Design and Creation of a Multi-illuminant Scene Image Dataset

Imtiaz Masud Ziko, Shida Beigpour, and Jon Yngve Hardeberg

The Norwegian Colour and Visual Computing Laboratory, Gjøvik University College
P. O. Box 191, N-2802 Gjøvik, Norway

Abstract. Most of the computational color constancy approaches are based on the assumption of a uniform illumination in the scene which is not the case in many real world scenarios. A crucial ingredient in developing color constancy algorithms which can handle these scenarios is a dataset of such images with accurate illumination ground truth to be used both for estimating the parameters and for evaluating the performance. Such datasets are rare due to the complexity of the procedure involved in capturing them. To this end, we provide a framework for capturing such dataset and propose our multi-illuminant scene image dataset with pixel-wise accurate ground truth. Our dataset consists of 6 different scenes under 5 illumination conditions provided by two or three distinctly colored illuminants. The scenes are made up of complex colored objects presenting diffuse and specular reflections. We present quantitative evaluation of the accuracy of our proposed ground truth and show that the effect of ambient light is negligible.

Keywords: Color Constancy, Multi-illuminant Dataset.

1 Introduction

It has been proven that there is a relative consistency in the colors of the objects perceived by humans regardless of the illuminant color. This characteristic is also present in some animals (e.g., bees and monkeys). In the case of computational methods and digital image processing this phenomena is a great challenge when the task is to achieve results that are comparable with human perception.

It is not a straightforward task to emulate the observed human color constancy exactly on computational model as the mechanism of this human ability is not yet totally understood [17,15]. The aim of the computational color constancy is to correct the effect of illuminant color to extract the actual object color value in the scene as it would appear under the canonical light [15]. Many of the theories try to model color constancy by computing invariant color descriptors for object recognition and surveillance application for e.g. in [11,13]. Other approaches try to estimate the illuminant chromaticity and transfer the image to the canonical illuminant such as in white balance applications [8,9,14,4,19].

There are several datasets available for illuminant estimation approaches. Bernard *et al.* [1] proposed a dataset of 30 constructed scenes with 11 different illuminants with known spectral distributions. For each illuminant per scene

A. Elmoataz et al. (Eds.): ICISP 2014, LNCS 8509, pp. 531–538, 2014.

they captured images at first placing a white reference perpendicularly to the illumination direction in order to adjust the exposure and calculate the ground truth. Then they moved the white reference and capture 50 successive images of the scene and finally the average image is used as the final image. However, the images do not represent the full variations of typical scenes as outdoor scenes are missing in that dataset. Ciurea and Funt [5] proposed another dataset extracting 11,000 frame images with various indoor and outdoor scenes from video clips. They used a matte gray ball connected to camera to be visible in the image which assumes to contain the illumination chromaticity of the scene. Images are quite correlated in that dataset as they were taken from the video clips. Further, a video based analysis was applied to extract the less correlated subset of 1,135 images [3]. Another dataset was introduced with 568 high quality images having indoor, outdoor scene and portrait images [12]. The argument to measure illuminant information preferably in standard co-ordinate systems such as CIE tristimulus values instead of camera RGB response due to the unavailability of complete camera calibration information from the manufacturer [5] lead to a dataset providing camera calibration for accurate mapping to cone activation space of human [18]. A dataset with HDR images are also proposed [10] but the ground truth of illuminant may not comply to the original illumination in the scene due to the non-linearity of **makehdr** function in matlab which actually implements the widely known algorithm by Debevec *et al.* [6] for high dynamic range image recovery from different exposured photographs.

In common, all of the above datasets ignore the realistic fact of having the presence of multiple light sources in scenes and assumed only uniform illumination over the whole scene. Thus, the requirement of the local illuminant information in the ground truth map lack on those datasets. To address these, Gijsenij *et al.* [16] proposed their local illuminant estimation method with a dataset of 59 laboratory images using two halogen lamps with four different color filters and 9 natural outdoor scene images. In their dataset, the illumination map creation for ground truth was done by manual annotation and segmentation which may be error prone. Beigpour *et al.* [2] addressed this problem and constructed a dataset with 58 laboratory images and 20 additional outdoor images. They have introduced a novel pixel wise ground truth calculation exploiting the linearity of light in camera sensor data.

We present a dataset with reliable ground truth calculation as well as adding combination of three light sources apart from having only two light sources in complex cluttered scene. We use the indoor light sources combined with outdoor light coming from the room window. For the ground truth calculation we follow the main idea proposed by Beigpour *et al.* [2]. In addition, we investigate the reliability of the ground truth by calculating the mean squared error between the model and the original image values and show that the error is negligible. In the following sections, we describe the design and setup of creating the dataset and the detailed procedure for calculating the ground truth images.

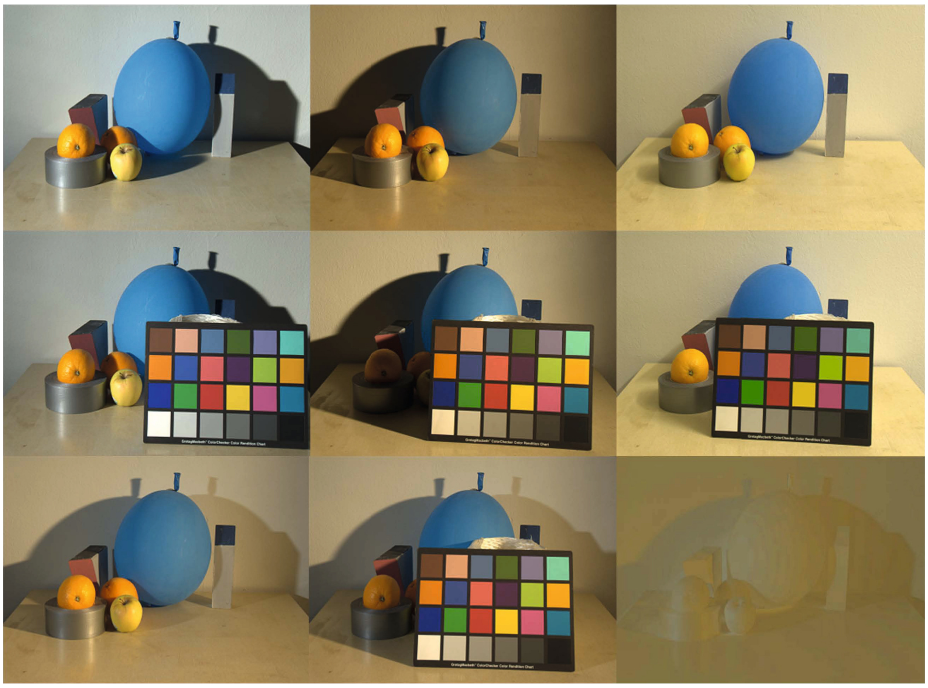

Fig. 1. Images taken for ground truth calculation for Scene 2 under L_5. Top row (left to right): scene under L_{BF}, L_{OF} and L_{RM}. Middle row (left to right): corresponding top row scenes only by placing a Macbeth color checker. Bottom row (left to right): multi-illuminant scene under L_5, same scene with Macbeth color checker and the ground truth of the multi-illuminant scene image.

2 The Setup and Image Acquisition

We used Nikon D200 camera which store image data in 12 bit RAW format and sensor image resolution is 3872×2592. Four different light sources are used which are: Two Photon Beared M200 (200 W) tungsten studio lamps with plastic blue (L_{BF}) and pale orange filter (L_{OF}), fluorescent light in the room from ceiling (L_{RM}) and outdoor light from the only window in the room (L_{OD}).

The room is made to be fully dark by covering and blocking all the light coming from the window. In each illumination condition the images are considered distinct for each scene i.e. we change the position of light sources for different illumination conditions and for different scenes. We assume the ambient light (cause by the light bouncing off the object surfaces in the scene) and inter-reflections are negligible (in Section 3 we discuss this in more detail). It is also assumed that the light coming from the window does not vary during the capturing of each scene. Five distinct illumination conditions are created from the four light sources:

1. L_1: L_{BF} from right and L_{OD} from left.

2. L_2: L_{OF} from right and L_{OD} from left.
3. L_3: L_{BF} from left and L_{RM} from top right of ceiling.
4. L_4: L_{BF} from left and L_{OF} from right.
5. L_5: L_{BF} from left, L_{RM} from top right of ceiling and L_{OF} from right.

For image acquisition, we set the camera on tripod with camera settings: white balance set to zero in auto mode, ISO 100, Aperture: f/8, and saved image in uncompressed *raw* format. This *raw* setting also ensures that the camera will not use a white balance setting and export the image exactly as it is recorded on the image sensor without any processing. Every scene is created by using the combination of specular,diffuse and colored objects. Then we proceed to capture the image same way as [2] under all the light sources together and then captured image under each single light source in each illumination condition per scene. All these images were captured once placing the Macbeth color checker on the scene and then without it. These additional images are needed for the ground truth calculation which is discussed in the following sections. One example of captured images along with the calculated ground truth for Scene 2 under L_5 are shown in Fig.1. All the images in this article are converted from raw to sRGB and enhanced for better visualization. To deal with saturation, we used three different setting of exposure (-0.7, 0 , +0.7) by auto-bracketing for each capture and choose the one which ensures not to have more than a very small number of saturated pixels.

In total, 6 images are captured for illumination conditions with two light sources and 8 images are captured for illumination condition with three light sources. We have 6 multi illuminant scene with 5 different illumination conditions

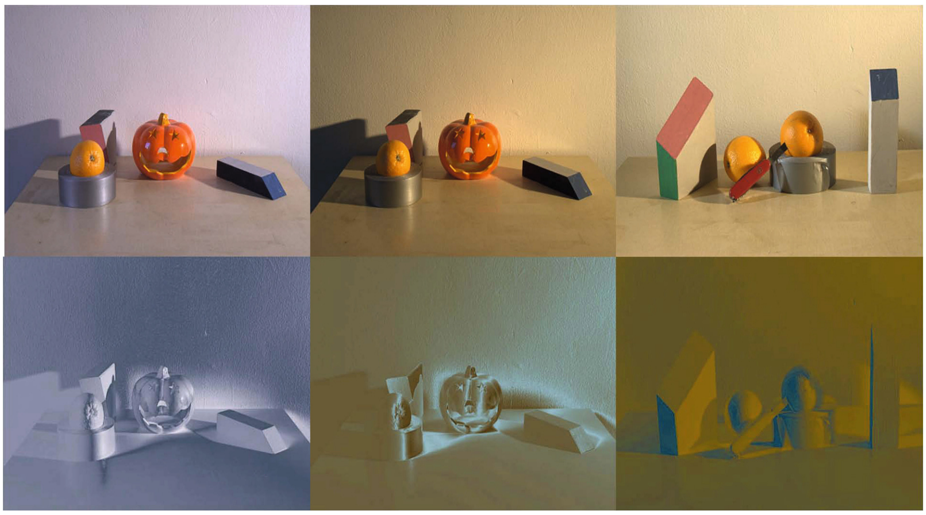

Fig. 2. Example of dataset images. Top row (left to right): Multi illuminant images for Scene 1 under L_1, Scene 1 under L_2 and Scene 5 under L_1. Bottom row (left to right): Corresponding ground truth for the top row multi-illuminant images.

which in total give 30 multi illuminant images. Due to some misalignment of the same scene images, we remove images for scene 3 in L_5 and for scene 5 in L_3. Finally in the dataset, we have 28 different multi-illuminant images for 6 different scenes.

3 Ground Truth Calculation

As ground truth we provide the RGB image with the known illuminant chromaticity in each pixel of the multi-illuminant images for each illumination condition. The main underlying idea is to exploit the linearity of light in sensor data i.e. scene image taken under multiple illuminant is equal to the summation of the raw images of same scene taken under each single illuminant in the multi-illuminant condition. For an example, if f_a is the image pixel under illuminant a and f_b is the image pixel under illuminant b then f_{ab}, the image pixel under the combination of illuminant a and b for the same scene can be found as (1).

$$f_{ab} = f_a + f_b \tag{1}$$

Now we need the chromaticity information of each illuminant in the scene. We extract that information from those additional images taken with Macbeth Color Checker as described in the previous section 2. We get three pixels sparsely from the middle gray patch of the color checker and use the average pixel (of each channel in RGB) as the illuminant chromaticity. Under Von Kries assumption [7], dividing each single illuminant image with its respective illuminant chromaticity, we get scene under white illumination as shown in (2) for illuminant a.

$$\begin{aligned} \hat{f}_{a,C} &= \frac{f_{a,C}}{I_{a,C}} \\ \text{where} \quad C &\in \{R, G, B\} \\ \text{and} \quad I &= \text{illuminant chromaticity} \end{aligned} \tag{2}$$

The per-pixel relative contribution of illuminant a in the multi-illuminant image pixel f_{ab} comprised of both illuminant a and b considering the green channel is calculated as per (3). Although any color channel from Red, Green and Blue would work, we considered working with green channel. The derivation of (3) can be found from [2]. Then the ground truth image, I_{ab} with contribution of individual illuminant chromaticity of I_a and I_b in the multi-illuminant image is found as the pixel wise linear interpolation of the both illuminant chromaticity as shown in (4). We provide the ground truth images color coded with the found color of individual illuminant in each pixel as per their corresponding weights. In case of illumination condition with three light sources such as a, b and c, we calculate the contribution of c quite the similar way. If s and r be the contribution of illuminant a and b then we calculate the ground truth I_{abc} using (5) and (6).

$$r = \frac{\hat{f}_{a,G}}{\hat{f}_{a,G} + \hat{f}_{b,G}} \tag{3}$$

$$I_{ab} = r.I_a + (1 - r).I_b \tag{4}$$

$$\begin{aligned} r &= \frac{\hat{f}_{a,G}}{\hat{f}_{a,G} + \hat{f}_{b,G} + \hat{f}_{c,G}} \\ s &= \frac{\hat{f}_{b,G}}{\hat{f}_{a,G} + \hat{f}_{b,G} + \hat{f}_{c,G}} \end{aligned} \tag{5}$$

$$I_{abc} = r.I_a + s.I_b + (1 - r - s).I_c \tag{6}$$

For the illumination conditions L_1 and L_2 having outdoor light mixed with other indoor light source , only the multi-illuminant image (e.g. image under L_1) and outdoor image (image under L_{OD}) are used to get the other image (image under L_{BF}) by simply taking the difference in between (i.e. image under $L_{BF} \approx$ image under L_1 $-$ image under L_{OD}). This was done assuming the possibility of having some outdoor light mixed with indoor individual light.

However, we could not directly work with Nikon Electronic Format (NEF) images in Matlab, we had to do some post processing to extract the raw linear RGB image data from Nikon format (NEF). For the post processing we at first convert images from Nikon format (NEF) to Adobe Digital Negative (DNG) format as uncompressed raw. For this task we used free version of Adobe DNG converter software. Then we can read the DNG format images in Tiff class of Matlab and could get the original EXIF metadata of raw images. Thereafter, we get the valid meaningful image area from the raw image using the metadata information. It is possible that the camera applied a non-linear transformation

Table 1. Calculated *MSE* for ground truth images

Illumination Conditions	*MSE*					
	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5	Scene 6
L_1	0.005	0.008	0.006	0.008	0.002	0.006
L_2	0.014	0.008	0.010	0.014	0.006	0.011
L_3	0.010	0.005	0.003	0.010	-	0.013
L_4	0.004	0.005	0.009	0.006	0.011	0.010
L_5	0.022	0.043	-	0.032	0.037	0.037

Table 2. Calculated *MSE_GT* for ground truth images in Scene 1, 2 and 3

Illumination Conditions	*MSE_GT*		
	Scene 1	Scene 2	Scene 3
L_1	0.048	0.000	0.004
L_2	0.054	0.007	0.000
L_3	0.000	0.000	0.000
L_4	0.022	0.000	0.002
L_5	0.001	0.000	-

to the sensor data for storage purposes. So we find linear data of raw images by mapping to the linearization table stored in image metadata. In the step, we take the *maximum* and *minimum* pixel value of the raw image and divide each *minimum* subtracted pixel by the difference between the *maximum* and the *minimum* value. In this way, the values are normalized in range of 0 to 1. Then we get the normal white balance data at the shot time from EXIF metadata and multiply to each CFA channel (For Nikon D200 the CFA pattern is RGGB). Finally, we apply default demosaicing algorithm, **demosaic** function implemented in Matlab to get linear RGB image and work with this linear RGB image for finding ground truth image.

4 Validation of Ground Truth Calculation

To measure the reliability of calculated ground truth images we calculate the Mean Squared Error (*MSE*) from the difference between the original multi-illuminant image and the summation of individual illuminant images. This value indicates the effect of ambient light in the calculation of ground truth. Again, we calculate another Mean Squared Error (denoted as MSE_GT) for the first three scenes taking the difference between the chromaticity of gray patch in Macbeth color checker in multi-illuminant image and illuminant chromaticity calculated in the ground truth image in the same pixel positions to check how close the ground truth is to the chromaticity found in the original multi-illuminant image. The calculated errors, *MSE* for all the scenes and MSE_GT for the first three scene images are shown respectively in Table 1 and in Table 2. As can be seen from the very small errors in Table 1 and Table 2, the effect of ambient light and inter-reflections is negligible and the calculated ground truth closely represents the chromaticity of the mixed illuminants in each pixel.

5 Conclusion

We created a novel dataset with images taken on scenes under non uniform illumination having multiple light sources (e.g. two or three) including outdoor light mixed with indoor light sources, creating complex scene with cluttered objects, overlapped shadow regions and with combination of specular, diffuse and differently colored object materials. We provide reliable ground truth images to have better evaluation of multi-illuminant estimation methods in computational color constancy. The provided precise ground truth images can be used efficiently to compare the color constancy algorithms and can also be utilized to have corrected white balanced images of the dataset from corresponding ground truth images. This CID:MI dataset (Colorlab Image Dataset: Multi-illuminant) can be downloaded from `http://www.colourlab.no/cid`. Our future work will be to extend the dataset by incorporating the HDR images for multi-illuminant context and creating the reliable ground truth associated with the HDR scene images, adding different light sources with the existing ones and making more lighting conditions from these added light sources along with the spectral power distribution (SPD) of the light sources to have more sharp idea about the nature of the light sources.

References

1. Barnard, K., Martin, L., Coath, A., Funt, B.: A comparison of computational color constancy algorithms. ii. experiments with image data. IEEE Transactions on Image Processing 11(9), 985–996 (2002)
2. Beigpour, S., Riess, C., van de Weijer, J., Angelopoulou, E.: Multi-illuminant estimation with conditional random fields. IEEE Transactions on Image Processing 23(1), 83–95 (2014)
3. Bianco, S., Ciocca, G., Cusano, C., Schettini, R.: Improving color constancy using indoor–outdoor image classification. IEEE Transactions on Image Processin 17(12), 2381–2392 (2008)
4. Brainard, D.H., Freeman, W.T.: Bayesian color constancy. JOSA A 14(7), 1393–1411 (1997)
5. Ciurea, F., Funt, B.: A large image database for color constancy research. In: Color and Imaging Conference, pp. 160–164. Society for Imaging Science and Technology (2003)
6. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: ACM SIGGRAPH 2008 Classes, p. 31. ACM (2008)
7. Fairchild, M.D.: Color appearance models. John Wiley & Sons (2013)
8. Finlayson, G.D., Hubel, P.M., Hordley, S.: Color by correlation. In: Color and Imaging Conference, pp. 6–11. Society for Imaging Science and Technology (1997)
9. Forsyth, D.A.: A novel algorithm for color constancy. International Journal of Computer Vision 5(1), 5–35 (1990)
10. Funt, B., Shi, L.: The rehabilitation of maxrgb. In: Color and Imaging Conference, vol. 2010, pp. 256–259. Society for Imaging Science and Technology (2010)
11. Funt, B.V., Finlayson, G.D.: Color constant color indexing. IEEE Transactions on Pattern Analysis and Machine Intelligenc 17(5), 522–529 (1995)
12. Gehler, P.V., Rother, C., Blake, A., Minka, T., Sharp, T.: Bayesian color constancy revisited. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8. IEEE (2008)
13. Gevers, T., Smeulders, A.W.: Color-based object recognition. Pattern Recognition 32(3), 453–464 (1999)
14. Gijsenij, A., Gevers, T.: Color constancy using natural image statistics. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8. IEEE (2007)
15. Gijsenij, A., Gevers, T., Van De Weijer, J.: Computational color constancy: Survey and experiments. IEEE Transactions on Image Processing 20(9), 2475–2489 (2011)
16. Gijsenij, A., Lu, R., Gevers, T.: Color constancy for multiple light sources. IEEE Transactions on Image Processing 21(2), 697–707 (2012)
17. Kraft, J.M., Brainard, D.H.: Mechanisms of color constancy under nearly natural viewing. Proceedings of the National Academy of Sciences 96(1), 307–312 (1999)
18. Parraga, C., Baldrich, R., Vanrell, M.: Accurate mapping of natural scenes radiance to cone activation space: a new image dataset. In: Conference on Colour in Graphics, Imaging, and Vision, pp. 50–57. Society for Imaging Science and Technology (2010)
19. Van De Weijer, J., Gevers, T.: Color constancy based on the grey-edge hypothesis. In: IEEE International Conference on Image Processing, ICIP 2005, vol. 2, pp. II–722. IEEE (2005)