COLOR CONSTANCY USING IMAGE REGIONS

A. Gijsenij and Th. Gevers

Intelligent Systems Lab Amsterdam
Faculty of Science, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
{gijsenij, gevers}@science.uva.nl

ABSTRACT

Color constancy is important for various applications such as image segmentation, object recognition and image retrieval where object color features are extracted invariant to the illumination conditions. Different color constancy methods have been proposed. These methods, in general, compute color constancy based on all image colors. However, not all pixels contain relevant information for color constancy. Eventually, biased pixel values may decrease the performance of color constancy methods.

To this end, in this paper, we propose a method based on low-level image features using subsets of pixels. Hence, instead of using the entire pixel set for estimating the illuminant, only relevant pixels in the image are used. Therefore, prior segmentation is performed to learn for different image categories (e.g. open country, street, indoor) which pixel set (i.e. image parts) is most appropriate for a reliable estimation.

Based on large scale experiments on real-world scenes, it can be derived that for certain categories, like *open country* and *street*, the estimation is far more accurate using image parts than when using the entire image.

***Index Terms*—** Color Constancy, Illuminant estimation

1. INTRODUCTION

Differences in illumination cause measurements of object colors to be biased towards the color of the light source. Fortunately, humans have the ability of color constancy: they perceive the same color of an object despite large differences in illumination. A similar color constancy capability is necessary for various computer vision tasks such as object recognition, video retrieval and scene classification. In this way, the extracted image features are only dependent on the colors of the objects. This is beneficial for the task at hand [1].

Many color constancy algorithms have been proposed, see [2] for a recent overview. In general, color constancy algorithms can be divided into two groups: algorithms based on low-level image features and algorithms that use information acquired in a learning phase to estimate the illuminant. Gamut-based methods [3, 4, 5] are examples of the latter group. Such methods are based on the assumption that *in real-world images, for a given illuminant, one observes only a limited number of colors*. Examples of methods using low-level features are the Grey-World algorithm [6], the White-Patch algorithm [7].

Focusing on low-level driven color constancy, one of the drawbacks of the White-Patch and the Grey-World algorithms is that they are highly dependent on the validity of their assumptions. For instance, if the average color in a scene is *not* achromatic, the Grey-World algorithm will most likely result in an estimate the is biased towards the actual average color of the scene. Methods to overcome this dependency have been introduced, for instance the Minkowski-norm [8]. However, one of the difficulties when using the Minkowski-norm lies in the choice of an appropriate value for this norm. Further, not all pixels contain relevant information for color constancy. Eventually, biased pixel values may decrease the performance of these color constancy methods. For example, the estimate of the illuminant of an image with a large portion of blue sky may be severely biased towards the blue.

Therefore, in this paper, we will introduce a color constancy method using image regions. In fact, instead of using the entire image to estimate the illuminant, we propose to use only *that* part of the image which contain the most valuable information for color constancy. Segmentation is performed to learn for a different category which pixel set (i.e. image part) is most appropriate for a reliable estimation.

The algorithm will be tested on a data set containing over $11,000$ images extracted from 2 hours of video for a wide variety of settings (including indoor, outdoor, desert, cityscape, and other settings).

2. COLOR CONSTANCY

Let's assume that an image $\mathbf{f}$ is composed of:

$$\mathbf{f}(\mathbf{x}) = \int_\omega e(\lambda)\mathbf{c}(\lambda)s(\mathbf{x},\lambda)d\lambda, \qquad (1)$$

where $e(\lambda)$ is the color of the light source, $s(\mathbf{x},\lambda)$ is de surface reflectance and $\mathbf{c}(\lambda)$ is the camera sensitivity function.

1-4244-1437-7/07/$20.00 ©2007 IEEE

Further, ω and $\mathbf{x}$ are the visible spectrum and the spatial coordinates respectively. Assuming that the observed color of the light source $\mathbf{e}$ depends on the color of the light source $e(\lambda)$ as well as the camera sensitivity function $\mathbf{c}(\lambda)$, then color constancy is equivalent to the estimation of $\mathbf{e}$ by:

$$\mathbf{e} = \int_{\omega} e(\lambda)\mathbf{c}(\lambda)d\lambda, \tag{2}$$

given the image values of $\mathbf{f}$, since both $e(\lambda)$ and $\mathbf{c}(\lambda)$ are, in general, unknown. This is an under-constrained problem and therefore it can not be solved without further assumptions.

There are two well-established algorithms which use low-level features. Both are based on the Retinex Theory proposed by Land [7]. The first is the White-Patch algorithm and is based on the White-Patch assumption, i.e. *the maximum response in the RGB-channels is caused by a white patch*. The second is the Grey-World algorithm [6] which is based on the Grey-World assumption, i.e. *the average reflectance in a scene is achromatic*. Often, a trade-off exists between taking the maximum response in the RGB-channels and taking the average of the RGB-channels. In [8], this trade-off was formalized by using the Minkowski-norm:

$$\mathcal{L}_p = \left(\frac{\int \mathbf{f}^p(\mathbf{x})d\mathbf{x}}{\int d\mathbf{x}} \right)^{\frac{1}{p}} = k\mathbf{e}. \tag{3}$$

When $p = 1$ is substituted, equation (3) is equivalent to computing the average of $\mathbf{f}(\mathbf{x})$, i.e. $\mathcal{L}_1$ equals the Grey-World algorithm. When $p = \infty$, equation (3) results in computing the maximum of $\mathbf{f}(\mathbf{v})$, i.e. $\mathcal{L}_\infty$ equals the White-Patch algorithm. When starting with $p = 1$, increasing this value boils down to assigning higher weights to higher pixel values, followed by computing a weighted average (with $p = \infty$ as extreme, where the maximum pixel value is assigned the highest weight possible and the other values are assigned weight 0). Another method of formalizing this trade-off is by first filtering the input image with a Gaussian filter with scale parameter σ followed by taking the maximum of the filtered image in all three channels:

$$\max_{\mathbf{x}} \mathbf{f}^\sigma(\mathbf{x}) = k\mathbf{e}, \tag{4}$$

where $\mathbf{f}^\sigma = \mathbf{f} \otimes \mathbf{G}^\sigma$ is a convolution of the image $\mathbf{f}$ with a Gaussian filter $\mathbf{G}$ with standard deviation σ. When $\sigma = 0$ is substituted (hence, no smoothing is performed), equation (4) is equivalent to the White-Patch algorithm. Similarly, if $\sigma \to \infty$, then every pixel will be assigned the same value, namely the average color in the image, and taking the maximum of this resulting image is equivalent to the Grey-World algorithm.

Recently, these two variations were incorporated into one framework [9], together with higher-order order statistics (i.e. image derivatives), resulting in one color constancy algorithm with three parameters:

$$\left(\int \left| \frac{\partial^n \mathbf{f}^\sigma(\mathbf{x})}{\partial \mathbf{x}^n} \right|^p d\mathbf{x} \right)^{\frac{1}{p}} = k\mathbf{e}^{n,p,\sigma}, \tag{5}$$

where n is the order of the derivative, p is the Minkowski-norm and $\mathbf{f}^\sigma(\mathbf{x}) = \mathbf{f} \otimes \mathbf{G}^\sigma$ is the convolution of the image with a Gaussian filter with scale parameter σ. However, since the Minkowski-norm p and the scale parameter σ actually try to model the same trade-off between the White-Patch algorithm and the Grey-World algorithm, effectively this framework consists of two parameters: the order of the derivative (n) and either the Minkowski-norm p or the scale parameter σ. Since the derivative also depends on the scale parameter σ, it seems a logical choice to keep the Minkowski-norm fixed.

3. USING IMAGE REGIONS

The main idea of using image parts or regions when estimating the illuminant is that certain parts of the image do not contribute to a robust estimate of the illuminant. Even worse, some pixel subsets may even harm the estimation of the light source. For instance, when a large part of the image consists of a blue sky (see the image in figure 1), then these pixel values will cause the estimation to be biased towards blue.

Therefore, instead of using the entire image, we propose to apply a segmentation prior to the estimate of the illuminant. In this way, parts of the image that negatively affect the estimation of the illuminant are ignored in the computation. Since segmentation itself is a very hard task to perform, we propose to incorporate scene knowledge in addition to the segmentation. By incorporating scene knowledge, the type of segmentation that is most suited for a certain image category can be learned in a supervised manner. After a model is learned for several scene categories, these models can be applied to images that are classified by a concept detection algorithm like [10].

For illustration purposes, images of the scene category *open country* are considered. Images in this category are likely to have some blue sky at the upper part of the image, which will cause a bias in the estimation process. In this case, the segmentation step would be to detect the horizon and subsequently ignore the sky part of the image to compute color constancy. Therefore, the main goal is to learn the best segmentation for any specific category. In this paper, a simple segmentation is used, but more complex segmentation algorithms can easily be incorporated. The segmentation that is proposed here is based on a grid: the image is divided into $p \times q$ regions (an 8×8 grid is used). After that, for each image region, the illuminant is estimated and evaluated. After determining the performance of all regions, those regions with the smallest estimation errors are merged together. This is done for a number of (training) images of the same category, and consequently the final segmentation is learned for every image category.

In figure 1 an example is given. This image is segmented into 8×8 regions and for every region, the illuminant is estimation using one single algorithm ($\mathbf{e}^{0,\infty,1}$), see figure 1(b). After that, regions with an accurate estimation (i.e. a low

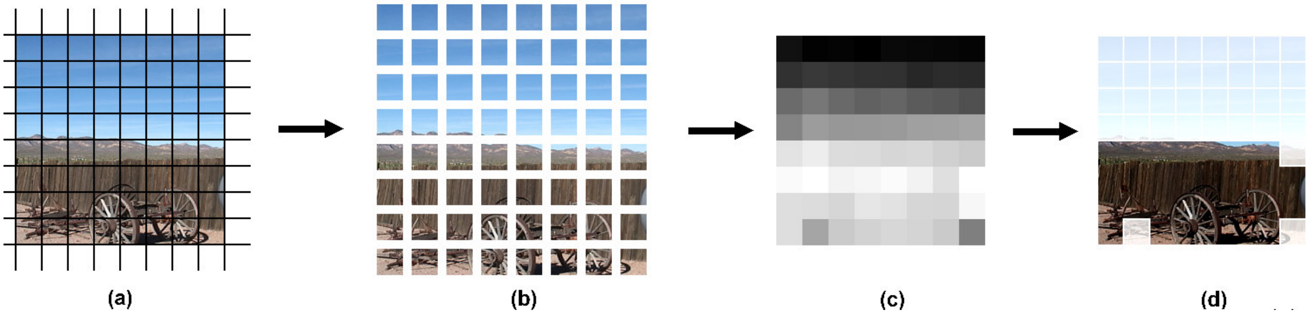

Fig. 1. An example of a basic segmentation into an 8×8 grid. First, the image is segmented into several image regions, figure (b). This segmentation can be done using any segmentation algorithm; in this paper, a grid-based segmentation is used. After that, the performance on every region is determined, figure (c). Finally, after analyzing which regions contain the most valuable information, the part with the blue sky is filtered out, figure (d).

angular error, see figure 1(c)) are merged together, see figure 1(d). Estimation of the illuminant on the combination of these regions results in a more accurate estimation than when using the entire image. By taking the average of several images of the same category, a final segmentation of this scene category can be learned and consequently be applied to other images of the same category.

To summarize, the algorithm is used to determine the final parameters to obtain color constancy for domain specific images, e.g. images that can be categorized into the same category. The algorithm consists of the following steps:

- Segment the image into a number of regions. This can be done using a simple method, but also using more complex methods could be used.
- Estimate the illuminant by using one of the algorithms based on low-level image features [9] for all image regions.
- Determine which region or combination of regions contains the most reliable information to estimate the illuminant. This can be done using one of the data sets with ground truth on the illuminant, like [11]. Alternatively, if one would use color constancy as preprocessing step for a scene recognition task, then the influence of the color constancy algorithms on the performance of the system can be used to determine which image region(s) should be used for estimating the illuminant for a certain category.

4. EXPERIMENTS

In this section, the hypothesis is tested that, for images from different categories, some parts of the images contain more valuable information than other parts for color constancy. To keep the segmentation simple, an 8×8 grid-based segmentation is performed, as shown in figure 1. The categories *open country*, *street* and *indoor* are used as examples, to show the correctness of the hypothesis.

Data set. The algorithm is tested on the data set introduced by [11]. This data set contains $11,000$ images, extracted from 2 hours of video for a wide variety of settings (including indoor, outdoor, desert, cityscape, and other settings). In total, the images are taken from 15 different clips taken at different locations. The main advantage of this data set is the availability of the ground truth of the color of the illuminant. This ground truth is acquired by making use of the small grey sphere in the bottom right corner of the images. Note that this grey sphere is masked while estimating the illuminant.

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

$$\epsilon = \cos^{-1}(\hat{\mathbf{e}}_l \cdot \hat{\mathbf{e}}_e), \tag{6}$$

where $\hat{\mathbf{e}}_l \cdot \hat{\mathbf{e}}_e$ is the dot product of the two normalized vectors representing the true color of the light source $\mathbf{e}_l$ and the estimated color of the light source $\mathbf{e}_e$. To measure the performance of an algorithm on a whole data set, the mean as well as the median angular error is considered [12].

Category-specific performance. To test if this simple segmentation suffices for different categories, a number of images were taken that were annotated as the same scene category. In total, 75 images from 5 clips of the complete data set (15 images per category) were annotated as *open country*, 70 images from 7 clips (10 per category) as *street* and also 70 images from 7 clips as *indoor* (10 per category).

In table 1, the results are shown when using the entire image and when using only learned image regions. The image region that is learned for the category *open country* is shown in figure 1(c), and for this category the performance is considerably better when using this region than when using the entire image. The performance for images from the

Open Country	Mean		Median	
Entire image	8.0°		7.2°	
Proposed method	6.0°	−25%	6.2°	−14%

Street	Mean		Median	
Entire image	5.7°		4.5°	
Proposed method	4.9°	−14%	3.5°	−22%

Indoor	Mean		Median	
Entire image	4.8°		4.0°	
Proposed method	4.7°	−2%	4.0°	±0%

Table 1. Performance of the color constancy framework eq. (5) with fixed parameter settings: $\mathbf{e}^{0,\infty,1}$.

category *street* is also improved by applying the learned segmentation (the segmentation is not shown here, but resembles the segmentation of the category *open country*): the mean angular error decreases from 5.7° when using the entire image to 4.9° when using the learned segmentation. Finally, images from the category *indoor* do not benefit from any segmentation. When learning the final segmentation, it becomes clear that there does not exist such a general segmentation as the other two categories; in fact, the final segmentation consists of nearly the entire image, and the gain in performance is marginal. The median angular error does not change, while the mean angular error goes from 4.8° to 4.7°, which can hardly be called an improvement. In conclusion, it can be derived that for certain categories, like *open country* and *street*, the estimation is far more accurate using image parts than when using the entire image. Note that for scene categories were the segmentation does not improve results, like *indoor*, the performance is similar as applying no segmentation. In such cases, it is learned that the best segmentation is to apply no segmentation at all.

5. CONCLUSION

In this paper, we proposed a method based on low-level image features using subsets of pixels. A simple segmentation is performed to learn for different categories which pixel set is most appropriate for a reliable estimation.

Experiments applied on real-world images show that for certain categories, like *open country* and *street*, the estimation is more accurate using only parts of the image than when using the entire image. For other types of categories (for instance *indoor*), the performance is similar to the performance when no segmentation is applied. Performance can probably be increased even further by learning a final segmentation for several color constancy algorithms. The different regions that are learned will have different properties (in terms of contrast and texture), and the different color constancy algorithms using low-level features are known to be more effective on images with specific natural image statistics[13].

6. REFERENCES

[1] Th. Gevers and A.W.M. Smeulders, "Pictoseek: combining color and shape invariant features for image retrieval," *IEEE Trans. Im. Proc.*, vol. 9, no. 1, pp. 102–119, 2000.

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

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

[4] G.D. Finlayson, S.D. Hordley, and I. Tastl, "Gamut constrained illuminant estimation," *Int. J. Compututer Vision*, vol. 67, no. 1, pp. 93–109, 2006.

[5] D.A. Forsyth, "A novel algorithm for color constancy," *Int. J. Compututer Vision*, vol. 5, no. 1, pp. 5–36, 1990.

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

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

[8] G.D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. of the Twelfth Color Imaging Conference*. 2004, pp. 37–41, IS&T - The Society for Imaging Science and Technology.

[9] J. van de Weijer, Th. Gevers, and A. Gijsenij, "Edge-based color constancy," Accepted for publication in Trans. on Image Processing, 2007.

[10] J. van Gemert, J. Geusebroek, C. Veenman, C. Snoek, and A. Smeulders, "Robust scene categorization by learning image statistics in context," in *SLAM, in conjunction with CVPR*, New York, USA, June 2006.

[11] F. Ciurea and B.V. Funt, "A large image database for color constancy research," in *Proc. of the Eleventh Color Imaging Conference*. 2003, pp. 160–164, IS&T - The Society for Imaging Science and Technology.

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

[13] A. Gijsenij and Th. Gevers, "Color constancy using natural image statistcs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, USA, 2007.