Using High-Level Visual Information for Color Constancy

Joost van de Weijer, Cordelia Schmid, Jakob Verbeek

Using High-Level Visual Information for Color Constancy

Joost van de Weijer, Cordelia Schmid, Jakob Verbeek
INRIA Rhône-Alpes, LEAR
655 Avenue de l'Europe, Montbonnot 38330, France
{Joost.van-de-Weijer, Cordelia.Schmid, Jakob.Verbeek}@inrialpes.fr

Abstract

We propose to use high-level visual information to improve illuminant estimation. Several illuminant estimation approaches are applied to compute a set of possible illuminants. For each of them an illuminant color corrected image is evaluated on the likelihood of its semantic content: is the grass green, the road grey, and the sky blue, in correspondence with our prior knowledge of the world. The illuminant resulting in the most likely semantic composition of the image is selected as the illuminant color. To evaluate the likelihood of the semantic content, we apply probabilistic latent semantic analysis. The image is modelled as a mixture of semantic classes, such as sky, grass, road, and building. The class description is based on texture, position and color information. Experiments show that the use of high-level information improves illuminant estimation over a purely bottom-up approach. Furthermore, the proposed method is shown to significantly improve semantic class recognition performance.

1. Introduction

Light reflected by an object which enters the eye, or a camera, is a product of the object reflectance properties and the illuminant spectrum. The task of color constancy is to disentangle the two, allowing to recognize the colors of objects independent of the color of the illuminant. Computational color constancy is relevant for many computer vision task such as object recognition, tracking, and surveillance [3, 4, 11]. In addition, it allows for illuminant correction of images, with the aim to present images consistent with human perception of the world.

Computational color constancy research can be roughly divided in two approaches. One line of research focusses on illuminant invariant representations, which are primarily based on color differences between different patches in the image [14, 16, 17, 30]. The second, and more prominent line of color constancy research aims at estimating the color of the illuminant, after which the image can be corrected to how it would appear under a canonical, usually white, illuminant [4, 6, 9, 10, 13, 29]. This second line of color constancy research has the advantage that it allows for correcting the image for deviations from a canonical light source.

Several color constancy methods return a set of possible illuminants, from which one is to be selected [9, 13]. The subsequent selection procedures are often based on a heuristic, such as taking the average color of all possible illuminants [1]. Tous [28] considers the low-level image information, on which these color constancy methods are based, insufficient to select between a set of possible illuminants. Consequently, he proposes to return a set of solutions to the computer vision application, leaving the selection of the actual illuminant to the application. Another approach has been proposed by Gijsenij and Gevers [18], who use image statistics to decide on the most appropriate color constancy method given an image. All these methods are similar in that the illuminant estimation is based purely on bottom-up information, and high-level top-down information is disregarded. In this paper, we will explore the use of high-level visual information to select the most likely illuminant of a scene.

A motivation for the use of high-level visual information for color constancy can be found in recent human vision research. The mechanisms underlying human color constancy are still poorly understood. Most research uses collages of color patches in a 2D plane, so called Mondrian images, to infer mechanism of human color constancy [23]. Experiments on more real world like settings were performed by Kraft and Brainard [21], in which they proved that bottom-up clues, such as inter-reflections, specularities, and the range of colors present in a scene, all contribute to human color constancy. However, the scene still consisted of abstract objects, such as colored squares, and specular cylinders. Only recently research investigated the use of high-level visual information to obtain color constancy. Hansen et al. [19] illuminated fruit objects with an adjustable light source. They asked human observers to adjust the color of the light source such that the natural fruit objects appeared

achromatic. When the illuminant was adjusted to the point that the physical reflectance of the object was achromatic, observers still perceived a color sensation. The fruit objects only looked achromatic when the illuminant was shifted further away from the grey point in the direction opposite to the fruit color. This implies that high-level information of the objects color plays a role in human color constancy.

The first contribution of this article is the use of high-level visual information to select the best illuminant out of a set of possible illuminants. We achieve this by restating the problem in terms of semantic interpretability of the image. Which of the illuminants results in a likely image interpretation, i.e., an image where the sky is blue and in the top of the image, and the road is grey and in the bottom can be considered more likely than an image with purple grass surrounding a reddish cow. Several color constancy methods are applied to generate a set of illuminant hypotheses. For each illuminant hypothesis, we correct the image, and evaluate the likelihood of the semantic content of the corrected image. Finally, the most likely illuminant color is selected.

As a second contribution, we extend the set of illuminant hypotheses with a set of top-down hypotheses based on the assumption that the average reflectance of semantic classes in an image is equal to the average reflectance of the semantic topic in the database. For each of the semantic classes present in the image we compute the illuminant which transforms the pixels assigned to this class in such a way that the average reflectance is in accordance with the average color of the class in the database. For example, a patch of grass which turned reddish in the evening light, will correctly hypothesize a red illuminant, since such an illuminant will transform it to green under white light.

In contrast with existing work on color constancy, which uses a purely bottom-up approach, we investigate to what extent top-down color constancy can improve results. Both contributions, the selection mechanism based on the semantic likelihood and the generation of top-down illuminant hypotheses, are derived from the idea that high-level information plays an important role in color constancy.

2. Probabilistic Color Constancy

In this section, we state the illuminant estimation problem in a probabilistic manner and give an overview of our method.

Probabilistic approaches compute the probability of an illuminant given the image data $P(\mathbf{c}|\,\mathbf{f})$. The illuminant of a scene is that illuminant which is most likely given the image data

$$c_{\max} = \underset{\mathbf{c}\in C}{\operatorname{argmax}}\ \log\left(P(\mathbf{c}|\,\mathbf{f})\right) \tag{1}$$

where $\mathbf{f} = (R, G, B)^T$, and C is the set of possible illuminants $\mathbf{c}$, which choice we will discuss later. Bold fonts are applied for vectors. Now assume that we have a function g which, if we know the illuminant of the scene, transforms the image as if it were taken under white light

$$g(\mathbf{f}^{\mathbf{c}}, \mathbf{c}) = \mathbf{f}^{\mathbf{w}}, \tag{2}$$

where superscript $\mathbf{c}$ denotes the image's illuminant and $\mathbf{w}$ indicates the white illuminant. Then, the probability that the image $\mathbf{f}$ is taken under illuminant $\mathbf{c}$ is equal to the probability that the transformed image $g(\mathbf{f}^{\mathbf{c}}, \mathbf{c})$ is taken under a white illuminant:

$$P(\mathbf{c}|\,\mathbf{f}) = P(\mathbf{w}|\,g(\mathbf{f}, \mathbf{c}))\,. \tag{3}$$

Applying this to Eq. 1 yields

$$\mathbf{c}_{\max} = \underset{c\in C}{\operatorname{argmax}}\ \log\left(P(\mathbf{w}|\,g(\mathbf{f}, \mathbf{c}))\right). \tag{4}$$

This equation will be applied to select the illuminant color. This equation selects that illuminant $\mathbf{c}_{\max}$ which maximizes the probability that the color corrected image $g(\mathbf{f}, \mathbf{c}_{max})$ was taken under white lighting.

Probabilistic color constancy is based on choosing the most likely illuminant given the image data. Methods very close to the formulation in Eq. 4 have been proposed in literature [5, 10]. However, these methods interpret the probability in a purely bottom-up way. They are based on the probability of an RGB value to occur under a particular light source. Here we will propose an integrated bottom-up and top-down approach, where both the pixel values in the image and the semantic interpretation of the image as a whole influence the probability of the illuminant given the image data.

The success of color constancy as derived from Eq. 4 depends on two points. Firstly, how do we compute the chance that an image is taken under white light $P(\mathbf{w}|\,\mathbf{f})$, and how can we incorporate high-level information in this probability. Secondly, since it is unfeasible to evaluate Eq. 4 for all possible illuminants c, how do we select a plausible set of color illuminants for a scene. An overview of our approach is given in Fig. 1. For an input image a set of bottom-up and top-down illuminant hypotheses are computed (explained in Section 4). For each of these hypotheses the image is corrected and subsequently evaluated on the likelihood of its semantic content (explained in Section 3). The illuminant which results in the most probable image content is considered to be the illuminant of the input image. In the depicted case, the method estimates the illuminant to be reddish, since after correcting for this light source the image could be interpreted as green grass under a blue sky.

For the function g, which transforms an image $\mathbf{f}^{\mathbf{c}}$ taken under illuminant $\mathbf{c}$ to an image $\mathbf{f}^{\mathbf{w}}$ taken under a white illuminant, we use a multiplication with a diagonal matrix.

$$g(\mathbf{f}^{\mathbf{c}}, \mathbf{c}) = \mathbf{D}^{\mathbf{c}}\mathbf{f}^{\mathbf{c}} = \mathbf{f}^{\mathbf{w}} \tag{5}$$

with

$$\mathrm{D} = \mathrm{diag}\left(\mathbf{w}\right)\left(\mathrm{diag}\left(\mathbf{c}\right)\right)^{-1} \tag{6}$$

This model is called the diagonal model, or von Kries model, and has been proven to sufficiently approximate reality [2, 8].

3. Images as a Mixture of Semantic Classes

In this section, we describe how to compute the probability of an image to occur under a white light source. For this purpose we will model images as a mixture of semantic classes, such as sky, grass, road and building. Each class is described by a distribution over visual words, which are described by three modalities texture, color and position. As an example, consider an image with sky and grass. This image will consist of visual words which are drawn form the distributions of sky and grass. Given these visual words, we will attempt to infer what classes are present in the image. Given the inferred classes and the visual words we compute a likelihood of the image, which we call the *semantic likelihood* of the image. For this purpose we use Probabilistic Latent Semantic Analysis (PLSA), a generative model introduced by Hofmann [20] for document analysis. Recently, PLSA models have shown good results for classification of pixels into semantic classes [27, 31].

Images are modelled as a mixture of latent topics. The topics are semantic classes in the image such as sky, grass, road, building, etc. They are described by a distribution over visual words. As visual descriptors we use 20x20 patches which are extracted on a regular grid from the image. Each patch, or visual word, is described by three modalities: 1. texture, which is described with the SIFT descriptor [24] , 2. color, which is described by the Gaussian averaged RGB value over the patch, and 3. position, which is described by imposing a 8x8 grid of regular cells on the image. Both the texture and color features are discretized by Kmeans clustering. We use a texture vocabulary of 750 words, and a color vocabulary of 1000 words. The position is described by 64 words, each referring to one of the 64 cells.

Given a set of images $F = \{\mathbf{f}_1, ..., \mathbf{f}_N\}$ each described in a visual vocabulary $V = \{v_1, ..., v_M\}$, the words are taken to be generated by latent topics $Z = \{z_1, ..., z_K\}$. In the PLSA model the conditional probability of a visual word v in an image $\mathbf{f}$ and an illuminant $\mathbf{c}$ is given by:

$$P\left(v|\,\mathbf{f},\mathbf{c}\right) = \sum_{z^{\mathbf{c}}\in Z^{\mathbf{c}}} P\left(v|\,z^{\mathbf{c}}\right)P\left(z^{\mathbf{c}}|\,\mathbf{f}\right). \tag{7}$$

where $z^{\mathbf{c}}$ indicates that the topic distribution has been computed from a data set which was taken under illuminant $\mathbf{c}$. Similar to the approach of Verbeek and Triggs [31], we assume the three modalities to be independent given the topics,

$$P\left(v|z\right) = P\left(v^T|z\right)P\left(v^C|z\right)P\left(v^P|z\right), \tag{8}$$

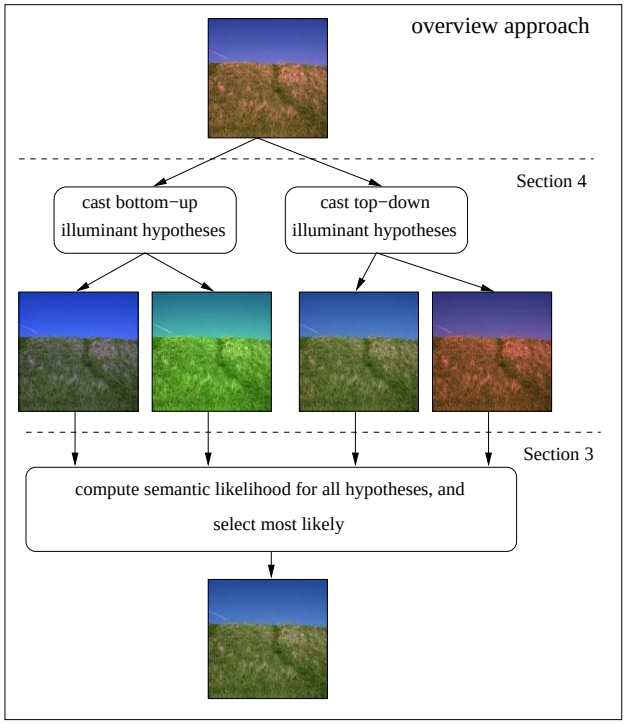

Figure 1. Overview of our approach. See text for details.

where v^T, v^C, v^P, are successively the texture, color and position word. The distributions $P(z|\mathbf{f})$ and the various $P(v|z)$'s are discrete, and can be estimated using an EM algorithm [20].

We set out to compute the chance that an image was taken under white light, which according to Bayes law is proportional to

$$P\left(\mathbf{w}|\mathbf{f}\right) \propto P\left(\mathbf{f}|\mathbf{w}\right)P\left(\mathbf{w}\right). \tag{9}$$

If we assume a uniform distribution over the illuminants $p\left(\mathbf{w}\right)$, this can be rewritten using Eq. 7 to,

$$\begin{aligned} P\left(\mathbf{w}|\mathbf{f}\right) \propto P\left(\mathbf{f}|\mathbf{w}\right) &= \prod_{m=1}^{M} P\left(v_m|\mathbf{f},\mathbf{w}\right) \\ &= \prod_{m=1}^{M}\sum_{z^{\mathbf{w}}\in Z^{\mathbf{w}}} P\left(v_m|z^{\mathbf{w}}\right)P\left(z^{\mathbf{w}}|\mathbf{f}\right), \end{aligned} \tag{10}$$

where $P\left(v_m|z^{\mathbf{w}}\right)$ means that the visual word topic distributions are learned from images taken under white light.

Let us consider what happens with Eq. 10 when we evaluate various illuminants. For the sake of simplicity we consider here that the texture descriptors do not change when varying the illuminant, although in the real implementation they are recomputed for each illuminant. By varying the illuminant color we change the color word v^C and via $P\left(v^C|z\right)$ both $P\left(v|z\right)$ and the topic distribution in the image $P\left(z|\mathbf{f}\right)$. The image will be more likely when

$P\left(v^C|z\right)$ corresponds with the combined distribution of $P\left(v^T|z\right)P\left(v^P|z\right)$. This means that illuminants become more likely when the color words they generate are in accordance with the texture and position information. Hence, color words representing green are more likely together with texture words describing grass, and a sky like texture in the top of the image is more likely to be blue.

The approach described here is related to the work of Manduchi [25], who uses the color similarity between a test image and labelled classes[1] in one training image taken under white light to estimate the illuminant color. The classes are described by a Gaussian color distribution. Each pixel is assigned to a class and an illuminant to optimize the likelihood of the image. The method has the advantage that multiple illuminants are allowed within an image. However, the methods is only demonstrated to succeed when a single training image, similar to the test image, is available. This might be due to the limited discriminative power of the class description, in which multi-modality in color space, as well as texture and position information are disregarded.

4. Casting Illuminant Hypotheses

Evaluating Eq. 4 for all possible illuminants is not feasible. Instead, we propose to evaluate only a subset of color illuminants, which we call illuminant hypotheses. From these illuminant hypotheses the illuminant which is most likely given the image is selected. We propose two ways to generate hypotheses: a bottom-up approach and a top-down approach.

Bottom-up hypotheses: We can use existing color constancy algorithms to generate a set of possible illuminant colors for a scene. We call this approach bottom-up because these color constancy methods do not use any high-level visual information in the image. Here we choose to use a set of color constancy methods based on low-level features. Finlayson and Trezzi [12] unified two simple, broadly used, color constancy methods, by proving that the two methods are actually two instantiations of the Minkowski norm of an image:

$$\left(\sum_{i=1}^{N}\left(\mathbf{f}_i\left(\mathbf{x}\right)\right)^p\right)^{\frac{1}{p}} = k\mathbf{c} \tag{11}$$

where i is counter over the N pixels $\mathbf{f}_i$, and k is a constant which is chosen such that the illuminant color $\mathbf{c}$ has unit length. The parameter p is the Minkowski norm. For $p = 1$ the illuminant estimate is equal to color constancy derived from the Grey-World hypothesis, which assumes the average reflectance in a scene to be grey [6]. Using $p = \infty$ the illuminant estimate is equal to the max-RGB method [22] which assumes the maximum responses of the separate channels to be equal to the illuminant color. Recently this framework was further extended to include edge-based color constancy [29]:

$$\left(\sum_{i=1}^{N}\left|\frac{\partial^n \mathbf{f}_i\left(\mathbf{x}\right)}{\partial \mathbf{x}^n}\right|^p\right)^{\frac{1}{p}} = k\mathbf{c} \tag{12}$$

where n indicates the order of differentiation. For $n = 1$ the method is equal to assuming that the average edge difference in a scene is grey.

In the experimental section we apply Eq. 12 to generate a set of illuminant color hypotheses. We will use $n = \{0, 1, 2\}$ and $p = \{2, 12\}$ to compute six illuminant estimates. These six hypotheses are subsequently evaluated with Eq. 4 to select the most probable bottom-up illuminant. In the literature [12, 29] these color constancy methods were found to achieve comparable results to more complex methods such as color-by-correlation [10] and gamut mapping [13]. We cast one extra hypothesis which states that the image was taken under white light and no color correction is required.

Top-down hypotheses: Bottom-up approaches typically fail when the statistics of the image values are insufficiently distributed. For such images ignoring color information for

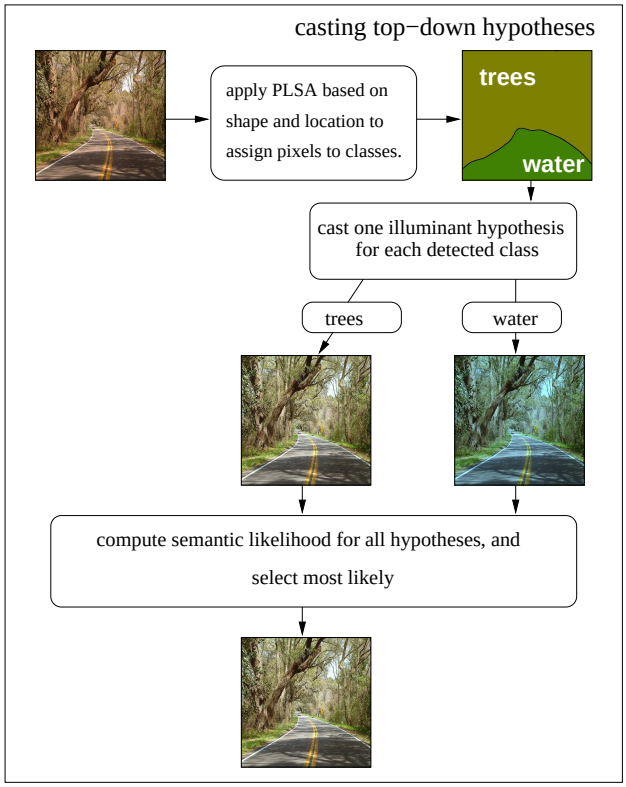

Figure 2. Overview of top-down casting of illuminant hypotheses. See text for details.

[1] These classes are not semantically meaningful as in this paper and are labelled "class I", "class II", etc.

recognition of semantic classes and relying instead on only position and texture information might yield a better interpretation of the image. We will here propose a method to exploit this information to compute a set of top-down color illuminant hypotheses. This is the second contribution of our paper.

In an extension to the Grey-World algorithm, which assumes the average reflectance in a scene to be achromatic, Gershon et al. [15] showed that for a coherent database, assuming the average of a scene to be equal to the average reflectance of the database, improves results over the Grey-World algorithm. As an example, they mention forest pictures full of green colors. In that case, most color constancy methods will predict illuminants biased towards the green color, whereas the database compensated algorithm resolves this problem. Since the eighties in which this algorithm has been proposed, the ability to extract the semantic information of an image has improved significantly. This increased semantic understanding of images can be used to precise Gershon's approach to semantic classes in the image. Therefore we propose the following color constancy hypothesis, which we call the *Green-Grass hypothesis*: the average reflectance of a semantic class in an image is equal to the average reflectance of the semantic topic in the database

$$\begin{aligned} \sum_{i \in T^s} \mathbf{f}_i(\mathbf{x}) &= k \operatorname{diag}(\mathbf{d}^s)\, \mathbf{c}^s \\ \mathbf{d}^s &= \sum_{i \in D^s} \mathbf{F}_i(\mathbf{x}), \end{aligned} \tag{13}$$

where T^s is the set of indexes to pixels in image $\mathbf{f}$ assigned to semantic topic s, $\mathbf{F}$ is the collection of all pixels in the training data set, D^s are the indexes to all pixels assigned to semantic topic s, and $\mathbf{c}^s$ is the estimate of the illuminant color based on topic s.

Fig. 2 presents an overview of the top-down casting of illuminant hypotheses. For each detected class in the image an illuminant hypothesis is casted. These hypotheses are subsequently evaluated based on the likelihood of their semantic content. In the above example the road is wrongly identified as water. The derived illuminant transforms the road pixels into blue which is the database average for the class water. The semantic likelihood, however, will prefer the hypothesis based on the tree-class, which considers the image to exist out of green trees above a grey road, thereby correctly estimating a reddish-yellow evening sun.

As depicted in Fig. 1 the bottom-up and top-down hypotheses are combined to compute a most likely illuminant for an image.

5. Experiments

In the experiments we evaluate the performance-gain of using high-level visual information. Firstly, we test our method on a traditional color constancy task, where the aim is to estimate the color of the illuminant and ground truth information is available. Secondly, we test the performance of the color constancy algorithm on a computer vision task, namely the classification of image pixels into a set of semantic classes.

5.1. Illuminant Estimation

In this experiment we apply our method to estimate the illuminant color of a scene. For evaluation the angular error between the estimated light source $\mathbf{c}_e$ and the actual light source $\mathbf{c}_l$ is used:

$$\text{angular error} = \cos^{-1}\left(\hat{\mathbf{c}}_l \cdot \hat{\mathbf{c}}_e\right), \tag{14}$$

where the $\hat{(\cdot)}$ indicates a normalized vector.

Data set: We test our approach on a data set assembled by Ciurea and Funt [7]. The database contains 11,000 images extracted from 2 hours of digital video. Both indoor and outdoor scenes from a wide variety of locations are represented, see Fig. 3. A small grey sphere was mounted onto the video camera, appearing in all images in the right bottom corner. The sphere is used to estimate the illuminant color in the scene. This color illuminant estimation is available with the database and is used as a ground truth. The original images were extracted from 15 different film clips taken at different locations. Because of the high correlation between the images in the database, the experiments are performed on a subset of 600 images taken at equal spacings form the set. We divide the set in 320 indoor images, of which 160 training and 160 test images, and 280 outdoor images of which 140 training and 140 test images. The pixels in the right bottom corner, which contains the grey sphere, are excluded from color constancy computation.

Training topic-word distribution: For all the images in the training data set the ground truth of the illuminant is given. We correct the images in the training data set for their illuminant using Eq. 6, and obtain a set of images under white light. Subsequently we compute the distribution of visual words over the topics $P(v|z^w)$ on this set. For these images no labels of the semantic content are available, therefore we apply PLSA to discover the topics from the unlabelled data, similarly as in [27, 31]. We found that for topic discovery it proved beneficial to only use the texture modality. The assignments of patches to topics based on texture $P(v^T|z)$ were then used to estimate the word-topic distributions for the other modalities $P(v^C|z)$ and $P(v^P|z)$. We used 20 topics for both the indoor and the outdoor set.

Results: The results for the indoor and the outdoor images are given in Table. 1[2]. For both sets we give the results without applying color constancy (i.e. assuming the illuminant to be white), and for the worst and the best of the

[2] See also the erratum appended after the ICCV paper

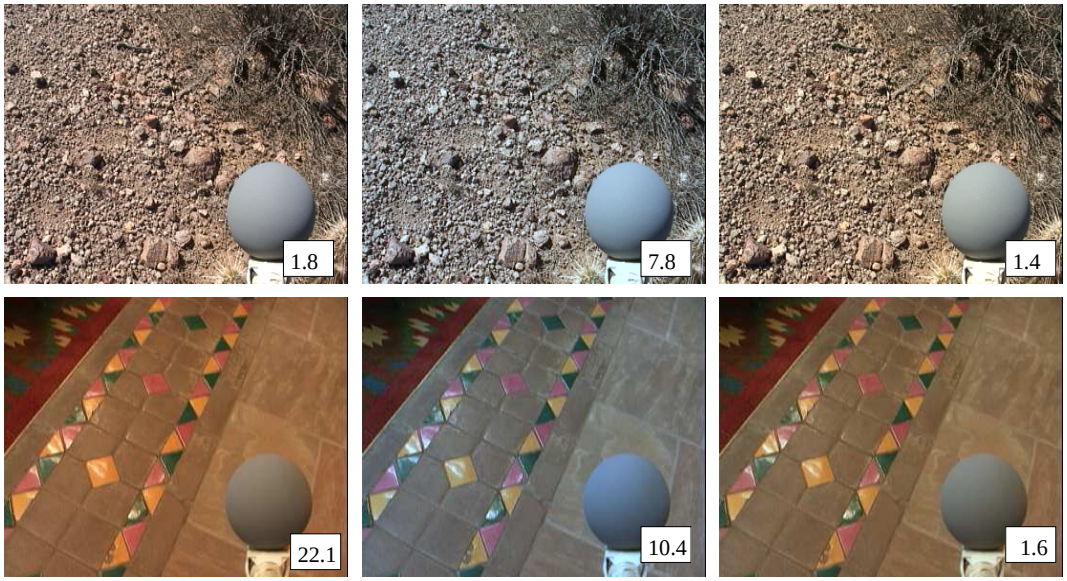

Figure 3. From left to right. Input image, Grey-World approach and the most likely top-down illuminant hypothesis. The angular error is indicated in the right bottom corner.

		standard color constancy		high-level selection using Eq.4		
	no cc	worst BU	best BU	BU	TD	BU & TD
indoor	10.2	8.6	4.8	4.8	4.8	4.8
outdoor	5.8	7.7	5.2	4.1	4.5	3.7

Table 1. Mean angular error for several color constancy methods. From left to right: without applying color constancy, worst and best result of Eq. 12, select the best estimate for only the bottom-up (BU) hypotheses, only the top-down (TD) hypotheses, or the combination of bottom-up and top-down hypotheses. The last three columns use the methods proposed in this paper.

bottom-up approach, when we would use a single approach on all images. Next we give results where we use the likelihood to select between only the bottom-up hypotheses, only the top-down hypotheses and both bottom-up and top-down hypotheses. On the indoor images the proposed approach is not able to perform better than the best of the bottom-up approaches (obtained with $n = 0$ and $p = 12$ in Eq. 12). This might be caused by the fact that in an indoor environment the semantic topics have a high variety of color appearances: doors, floors, chairs, clothes, all change color from one setting to another. On the outdoor set our approach obtains significantly better results than any of the bottom-up approaches. Here the best bottom-up approach achieves an angular error of 5.2, (obtained with $n = 1$ and $p = 12$). Combining the bottom-up approaches yields a performance gain of 20%. If we also consider top-down hypothesis we improve results by almost 30% to an angular error of 3.7.

Fig. 3 shows two images on which the bottom-up approaches fail and the top-down approach finds a reasonable illuminant estimate. The bottom-up results are computed with the Grey-World algorithm. Assuming an average grey reflectance yields for both images an unsatisfying illuminant estimation. For example, the reddish sand image in the first row is turned grey by the bottom-up approach. The top-down method succeeds, because one of the topics describes brown sand like structures, which resulted in a good top-down hypothesis with a high semantic likelihood.

In conclusions, the results show that selecting color constancy methods based on the likelihood that an image is generated by a mixture of topics learned under white lighting, improves color constancy results significantly for outdoor data. On indoor data, results are comparable to the best bottom-up approach.

5.2. Image Pixel Classification

In this experiment we will test the proposed approach on pixel classification. Pixels are to be classified as one of nine classes: building, grass, tree, cow, sheep, sky, water, face, and road. Because we already computed $P(v^m|\mathbf{f}, \mathbf{w})$ for each illuminant, pixel classification is only one step away. It is simply obtained by taking the most likely topic for each visual word.

Data set: To learn the nine semantic classes we use the labelled images of the Microsoft Research Cambridge (MSRC) set [26]. We remove images which we consider to be taken under non-white light, and those which did not contain any of the nine semantic classes (resulting in 240 training images). To extend the variability of the training

	standard color constancy		high-level selection using Eq.4		
no cc	worst BU	best BU	BU	TD	BU & TD
39.6	41.4	52.2	53.4	59.5	64.2

Table 2. Percentage of correctly classified pixels.

data we labelled another ten images collected from Google Image for each class. As a test set we selected four images per class from Google Image. These images were not present in the training set, and contained varying lighting conditions. The total test set contained 36 hand-labelled images (see Fig. 4).

Training Topic-Word Distribution: In this case the training data set is pixel labelled. The distributions of the visual words over the topics $P(v|z^w)$ are then obtained by assigning the visual words in the training data set to the topic distribution of their label. We did not have a ground truth of the illuminant for these images, and there exist many small deviations from white light. We assume, however, that all classes occur most often under white lighting.

Results: In Table 2 the results of the pixel classification is given. Not applying color constancy, as is done in most current state-of-the-art pixel classification systems [27, 31], obtains unsatisfying results on images with varying lighting conditions, with only 40% of the pixels correctly classified. The best bottom-up color constancy method correctly classifies 52% of the pixels. The top-down hypotheses obtain a very good score indicating that hypotheses based on the semantic content often yield reasonable estimates. These hypotheses often differ from the bottom-up hypotheses, as shown by the gain in performance when combining bottom-up and top-down hypotheses.

In Fig. 4 we show illustrations of images for which the top-down approach improved classification results. For all four images, classification without the use of any color constancy on the input image completely failed, except for the face image where the grass was recognized but not the face. For all images a number of top-down hypotheses were casted. We only show results of the hypotheses which resulted in the most likely image content. Although the classification results (see row 3 Fig. 4) still contain wrongly classified pixels, the results are good considering the difficult input images. The fourth column shows an example of the danger of top-down hypotheses. Based on the pixels which were identified as tree pixels, the illuminant is chosen, such that these tree pixels turn green. Although this improved pixel classification, the illuminant estimation is false, because the image depicts a reddish-brown tree in autumn.

In conclusions, using the likelihood of images to select the best illuminant to use for pixel classification is proven to be beneficial. The proposed method significantly improved results over standard color constancy methods.

6. Conclusions

This paper has presented a method to exploit high-level visual information for color constancy. Existing color constancy methods, as well as a new method based on prior knowledge of semantic classes in the world, are used to cast illuminant hypotheses. For each of the hypotheses we analyze the semantic likelihood based on a PLSA algorithm. The illuminant resulting in the most likely semantic composition of the image is selected as the illuminant color of the image. Results for both illuminant estimation and pixel classification into semantic classes demonstrate that using high-level image information improves results significantly.

7. Acknowledgements

This work is supported by the Marie Curie Intra-European Fellowship Program of the European Commission.

References

[1] K. Barnard. Improvements to gamut mapping colour constancy algorithms. In *Proc. European Conf. on Computer Vision*, pages 390–403, 2000.

[2] K. Barnard and B. Funt. Experiments in sensor sharpening for color constancy. In *Proc. IS&T/SID's Color Imaging Conf.*, 1998.

[3] K. Barnard and P. Gabbur. Color and color constancy in a translation model for object recognition. In *Proc. IS&T/SID's Color Imaging Conf.*, 2003.

[4] K. Barnard, L. Martin, A. Coath, and B. Funt. A comparison of computational color constancy algorithms-part ii: Experiments with image data. *IEEE Trans. on Image Processing*, 11(9):985–996, 2002.

[5] D. Brainard and W. 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 object colour perception. *Journal of the Franklin Institute*, 310, 1980.

[7] F. Ciurea and B. Funt. A large image database for color constancy research. In *Proc. IS&T/SID's Color Imaging Conf.*, pages 160–64, 2004.

[8] G. Finlayson, M. Drew, and B. Funt. Color constancy: Generalized diagonal transforms suffice. *Journal of the Optical Society of America A.*, 11:3011–3022, 1994.

[9] G. Finlayson and S. Hordley. Gamut constrained illumination estimation. *Int. Journal of Computer Vision*, 67(1):93–109, 2006.

[10] G. Finlayson, S. Hordley, and P. Hubel. Color by correlation: A simple, unifying framework for color constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, 2001.

[11] G. Finlayson, S. Hordley, and P. Hubel. Illuminant estimation for object recognition. *Color research and application*, 27:260–270, 2002.

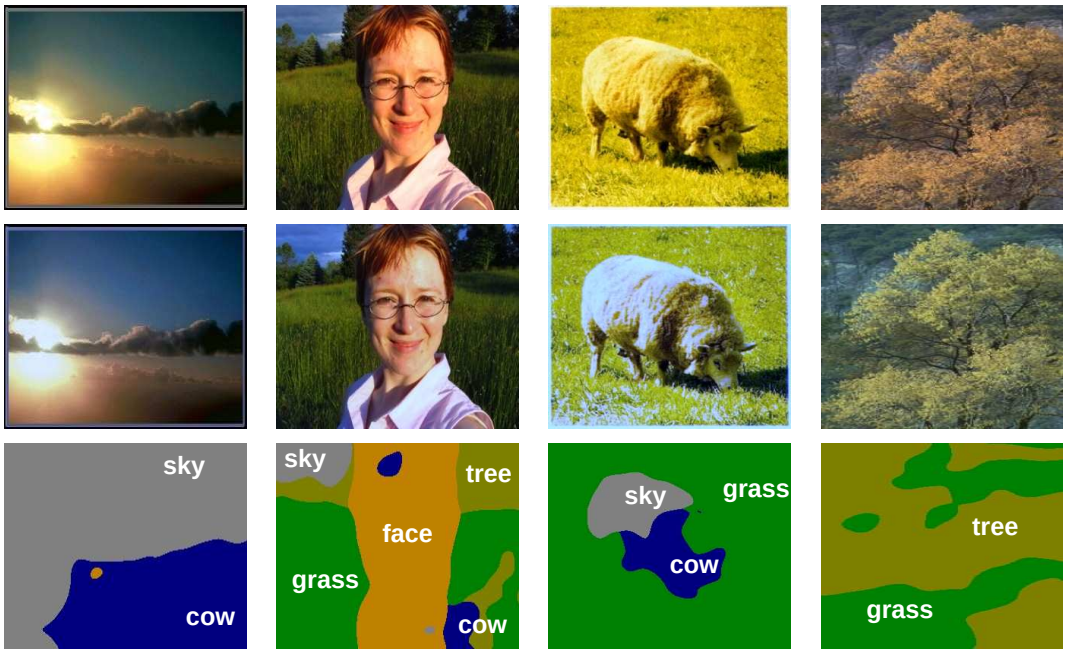

Figure 4. Example of pixel classification. First row: input images. Second row: illuminant corrected images selected by our approach. All four are based on a top-down illuminant hypothesis. The hypotheses are based from left to right on the classes sky, face, grass, and tree. Third row: pixel classification results based on the images in the second row.

[12] G. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Proc. IS&T/SID's Color Imaging Conf.*, pages 37–41, 2004.

[13] D. Forsyth. A novel algorithm for color constancy. *Int. Journal of Computer Vision*, 5(1):5–36, 1990.

[14] B. Funt and G. Finlayson. Color constant color indexing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.

[15] R. Gershon, A. D. Jepson, and J. K. Tsotsos. From [r,g,b] to surface reflectance: Computing color constant descriptors in images. In *Proc. of the 10th IJCAI*, 1987.

[16] J. Geusebroek, R. van den Boomgaard, A. Smeulders, and H. Geerts. Color invariance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(12):1338–1350, 2001.

[17] T. Gevers and A. Smeulders. Color based object recognition. *Pattern Recognition*, 32:453–464, 1999.

[18] A. Gijsenij and T. Gevers. Color constancy using natural image statistics. In *Proc. Computer Vision and Pattern Recognition*, 2007.

[19] T. Hansen, M. Olkkonen, S. Walter, and K. R. Gegenfurtner. Memory modulates color appearance. *Nature Neuroscience*, 9(11):1367–1368, October 2006.

[20] T. Hofmann. Probabilistic latent semantic indexing. In *Proc. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 50–57, 1999.

[21] J. Kraft and D. Brainard. Mechanisms of color constancy under nearly natural viewing. *Proc. Natural Acadamy of Science USA*, 96:307–312, 1999.

[22] E. Land and J. McCann. Lightness and retinex theory. *The Journal of the Optical Society of America A.*, 61(1):1–11, 1971.

[23] E. H. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–128, 1977.

[24] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 60(2):91–110, 2004.

[25] R. Manduchi. Learning outdoor color classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1713–1723, 2006.

[26] J. Shotton, J. M. Winn, C. Rother, and A. Criminisi. Textonboost:joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *Proc. European Conf. on Computer Vision*, pages 1–15, 2006.

[27] J. Sivic, B. Russell, A. Efros, A. Zisserman, and B. Freeman. Discovering objects and their location in images. In *Proc. IEEE Int. Conf. on Computer Vision*, 2005.

[28] F. Tous. *Computational framework for the white point interpretation based on colour matching*. PhD thesis, Universitat Autonoma de Barcelona, 2006.

[29] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Transactions on Image Processing*, 16(9):2207–2214, 2007.

[30] J. van de Weijer and C. Schmid. Blur robust and color constant image description. In *International Conference on Image Processing*, 2006.

[31] J. Verbeek and B. Triggs. Region classification with markov field aspect models. In *Proc. Computer Vision and Pattern Recognition*, 2007.

Erratum (23 September 2008)

A bug occurred in our implementation of experiment 5.1. For the bottom-up approaches we included the whole image in the illuminant calculation. The grey ball should have been excluded.

The correct results of the experiment are shown in Table. 3. Selection of the color constancy method based on the semantic likelihood of the images is shown to improve results. For both indoor and outdoor the selected bottom-up approach outperforms the best hand-picked bottom-up approaches (obtained with $n = 0$ and $p = 2$ for indoor, and $n = 2$ and $p = 2$ for outdoor). Combining the bottom-up and top-down cues is shown to help in the case of outdoor images. In conclusion, using semantic likelihood to select the color constancy method obtains a improvement of 10% on the outdoor set and of 20% on the indoor set against the best *hand-picked* bottom-up approach.

		standard color constancy		high-level selection using Eq.4		
	no cc	worst BU	best BU	BU	TD	BU & TD
indoor	12.8	12.3	6.1	5.3	5.6	5.3
outdoor	5.5	7.4	4.9	4.7	4.7	4.5

Table 3. Mean angular error for several color constancy methods. From left to right: without applying color constancy, worst and best result of Eq. 12, select the best estimate for only the bottom-up (BU) hypotheses, only the top-down (TD) hypotheses, or the combination of bottom-up and top-down hypotheses. The last three columns use the methods proposed in this paper.

We thank both Peter Gehler and Mark Everingham for bringing this error to our attention.