Large Scale Multi-Illuminant (LSMI) Dataset for Developing White Balance Algorithm under Mixed Illumination

Dongyoung Kim[1], Jinwoo Kim[1], Seonghyeon Nam[2], Dongwoo Lee[1], Yeonkyung Lee[3], Nahyup Kang[3], Hyong-Euk Lee[3], ByungIn Yoo[3], Jae-Joon Han[3], Seon Joo Kim[1]

[1]Yonsei University, [2]York University, [3]Samsung Advanced Institute of Technology

Abstract

We introduce a Large Scale Multi-Illuminant (LSMI) Dataset that contains 7,486 images, captured with three different cameras on more than 2,700 scenes with two or three illuminants. For each image in the dataset, the new dataset provides not only the pixel-wise ground truth illumination but also the chromaticity of each illuminant in the scene and the mixture ratio of illuminants per pixel. Images in our dataset are mostly captured with illuminants existing in the scene, and the ground truth illumination is computed by taking the difference between the images with different illumination combination. Therefore, our dataset captures natural composition in the real-world setting with wide field-of-view, providing more extensive dataset compared to existing datasets for multi-illumination white balance. As conventional single illuminant white balance algorithms cannot be directly applied, we also apply per-pixel DNN-based white balance algorithm and show its effectiveness against using patch-wise white balancing. We validate the benefits of our dataset through extensive analysis including a user-study, and expect the dataset to make meaningful contribution for future work in white balancing.

Figure 1. Two and three-illuminant scene samples from LSMI dataset (left) and illuminant coefficient maps that show the ratio of how illuminants are mixed per pixel (right). Raw images are converted to the sRGB space with an auto white balance for visualization purposes.

1. Introduction

White balance (WB) is a key feature in cameras that estimates the color of the illumination in the scene, in order to remove the color cast by the illumination. WB imitates the color constancy in human visual system, and it is one of the core components of the in-camera imaging pipeline for developing visually pleasing photographs.

White balancing, or computational color constancy, is a long-standing problem in computer vision and most prior works have focused on scenes with single illumination [28, 9, 34]. As with other areas in computer vision, recent WB algorithms have been developed in a data-driven manner [33, 5, 3] that requires large WB datasets. There are many good datasets for developing learning based WB algorithms for single illumination scenes [12, 31, 17, 11, 22].

However, many real-world scenes contain multiple illuminants. A typical example is taking a photograph of an indoor scene with windows. There are lights originating from indoor light sources as well as the sunlight coming in from windows. Compared to the single illumination problem, only a few studies [21, 4, 6, 32] have addressed the multi-

illumination WB problem, mainly due to the difficulty of collecting datasets. Unlike the single-illumination task, illumination is spatially varying due to multiple light sources, making the acquisition of the ground-truth very challenging. In existing multi-illumination datasets, a limited number of images were collected either by synthesis [29] or by controlling the capturing environment [8, 4, 21]. Despite the tremendous effort by these works, there is still a need for larger and more realistic dataset for multi-illumination WB to support future work in developing learning based WB algorithms under multiple illumination.

In this paper, we introduce a new Large Scale Multi-Illuminant dataset (LSMI) for multi-illumination white balancing. The dataset contains a total of 7,486 images, 2,762 multi-illumination scenes taken with three different cameras (Samsung Galaxy Note 20 Ultra, Sony α9, Nikon D810). As shown in Fig. 1, our dataset provides images of a variety of realistic scenes with multiple illuminants with the pixel-level ground truth illumination maps. In addition, we also provide the chromaticity of all illuminants of the scene as well as the ratio of how the illuminations are mixed per pixel (mixture ratio). Using the mixture ratio, we can synthetically generate more data with arbitrary illumination, which helps to easily augment our dataset for training deep convolutional neural networks (CNNs). All images and data, including ground truth illumination maps and mixture ratios will be made available to the public[1].

Our new dataset can serve as a catalyst for encouraging more research on multi-illumination white balancing. In the single illumination case, the output of the WB algorithm is simply the color of the illuminant. For the multi-illumination case, more sophisticated algorithm is necessary as it has to output the illumination color per pixel. Instead of applying a patch-based algorithm as previously done, we formulate the multi-illumination WB problem as an image-to-image problem in which the input image is transformed to an white balanced image after passing through a deep CNN. We show the effectiveness of this framework through extensive experiments including a user study.

2. Related Work

2.1. White Balance Algorithms

Most computational color constancy or white balance algorithms assume a uniform illumination, and can be divided into two major categories: statistic-based and learning-based. Statistic-based methods make their own assumptions about the characteristics of the light [9, 14, 15, 19, 20, 28, 34]. Learning-based methods are trained on a given dataset. [2, 25] regard the color constancy as a discriminative task, and train a model to classify white-balanced images and non-white-balanced images. [16, 33, 5] use various neural networks, especially CNNs to directly predict the illuminations of the scenes in images.

Recently, studies on more complex multi-illumination have been conducted. [23, 21, 7] proposed white balance algorithms under mixed illumination, but their methods require some prior knowledge such as the chromaticity and the number of illuminants, or faces in the image. The work of [26] used flash photography to perform white balancing under mixed illumination, but the performance is limited for the objects in distance that the flash cannot reach. [4] formulated the white balance problem as an energy minimization task within a conditional random field over a set of local illuminant estimates. [6] proposed patch-based illumination inference CNN model. In [32], generative adversarial networks (GANs) based approach was proposed to correct images using a model trained on synthetic data without illumination estimation.

	Scenes	Illuminants per image	Number of cameras	Light sources
[21]	4	1-2	1	Reuter lamp
[8]	68	2	1	Natural light, Indoor light
[4]	30	2	1	Natural light, Indoor light
[29]	1,015	1	1	Flash light
Ours	**2,762**	**1 - 3**	**3**	**Natural light, Indoor light**

Table 1. Comparison of multi-illuminant datasets.

2.2. White Balance Datasets

Single illumination datasets. In the single light source setting, the chromaticity of the light is easily acquired by computing the color cast of a gray patch in a scene. With this approach, many single image illumination datasets have been introduced. In [12], a collection of 11,000 images of scenes with a gray ball captured with a video camera was introduced. 568 images with Macbeth color chart was released in [17], which was later reprocessed and released as Gehler-Shi [31] and ColorChecker RECommended [22] datasets. In [11], NUS-8 dataset was introduced by capturing a total of 1,736 images using 8 cameras, and the authors also provided a benchmark for existing color constancy algorithms. The cube, cube+ [1], and cube++ [13] datasets were also released recently, which include various types of scenes with spydercube and ground truth illumination in various directions.

Multiple illumination datasets. Few works have addressed the problem of collecting datasets for multiple illu-

[1] https://github.com/DY112/LSMI-dataset

Multi Illuminant [21] | Multiple Light Sources [8] | MIMO [4] | Multi Illumination in the Wild [29] | LSMI (Ours)

Figure 2. Image samples from various multi-illuminant datasets. Compared to other datasets, our dataset includes more diverse and realistic set of images.

mination. A dataset composed of 4 laboratory scenes with 17 illumination conditions using Reuter lamps and LEE filters were introduced in [21]. Multiple Light Source dataset, consisting of 59 laboratory and 9 outdoor images under the multi-illuminant setting was proposed in [8]. MIMO dataset [4] released 80 images captured under 10 laboratory settings and 6 illuminations, as well as 20 real world images. Most of the images in the above datasets were taken in a laboratory where all light sources were completely controlled. While such environment enables accurate recovery of the ground truth illumination, it is limited in capturing realistic and natural scenes. Moreover, the number of scenes in these datasets is insufficient for training a learning-based white balancing algorithm as shown in Table 1.

Millan Portrait Dataset composed of 1,145 images with human faces in natural scenes was introduced in [7], however, the dataset is not publicly available to the best of our knowledge. A dataset of multi-illumination images in the wild has been introduced recently in [29], which includes images of 1,015 scenes captured by moving the direction of a flash unit. Since they captured 25 single flash illuminated images, multi-illuminant images are synthesized by combining multiple images with different light directions after the images are relighted using different chromaticity. Three applications including single-image illumination estimation, image relighting, and mixed illumination white balance were demonstrated on this dataset.

We propose a new large scale multi-illuminant dataset, to solve the white balancing problem under multiple illumination. Fig. 2 and Table 1 compares images and characteristics of different datasets, respectively. Our dataset is much larger in scale and more natural compared to other datasets. While the dataset in [29] is also a large scale dataset, the scene is limited to a small area as shown in Fig. 2 as the dataset was not designed only for the white balancing. In addition, multi-illuminant images in [29] are synthesized by mixing multiple images with different illuminations. In comparison, the images in our dataset cover various ranges, from small to large areas, and look natural due to the design of the data acquisition. Many scenes that are likely to be captured by a consumer camera are collected in a real illumination setting. Moreover, each multi-illuminant image in our dataset is captured with one camera shot, without requiring additional synthesis.

3. LSMI Dataset

3.1. Image Model

We use the following imaging model for designing our dataset,

$$\mathbf{I}(x) = \mathbf{r}(x) \odot \eta(x)\boldsymbol{\ell}, \tag{1}$$

where $\mathbf{I}$ is an RGB image, $\mathbf{r}$ represents the surface reflectance in RGB, η is the scaling term that includes the intensity of the illumination and shading, $\boldsymbol{\ell}$ denotes an RGB illuminant chromaticity vector, and x is the pixel location. In addition, $\odot$ represents element-wise multiplication. We assume that the value of the green channel in $\boldsymbol{\ell}$ is normalized to 1.

If there are two illuminants a, b in the scene, Eq. (1) can

Figure 3. Example illustration of capturing environment for a three-illuminant scene.

be extended as follows:

$$\mathbf{I}_{ab}(x) = \mathbf{r}(x) \odot (\eta_a(x)\boldsymbol{\ell}_a + \eta_b(x)\boldsymbol{\ell}_b). \tag{2}$$

The white balancing problem can be interpreted as making all illuminants to have a canonical chromaticity, e.g. white illuminant $\mathbf{1}$. Let $\hat{\mathbf{I}}$ denote a properly white balanced image, which can be described as follows:

$$\hat{\mathbf{I}}_{ab}(x) = \mathbf{r}(x) \odot (\eta_a(x)\mathbf{1} + \eta_b(x)\mathbf{1}). \tag{3}$$

The relationship between $\mathbf{I}_{ab}$ and $\hat{\mathbf{I}}_{ab}$ under two-illuminant scene is then as follows (pixel location x omitted):

$$\begin{aligned}\mathbf{I}_{ab} &= \mathbf{r} \odot (\eta_a\boldsymbol{\ell}_a + \eta_b\boldsymbol{\ell}_b) \\ &= \mathbf{r} \odot (\eta_a\mathbf{1} + \eta_b\mathbf{1}) \odot (\frac{\eta_a\boldsymbol{\ell}_a}{\eta_a + \eta_b} + \frac{\eta_b\boldsymbol{\ell}_b}{\eta_a + \eta_b}) \\ &= \hat{\mathbf{I}}_{ab} \odot (\alpha\boldsymbol{\ell}_a + (1-\alpha)\boldsymbol{\ell}_b) \\ &= \hat{\mathbf{I}}_{ab} \odot \boldsymbol{\ell}_{ab}, \quad \text{where} \quad \alpha = \frac{\eta_a}{\eta_a + \eta_b}.\end{aligned} \tag{4}$$

The above equation shows that the pixel-level illumination under multi-illuminant scene can be formulated as the weighted combination of two illuminant chromaticity vectors $\boldsymbol{\ell}_a$ and $\boldsymbol{\ell}_b$ using $\alpha = \frac{\eta_a}{\eta_a+\eta_b}$ as the weight. Since both η_a and η_b are varying according to pixels, α also has different values spatially.

3.2. Dataset Acquisition

We use three types of cameras with different maximum sensor values (10bit and 14bit) to cover a wide range of raw sensor data – Samsung Galaxy Note 20 Ultra, Sony α9 (ILCE-9) with SEL24105G lens, and Nikon D810 with Nikon24-70vr lens. As shown in Fig. 3, the scenes are captured under the natural configuration of multiple light sources including both the sunlight and artificial lamps. For the artificial lights, we use the indoor lighting installed in the scene, and a portable lamp. To acquire the ground truth illumination map, we capture multiple images of the same scene under different combination of the lights. Specifically, we take images by turning the controllable lights on-and-off according to the combination described in Table 2. Details on the number of scenes and the number of images for different cameras are shown in Table 3. For the scene diversity, we captured various real-world places such as offices, studios, living rooms, bedrooms, restaurants, cafes, etc. For each scene, 3 Macbeth color charts were arranged in places that are well affected by each light source in the scene. The charts are used to extract the chromaticity of each light source, which is described in the following subsection. All multiple images of the same scene are taken under the same camera settings. We also made an effort to firmly fix the camera with a tripod and used remote capturing to maintain the pixel correspondence between images.

	two-illuminant		three-illuminant		
	light 1	light 2	light 1	light 2	light 3
shot 1	on	off	on	off	off
shot 2	on	on	on	on	off
shot 3			on	off	on
shot 4			on	on	on

Table 2. Light source on&off compositions of two and three-illuminant scenes.

	2-illum Scenes	3-illum Scenes	Total Scenes	Total Images
Samsung Galaxy Note 20 Ultra	1,000	125	1,125	2,500
Nikon D810	916	39	955	1,988
Sony α9	1,135	182	1,317	2,998

Table 3. Dataset subset compositions captured with different cameras. Since there is scene overlap between camera subsets, the total number of unique scenes is 2,762. There is a total of 7,486 images in our dataset.

3.3. Ground Truth Labelling

Since the combination of multiple illuminants is linear in the unprocessed RAW space, we can decompose each illuminant from the scene using our captured images. For simplicity, we describe our method for calculating per-pixel illuminants and their mixture coefficients under the two-illuminant setting.

Fig. 4 depicts the overview of the method. Let $\mathbf{I}_a$ and $\mathbf{I}_{ab}$ denote an image captured under the illuminant a and an image captured under both illuminants a and b, respectively. According to Eq. (2), we can get an image under the illuminant b by subtracting $\mathbf{I}_a$ from $\mathbf{I}_{ab}$, which is formally

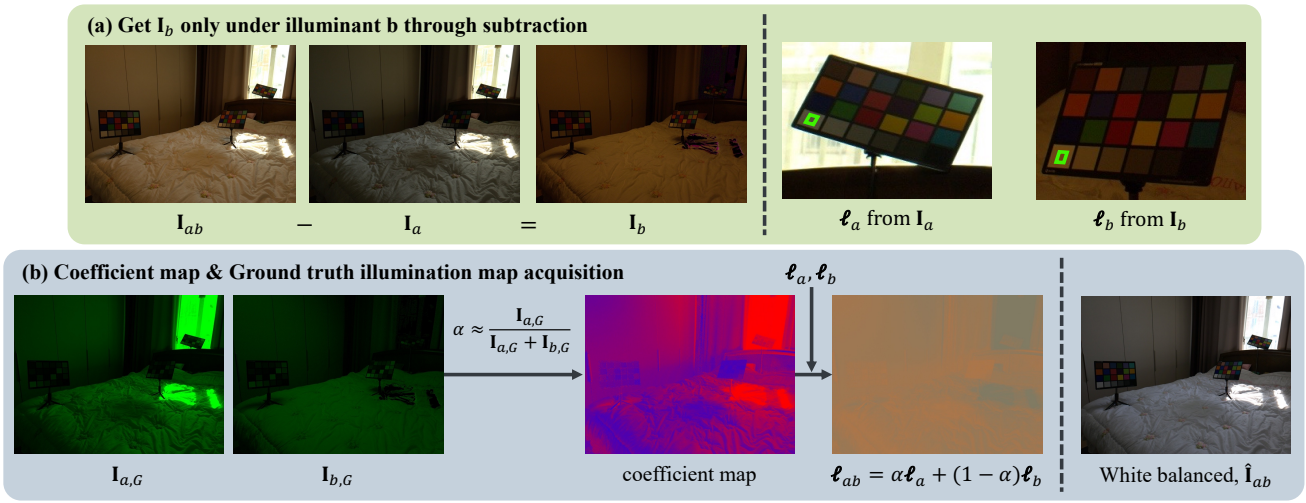


Figure 4. Visualization of the generation process of illuminant coefficient map and ground truth illumination map. (a) We get image $\mathbf{I}_b$ only under illuminant b, through image pair subtraction. Next, inspect the Macbeth color chart in each image $\mathbf{I}_a$ and $\mathbf{I}_b$, to get the chromaticity of each illuminant. (b) Here, we derive the coefficient map through approximation of scaling term η_a and η_b, utilizing the green channel value of each image, $\mathbf{I}_{a,G}$ and $\mathbf{I}_{b,G}$. Now we can calculate the ground truth illumination map $\boldsymbol{\ell}_{ab}$, through linear combination of $\boldsymbol{\ell}_a$ and $\boldsymbol{\ell}_b$. By using $\boldsymbol{\ell}_{ab}$, properly white balance image, $\hat{\mathbf{I}}_{ab}$ is obtained. Daylight white blance is applied to $\mathbf{I}_a$, $\mathbf{I}_b$, $\mathbf{I}_{ab}$, and $\boldsymbol{\ell}_{ab}$, to increase visibility.

described as

$$\begin{aligned}\mathbf{I}_b &= \mathbf{r} \odot (\eta_b \boldsymbol{\ell}_b) \\ &= \mathbf{r} \odot (\eta_a \boldsymbol{\ell}_a + \eta_b \boldsymbol{\ell}_b) - \mathbf{r} \odot (\eta_a \boldsymbol{\ell}_a) \\ &= \mathbf{I}_{ab} - \mathbf{I}_a.\end{aligned} \tag{5}$$

We find a chromaticity of each illuminant, $\boldsymbol{\ell}_a$ and $\boldsymbol{\ell}_b$, using the color chart in $\mathbf{I}_a$ and $\mathbf{I}_b$, respectively. We average the pixels of the brightest achromatic patch among the charts without saturation in an image to compute the chromaticity.

To compute the illuminant coefficient α in Eq. (4), we use the pixel intensity of $\mathbf{I}_a$ and $\mathbf{I}_b$ as an approximation of η_a and η_b respectively, since it is proportional to η according to Eq. (1). Among RGB channels, we use the green channel to compute the coefficient because the sensitivity of the Bayer pattern sensors of digital cameras is highest in the green channel and the intensity of the green channel is typically normalized to 1 in white balancing. Our approximated coefficient is formally described as

$$\begin{aligned}\alpha(x) &= \frac{\eta_a(x)}{\eta_a(x) + \eta_b(x)} \\ &\approx \frac{\mathbf{I}_{a,G}(x)}{\mathbf{I}_{a,G}(x) + \mathbf{I}_{b,G}(x)},\end{aligned} \tag{6}$$

where $\mathbf{I}_{a,G}(x)$ and $\mathbf{I}_{b,G}(x)$ are the intensity of the green channel. With this procedure, pixel-level α, $\boldsymbol{\ell}_a$, and $\boldsymbol{\ell}_b$ are obtained. The per-pixel ground-truth illumination map $\boldsymbol{\ell}_{ab}$ is computed using these variables following Eq. (4).

Figure 5. Pixel-level relighting example of two-illuminant scene from LSMI.

3.4. Pixel-level Relighting

Even though LSMI contains many images with various lighting settings, the diversity is still limited compared to the entire space of real-world lighting conditions. Our dataset has the flexibility to augment the diversity of lighting by adjusting the chromaticity of illuminants. Since the chromaticity and the mixture of each illuminant are decomposed, we can freely manipulate the color of the lighting while maintaining the scene geometry as shown in Fig. 5.

To perform pixel-level relighting on our LSMI data, we sample HSV color vectors within the range H[0,1], S[0.2,0.8], and V=1. The sampled color vector is converted to RGB space and normalized so that $G = 1$. And they are linearly combined using the original pixel-wise illuminant coefficient α, and then multiplied to original raw image. We provide more details about the pixel-level relighting process in the supplementary material.

4. Pixel-level White Balancing Models

The large amount of scenes provided by our dataset and its pixel-level ground truth allow for the training of

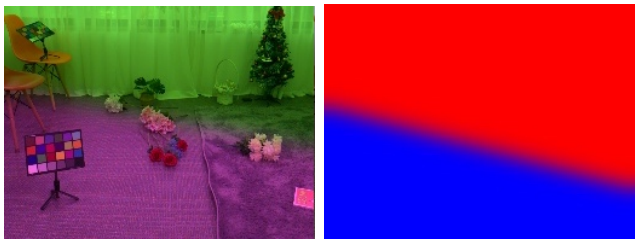
pixel-level white balancing models, which is different from previous patch-based methods used for multi-illumination scenes. We train two types of pixel-level inference model, HDRnet [18] and U-Net [30]. These models were trained to output white balanced images.

HDRnet. HDRnet [18] is proposed as a lightweight CNN for image enhancement on mobile devices. To train HDRnet, we use 256×256 and 128×128 as the size of a high resolution and a low resolution input, respectively. We use the default hyperparameters of the network and the model outputs the RGB values of each pixel of images as the original work. For optimization, we use both a cosine similarity loss and a mean squared error (MSE) loss to enforce the model to learn about the chromaticity of each pixel of images.

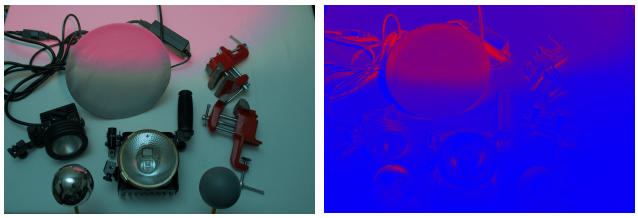
U-Net. U-Net [30] is originally designed to be used in the field of image segmentation, but it is widely used in various pixel-level tasks such as image processing and image-to-image translation [24, 27]. We train a U-Net with 7 downsampling steps and the input size of 256×256. The input image is transformed to a single luminance channel l and two chrominance channels u and v before being fed into the network since the chromaticity information is more useful than RGB values in our task [29, 2]. Our transformation is formulated as

$$\begin{aligned} l &= log(\mathbf{I}_G + \epsilon), \\ u &= log(\mathbf{I}_R + \epsilon) - log(\mathbf{I}_G + \epsilon), \\ v &= log(\mathbf{I}_B + \epsilon) - log(\mathbf{I}_G + \epsilon), \end{aligned} \tag{7}$$

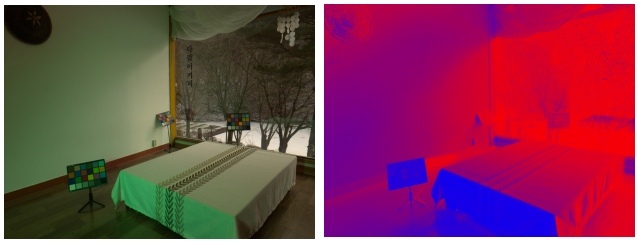
and the model output is two channel chrominance of ground truth white balanced image. Since our U-Net utilizes the chrominance values as input and output, not the RGB values, the MSE loss between chrominance vectors works in a similar way to the mixture of MSE and cosine similarity loss of HDRnet.



(a) Conventional relighting method [21, 32, 6]



(b) Multi Illumination in the Wild [29]



(c) Ours

Figure 6. Multi-illumination images and their coefficient maps from different datasets.

	LSMI (ours)	Multi Illumination in the Wild [29]
two-illuminant	0.1504	0.0838
three-illuminant	0.1633	0.1118

Table 4. Mean standard deviation of illuminant coefficient for our dataset and Multi Illumination in the Wild [29]. The results indicate that our data contains more spatially-diverse illumination.

5. Experiments

5.1. Multi-illumination Dataset Analysis

In addition to the statistic of the datasets provided earlier, we provide deeper analysis on image quality of different multi-illumination datasets. Fig. 6 (a) is an example of applying multi-illumination synthesis method [6, 21, 32] on our dataset. The images are synthesized by applying different color casts to parts of an image, which are arbitrarily divided. The boundaries are smoothed by a Gaussian kernel. The resulting image does not look natural as the scene geometry is ignored during the synthesis. Additionally, the actual mix of multiple illumination only occurs in a small region around the boundaries.

Fig. 6 (b) shows an example from [29], where images with multiple-illumination are synthesized by combining multiple images captured under different directions of a flash light. We have observed that this synthesis can result in images with almost uniform lighting, especially when using the indirect lighting that is bounced by the wall or the ceiling. When using directional flash light, we have also observed that it tends to saturate pixels too much, so a careful engineering was required when selecting lights to be used for the synthesis.

In comparison, our dataset is captured in a real-word illumination setting without further synthesis. As shown in Fig. 6 (c), our dataset provides more natural images with wide field-of-view that include larger variation of mixture ratio of illuminants.

We also compare the mean standard deviation of illuminant coefficients of datasets, to further show the difference of [29] and our data quantitatively. Since the original dataset in [29] does not provide synthesized images for

MAE		Single		Multi		Mixed	
		mean	median	mean	median	mean	median
Patch-based	statistical [20]	7.49	6.04	12.38	9.57	10.09	7.43
	learning [6]	4.15	3.30	5.56	4.33	4.89	3.83
Pixel-level	HDRnet	**2.85**	2.20	3.13	2.70	3.06	2.54
	U-Net	2.95	**1.86**	**2.35**	**2.00**	**2.63**	**1.91**
PSNR		Single		Multi		Mixed	
		mean	median	mean	median	mean	median
Patch-based	statistical [20]	29.4	30.0	25.4	25.2	27.3	26.8
	learning [6]	34.0	33.4	28.6	28.3	31.1	30.4
Pixel-level	HDRnet	**45.0**	**44.5**	38.3	37.6	41.1	40.1
	U-Net	44.6	43.9	**39.1**	**39.5**	**41.7**	**41.4**

Table 5. Mean Angular Error (MAE) and Peak Signal-to-Noise Ratio (PSNR) values of patch-based infernce model and pixel-level inference model.

multi-illuminant scenes, we generated the dataset from the provided images. Note that we excluded highly saturated images while synthesizing images, then followed the procedure described in the paper. We calculated the standard deviation of illuminant coefficients for each image, and averaged those for all images in the dataset. High values of the mean standard deviation indicate that the mixture ratio of illuminants is changing across the scene, which is common in typical multi-illuminant scenes. In contrast, low values indicate that the lights are mixed in similar proportions and the scene looks almost like a single light illuminated scene. As can be seen in Table 4, the values of our dataset are higher than those of [29], which demonstrate that our dataset contains more realistic and challenging examples of multi-illuminant scenes.

5.2. Pixel-level White Balance Results

Comparison with patch-based methods

To demonstrate the effectiveness of the pixel-level white balancing approach for multi-illumination scenes, we compare pixel-level models U-Net and HDRnet in Sec. 4 with two existing patch-based methods which we modified from patch-based model [6], and statistic-based model [20]. For the patch-based methods, training on multi-illuminant images is technically not feasible. Following the original works, we train them only with single-illuminated images using the relighting augmentation technique, and applied them to multi-illuminant scenes in a patch-wise manner. All images are resized to 256×256 images as used in U-Net and HDRnet, and we set the patch size to 16×16. In all experiments, we use a subset of our LSMI dataset captured with *Galaxy Note 20 Ultra*, which is split into train, validation, and test set with the ratio of 0.7, 0.2, and 0.1. Results on other cameras are provided in the supplementary material.

Fig. 7 shows a qualitative comparison of results. It can be seen that the pixel-level estimation models perform better white balancing compared to patch-based methods. The patch-based algorithms have the disadvantage that the information available to the model is limited by the patch, not the entire scene. Consistency between patches is also a major drawback. On the other hand, pixel-level methods can be trained to estimate the pixel-level illumination with the context of the whole image, resulting in better white balanced images.

MAE		Single		Multi		Mixed	
		mean	median	mean	median	mean	median
HDRnet	Original set	**3.13**	**2.53**	3.57	3.09	3.42	2.98
	Augmented set	3.16	2.56	**3.43**	**2.73**	**3.30**	**2.64**
U-Net	Original set	3.64	2.63	3.43	2.99	3.53	2.83
	Augmented set	**2.95**	**1.86**	**2.35**	**2.00**	**2.63**	**1.91**

Table 6. Mean Angular Error (MAE) values of HDRnet and U-Net trained by the original train set and the augmented train set of LSMI.

Quantitative results are shown in Table 5 . The test set is divided into three subsets – single illumination images (99), multiple illuminating images (112), and mixture of both (211). While the mean angular error (MAE) is a conventional evaluation metric for white balancing, we also provide PSNR as the objective is to recover white balanced images close to the ground truth. As expected, the pixel-based algorithms using DNNs perform better with larger gap in the case of multi-illumination.

Effect of data augmentation by relighting

To show the effectiveness of the data augmentation with relighting, we additionally compare results of U-Net and HDRnet on the original and the augmented dataset. We generated 4 more images for each scene using the pixel-level relighting, and the total of 7,345 images are used for the augmented training set. The quantitative results are shown in Table 6, and there is a clear improvement in performance using the augmented training set. It demonstrates that our dataset provides a way to further boost the performance of multi-illuminant white balancing models through data augmentation with relighting.

5.3. User study

An interesting study on white balancing under multiple illumination was presented in [10]. Under two illumination – indoor and outdoor illumination – they conducted a user study to find the preference between correcting the illumination with either indoor or outdoor lighting. According to their investigation, images corrected by the outdoor illumination were preferred by almost 80%.

We conducted a similar user study through Amazon Mechanical Turk to learn about users' preference on white balancing. On images with two illumination, we provided users with four white balanced images – white balanced with respect to outdoor illumination, indoor illumination, auto white balanced using LibRaw library, and pixel-wise white balance using the ground truth illumination map. Two

Figure 7. Visualization of various white balanced results using two patch-based models and two pixel-level models. Illumination mean angular errors provided for the reference.

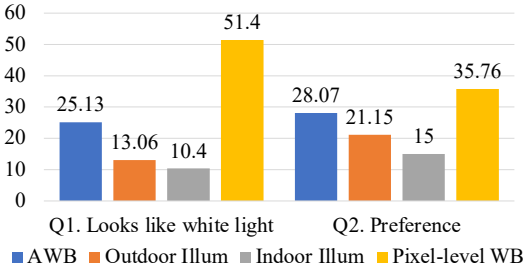


Figure 8. User response of our Amazon Mechanical Turk survey. All labels are expressed in percentage.

different studies were conducted. In the first task, we asked the users which photos were likely to be captured under the white light. For the second task, we asked the users to choose the picture they liked the most, paying attention to the color tone of the image. A total of 30 multi-illuminant scenes were provided, and for each scene, 50 answers were recorded in the first study, and 26 answers in the second study. Fig. 8 shows the results of the studies. According to the survey, pixel-level white balanced images received the most votes by a large margin. These results show the necessity of the pixel-level white balance algorithm for multi-illumination environments. Due to the lack of data, we have not seen much progress in this direction. Our new dataset opens the door for more efforts to be made in developing multi-illumination white balance algorithms.

6. Conclusion

In this paper, we introduced a new large scale multi-illuminant dataset for data-driven mixed illumination white balance algorithm. Our dataset provides the large amount of multi-illuminant scenes captured under the real-world setting, pixel-level labels of the chromaticity of illuminants and their mixture ratio. Our experiments show that CNN-based pixel-level methods outperform existing patch-based methods, and pixel-level white balance is mostly preferred by human observers compared with single illuminant white balance. Both the dataset and insightful experiments are expected to bring more attention on multi-illuminant white balance in future works, especially for pixel-level approaches. estimation. There is still a room for improvement in CNN-based methods used in this work, which may include incorporating domain knowledge of multi-illuminant white balance to the design of deep networks.

Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT), Artifitial Intelligence Graduate School Program, Yonsei University, under Grant 2020-0-01361

References

[1] Nikola Banić, Karlo Koščević, and Sven Lončarić. Unsupervised learning for color constancy. *arXiv preprint arXiv:1712.00436*, 2017. 2

[2] Jonathan T Barron. Convolutional color constancy. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 379–387, 2015. 2, 6

[3] Jonathan T Barron and Yun-Ta Tsai. Fast fourier color constancy. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–894, 2017. 1

[4] Shida Beigpour, Christian Riess, Joost Van De Weijer, and Elli Angelopoulou. Multi-illuminant estimation with conditional random fields. *IEEE Transactions on Image Processing*, 23(1):83–96, 2013. 1, 2, 3

[5] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Color constancy using cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 81–89, 2015. 1, 2

[6] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Single and multiple illuminant estimation using convolutional neural networks. *IEEE Transactions on Image Processing*, 26(9):4347–4362, 2017. 1, 2, 6, 7

[7] Simone Bianco and Raimondo Schettini. Adaptive color constancy using faces. *IEEE transactions on pattern analysis and machine intelligence*, 36(8):1505–1518, 2014. 2, 3

[8] Michael Bleier, Christian Riess, Shida Beigpour, Eva Eibenberger, Elli Angelopoulou, Tobias Tröger, and André Kaup. Color constancy and non-uniform illumination: Can existing algorithms work? In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 774–781. IEEE, 2011. 2, 3

[9] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980. 1, 2

[10] Dongliang Cheng, Abdelrahman Abdelhamed, Brian Price, Scott Cohen, and Michael S Brown. Two illuminant estimation and user correction preference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 469–477, 2016. 7

[11] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, 31(5):1049–1058, 2014. 1, 2

[12] Florian Ciurea and Brian Funt. A large image database for color constancy research. In *Color and Imaging Conference*, volume 2003, pages 160–164. Society for Imaging Science and Technology, 2003. 1, 2

[13] Egor Ershov, Alexey Savchik, Illya Semenkov, Nikola Banić, Alexander Belokopytov, Daria Senshina, Karlo Koščević, Marko Subašić, and Sven Lončarić. The cube++ illumination estimation dataset. *IEEE Access*, 8:227511–227527, 2020. 2

[14] Graham D. Finlayson, Steven D. Hordley, and Paul M. Hubel. Color by correlation: A simple, unifying framework for color constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, 2001. 2

[15] David A Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–35, 1990. 2

[16] Brian Funt, Vlad Cardei, and Kobus Barnard. Learning color constancy. In *Color and Imaging Conference*, volume 1996, pages 58–60. Society for Imaging Science and Technology, 1996. 2

[17] Peter Vincent Gehler, Carsten Rother, Andrew Blake, Tom Minka, and Toby Sharp. Bayesian color constancy revisited. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008. 1, 2

[18] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics (TOG)*, 36(4):1–12, 2017. 6

[19] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Generalized gamut mapping using image derivative structures for color constancy. *International Journal of Computer Vision*, 86(2-3):127–139, 2010. 2

[20] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Improving color constancy by photometric edge weighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):918–929, 2011. 2, 7

[21] Arjan Gijsenij, Rui Lu, and Theo Gevers. Color constancy for multiple light sources. *IEEE Transactions on Image Processing*, 21(2):697–707, 2011. 1, 2, 3, 6

[22] Ghalia Hemrit, Graham D Finlayson, Arjan Gijsenij, Peter Gehler, Simone Bianco, Brian Funt, Mark Drew, and Lilong Shi. Rehabilitating the colorchecker dataset for illuminant estimation. In *Color and Imaging Conference*, volume 2018, pages 350–353. Society for Imaging Science and Technology, 2018. 1, 2

[23] Eugene Hsu, Tom Mertens, Sylvain Paris, Shai Avidan, and Frédo Durand. Light mixture estimation for spatially varying white balance. In *ACM SIGGRAPH 2008 papers*, pages 1–7. 2008. 2

[24] Xiaodan Hu, Mohamed A Naiel, Alexander Wong, Mark Lamm, and Paul Fieguth. Runet: A robust unet architecture for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 6

[25] Yuanming Hu, Baoyuan Wang, and Stephen Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4085–4094, 2017. 2

[26] Zhuo Hui, Aswin C Sankaranarayanan, Kalyan Sunkavalli, and Sunil Hadap. White balance under mixed illumination using flash photography. In *2016 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2016. 2

[27] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 6

[28] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 1, 2

[29] Lukas Murmann, Michael Gharbi, Miika Aittala, and Fredo Durand. A dataset of multi-illumination images in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4080–4089, 2019. 2, 3, 6, 7

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 6

[31] Lilong Shi. Re-processed version of the gehler color constancy dataset of 568 images. *http://www. cs. sfu. ca/˜ color/data/*, 2000. 1, 2

[32] Oleksii Sidorov. Conditional gans for multi-illuminant color constancy: Revolution or yet another approach? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 1, 2, 6

[33] Rytis Stanikunas, Henrikas Vaitkevicius, and Janus J Kulikowski. Investigation of color constancy with a neural network. *Neural Networks*, 17(3):327–337, 2004. 1, 2

[34] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 16(9):2207–2214, 2007. 1, 2