing) (a) Original image (b) ground truth (c) Result of gray world (d) Result of 2D Gamut Mapping

algorithms, the tested database was unknown beforehand. Similarly, also the "Average Illuminant" refers to the average of the ground truth of the database by Gehler *et al.* In this experiment, regression by boosting did not improve the results. However, regression based on Random forests could lower the median error to 4.1 degrees. The improved estimates yield an error level that again clearly improves over the single illuminant estimates. This is a surprising result, as the underlying estimates by themselves are comparably weak, and no additional image features are used for guidance of the fusion process. As a consequence, we conclude that to some extent, color constancy under non-uniform illumination can be addressed by ensembles of locally applied single-illuminant estimators.

6. Discussion

We evaluated single illuminant estimation algorithms and fused versions of these algorithms in three scenarios. The two cases on the database by Gehler *et al.* can be used as a baseline reference for the performance of our implementation. Among these, the indoor pictures exhibit a greater variability. Thus, when evaluating on the newly captured data, we trained our algorithms on the Gehler indoor images. In general, machine learning algorithms have difficulties when training on one database and evaluating on another. During our experiments, we also noted this behavior. Nevertheless, results show that we managed to avoid overfitting to some extend. The methods generalize sufficiently, such that their good performance can be confirmed on the unknown data.

We consider several aspects of the evaluation section worth to discuss in greater detail. At first, it is counter-intuitive that the statistical methods, *i.e.* Gamut Mapping and Bayesian Color Constancy, yield at all useful results on the superpixel estimation. Indeed, we observed that a drop color variation yields to several problems with the statistical algorithms. However, note that in the present case, the superpixels are still relatively large with respect to the level of detail in the image. As a consequence, at least two surfaces, and some shading artifacts, are typically contained within one superpixel.

Another particularity of this approach is the fact, that no additional image features are used to support the fusion of the estimates. Thus, the fusion methods can be seen as a way of integrating brute force results on the multi-illuminant problem. Thus, the results suggest – although this should be confirmed in more extensive experiments – that the error behavior of existing algorithms contains useful patterns to estimate the true illuminant.

As a consequence, we see in these results two main points. First, evaluating a large number of existing algorithms for every superpixel in the image dramatically increases the computational cost. Additionally, methods

like 2D-Gamut Mapping can become slow if the observed Gamut is very small. More efficient algorithms are necessary for a more practical solution to the multi-illuminant problem. At the same time, we are surprised how well the lack of color information can be compensated by adding more estimators. Looking at the error rates, we eventually reached a median error of 4.1° on non-uniform illumination, which is even better than the best performing estimator on the single-illuminant indoor dataset by Gehler *et al*.

7. Conclusion

We addressed the problem of recovering a relatively dense field of localized illuminant estimates for color constancy under non-uniform illumination. We implemented several variants of Gray World, Gamut Mapping and Bayesian color constancy as base illuminant estimators. For the ground-truth data, we captured 4 scenes for a total of 36 multi-illuminant images. Then, we trained the base algorithms on the reprocessed real-world dataset by Gehler *et al*. and obtained illuminant estimates over small superpixels per image. As expected, the error of these estimates was much higher compared to the errors under uniform illumination. Surprisingly, regression on these weak estimates from inhomogeneous illumination lead to results that are comparable to the performance of single estimators on homogeneous illumination. Thus, we conclude that combinations of single-illuminant estimators yield promising results to address the recovery of illumination color under non-uniform illumination. We presented preliminary results. Note that the individual error rates require closer investigation, and that the ground truth dataset can be further improved. In future work, we plan to extend our database of non-uniformly illuminated scenes.

References

[1] K. Barnard. *Practical Color Constancy*. PhD Thesis, School of Computing, Simon Fraser University, Burnaby, Canada, 1999.

[2] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini. Automatic Color Constancy Algorithm Selection and Combination. *Pattern Recognition*, 43(3):695–705, Mar. 2010.

[3] Y. Boykov and V. Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, Sept. 2004.

[4] Y. Boykov, O. Veksler, and R.Zabih. Efficient Approximate Energy Minimization via Graph Cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1222–1239, Nov. 2001.

[5] D. H. Brainard and W. T. Freeman. Bayesian Color Constancy. *Journal of the Optical Society of America A*, 14(7):1393–1411, 1997.

[6] G. Buchsbaum. A Spatial Processor Model for Color Perception. *Journal of the Franklin Institute*, 310(1):1–26, July 1980.

[7] M. Ebner. Color Constancy Using Local Color Shifts. In *European Conference in Computer Vision*, pages 276–287. Springer-Verlag, 2004.

[8] M. Ebner. *Color Constancy*. John Wiley & Sons, Chichester, England, 2007.

[9] G. Finlayson and S. Hordley. Improving Gamut Mapping Color Constancy. *IEEE Transactions on Image Processing*, 9(10):1774–1783, Oct. 2000.

[10] G. D. Finlayson. Color in Perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):1034–1038, Oct. 1996.

[11] G. D. Finlayson, S. D. Hordley, and I. Tastl. Gamut Constrained Illuminant Estimation. *International Journal of Computer Vision*, 67(1):93–109, 2006.

[12] G. D. Finlayson and E. Trezzi. Shades of Gray and Colour Constancy. In *Twelfth Color Imaging Conference*, pages 37–41, 2004.

[13] D. Forsyth. A Novel Algorithm for Color Constancy. *International Journal of Computer Vision*, 5(1):5–36, 1990.

[14] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian Color Constancy Revisited. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2008.

[15] A. Gijsenij, T. Gevers, and J. van de Weijer. Generalized Gamut Mapping using Image Derivative Structures for Color Constancy. *International Journal of Computer Vision*, 86(2–3):127–139, Jan. 2010.

[16] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand. Light Mixture Estimation for Spatially Varying White Balance. *ACM Transactions on Graphics*, 27(3):70:1–70:7, Aug. 2008.

[17] R. Kawakami, K. Ikeuchi, and R. T. Tan. Consistent Surface Color for Texturing Large Objects in Outdoor Scenes. In *IEEE International Conference on Computer Vision*, pages 1200–1207, 2005.

[18] V. Kolmogorov and R.Zabih. What Energy Functions can be Minimized via Graph Cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, Feb. 2004.

[19] E. H. Land. Lightness and the Retinex Theory. *Scientific American*, 237(6):108–129, Dec. 1977.

[20] P. M. O. Veksler, Y. Boykov. Superpixels and Supervoxels in an Energy Optimization Framework. In *European Conference on Computer Vision*, Sept. 2010.

[21] C. Rosenberg, T. Minka, and A. Ladsariya. Bayesian Color Constancy with Non-Gaussian Models. In *Advances in Neural Information Processing Systems*, volume 16, 2003.

[22] L. Shi and B. Funt. Re-processed version of the gehler color constancy dataset of 568 images. `http://www.cs.sfu.ca/~colour/data/shi_gehler/`, Jan 2011.

[23] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge-Based Color Constancy. *IEEE Transactions on Image Processing*, 16(9):2207–2214, Sept. 2007.