Estimating Illumination Chromaticity Based on Structured Support Vector Machine

Zheng TANG, Hong-zhe LIU*, Jia-zheng YUAN, Chao LI and Yong-rong ZHENG

Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China
*Corresponding author

Keywords: Color constancy, Multi-illuminant, White balance.

Abstract. Illumination estimation is crucial to the color constancy problem. Most existing color constancy algorithms assume uniform illumination. However, in real-world scenes, this is not often the case. Thus, we propose a novel framework for estimation the colors of multiple illuminants and their spatial distribution in the scene. We formulate this problem as an illuminant color estimate is obtained independently from distinct image main-regions. We think there is only one kind of light in each main-regions. From these main-regions a robust local illumination color is computed by consensus. So we can solve the problem of multiple illumination estimation using single illuminant color constancy method. The performance if our method is evaluated on multiple data sets. Experimental results show that we can reliably process mixed illuminant scenes.

Introduction

At present, most of the color constancy algorithm assumes that the scene illumination is evenly distributed in the whole scene; there is only one light value. In real life, however, there are often multiple illuminations in the scene, as shown in Figure 1. In indoor scenes, as shown in Figure 1(a), some areas are mainly indoor tungsten light source, and some areas are both indoor tungsten light and outdoor light source. As shown in Figure 1(b), an outdoor scene, where parts of the image are in direct sunlight, while others are in shadow which is illuminated by blue skylight. Color is one of the most basic features of machine vision, in the field of image processing and computer vision have been widely used, such as image segmentation, object recognition, image matching, video retrieval and others. The key to cancel out the effect of incident light is to estimate the color of the overall scene illumination. Unfortunately, color constancy is an ill-posed problem, so light color cannot be gained without any assumptions [1]. Because of the multi-illumination scene, the change of the illumination value will change with the change of its spatial position. In this case, if the assumption of scene illumination is uniform, and the use of single color under illumination condition estimation those. And based on the illumination value to correct the image color, will produce a lot of deviation, is not accurate to restore the surface color of the object. Compared with the single illuminant Chromaticity estimation, multiple color illumination estimation is more challenging. Extending existing color constancy methods to successfully compute multi-illuminant estimates is a challenging problem [14].

(a) indoor scene  (b) outdoor scene

Figure 1. Different illuminant scenes.
Consider two of the most popular branches of existing color constancy approaches: statistic-based methods and machine learning-based methods [14]. The success of statistics-based techniques [1], [2], [3], [4] depends on the size of the statistical sample. Applying these methods to small image regions introduces inaccuracies [5] and is unlikely to yield stable results. This kind of method originates from the Retinex theory [7], the color of objects is reflected by the object of long wave, medium wave, short wave to decide, rather than by the intensity of light. Based on the Retinex theory and algorithm, the researchers proposed the hypothesis of Grey World and Patch White hypothesis [6]. Patch White algorithm is based on the assumption of Patch White [7]. The maximum response in a scene is caused by a perfect reflectance. Because the white surface can completely reflect the color of the scene, so the maximum value of the RGB three color channels will be used as the image of the light color value [7]. Finlayson et al.[8] based on Grey World algorithm, $P$ (Minkowski-norm) is introduced into the Grey-World algorithm, using $P$ distance instead of simple averaging methods RGB color channels, and presents a general algorithm framework of Shades-of-Grey (SoG) algorithm, through changing the $p$ value ($0<p<\infty$), which contains White Patch and Grey World algorithm. Gamut mapping algorithm Forsyth proposed by [9], the assumption that any one image type light source color is limited, color channel RGB value after normalization in the color space is formed on a closed convex hull (convex hull) also known as color (Gamut), schematic diagram as shown in Figure 2 color gamut mapping:

![Color gamut mapping space.](image)

D.H. Brainard [4] proposed a color constancy algorithm based on Bayesian inference, after Rosenberg and Gehler made a series of improvements. Finlayson proposed a color constancy algorithm based on correlation (Color by Correlation) [9], the algorithm is more practical.

Based on machine learning for estimation the colors of multiple illuminants. Study the original method of color constancy machine is based on the neural network model [9], compared with the previous color constancy algorithm, neural network algorithm has superior performance [12], parallel processing speed is higher than the traditional algorithm with adaptive function [10]. According to the data provided by the sample to learn and find out the intrinsic link between the output data and the implicit law gradually come out. Methods Cardei and Funt [11] BP neural network is introduced to estimate the illumination, the BP neural network is a multiple layer feed forward neural network with back propagation algorithm, nonlinear mapping relationship, have good generalization ability. W. Xiong [18] in order to solve the problem of the illumination estimation algorithm based on neural network, a new algorithm of illumination estimation based on support vector regression is proposed. Li [13] introduced Extreme Learning Machine (ELM), and presents the color constancy algorithm is a fast learning, the algorithm is based on Grey-World framework.

We propose an illuminant color estimation method which can explicitly handle multiple illuminants. That method based on Structured Support Vector Machine to solve it. Since our method is physics-based, its effectiveness is independent of the quality of training data, a known weakness of the very successful learning-based techniques. We identify regions of different illuminants by creating local estimates and merging them together to major areas of highly similar illumination. Finally, we end up with a small number of dominant illuminants in the scene.
**Color Constancy**

The computational color constancy problem is generally formulated as: given an image under illumination of unknown color, predict what the image under illumination of unknown color, predict what the image of the same scene would be if taken under a canonical illuminant of known color [3]. The purpose of color constancy computation is to transform an image under unknown illumination to the standard white light image [3], [7]. This process generally can be divided into two major steps: (1) estimating the color of the illumination, and (2) adjusting the image colors based on the difference between the estimated and canonical light sources [4]. The latter step is usually addressed by a scaling of the R, G, and B channels that is often referred to as a von Kries or a diagonal transformation [5].

**Structured Support Vector Machine (SSVM)**

Suppose we are given a training set of input-output structure pairs \( s = \{(x_i, y_i), \ldots, (x_n, y_n)\} \in (\mathbb{Z} \times \mathbb{Y})^n \). We want to learn a linear prediction rule of the form

\[
 f_{\omega}(x) = \arg \max_{y \in \mathbb{Y}} [\omega \cdot \phi(x, y)]
\]  

where \( \phi \) is a joint feature vector that describes the relationship between input \( x \) and structured output \( y \), with \( \omega \) being the parameter vector. The optimization problem of computing this argmax is typically referred to as the “inference” or “prediction” problem [19], [24]. Risk is measured through a user-supplied loss function \( L(y_i, \hat{y}_i) \) that quantifies how much the prediction \( \hat{y}_i \) differs from the correct output \( y_i \). For supervised learning, the input and output pair \((x, y)\) is associated with a certain distribution of \( P(x, y) \), so the purpose of supervised learning is to minimize the expected risk for finding the decision function \( f \):

\[
 R^p(f) = \int_{x \in \mathbb{X}} \int_{y \in \mathbb{Y}} L(y, f(x)) dP(x, y) 
\]  

(2)

Because \( P(x, y) \) is unknown, it is impossible to calculate the expected risk of a decision function for a decision function [20], [23]. We need to find a number of indicators to replace the expected risk, it must be able to calculate, but also reflects the merits of a decision function:

\[
 R^s(f) = \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i)) 
\]  

(3)

Training a structural SVM (SSVM) is framed as the problem of learning a real-valued weight vector \( \omega \) by solving the following optimization problem [21], [22]. When training Structural SVMs, the parameter vector \( \omega \) is determined by minimizing the risk on the training set \( (x_i, y_i), \ldots, (x_n, y_n) \), \( \omega \) weight vector can be obtained by training SSVM

\[
 \forall i, \forall y \in \mathbb{Y} \setminus y_i : \langle w, \delta y_i(y) \rangle > 0 \quad \delta y_i(y) = \Psi(x_i, y_i) - \Psi(x, y) 
\]  

(4)

We use the maximum interval method, and in order to allow a certain error in the training set. Slack variables are introduced,

\[
 \frac{1}{2} \|w\|^2 + \frac{C}{n} \sum_{i=1}^{n} \xi_i, \text{ s.t } \forall i, \xi_i \geq 0, \forall i, \forall y \in \mathbb{Y} \setminus y_i : \langle w, \delta y_i(y) \rangle \geq 1 - \xi_i, 
\]  

(5)

Due to the fact that the actual problems \( |\mathbb{Y}| \) is often very large, the above form is not suitable to solve such problems. So the loss function is introduced in the constraint condition:

\[
 \frac{1}{2} \|w\|^2 + \frac{C}{n} \sum_{i=1}^{n} \xi_i, \text{ s.t } \forall i, \xi_i \geq 0, \forall i, \forall y \in \mathbb{Y} \setminus y_i : \langle w, \delta y_i(y) \rangle \geq \Delta(y_i, y) - \xi_i 
\]  

(6)

when the size of the constraint is very large, the traditional two optimization has been difficult to
solve. Therefore, the optimization problem is rewritten as dual form. In the specific optimization process, only to consider a small part of the constraints set. The following is the formula (5) dual form:

$$
\max_{\alpha} \sum_{i,y} \alpha_{i,y} - \frac{1}{2} \sum_{i,y:j \neq y} \alpha_{i,y} \alpha_{j,y} \left( \delta \psi_i(y), \delta \psi_j(y) \right)
$$

s.t. \( \forall i, \forall y \neq Y \setminus y_i : \alpha_{i,y} \geq 0. \) (7)

During \( \alpha_{i,y} \) is Lagrange multiplier. Here, the Slack variable \( \xi \) is introduced to increase the generalization ability of SSVM compared with the conventional SVM, and it can be found that the SSVM takes into account the data of several sources rather than the isolated consideration [25].

**SSVM for Illumination Chromaticity Estimation**

In this section, we discuss how the SSVM technique can be applied to analyze the relationship between the image of a scene and the chromaticity of the illumination chromaticity incident upon it.

In this paper, we use the method of color constancy to process the real-world mixed-illuminant scenes image.

**Multi Feature Extraction**

According to the Lambertian reflectance model, the image \( f = (R, G, B)^T \) can be computed as follows:

$$
f(X) = \int d(\lambda)a(X,\lambda)g(\lambda)d\lambda,
$$

where \( X \) is the spatial coordinate, \( \lambda \) is the wavelength, and \( \omega \) represents the visible spectrum, \( a(\lambda) \) is the spectral power distribution of the light source, \( g(X,\lambda) \) is the surface reflectance, and \( a(\lambda) = (R(\lambda), G(\lambda), B(\lambda))^T \) is the camera sensitivity function [26]. As shown in figure 3, Weijer et al. [4] proposed an edge-based color constancy algorithm by incorporating higher-order derivatives of an image based on grey-edge hypothesis, which assumes that the average of reflectance differences in a scene is achromatic. By introducing Minkowski-norm and Gauss filter; it becomes

$$
(\int (\frac{\partial^n f}{\partial X^n} \delta(X))^2 dx)^{\frac{1}{2}} = k e^{n,p,\sigma}
$$

where \( f = (R, G, B)^T \) represents the image color, \( X \) is the spatial coordinate again, and \( f^i(x) = f \ast \mathcal{G} \) is the convolution of the image with a Gaussian filter \( \mathcal{G} \) with a standard deviation \( \sigma \). \( \frac{\partial^n}{\partial X^n} \) is an n-order derivative for \( f^i \ast \mathcal{G} \) is the Minkowski-norm. Equation (9) is obviously a general framework that can reformulate many existing algorithms by changing the parameters \( n, p, \delta \) reflect the assumptions of the relationship between illumination and image colors. A. Gijisenij et al. points out that the different image texture feature is suitable to different color constancy algorithms. So the color constancy algorithm can be used as a global image features extraction method. The algorithm candidate set \( M = \{e^{1,0}, e^{0,-1}, e^{0,1}, e^{1,2}, e^{2,1,1} \} \) did well in the study of constancy, and development mature in the research of feature fusion.

---

**Figure 3. Imaging System.**

**Figure 4. (a) Original image, (b) Local illuminant estimation.**
**Von Kries**

Because the surface reflectivity and the color of the light source are unknown, this is an ill-posed problem and cannot be solved without additional constraints or assumptions [7]. Research on color constancy computation the goal is to eliminate the influence of scene illumination. The image of the unknown image \( f_e \) illuminant color value \((R_e, G_e, B_e)\), which will deviate from the actual ambient light source, is multiplied by the diagonal matrix \( D \) to process the image \( f_e \) color value \((R_e, G_e, B_e)\) of the standard white illuminant source [15], [16]. Its expression is shown in (10)

\[
f_a = f_eD = f_e \begin{bmatrix} R_e/R_e & 0 & 0 \\ 0 & G_e/G_e & 0 \\ 0 & 0 & B_e/B_e \end{bmatrix}
\]

Diagonal matrix \( D \) is a transformation coefficient matrix. J. von Kries et al. [14] proposed a von Kries model (also known as diagonal model); respectively, the RGB three channel color values were corrected.

**Multiple Illuminants**

Once local illuminant estimates are obtained per super-pixel, the local information can be combined as follows for the final computation of the number and color of the dominant illuminants in the scene. An example of this process is shown in Figure 4. The details of each of the aforementioned steps are provided in the following subsections. Group local estimates into regions with similar illuminant color. The downscaling suppresses a large amount of relatively small noisy regions. The same color in the scene area is considered to be the same kind of light source. In this relatively large connection area that considered only a light value and obtain a new estimate per illuminant region [17]. The area is used to estimate illuminant and color correction, the purpose is to multi-illuminant translate into a single illuminant problems.

**Experiments**

In this paper, the test is with the SFU subset [30]. The SFU subset contains 15 groups of images taken in different places. Following the scheme of Gijsenij et al.,[31], to ensure that the training and testing subsets are truly distinct, the 1,135 images are partitioned into 15 subsets based on geographical location. One subset is used for testing and the other 14 are used for training. This procedure is repeated 15 times with different test set selection.

**Error Measures**

Several different error measures are used to evaluate performance. The first is the distance between the actual \((r_a, g_a)\) and estimated chromaticity of the illuminant \((r_e, g_e)\):
For the distance error, we also compute the root mean square (RMS), mean, and median errors over a set of N test images. It has been argued that the median is the most appropriate metric for evaluating color constancy [24]. The standard RMS is defined as:

\[ \text{RMS}_{\text{dist}} = \frac{1}{N} \sqrt{\sum_{i=1}^{N} E_{i-\text{dist}}^2} \]  

(12)

The second error measure is the angular error between the chromaticity 3-vectors when the b-chromaticity component is included. Given \( r \) and \( g \), \( b = 1 - r - g \). Thus, we can view the real illumination and estimated illumination as two \( \langle r,g,b \rangle \) vectors in 3D chromaticity space and calculate the angle between them. The angular error represented in degrees is:

\[ E_{i-\text{angular}} = \cos^{-1} \left[ \frac{(r_g g_a, b_a) \cdot (r_e g_e, b_e)}{\sqrt{r_a^2 + g_a^2 + b_a^2} \times \sqrt{r_e^2 + g_e^2 + b_e^2}} \right] \times \frac{2\pi}{360} \]  

(13)

We also compute the RMS, mean, and median angular error over a set of images. Even if the median angular error for one method is less than for another, the difference may not be statistically significant. To evaluate whether a difference is significant, we use the Wilcoxon signed-rank test [24]. In the experiments, the error rate for accepting or rejecting null hypothesis is always set to 0.01.

**Experimental Result Analysis**

SSVM algorithm is used to fuse the multi-features extracted from the image, which to evaluate the image of the scene illuminant. In the SSVM algorithm, RBF kernel function is chosen to deal with the complex nonlinear problem of color constancy. Based on the realization of k-fold cross-validation, we find the optimal penalty parameter c and the parameter sig2 in RBF kernel function. All the training samples are divided into k groups, then k-1 group is selected as the training sample, and k group is used as the test [28]. There are 1,135 images in the dataset, of which 900 are used to build the learning model of the SSVM algorithm, and the remaining 235 are used to test the effect of color constancy. The error analysis is shown in Table 1 and Table 2:

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Mean</th>
<th>Max</th>
<th>Min</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSVM</td>
<td>0.043</td>
<td>0.301</td>
<td>0.003</td>
<td>0.056</td>
</tr>
<tr>
<td>( \varepsilon^{0.1,0} )</td>
<td>0.125</td>
<td>0.367</td>
<td>0.006</td>
<td>0.164</td>
</tr>
<tr>
<td>( \varepsilon^{0.1,2} )</td>
<td>0.049</td>
<td>0.321</td>
<td>0.004</td>
<td>0.076</td>
</tr>
<tr>
<td>( \varepsilon^{1,1,2} )</td>
<td>0.045</td>
<td>0.101</td>
<td>0.026</td>
<td>0.052</td>
</tr>
<tr>
<td>( \varepsilon^{2,1,4} )</td>
<td>0.062</td>
<td>0.126</td>
<td>0.014</td>
<td>0.067</td>
</tr>
</tbody>
</table>

Table 1. Chromaticity error of SSVM and candidate algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Mean</th>
<th>Max</th>
<th>Min</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSVM</td>
<td>3.564</td>
<td>14.437</td>
<td>0.005</td>
<td>4.988</td>
</tr>
<tr>
<td>( \varepsilon^{0.1,0} )</td>
<td>9.487</td>
<td>26.673</td>
<td>0.378</td>
<td>13.232</td>
</tr>
<tr>
<td>( \varepsilon^{0.1,2} )</td>
<td>4.325</td>
<td>12.234</td>
<td>0.043</td>
<td>4.643</td>
</tr>
<tr>
<td>( \varepsilon^{1,1,2} )</td>
<td>5.123</td>
<td>13.983</td>
<td>0.037</td>
<td>6.383</td>
</tr>
<tr>
<td>( \varepsilon^{2,1,4} )</td>
<td>3.903</td>
<td>11.159</td>
<td>0.343</td>
<td>4.057</td>
</tr>
</tbody>
</table>

Table 2. Angle Error of SSVM and Candidate Algorithm.

From Table 1 on the 11 kinds of illuminant conditions, the chromaticity difference analysis shows that the SSVM fusion algorithm to obtain the smallest error, but with the algorithm \( \varepsilon^{1,1,2} \) results
are similar. While the results of algorithm $e^{0,1,2}$ and $e^{2,1,4}$ are similar, and the constancy of the effect is also very good. The effect of the gray world algorithm $e^{0,1,0}$ is the worst, and the error is two times higher than that of the SSVM algorithm. The results show that the fusion algorithm can get closer to the true image of illuminant, and the gray world algorithm for measuring image due to the dark, unable to meet the corresponding assumptions and the constancy of poor effect. The error value of SSVM is $3.564^\circ$ degrees from the angle of Table 2, the effect better than the $e^{1,1,2}$ algorithm. The gray world algorithm results is the worst, the other three algorithms were similar. The constancy effect than fusion algorithm almost. In general, the fusion algorithm of error analysis is SSVM to obtain the color value of the light source which is close to the real scene.

Conclusion

Many previous methods of estimating the aromaticity of the scene illumination have been based in one way or another on statistics of the RGB colors arising in an image, independent of their spatial location or frequency of occurrence in the image [27]. Structured Support Vector Machine is a relatively new tool developed primarily for machine learning that can be applied in a similar way. We have tried it here, with good results, to the problem of learning the association between color histograms and illumination chromaticity. Under almost the same experimentation conditions as those used by Barnard [29], tests of the shades-of-grey, neural-network, color-by-correlation, Max RGB, and Gray-world methods, show that SSVM performance generally is comparable to or better than these other methods.

Acknowledgement

This paper is supported by the following projects: The National Natural Science Foundation of China (No. 61372148, No. 61571045); Beijing Natural Science Foundation (No. 4152016, No. 4152018); Beijing Advanced Innovation Center for Imaging Technology (BAICIT-2016002); The National Key Technology R&D Program (2014BAK08B02, 2015BAH55F03).

References


