Illumination Estimation based on Bilayer Sparse Coding

Bing Li Weihua Xiong Weiming Hu Houwen Peng
National Laboratory of Pattern Recognition (NLPR), Institute of Automation,
Chinese Academy of Sciences, Beijing 100190, China
bli@nlpr.ia.ac.cn, wallace.xiong@gmail.com, wmhu@nlpr.ia.ac.cn, houwen.peng@nlpr.ia.ac.cn

Abstract

Computational color constancy is a very important topic in computer vision and has attracted many researchers' attention. Recently, lots of research has shown the effects of using high level visual content cues for improving illumination estimation. However, nearly all the existing methods are essentially combinational strategies in which image's content analysis is only used to guide the combination or selection from a variety of individual illumination estimation methods. In this paper, we propose a novel bilayer sparse coding model for illumination estimation that considers image similarity in terms of both low level color distribution and high level image scene content simultaneously. For the purpose, the image's scene content information is integrated with its color distribution to obtain optimal illumination estimation model. The experimental results on real-world image sets show that our algorithm is superior to some prevailing illumination estimation methods, even better than some combinational methods.

1. Introduction

The color signals of any object from an imaging device are determined by three factors: the color of light incident on the scene, the surface reflectance of the object, and sensor sensitivity function of the camera [7] [8]. Therefore, the color of same surface will usually appear differently under varying light sources. In contrast, the human beings have the ability to "see" a surface as having the same color independent of variations of the illumination, which is called "Color Constancy" [16]. Computational color constancy is targeted for providing the same sort of color stability in the context of computer vision [1], and its central issue is to build up an optimal illumination estimation model.

1.1. Related Work

Illumination estimation is actually an ill-posed problem and cannot be solved without any assumption. It has been an active research topic in both scientific community and imaging industry for several decades. Most early studies treat an image as a bag of pixels with RGB values and give out the illumination estimation model without considering the underlying semantic content expressed by the pixels' arrangement. We name these methods as "Data Driven Estimation Methods"(DD). The DD methods can further be classified into unsupervised DD methods and supervised DD methods. The unsupervised DD methods, such as Grey World (GW)[11], maxRGB [24], Shades of Grey (SoG)[18], and Edge-based method [29] (also called Grey Edge, GE), etc, predefine fixed illumination estimation models based on certain hypotheses for all images. On the other hand, the supervised DD methods, including Color-by-Correlation (C-by-C) [17], Spatio-Spectral statistics-based method (Spatio-Spectral) [13], Neural Networks-based method (NN) [12], Support Vector Regression-based method (SVR) [32], Gamut Mapping [22], edge-based Gamut Mapping [22] etc, learn the estimation models on the color distribution or related features of training data through the training procedures. Although the DD methods are simple and have much lower complexities; the fixed estimation models embedded in them result in lower generalizations. Once the model is fixed in a DD method, the illumination colors of all the test images are computed out using the same model. Therefore, the DD methods are effective only when the distribution of colors of the test image fits the assumed model very well.

In order to avoid the fixed model problem, many researchers focus on the model selection or combination for illumination estimation. Recent years have witnessed a rise in applying image content analysis to guide illumination estimation. We name these methods as "Content Driven Estimation Methods" (CD). All the existing CD methods essentially are combinational methods [25] that generally contain two steps: (1) applying several DD estimation models (rather than only one) on the same image, (2) then selecting the best estimate or combining their outputs based on the image's content characteristics. Previous efforts in this area include the work of Gijsenij et al. [20], which selects the most appropriate unsupervised DD method based on natu-

ral texture statistics and scene semantics of the test image (NIS). Lu et al. [27] use 3D stage geometry model (SG) to divide images into different geometrical regions, and select appropriate estimations per depth layer or geometrical section. Bianco et al. [10] propose to use the indoor/outdoor scene classification for choosing the most appropriate estimation method (IO). Weijer et al. [30] use high level visual information for improving illumination estimation (HVI), in which an image is modeled as a mixture of semantic classes, such as sky, grass, road, and building. Then they evaluate several different illumination estimation models on the likelihood of its semantic content in correspondence with prior knowledge of the world, and produce the final output that results in the most likely semantic composition of the image. According to the analysis on the CD methods, we obtain the following observations:

- Since most existing CD methods are combinational methods, their performance is inevitably affected by the DD methods used for combination.
- Nearly all the existing CD methods combine the unsupervised DD methods. The supervised DD methods, which generally have better performance [21][25], have not been considered.
- Although the high level scene content is useful for illumination estimation, automatical scene content classification, such as 3D stage classification or indoor/outdoor classification, is another difficult and unsolved computer vision problem.

1.2. Our work

According to the observations above, this paper proposes a novel bilayer sparse coding model (BSC) for illumination estimation that integrates the high level content cues and low level color features into a unified supervised framework. The proposed BSC method models illumination estimation as an image similarity problem and considers low level color distribution and high level scene category simultaneously. Our work is primarily inspired by two hypotheses: (1) The images with similar color distributions are preferable to be captured under the similar light colors; and (2) the scenes belonging to the same high level category have the similar illumination conditions [10]. This is because the varying range of light colors in a certain type of scene is often limited. For example, indoor lights tend to be red; while outdoor lights are mostly bluish. The first hypothesis has been validated in many supervised DD methods. The second one has also been shown to be effective in some CD methods [20][27][10].

The proposed BSC method sounds similar to the IO algorithm [10], but they are completely different in essence:

- The BSC method is an individual supervised methods rather than a combinational method. It is not based on any other DD methods. So the BSC method is to directly estimate the illumination color of the test image based on the training images that are similar to the test image from both color and scene viewpoints.
- The BSC method need not explicitly classify the scene into predefined indoor/outdoor or other scene categories. Instead, it integrates the high level scene content similarity into the supervised illumination estimation procedure so as to avoid negative impact of incorrect hard scene classification. In addition, the scene categories are far more than indoor/outdoor.
- Another contribution in the proposed method lies in unfixed model. Since the sparse coding is a no-model learning algorithm. Compared with most existing methods that always use a prefixed model(or a limited model set for selection) for all the test images, our BSC algorithm adaptively learns a individualized model for each test image according to its color and scene cues.

2. Sparse Coding Preliminaries

Before introducing the details of our model, we start with a brief overview of sparse coding that is the basis of the proposed algorithm. Recently, much interest has been shown in computing linear sparse representation with respect to an overcomplete dictionary of the basis elements. The goal of sparse coding is to sparsely represent input vectors approximately as a weighted linear combination of a number of "basis vectors". Given an input vector $x \in R^k$ and basis vectors $\mathbf{U} = [u_1, u_2, ..., u_n] \in R^{k \times n}$, sparse coding aims to find a sparse vector of coefficients $\alpha \in R^n$, such that $x \approx \mathbf{U}\alpha = \sum_j u_j \alpha_j$. It equals to solving the following objective.

$$\min_{\alpha} \|x - \mathbf{U}\alpha\|_2^2 + \lambda \|\alpha\|_0 , \tag{1}$$

where $\|\alpha\|_0$ denotes the ℓ_0-norm, which counts the number of nonzero entries in a vector α. It is well known that the sparsest representation problem is NP-hard in general case. Fortunately, recent results [31] show that, if the solution is sparse enough, the sparse representation can be recovered by the following convex ℓ_1-norm minimization [31] as:

$$\min_{\alpha} \|x - \mathbf{U}\alpha\|_2^2 + \lambda \|\alpha\|_1 , \tag{2}$$

where the first term of Eq(2) is the reconstruction error, and the second term is to control the sparsity of the coefficients vector α with the ℓ_1-norm. λ is regularization coefficient to control the sparsity of α. The larger λ implies the sparser solution of α. The sparse coding technique based on ℓ_1-norm has been widely applied in many applications, including face recognition, image classification, etc [31].

3. Bilayer Sparse Coding for Illumination Estimation

In this section, we firstly propose bilayer sparse coding model (BSC) for illumination estimation; then discuss color feature and scene feature used in BSC; and finally give out an optimization algorithm for BSC.

3.1. Bilayer Sparse Coding Model (BSC)

3.1.1 Sparse Coding for Color Similarity

Given N training images $I_1, ..., I_N$, the color feature vector of the image I_i is $C_i \in R^d$. Here, color feature C_i can be binarized 2D/3D chromaticity histogram that has been proved to be effective for many supervised color constancy algorithms [17][12][32]. For any test image I_y with color feature $C_y \in R^d$, we can linearly reconstruct its color feature using the training images under the sparse coding framework, as:

$$\min_{\gamma} \|C_y - \mathbf{C}\gamma\|_2^2 + \lambda \|\gamma\|_1 , \tag{3}$$

where $\mathbf{C} = [C_1, C_2, ..., C_N]$; $\gamma = [\gamma_1, \gamma_2, ..., \gamma_N]^T$ is a N-dimensional coefficient vector that indicates the reconstruction weight associated with each training image. From viewpoint of color gamut, the Eq(3) is actually to reconstruct the color gamut of the test image using color gamut of all the training images. The sparse code γ can also be viewed as the color correlation coefficient between I_y and each training image.

3.1.2 Sparse Coding for Scene Category Similarity

Generally speaking, a typical type of scene is determined by a bag of certain objects and their co-occurrence relationships [23]. For example, the 'street' scene sometimes contains roads and buildings.

To model appearances of different objects in the scene, we segment each training image I_i into n_i objects, denoted as $I_i^1, I_i^2, ..., I_i^{n_i}$, then we have $n_1 + n_2 + ... + n_N$ objects in the training image set in all. Each object I_i^k is represented by visual vocabulary histogram $v_i^k \in R^m$ that is gained from Bag-of-Words model (BOW) [2]; and all the objects in I_i are denoted as $V_i = [v_i^1, v_i^2, ..., v_i^{n_i}] \in R^{m \times n_i}$. The test image I_y is also segmented into n_y objects $I_y^1, I_y^2, ..., I_y^{n_y}$, their corresponding vocabulary histograms are represented as $v_y^1, v_y^2, ..., v_y^{n_y} \in R^m$. The scene category similarity analysis here is to reconstruct the n_y objects in the test image by using the $n_1 + n_2 + ... + n_N$ objects in the training images, as show in Figure 1.

Considering co-occurrence property of objects in the same image, we should try to reconstruct the objects in the test image using those objects from the same training image. Therefore, we introduce the multi-task joint sparse representation based on $\ell_{2,1}$ norm [33]. The multi-task joint sparse representation can be regarded as a combinational model of group Lasso and multi-task Lasso by penalizing the sum of ℓ_2 norms of the blocks of coefficients associated with each covariate group (objects in each training image) across different reconstruction tasks (object reconstruction in the test image)[33].

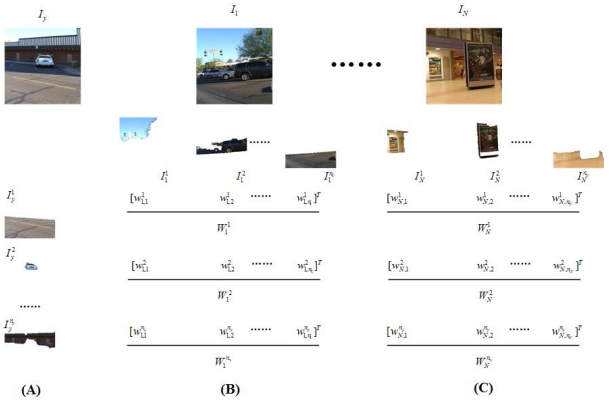

Figure 1. Sparse reconstruction of image's scene content: (A) test images I_y and its segmented objects. (B) Training image I_1 and its segmented objects, $W_1^j, (j = 1, 2, ..., n_y)$ is a reconstruction coefficient vector of the j^{th} object in I_y associated with all the objects in I_1. (C) Training images I_N and its segmented objects, $W_N^j, (j = 1, 2, ..., n_y)$ is reconstruction coefficient vector of the j^{th} object in I_y associated with all the objects in I_N.

For any test object I_y^j in the test image I_y, if $W_i^j \in R^{n_i}$ denotes the reconstruction coefficient associated with the objects $I_i^1, I_i^2, ..., I_i^{n_i}$ in the image I_i, we can use $W_i = [W_i^1, W_i^2, ..., W_i^{n_y}] \in R^{n_i \times n_y}$ to represent the reconstruction coefficient matrix of all the objects in I_y associated with all the objects in the image I_i . The details of corresponding relationship between objects and coefficient are shown in Figure 1. The joint sparse representation of all the objects in the test image can be formulated as [33]:

$$\min_{\mathbf{W}} \sum_{j=1}^{n_y} \left\| v_y^j - \sum_{i=1}^{N} V_i W_i^j \right\|_2^2 + \beta \sum_{i=1}^{N} \|W_i\|_2^1, \tag{4}$$

where $\mathbf{W} = [W_1, W_2, ..., W_N]^T$ is the sparse coefficient matrix for all the objects in the test image; β is the regularization coefficient. The optimization problem in Eq(4), which is known as multi-task joint covariant selection in Lasso related research, can be effectively solved by $\ell_{2,1}$ mixed-norm Accelerated Proximal Gradient (APG) algorithm proposed by Yuan et al [33].

3.1.3 Bilayer Sparse Coding for Illumination Estimation(BSC)

In order to integrate scene category information into illumination estimation model, a bilayer sparse coding model for

illumination estimation is formulated to include similarity analysis on both color distribution and scene category, as:

Color Layer:

$$\begin{aligned} &\min_{\gamma} \|C_y - \mathbf{C}\gamma\|_2^2 + \lambda \|\mathbf{D}\gamma\|_1, \\ &\mathbf{D} = diag(f(\|W_1\|_2^1), f(\|W_2\|_2^1), ..., f(\|W_N\|_2^1)), \end{aligned} \tag{5}$$

Scene Layer:

$$\min_{\mathbf{W}} \sum_{j=1}^{n_y} \left\| v_y^j - \sum_{i=1}^{N} V_i W_i^j \right\|_2^2 + \beta \sum_{i=1}^{N} g(\|\gamma_i\|_1) \|W_i\|_2^1, \tag{6}$$

where

$$f(\|W_i\|_2^1) = \frac{\max\limits_{k=1..N}(\|W_k\|_2^1) - \|W_i\|_2^1 + \varepsilon}{\max\limits_{k=1..N}(\|W_k\|_2^1) - \min\limits_{k=1..N}(\|W_k\|_2^1) + \varepsilon}, \tag{7}$$

$$g(\|\gamma_i\|_1) = \frac{\max\limits_{k=1..N}(\|\gamma_k\|_1) - \|\gamma_i\|_1 + \varepsilon}{\max\limits_{k=1..N}(\|\gamma_k\|_1) - \min\limits_{k=1..N}(\|\gamma_k\|_1) + \varepsilon}. \tag{8}$$

From the formulation above, we can find that the function $f(\|W_i\|_2^1)$ and $g(\|\gamma_i\|_1)$ are the monotone decreasing functions, in which ε is used to avoid 0. Their outputs are between $(0, 1]$ and can be viewed as the costs in sparse color reconstruction and sparse scene content reconstruction. In the color layer, it tends to select the images with lower $f(\|W_i\|_2^1)$ values, which is corresponding to the higher $\ell_{2,1}$ norm $\|W_i\|_2^1$ of the scene reconstruction coefficient W_i, to reconstruct the test image's color feature. Similarly, in the scene layer, it tends to select the images with lower $g(\|\gamma_i\|_1)$, which is corresponding to the higher ℓ_1 norm $\|\gamma_i\|_1$ of the color reconstruction coefficient γ_i, to reconstruct the test image's scene content. Comparing Eq(5) with Eq(3) can tell us that the γ in BSC model contains not only color correlation but also scene content correlation information. The optimization of the BSC model will be discussed in section 3.3.

3.1.4 Illumination Estimation

The coefficient γ in Eq(5), which represents the correlation between the test image and all training images, is used for illumination estimation. To remove the shading effect, the ground truth illumination color value $e_i = (R_i, G_i, B_i)^T$ of the training image I_i is mapped into 2D chromaticity space through: $l_i = \left(\frac{R_i}{R_i+G_i+B_i}, \frac{G_i}{R_i+G_i+B_i}\right)^T$. And the coefficient vector γ is also normalized by ℓ_1 norm as: $\hat{\gamma} = \frac{|\gamma|}{\|\gamma\|_1}$. So the final illumination chromaticity $l_y = (r_y, g_y)^T$ of the test image can be estimated as the weighted average of the illumination values of all the training images as:

$$l_y = \mathbf{L}\hat{\gamma}, \qquad \mathbf{L} = [l_1, l_2, ..., l_N] \tag{9}$$

3.2. Feature Extraction

This section discusses the feature extraction in the B-SC model. In the color reconstruction layer, we consider 3D color space as [32]: two chromaticity values, defined as $(r, g)^T = \left(\frac{R}{R+G+B}, \frac{G}{R+G+B}\right)^T$, and one intensity value, defined as $L = (R+G+B)$. The chromaticity space $(r, g)^T$ is equally partitioned along each component into 50 equal parts yields 2500 bins. The intensity L is quantized into 25 equal steps [32][9], so the 3D color histograms consist of 62,500 ($50 \times 50 \times 25$) bins [32]. Each image is represented as a binarized 3D chromaticity histogram, in which '1' or '0' indicates the presence or absence of the corresponding chromaticity and intensity in the image.Since $0 \leq r+g \leq 1$, a compact 3D chromaticity histogram can be obtained by discarding the space with $r+g > 1$.

In the scene layer, the SIFT descriptor [26] that is widely applied to scene classification, is used as object's visual feature under the Bag-of-Word (BoW) model [2]. Considering that the scene layer is to find the training images with both similar scene contents and similar illumination conditions to the test image, color SIFT descriptor on r-g chromaticity space is used as scene feature. The dense SIFT descriptors are extracted with 8-pixel step for each image. Then all the SIFT descriptors in the training images are clustered as m visual words via Kmeans scheme. Finally, each segmented region with the corresponding SIFT descriptors in it is represented as a m-dimensional visual words histogram $v_i^k \in R^m$ via BoW model.

3.3. Optimization for Bilayer Sparse Coding Model

The optimization in Eq(5) and Eq(6) is not straightforward. However, if the value of γ is fixed, the optimization in scene layer is just a multi-task joint sparse coding, which can be effectively solved via the $\ell_{2,1}$ mixed-norm Accelerated Proximal Gradient (APG) algorithm [33]. On the other hand, if the coefficient matrix $\mathbf{W}$ is given, the optimization in color layer is just a general sparse coding with a cost constrain that can also be solved by the $\ell_{2,1}$ mixed-norm APG. Consequently, we give an approximate iterative $\ell_{2,1}$ mixed-norm APG algorithm to optimize the bilayer sparse coding as shown in Algorithm 1.

At each iteration, the new values of γ or $\mathbf{W}$ is obtained for the next iteration. The $\|\hat{p} - \hat{\gamma}\| \leq \delta$, which indicates the distance between successive solutions of γ, is the stopping condition of the iterations.

4. Experiments

We evaluate the proposed BSC algorithm on two real-world image sets. The first one is provided by Gehler et al.[19][3] and subsequently reprocessed by Shi et al.[4] (denoted as Gehler-Shi set). The second one includes the real-

Algorithm 1 Pseudo-code for bilayer sparse coding optimization.

Input: The color feature C_i and scene feature $V_i = [v_i^1, v_i^2, ..., v_i^{n_i}] \in R^{m \times n_i}$ of each training image, the color feature C_y and scene feature $V_y = [v_y^1, v_y^2, ..., v_y^{n_y}]$ of the test image,the regularization coefficient λ and β, the threshold ε.

1: Initialize $\mathbf{D} = diag(1, 1, ..., 1)$, solve γ in Eq(5) via the $\ell_{2,1}$ mixed-norm APG.
2: **repeat**
3: Set $p = \gamma$.
4: **for** $i = 1 \rightarrow N$ **do**
5: Compute $g(\|\gamma_i\|_1)$.
6: **end for**
7: Solve $\mathbf{W}$ in Eq(6) with $g(\|\gamma_i\|_1)$ via the $\ell_{2,1}$ mixed-norm APG algorithm.
8: **for** $i = 1 \rightarrow N$ **do**
9: Compute $f(\|W_i\|_2^1)$..
10: **end for**
11: Solve γ in Eq(5) with $f(\|W_i\|_2^1)$ via the $\ell_{2,1}$ mixed-norm APG algorithm.
12: **until** ($\|\hat{p} - \hat{\gamma}\| \leq \delta$ or max iteration times are arrived)

Output: γ and $\mathbf{W}$.

world images captured from a digital video provided by Ciurea et al [14]; and then is linearized by Gijsenij et al [21] to remove the gamma-correction (denoted as Linear SFU set).

The BSC method is compared with some leading illumination estimation methods, including GW [11], maxRGB [24], Grey Edge (0^{th}, 1^{st}, 2^{nd}-order)[29], Gamut Mapping [22], Spatio-Spectral [13], SVR[32], HVI [30] and NIS[20]. All the parameter settings for each algorithm are determined according to the settings in the excellent survey [21][6].The binarized 3D color histogram is also used in the SVR method. There are three parameters that are regularization coefficients λ, β and number of visual words m in BoW need to be set in advance in the BSC algorithm. The optimal parameters λ, β and m are selected out from $\lambda, \beta \in \{0.01, 0.1, 1, 10\}, m \in \{500, 1000, 1500\}$ through 3-fold cross validation on training set in each experiment. The JSEG algorithm[15] is used to segment each object in the image due to its flexibility in adjusting the number of regions. In order to further validate the effect of the scene category for illumination estimation, the single color layer in BSC (denoted as SSC) excluding any scene cue is also used in comparison. The matrix $\mathbf{D}$ in SSC is always fixed as $\mathbf{D} = diag(1, 1, ..., 1)$.

4.1. Error Measurement

The error measurements is one of the most important issues in experiments. For each image in the image sets, the ground truth chromaticity of the light source $e_a = (r_a, g_a, b_a)$ is known. To measure how close the estimated illumination resembles the true color of the light source, the angular error measurement, which is the angular distance between the estimated illumination chromaticity $e_y = (r_y, g_y, 1-r_y-g_y)^T$ and the ground truth chromaticity e_a, is adopted to evaluate the performance of diverse algorithms. The angular error function $angular(e_y, e_a)$ is defined as

$$angular(e_y, e_a) = \frac{180^o}{\pi} \cos^{-1}\left(\frac{e_y \bullet e_a}{\|e_y\| \, \|e_a\|}\right), \quad (10)$$

where $e_y \bullet e_a$ is the dot product of e_y and the e_a ; and $\|\bullet\|$ indicates the Euclidean norm. The mean, median, trimean, best-25% and worst-25% angular errors are used to measure the performance of each algorithm on a data set [21]. The worst-25% (or best-25%) indicates the mean angular error of the largest (or smallest) 25% angular errors on test images. In addition, to provide more insight into the complete distribution of errors on an image set, we also compute the on a data set.

4.2. Experimental Results on the Gehler-Shi Set

The Gehler-Shi image set contains 568 images that are taken using two high quality DSLR cameras (Canon 5D and Canon1D) and includes a wide variety of indoor and outdoor shots. All the images were saved in Canon RAW format. Because the tiff images provided by Gehler et al [19] in this set were produced automatically, they contain clipped pixels that are non-linear (i.e., have gamma or tone curve correction applied) and include the effect of the camera's white balancing. To avoid these problems, Shi et al. [4] reprocessed the raw data and created almost-raw 12-bit Portable Network Graphics (PNG) format images. This results in a 2041×1359 (for Canon 1D) or 2193×1640 (for Canon 5D) linear images (gamma=1) in camera RGB space. Consequently, the reprocessed Gehler-Shi set is used in the following experiments.

The same as the setting in [21], the 3-fold cross-validation strategy is conducted on this set. The three folds are provided by the authors of the data set and to ensure repeatability of the results. During the experiment, one subset is picked as test set; the other two are used as training set. This procedure is repeated 3 times with different test set selection, the overall performance is used as the final result. The final experimental results are shown in Table 1.The Do Nothing method always estimates the illuminate as being white ($r = g = b$).

According to the results in Table 1, the proposed BSC, Gamut Mapping, and HVI methods, which outperform all the other methods, are comparable to each other. The BSC method achieves much lower median, trimean and best-25% angular errors. The low worst-25% error of the BSC method

Table 1. Comparison of performance on the Gehler-Shi image set. The performance of other algorithms is from [21]

Algorithm	Mean	Median	Trimean	Best-25%	Worst-25%
Do Nothing	13.7	13.6	13.5	10.4	17.2
maxRGB	7.5	5.7	6.4	1.5	16.2
GW	6.4	6.3	6.3	2.3	10.6
general GW ($\mathbf{e}^{0,p,\sigma}$)	4.7	3.5	3.8	1.0	10.1
1^{st}-order GE ($\mathbf{e}^{1,p,\sigma}$)	5.4	4.5	4.8	1.9	10.0
2^{nd}-order GE ($\mathbf{e}^{2,p,\sigma}$)	5.1	4.4	4.6	1.9	10.0
Gamut Mapping	4.1	2.5	3.0	0.6	10.3
Edge-based Gamut Mapping	6.7	5.5	5.8	2.1	13.7
Spatio-Spectral	5.9	5.1	5.4	2.4	10.8
SVR	8.1	6.7	7.2	3.3	14.9
HVI	**3.5**	2.5	2.6	0.8	**8.0**
NIS	4.2	3.1	3.5	1.0	9.2
SSC	4.8	3.8	4.0	1.0	10.8
BSC	4.0	**2.5**	**2.8**	**0.6**	9.6

implies the stableness of our algorithm. The HVI and NIS also have good performance. In addition, the performance of the BSC method is much better than SSC method. The two facts imply that high level scene category cues can indeed improve the illumination estimation. Furthermore, the SSC outperforms SVR method, which shows that the sparse coding technique is a good alternative learning tool for illumination estimation.

4.3. Experimental Results on the Linear SFU Set

The second image set is introduced by Ciurea et al. [14] which consists of more than 11,000 frames from videos. Since a matte grey sphere ball is mounted onto the video camera to obtain the ground truth illumination of each image; in order to ensure that the grey ball has no effect on our results, the grey sphere is masked during experiments. Another issue of this data set is that an unknown post-processing procedure is applied to the images by the camera, including gamma-correction and compression. Gijsenij et al [21] created a modification of this set by applying gamma-correction (with $\gamma = 2.2$). For consistency, the ground truth is also recomputed on the linear images [6]. The recomputed linear image set is also used in this experiment. Since the SFU set contains 15 subcategories from which images are taken in different places, the 15-fold cross-validation is adopted here to ensure that the correlated images of the same scene in the same group [21]. Then one subset is used for testing and the other 14 ones are used for training. This procedure is repeated 15 times. The overall performance is shown in Table 2.

The proposed BSC method outperforms all other methods, even better than the combinational method NIS and HVI. Some similar conclusions to previous experiment can also be obtained. The SSC method also achieves much better performance than all the other methods except NIS and HVI, which again implies the effect of the sparse code technique for illumination estimation. The fact that the BSC method outperforms SSC further confirms the effect of scene category cues for illumination estimation.

5. Conclusion

Image's high level content cue has been evidenced to be helpful for improving the illumination estimation. However, most prevailing methods using high level content cues can be viewed as combinational methods. In this paper, we integrate image's color distribution and scene content analysis into a unified bilayer sparse coding framework for illumination estimation. The experiments on real-world image sets show that the mutually constrained combination can improve the accuracy of illumination estimation.

Acknowledgements

This work is partly supported by the National Nature Science Foundation of China (No. 61005030, 60935002, 60825204, 61100142, 61272352, and 61202245) and Chinese National Programs for High Technology Research and Development (863 Program) (No. 2012AA012503 and No. 2012AA012504).

References

[1] K. Barnard. Practical Colour Constancy. PhD thesis, Simon Fraser University, Canada, 1999.
[2] Fei-Fei Li. Bag-of-Words model. Tutorial. In CVPR, 2007.
[3] http://www.kyb.mpg.de/bs/people/pgehler/colour/.
[4] L. Shi, B. Funt. Re-processed Version of the Gehler Color Constancy Dataset of 568 Images. http://www.cs.sfu.ca/∼colour/data/.
[5] http://www.cvc.uab.es/color_calibration/Database.html.
[6] http://www.colorconstancy.com.
[7] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms-part 1: Methodology and

Table 2. Comparison of performance on the Linear SFU image set. The performance of other algorithms is from [21]

Algorithm	Mean	Median	Trimean	Best-25%	Worst-25%
Do Nothing	15.6	14.0	14.6	2.1	33.0
maxRGB	12.7	10.5	11.3	2.5	26.2
GW	13.0	11.0	11.5	3.1	26.0
general GW ($\mathbf{e}^{0,p,\sigma}$)	12.6	11.1	11.6	3.8	23.9
1^{st}-order GE ($\mathbf{e}^{1,p,\sigma}$)	11.1	9.5	9.8	3.2	21.7
2^{nd}-order GE ($\mathbf{e}^{2,p,\sigma}$)	11.2	9.6	10.0	3.4	21.7
Gamut Mapping	11.8	8.9	10.0	2.8	24.9
Edge-based Gamut Mapping	13.7	11.9	12.3	3.7	26.9
Spatio-Spectral	12.7	10.8	11.5	2.4	26.0
SVR	13.1	11.2	11.8	4.4	25.0
HVI	9.7	7.7	8.2	2.3	20.6
NIS	9.9	7.7	8.3	2.4	20.8
SSC	10.8	8.5	9.1	2.4	22.8
BSC	**9.2**	**7.3**	**7.8**	**2.1**	**19.6**

experiments with synthesized data. *IEEE TIP*, 11(9):972–983, 2002.

[8] K. Barnard, L. Martin, A. Coath, and B. Funt. A comparison of computational color constancy algorithms-part 2: Experiments with image data. *IEEE TIP*, 11(9):985–996, 2002.

[9] K. Barnard, L. Martin, and B. Funt. Color by correlation in a three-dimensional color space. In *Proc. of ECCV*, pages 375–389, 2000.

[10] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini. Improving color constancy using indoor-outdoor. *IEEE TIP*, 17(12):2381–2392, 2010.

[11] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):337–350, 1980.

[12] V. Cardei, B. Funt, and K. Barnard. Estimating the scene illumination chromaticity using a neural network. *JOSA A*, 19(12):2374–2386, 2002.

[13] A. Chakrabarti, K. Hirakawa, and T. Zickler. Color constancy with spatio-spectral statistics. *IEEE TPAMI*, 34(8):1509–1519, 2012.

[14] F. Ciurea and B. Funt. A large image database for color constancy research. In *Proc. of IS&T/SID Color Imaging Conference*, pages 160–164, 2003.

[15] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE TPAMI*, 23(8):800–810, 2001.

[16] M. Ebner. *Color Constancy*. John Wiley & Sons, 2006.

[17] G. Finlayson, S. Hordley, and P. Hubel. Color by correlation: A simple, unifying framework for color constancy. *IEEE TPAMI*, 22(11):1209–1221, 2001.

[18] G. Finlayson and E. Trezzi. Shades of gray and color constancy. In *Proc. of IS&T/SID Color Imaging Conference*, pages 37–41, 2004.

[19] P. V. Gehler, C. Rother, A. Blake, and T. Minka. Bayesian color constancy revisited. In *Proc. of CVPR*, pages 1–8, 2008.

[20] A. Gijsenij and T. Gevers. Color constancy using natural image statistics and scene semantics. *IEEE TPAMI*, 33(4):687–698, 2011.

[21] A. Gijsenij, T. Gevers, and J. van de Weijer. Computational color constancy: Survey and experiments. *IEEE TIP*, 20(9):2475–2489, 2011.

[22] A. Gijsenij, T. Gevers, and J. V. Weijer. Generalized gamut mapping using image derivative structuresfor color constancy. *IJCV*, 86(2):127–139, 2010.

[23] D. Gokalp and S. Aksoy. Scene classification using bag-of-regions representations. In *Proc. of CVPR*, pages 1–8, 2007.

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

[25] B. Li, W. Xiong, and W. Hu. Evaluating combinational color constancy methods on real-world images. In *Proc. of CVPR*, pages 1929–1936, 2011.

[26] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[27] R. Lu, A. Gijsenij, and T. Gevers. Color constancy using 3d scene geometry. In *Proc. of ICCV*, pages 1749–1756, 2009.

[28] J. Vazquez-Corral, C. A. Parraga, M. Vanrell, and R. Baldrich. Color constancy algorithms: psychophysical evaluation on a new dataset. *JIST*, 55(3):31105–31109, 2009.

[29] J. V. Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE TIP*, 16(9):2207–2214, 2007.

[30] J. V. Weijer, C. Schmid, and J. Verbeek. Using high-level visual information for color constancy. In *Proc. of ICCV*, pages 1–8, 2007.

[31] J. Wright, Y. Ma, J. Mairal, and G. Sapiro. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044, 2010.

[32] W. Xiong and B. Funt. Estimating illumination chromaticity via support vector regression. *Journal of Imaging Science and Technology*, 50(4):341–348, 2006.

[33] X. Yuan and S. Yan. Visual classification with multi-task joint sparse representation. In *Proc. of CVPR*, pages 3493–3500, 2010.