IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 January 2023, accepted 23 February 2023, date of publication 27 February 2023, date of current version 2 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3249779

RESEARCH ARTICLE

Deep Highlight Removal Using Temporal Dark Prior in High-Speed Domain

JEONG-WON HA [1], **KANG-KYU LEE** [1], **JUN-SANG YOO** [2], **AND JONG-OK KIM** [1], **(Member, IEEE)**
[1]School of Electrical Engineering, Korea University, Seoul 02841, South Korea
[2]Computer Vision Laboratory, Samsung Advanced Institute of Technology, Suwon, Gyeonggi-do 16678, South Korea

Corresponding author: Jong-Ok Kim (jokim@korea.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant by the Korea Government through the Ministry of Science and Information & Communications Technology (MSIT) under Grant 2020R1A4A4079705; and in part by MSIT, South Korea, through the Information Technology Research Center (ITRC) by the Institute of Information and Communications Technology Planning and Evaluation (IITP) under Grant IITP-2020-0-01749.

ABSTRACT This paper proposes a deep highlight removal method based on the dichromatic model under alternating current (AC) light sources with a new diffuse prior on temporal domain, temporal dark prior. An input image is decomposed into specular and diffuse components using a deep network. Due to AC powered lights, both incident and reflected lights are time-varying. We exploit the periodic variation property of the specular and diffuse reflections as a prior for dichromatic model based image decomposition. In addition, we propose a new temporal dark prior as a pseudo-diffuse reflection. Unlike the conventional prior in the spatial domain, to the best of our knowledge, this is the first study to utilize a diffuse prior on the temporal domain for highlight removal. The blurred version of the temporal dark prior is additionally fed to the network to alleviate hole artifacts. It is demonstrated through diverse experiments that these temporal priors can strongly contribute to accurate image decomposition, leading to better highlight removal.

INDEX TERMS Dichromatic model, highlight removal, high-speed camera.

I. INTRODUCTION

The dichromatic reflection model assumes that the reflection of an object is represented as a linear combination of specular and diffuse reflections. This means that the pixel intensity of a captured image is the sum of the specular and diffuse reflection, as illustrated in Fig. 1. Shafer [1] defined the specular and diffuse reflection as interface and body reflection. When the light strikes a surface, it first passes through the interface between the air and the surface medium. Because the medium and the air have different refraction index, some of the light is reflected at the interface and produces interface reflection. The amount of interface reflection is determined by Fresnel's law that relates interface reflection to the angle of incidence, the refraction index of the material, and the polarization of the incoming illumination. Interface reflection is assumed to be constant with respect to wavelength, and it has the same color as the illuminant. The light that penetrates through the interface passes through the medium, and it undergoes scattering from the colorant. It is absorbed by the colorant or re-emitted through the interface, producing body reflection.

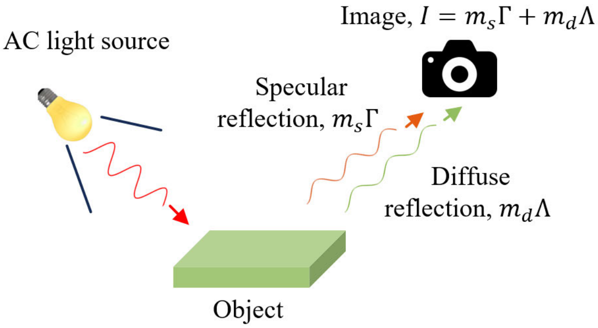
Image, $I = m_s\Gamma + m_d\Lambda$
AC light source
Specular reflection, $m_s\Gamma$
Diffuse reflection, $m_d\Lambda$
Object

FIGURE 1. Image formation scenario based on the dichromatic model under alternating current(AC) light sources.

The associate editor coordinating the review of this manuscript and approving it for publication was Yongjie Li.

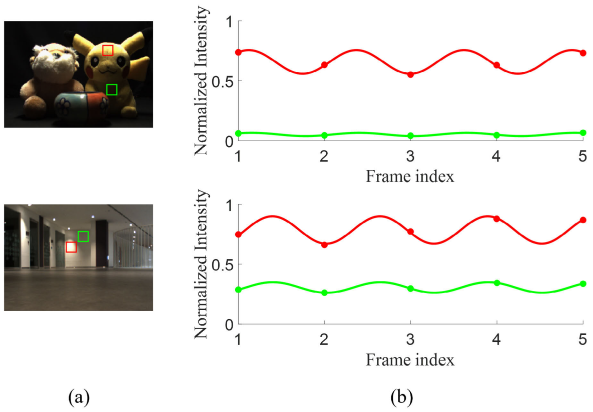

FIGURE 2. **(a) input image, (b) sine fitted curves for highlight and highlight-free regions.**

The color of the body reflection is generally different from the illumination, because interactions with colorant particles result in absorption with a probability depending on wavelength. The dichromatic reflection model has been widely used for estimating illumination chromaticity and highlight removal. Estimating the model from image signals, however, is a highly ill-posed problem with four unknown parameters (chromaticity and intensity of specular and diffuse reflection). Thus, it is frequently assumed in many conventional methods that the illuminant is known in advance.

The specular component can be an important clue for both illumination estimation and separation. There have been a number of studies on color constancy and highlight removal using the dichromatic model. Since most vision tasks exploit the color information, it is important to restore the original reflectance color from highlight and saturated regions [2], [3], [4]. Highlights can cause failure in stereo matching, object recognition (e.g., face), and road segmentation which plays a significant role in advanced driver assistance systems (ADAS) [2], [5]. [2] showed the performance improvement with highlight removal in road segmentation. Reference [3] removed highlight especially on text image, and improved the performance of text detection and recognition. Also, in face recognition and iris segmentation, highlight removal is an important preprocess [6], [7].

Since the invention of electric bulbs, we have been living under various electric light sources. Because these light bulbs are supplied with alternative current (AC) power, their intensity varies sinusoidally with double the AC standard frequency [8], [9], [10]. Because the variations are faster than a capability of human eyes to capture the flicker, we cannot observe the intensity fluctuations of AC light sources. However, with a high-speed camera, the intensity variations can be easily captured. These variations can be beneficial temporal features, in addition to spatial features. Fig. 2 shows high speed frames, and the red and green boxed regions belong to highlight and highlight-free regions, respectively. The fitted sine curve with average intensity of each region is shown. Because of AC-power variation, the intensities of both highlight and highlight-free regions vary sinusoidally. The amplitude of the highlight region is larger than the highlight-free region in both closed and open scenes. We propose leveraging this temporal fluctuation for highlight removal.

This paper proposes a novel deep highlight removal method based on a dichromatic model under AC light sources. A frame captured with high-speed camera is decomposed into specular and diffuse components using a deep network, and highlight removal is easily achieved by simply discarding the specular component in the dichromatic model. Fig. 1 shows the scenario of image formation considered for highlight removal under AC powered light. We exploit the periodic variation property of specular and diffuse reflections as a prior for dichromatic based image decomposition. However, it is difficult to get ground truths for the specular and diffuse components. This makes it significantly challenging to train an image decomposition network.

Several previous studies on highlight removal have used a prior for specular reflection. To determine pseudo specular reflection, they studied a threshold for Value or intensity [11], [12], [13], [14] and used the minimum of RGB channels [15], [16]. In this paper, a new prior for the diffuse component is proposed for stable convergence and accurate model estimation. The minimum luminance is taken on high-speed temporal frames, named as temporal dark prior (TDP). It can be assumed that the TDP is less affected by the light source, and thus, it is close to a specular free image. In other words, it can be regarded as a pseudo diffuse reflection. The novel temporal prior can contribute to more accurate diffuse component generation at the network output, leading to higher dichromatic decomposition.

The blurred version of the TDP is also proposed as a prior for diffuse chromaticity under the assumption that the chromaticity of the highlight would be similar to that of its surroundings. It can play an important role in filling the holes in highlight regions, which often occur at highlight removal.

In this paper, we propose a novel highlight removal method that exploits temporal features. The main contributions of the paper are summarized as follows:

- In our previous works [8], [10], [17], we firstly proposed to exploit the variation of AC lights for illuminant estimation and highlight removal. This paper extends the previous AC light based dichromatic decomposition by adding TDP for highlight removal.
- We propose TDP which is extracted from the temporal variations of AC light images, and it is a temporal extension of the conventional spatial dark channel prior for highlight removal [15], [16].
- We built a new highlight removal dataset which contains high-speed video frames acquired under AC lights (indoor and closed laboratory environments).

The rest of the paper is organized as follows. Section II describes the previous studies of highlight removal. The details of the proposed method are presented in Section III.

Experiment results are shown in Section IV. Finally, Section V concludes the paper.

II. RELATED WORKS

The highlight removal is based on the dichromatic reflection model. This model assumes that the reflected light consists of specular and diffuse reflections as their linear combination:

$$i = m_d \Lambda + m_s \Gamma \tag{1}$$

where i denotes the luminance of a pixel, Λ and Γ are diffuse and specular chromaticity, respectively. Also, m_d and m_s are weight coefficients, and are defined as:

$$m_d = w_d \sum B_c, m_s = w_s \sum G_c \tag{2}$$

where w_d and w_s are geometrical parameters that depend on the geometric structure. Both $\sum B_c$ and $\sum G_c$ depend on the intensity of incident light, and they represent diffuse albedo and Fresnel reflectance, respectively [18]. So, the intensity of diffuse and specular reflection vary by the fluctuation of AC light source. Highlight removal is equivalent to separating the specular reflection from the input image. Previous studies on highlight removal can be divided into two groups according to the number of input images: single image and multiple images.

A. A SINGLE-IMAGE APPROACH

Because finding diffuse and specular reflection is a highly ill-posed problem, most single image methods rely on constraints, including the white illuminant assumption. Many single-image methods start with pseudo specular reflection and iteratively solve the inverse problem using optimization techniques. Xu and Zhou [11] and Xia et al. [13] obtain a pseudo specular reflection by thresholding input pixels with the Value in the HSV color space. Kim et al. [15] and Ramos et al. [16] proposed that the minimum intensity among the RGB channels for each pixel can be used as a prior of specular reflection. However, these methods detect gray pixels as fully specular reflection and fail to reconstruct the diffuse color. The proposed method generates pseudo diffuse reflection to help the reconstruction of intrinsic color on specular regions. Tan and Ikeuchi [19] proposed the relation between the diffuse reflection and maximum diffuse reflection chromaticity, and Shan et al. [12] used it to obtain pseudo diffuse component. However, the prior still shows artifacts on saturated regions and it leads to failure of diffuse reconstruction. The priors are compared in Fig. 13. Several methods use clustering to recover the chromaticity of the diffuse components. Suo et al. [20] defined the l_2 chromaticity and used material clustering on an illuminant orthogonal subspace to find pixels with the same diffuse reflectance. Yang et al. [21] and Shen and Zheng [22] obtained a specular free image by pixel clustering based on hue and intensity ratio. The clustering based methods [20], [21], [22] have limitation in recovering various colors. In these works, all the previous priors are extracted in the spatial domain. However, we propose a new temporal prior, which is the first trial in the temporal domain.

Deep-learning based single-image methods [23], [24] require a ground truth specular or diffuse reflection for supervised learning. These methods contain structural similarity loss with a ground truth image in the loss function. They usually use synthetic images with ground truths for training. However, it is difficult to obtain ground truth specular and diffuse reflection components for real images. Although ground truth reflection components can be obtained with images captured with different degrees of polarization filter [25], [26] or controlling illumination and a sensor mask [27], it is complicated to acquire additional polarization images or masked images for ground truth. Also, Unlike previous methods, the proposed network is trained in an unsupervised way without ground truth. The TDP and AC light variations make it possible to learn it using real images.

B. MULTI-IMAGE APPROACH

There are two types of previous studies using multiple images: different directions of a light source [28], [29], [30] and different viewpoints [31], [32]. These methods focus on the spatial information of the images. References [28], [29], [30] used images taken from the same viewpoint with different illuminant positions. The datasets of [31] and [32] were obtained with different vantage points. These methods require additional experimental settings for different positions of a light source and camera that is not suitable for real world situations. However, our proposed method does not require additional constraints for position settings, and the only constraint for the proposed method (AC light sources) is common in indoor environments. Therefore, our proposed method is more practical than previous studies [28], [29], [30], [31], [32]. In addition, it utilizes temporal correlation, while the previous methods with multiple images still rely on spatial information.

Few studies have exploited temporal information for highlight removal [17], [33], [34]. Tsuji [33] also exploited the fluctuation of AC lights. The specular reflection is removed by assuming that the min/max luminance of a high-speed video is a linear prior for the diffuse reflection. However, it has a limitation that the result is highly dependent on the configuration of a parameter (α in [33]) that determines the amount of specular reflection to be removed. Prinet et al. [34] proposed a temporal prior with normal speed frames for the enhancement of specular highlight, unlike specular removal and high-speed in this paper. Yoo et al. [17] proposed a deep network for estimating the dichromatic model under AC light sources. It aims to obtain all the parameters of the dichromatic model and the estimated parameters are exploited for color constancy and highlight removal. It is similar to our proposed method in that it considers AC lights. However, it estimates the dichromatic model itself (i.e., four model parameters) by leveraging the property of AC lights, which is a significantly ill-posed problem. In this paper, this strict model estimation problem is alleviated, primarily aiming at highlight removal. The dichromatic model is regarded as

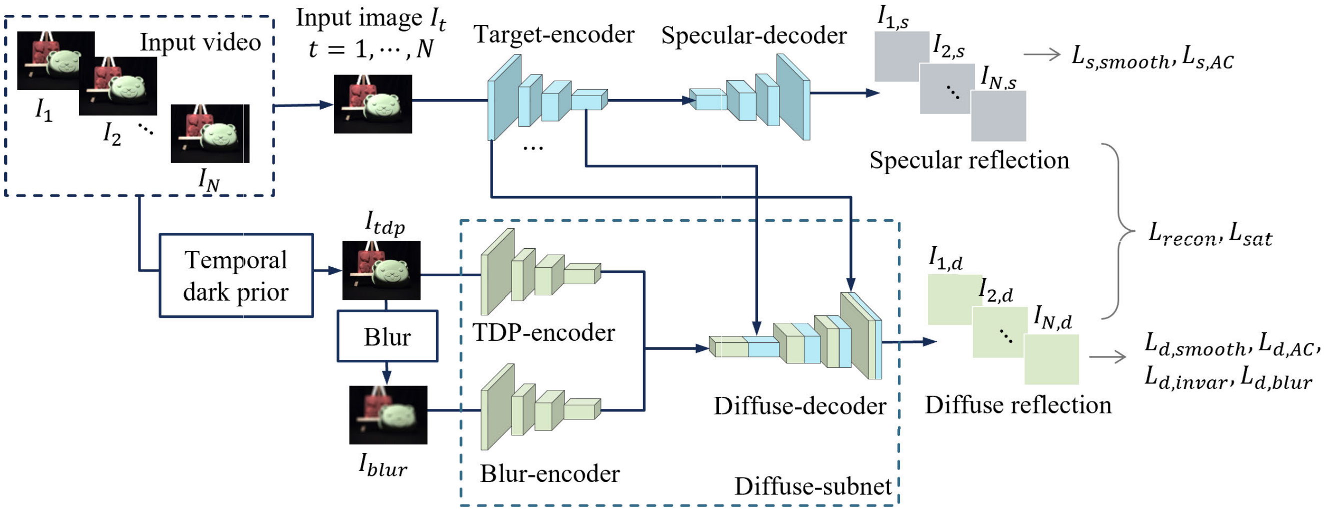

FIGURE 3. The proposed network architecture. The proposed network consists of three encoders (Target, TDP and Blur encoder) and two decoder that generate specular and diffuse reflection.

an image decomposition problem. In other words, an input image is decomposed into the specular and diffuse components, instead of the estimation of the four dichromatic parameters.

III. THE PROPOSED METHOD

The proposed network uses a high-speed video as an input and generates specular and diffuse reflection components at its output. Fig. 3 illustrates the overall network structure of the proposed network. The input image I_t is the t^{th} frame of the input video, and all frames are sequentially fed into the target encoder one by one. The diffuse reflection $I_{t,d}$, and specular reflection $I_{t,s}$ are the network output for the input I_t. N denotes the number of input frames. The proposed network consists of two subnets designed for both specular and diffuse generation. The lower diffuse subnet in Fig. 3 includes two encoders that accept important prior information useful for the estimation of diffuse reflection. The inputs to the two diffuse encoders are the temporal dark prior and its blurred version. The former corresponds to a pseudo diffuse component for image decomposition, while the latter provides the prior to alleviate the hole artifact that commonly occurs in highlight regions. Unlike conventional works that propose specular prior on the spatial domain, this study is the first to utilize diffuse priors in the temporal domain. Leveraging the colors of surrounding pixels for hole-filling of highlight removal has already been studied in the non-deep-learning approach [6], [13]. The blur encoder was originally inspired by this conventional method. As demonstrated in the ablation study of the experimental results section, these two inputs contribute significantly to image decomposition.

The encoder of each subnet is VGG 16 network [35] without fully connected layers. Convolution and max pooling layers encode the input image ($H \times W \times 3$) to $H/32 \times W/32 \times 512$ features. The decoder is composed of '1 × 1 Conv + 4 × 4 Deconv' and skip connection that transfers the feature of each level in the encoder to the decoder. The subnets adopt a convolutional auto-encoder based on VGG 16 network [36]. The features of the target encoder are transferred to the diffuse decoder as well as the specular decoder. Each hierarchical spatial feature of the target encoder is concatenated with the corresponding feature of the diffuse decoder, as shown in sky blue in Fig. 3. To accurately reconstruct the diffuse reflection, the features of the TDP and blur encoders are combined, and the results are decoded by concatenating hierarchically with the features of the target encoder.

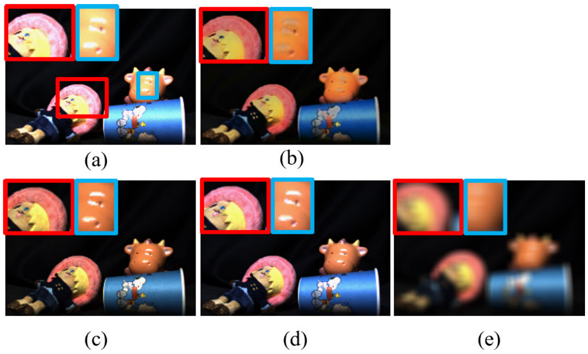

FIGURE 4. (a) a target frame, (b) its estimated diffuse reflection with the proposed method, (c) temporal dark prior (TDP) from five frames, (d) TDP from three frames, (e) blurred TDP.

A. TEMPORAL DARK PRIOR

The proposed method can improve the reconstruction performance of diffuse reflection using the proposed temporal dark prior (TDP). We considered a high-speed video under AC light sources, whose brightness variations over time can be observed easily. Due to the sinusoidal intensity variations

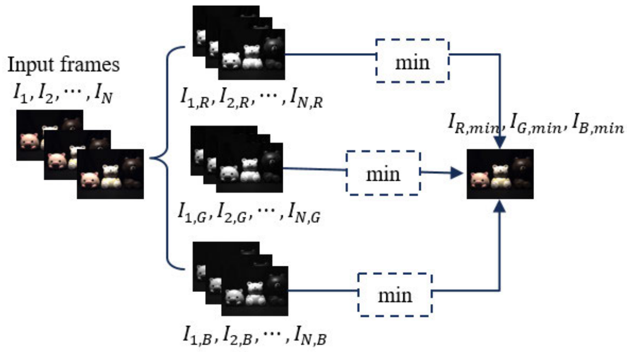
Input frames
$I_1, I_2, \cdots, I_N$
$I_{1,R}, I_{2,R}, \cdots, I_{N,R}$
min
$I_{R,min}, I_{G,min}, I_{B,min}$
$I_{1,G}, I_{2,G}, \cdots, I_{N,G}$
min
$I_{1,B}, I_{2,B}, \cdots, I_{N,B}$
min

FIGURE 5. **The generation of TDP.**

of AC power, there will probably be a frame that is captured when the power of the light sources is minimum. Since this frame is generated by weaker illumination than any other frame, it is assumed that the frame is the closest to the true diffuse reflection in high-speed video. It is regarded as a pseudo diffuse reflection, and is named as temporal dark prior (TDP) in this paper. It is practically difficult to find a video frame captured at the minimum AC power due to severe noises in high-speed video and small AC signal amplitudes. Thus, the minimum was taken from the temporal signals. Fig. 4 (a) is a high-speed video frame, (b) is the diffuse reflection of (a) estimated using the proposed method, and (c) shows the TDP of the input video. As shown in Fig. 4, the TDP is similar to the diffuse reflection, and it is reasonable to use the TDP as a pseudo diffuse reflection. In addition, TDP can be a clue for restoring diffuse chromaticity, which is already lost by dominant specularity, as shown in the red boxed region.

Fig. 5 shows how the TDP was obtained. The minimum at a fixed location of the input video is taken as the TDP (denoted by I_{tdp}), which is given by

$$I_{tdp,c}(i) = min\{I_{1,c}(i), I_{2,c}(i), \cdots, I_{N,c}(i)\}, \quad where\ c \in \{r, g, b\} \tag{3}$$

where c and i denote the r, g, b channels and i^{th} pixel of the input frames, respectively. N is the total number of frames in the input video. The $min\{\cdot\}$ operation is applied for every channel, and consequently, we can obtain a temporal dark prior image.

The TDP can provide image details on saturated regions where the diffuse colors of a surface are lost by excessive illumination. These regions are highly challenging to estimate intrinsic colors. According to the time-varying intensity of AC light sources, it could be possible to extract the information of intrinsic colors from the high-speed frames. For example, for the red boxed region in Fig. 4, the target frame (a) lost image details because of strong illumination, while the details are still preserved to some extent in the TDP (c). Therefore, the TDP can be a useful prior for the estimation of diffuse chromaticity, and this is experimentally confirmed in the ablation study in the next section.

In addition, the blurred TDP is fed into the proposed network. The saturated pixels of an object are sparse, and diffuse chromaticity is locally constant [37], [38]. Therefore, it can be assumed that the chromaticity of the surrounding pixels is similar to that of the saturated pixels. In the blue boxed region of Fig. 4, it is confirmed that the chromaticity of the saturated region is recovered with the blurred TDP. Reference [6] claims that Gaussian blurred prior can be a helpful cue for the reduction of saturation artifacts and noises in the image. These previous studies show that the surrounding pixels contribute to alleviating hole artifacts. Also, [39] mentioned the importance of the receptive field size. The large highlight region cannot be filled with small receptive field, while small highlight region may not be detected with large receptive field. By using the blurred image as an input, the larger receptive field can be used in the former layer, and this can contribute to improve the hole-filling. Based on this property, the blurred TDP is obtained by applying average filtering to the TDP. It is expected to fill holes in highlight regions with their surrounding pixel colors.

B. LOSS FUNCTIONS

To train the network, we design a couple of losses that reflect the characteristics of the specular and diffuse components. The network is trained to minimize the weighted sum of losses as follows:

$$L_{tot} = L_{recon} + w_1 L_{sat} + w_2 L_{diff} + w_3 L_{spec}. \tag{4}$$

The sub-losses $L_{recon}, L_{sat}, L_{diff}$, and L_{spec} represent the reconstruction, saturation, diffuse reflection, and specular reflection losses, respectively.

The reconstruction loss L_{recon} is L1 loss between the input frame and the reconstructed frame with diffuse and specular components. Based on the dichromatic model, the reflectance is represented as the sum of two reflection components. For our proposed network, a target frame I_t should be equal to the sum of the specular $I_{s,t}$ and diffuse $I_{d,t}$ generated as the network outputs of I_t, and is given by

$$I_t = I_{t,d} + I_{t,s}, \ t = 1, \cdots, N. \tag{5}$$

For saturated pixels, the sum of the specular and diffuse reflection components is greater than 255. This causes the network to be trained with an inaccurate reconstruction loss. Therefore, the reconstruction loss is calculated only on non-saturation regions as follows:

$$I_{t,recon}(i) = \begin{cases} I_{t,d}(i) + I_{t,s}(i), & I_{t,d}(i) + I_{t,s}(i) < 255 \\ 255, & otherwise. \end{cases} \tag{6}$$

where i denotes the pixel index of the image. The reconstruction loss, L_{recon}, is given by

$$L_{recon} = \frac{1}{N} \sum_{t=1}^{N} \| I_t - I_{t,recon} \|_1 . \tag{7}$$

It is expected that for saturated pixels, specular reflection is much stronger than diffuse reflection. This relation is formulated as the saturation loss, L_{sat}. We define the ratio of the diffuse coefficient to the specular coefficient as the saturation loss for saturated pixels, I_{sat}, as follows:

$$L_{sat} = \frac{1}{N}\sum_{t=1}^{N}\sum_{i\in I_{sat}} \frac{\| I_{t,d}(i) \|_1}{\| I_{t,s}(i) \|_1}. \tag{8}$$

To find a saturated pixel, previous studies [11], [12], [13], [14] threshold the intensity of a pixel, while the proposed method imposes an additional temporal constraint on it. Under AC light sources, the intensity in the non-saturated regions is sinusoidally varying with time, but the intensity in the saturated region is constant. Thus, these saturated pixels have zero temporal gradients. In this study, pixels within a prescribed threshold of the temporal gradient $TG(i)$ were selected as saturated pixels. In addition, a saturated pixel is strongly illuminated, and its intensity should be high. Thus, a saturated pixel is determined by

$$I_{sat}(i) = \{i | I(i) > Th_1,\ TG(i) < Th_2\}, \tag{9}$$

where Th_1 and Th_2 are threshold values of intensity and temporal gradient, respectively.

The loss for the diffuse reflection, L_{diff}, consists of invariant loss $L_{d,invar}$, blur loss $L_{d,blur}$, smooth loss $L_{d,smooth}$, and AC fitting loss $L_{d,AC}$:

$$L_{diff} = L_{d,invar} + \alpha_1 L_{d,blur} + \alpha_2 L_{d,smooth} + \alpha_3 L_{d,AC} \tag{10}$$

If an object and the camera are static in the input video, the diffuse chromaticity along all frames should be constant. This property is reflected as an invariant loss, which is expressed as follows:

$$L_{d,invar} = \sum_{t=1}^{N-1}\sum_{t'=t+1}^{N} \left\| \frac{I_{t,d}}{\sum_{c\in\{r,g,b\}} I^c_{t,d}} - \frac{I_{t',d}}{\sum_{c\in\{r,g,b\}} I^c_{t',d}} \right\|_1 \tag{11}$$

Saturated pixels commonly lead to hole artifacts in the diffuse reflection component, which makes highlight removal more challenging. The blurred TDP can provide a clue for the intrinsic color of a saturated pixel, and the similarity between the blurred TDP, I_{blur}, and the estimated diffuse reflection is used as the blur loss:

$$L_{d,blur} = \frac{1}{M_{sat}}\sum_{t=1}^{N}\sum_{i\in I_{sat}} \| I_{d,t}(i) - I_{blur}(i) \|_2^2, \tag{12}$$

where M_{sat} means the number of saturated pixels.

The diffuse chromaticity is piecewise constant [38] and edge preserving, and it is applied as the TV-L1 loss for diffuse reflection:

$$L_{d,smooth} = \sum_{t=1}^{N} \| \nabla I_{t,d} \|_1 \tag{13}$$

This loss can contribute to restore diffuse chromaticity on highlight regions.

Under AC light sources, the intensity of the reflected light varies sinusoidally [8], [17]. The periodic variation is fit with a sine curve, and the regression errors are measured using the AC fitting loss. Instead of the regression of temporal pixels, the mean values of all the frames are fit with $g(t, \Theta)$:

$$g(t, \Theta) = A\ sin(4\pi f_0 t / f_{cam} + \phi) + off, \tag{14}$$

where A denotes the amplitude of a fitting function, f_0 is the standard frequency of the AC current, $f_{cam}$is the video frame rate, and off is an offset value. The diffuse reflection estimated with the proposed network should be fit with $g(t, \Theta)$, and the regression errors are defined as the AC fitting loss:

$$L_{d,AC} = \sum_{t=1}^{N} \left(\overline{I_{d,t}} - g(t, \Theta)\right)^2 \tag{15}$$

For specular reflection, there exist smooth loss $L_{s,smooth}$ and AC fitting loss $L_{s,AC}$:

$$L_{spec} = L_{s,smooth} + w_1 L_{s,AC} \tag{16}$$

The specular reflection is spatially smooth on smooth surfaces [38], and this property is reflected as the TV-L2 loss:

$$L_{s,smooth} = \sum_{t=1}^{N} \| \nabla I_{t,s} \|_2^2 \tag{17}$$

With this smooth loss, we can extract the specular reflection closer to the ground truth. Identical to the diffuse reflection, the specular reflection also varies sinusoidally; accordingly, the AC fitting loss is also used for the specular reflection:

$$L_{s,AC} = \sum_{t=1}^{N} \left(\overline{I_{s,t}} - g(t, \Theta)\right)^2 \tag{18}$$

C. NETWORK TRAINING

Among N frames of the input high-speed video, each frame is sequentially fed into the proposed network, consequently generating N frames of diffuse and specular components. Our proposed network exploits two types of losses for training: temporal loss and spatial loss. The temporal loss (e.g., $L_{d,AC}$ and $L_{s,AC}$) is calculated from all frames of the input video, and the spatial loss is calculated from each input frame. Therefore, the network is updated at every scene (N frames in total), not by a single frame during training. For spatial loss, the average from N frames is used for training. So, the total loss used for updating the network (the weighted sum of the average spatial losses of N frames and temporal losses) is calculated every N frames. Algorithm 1 shows the process of calculating loss and updating the proposed network.

IV. EXPERIMENTAL RESULTS

The proposed network was trained with our proprietary dataset captured with a high-speed camera. The Adam optimizer was used for training with a batch size of 16. The initial

Algorithm 1 Network Training

```
Input video, $I = \{I_1, \cdots, I_N\}$
Generate $I_{tdp}$ and $I_{blur}$
Initialize $L_{tot} = 0$
**for** $t$ in $(1,N)$ **do**
    $I_{t,d}, I_{t,s} = model(I_t, I_{tdp}, I_{blur})$
    $L_{tot} = L_{tot} + L_{sat}(I_{t,d}, I_{t,s})$
    $L_{tot} = L_{tot} + L_{d,blur}(I_{t,d}, I_{tdp})$
    $L_{tot} = L_{tot} + L_{d,smooth}(I_{t,d}) + L_{s,smooth}(I_{t,s})$
**end for**
$D = \{I_{1,d}, \cdots, I_{N,d}\},\ S = \{I_{1,s}, \cdots, I_{N,s}\}$
$L_{tot} = L_{tot} + L_{recon}(I, D, S)$
$L_{tot} = L_{tot} + L_{d,invar}(D)$
$L_{tot} = L_{tot} + L_{d,AC}(D) + L_{s,AC}(S)$
update the network with $L_{tot}$
```

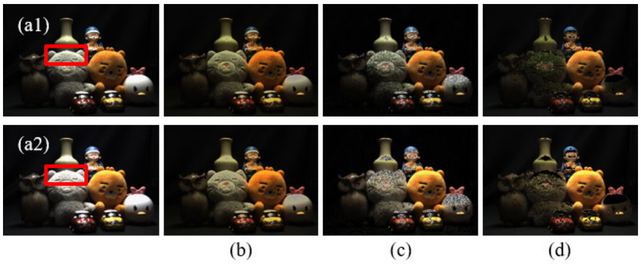

FIGURE 6. (a1) the darkest input image and (a2) the brightest one among input video, (b-d) the diffuse reflection of (b) the proposed method, (c) Akashi and Okatani [41], (d) Yamamoto and Nakazawa [42].

learning rate was 1×10^{-3}, and the learning rate is decayed with epochs. The number of frames (N in Section III) used for training was 5. The performance of highlight removal was evaluated qualitatively from the aspects of hole artifacts and the separation of diffuse and specular reflection components. Since we do not have ground truth reflection components of our dataset, several no-reference image quality assessments are compared. Also, we conducted evaluation with WHU-Specular [40] and SHIQ [39] datasets.

A. DATASET

Because there is no public dataset built under AC light sources, we used our own dataset, which contains diverse and general scenes in both closed and open environments. The scenes are captured with Sentech STC-MCS43U3V high-speed vision camera. The camera frame rate was 150 FPS, and the exposure time was 1/300 sec. The resolution of the video frames is 740 × 540. The closed scenes are captured under incandescent and fluorescent light sources in a laboratory environment that can control external lighting. The open scenes are taken in public indoor spaces such as hospitals, schools and library. The scenes were additionally captured with a color checker to obtain ground truth illuminant. The images in indoor spaces occupy 66% of the dataset, and the number of closed scenes is 33%. A total of 150 and 75 scenes were used for training and testing, respectively. To evaluate our method for diverse materials, our dataset contains various object colors and surfaces such as ceramic, wood, and plastic. We captured static scenes with a fixed camera, and assumed there is no movement in the scenes.

B. COMPARISONS WITH CONVENTIONAL METHODS

The performance was compared with several conventional methods. Akashi and Okatani [41], Yamamoto and Nakazawa [42], Yang et al. [43], Fu et al. [44] and JSHDR [39] are single-image approaches that exploit only spatial features, and Tsuji [33] and Yoo et al. [17] are a multiple-image approach using a high-speed video as an input. As shown in Fig. 6, the performance of the single-image method varies according to the intensity of the input image. Although our proposed method performs well for both bright and dark frames, single-image methods show better performance with darker images. For this reason, TDP was used as the input of Akashi and Okatani [41], Yamamoto and Nakazawa [42], Yang et al. [43] and Fu et al. [44] for fair comparison. Since the result of Tsuji [33] varies by parameter α, the optimal result is chosen experimentally. Since JSHDR [39] and JSHDR-Trans [45] are supervised methods and there is no ground-truth in the proposed dataset, the model is trained with the same loss as the proposed method in an unsupervised manner. Network structure in [39], [45] was not changed. Note that the learning-based models (JSHDR, JSHDR-Trans [45] and the proposed method) are trained with the proposed dataset.

Fig. 7 compares the proposed method with the conventional methods, whose results are noisy, particularly for saturated regions. Our proposed method performs well in separating the diffuse and specular reflections and obtains accurate diffuse chromaticity. As shown in the input image (a1) (with outstanding highlight) in Fig. 7, the proposed method greatly reduces hole artifacts, while the highlight still remains for Tsuji [33], Fu et al. [44] and Yoo et al. [17]. Also, Akashi and Okatani [41], Yamamoto and Nakazawa [42] and Yang et al. [43] extravagantly removed the specular reflection, and failed to reconstruct clean diffuse chromaticity, resulting in severe color distortion. Also, the proposed method performed better than the conventional methods for the complex colored scene as (a2) in Fig. 7. JSHDR [39] extracts a feature with a single encoder-decoder, which is used to estimate both specular and diffuse reflection. The proposed method, however, exploits different sub-networks for specular and diffuse reflection. Our proposed method works well for public indoor scenes, as shown in (a3) and (a4) of Fig. 7. For the boxed regions, our proposed method successfully separates diffuse chromaticity, whereas the conventional methods have hole artifacts. There is no ground truth for the diffuse and specular reflection of dataset. To confirm the performance of our proposed method, the TDP is shown in Fig. 7 (b1-b4). The TDP is less affected by illuminant and contributes to the realistic and clean visual quality of the proposed method. As shown in Fig. 7, the diffuse reflection of the proposed method has similar chromaticity with the TDP.

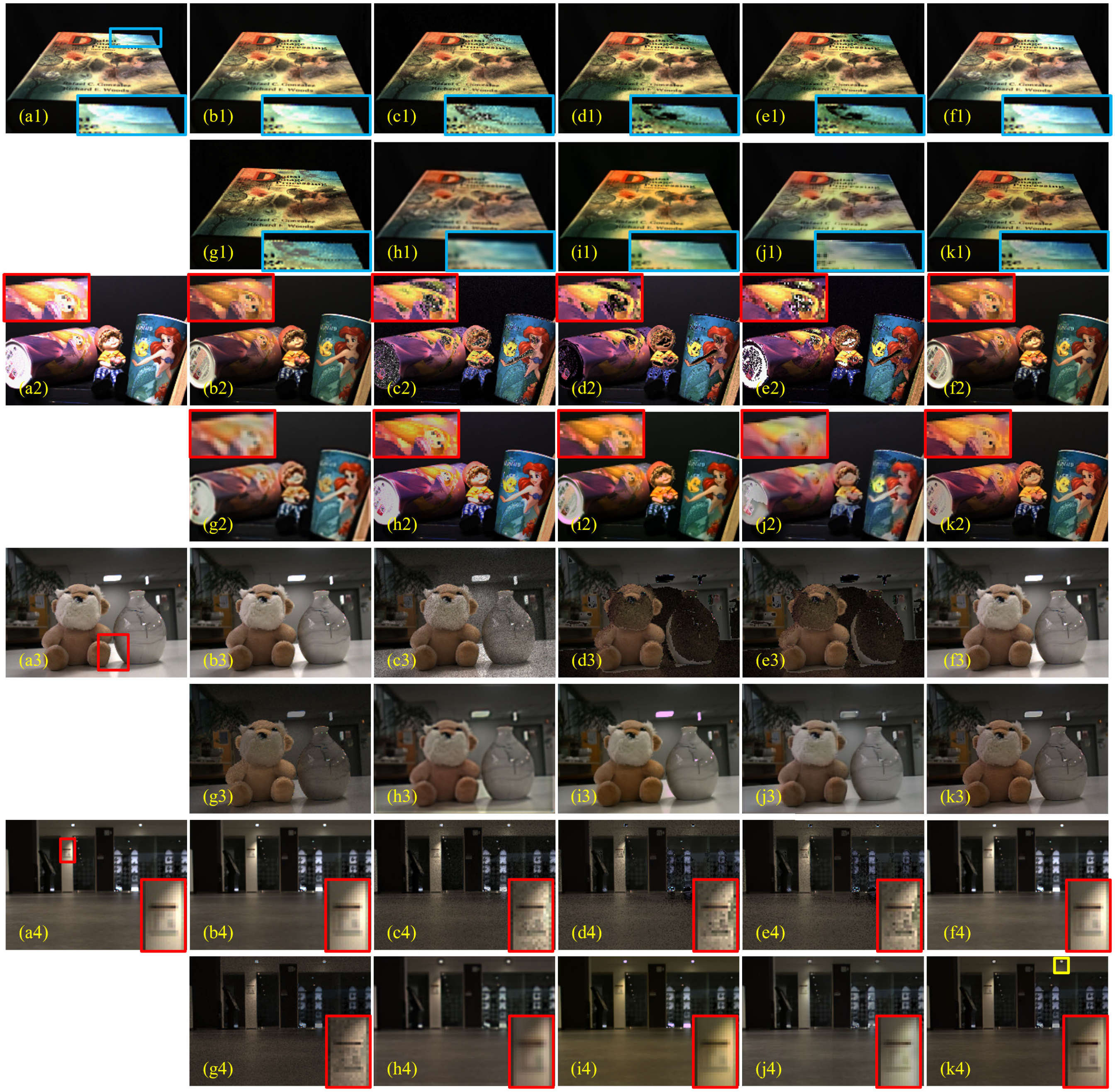

FIGURE 7. **(a1-a4) input image, (b1-b4) TDP, (c-j) the diffuse reflection of (c) Akashi and Okatani [41], (d) Yamamoto and Nakazawa [42], (e) Yang et al. [43], (f) Tsuji [33], (g) Fu et al. [44], (h) JSHDR [39], (i) Yoo et al. [17], (j) JSHDR-Trans [45], (k) the proposed method.**

TABLE 1. **Non-reference image quality assessment metrics comparison. Smaller BRISQUE, NIQE, and PIQUE indicates better image quality.**

	Akashi *et al.* [41]	Yamamoto *et al.* [42]	Yang *et al.* [43]	Tsuji [33]	Fu *et al.* [44]	JSHDR [39]	Yoo *et al.* [17]	JSHDR-Trans [45]	Proposed
BRISQUE↓	46.77	45.42	46.01	46.48	45.64	39.10	47.39	31.90	**30.38**
NIQE↓	30.78	26.55	26.95	26.82	33.26	7.42	23.64	**7.27**	7.43
PIQUE↓	87.37	83.41	84.56	90.37	88.37	77.65	91.20	60.30	**49.74**

The proposed method performs well at reconstructing image details in the diffuse reflection, as demonstrated in (a1) of Fig 7. The image details of the blue boxed region in the input frame are lost because of strong illumination. The proposed method restores the image details by successfully separating the specular component from the input. In contrast,

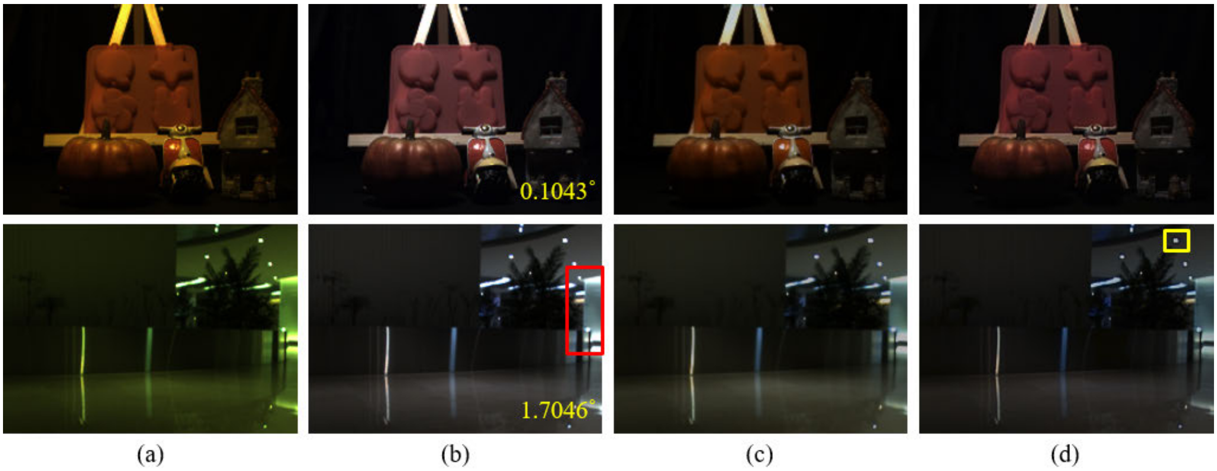

FIGURE 8. (a) color illuminant image, (b) white balanced image with the estimated illuminant of Yoo et al. [17] and its angular error, the diffuse reflection of (c) Yoo et al. [17] and (d) the proposed method.

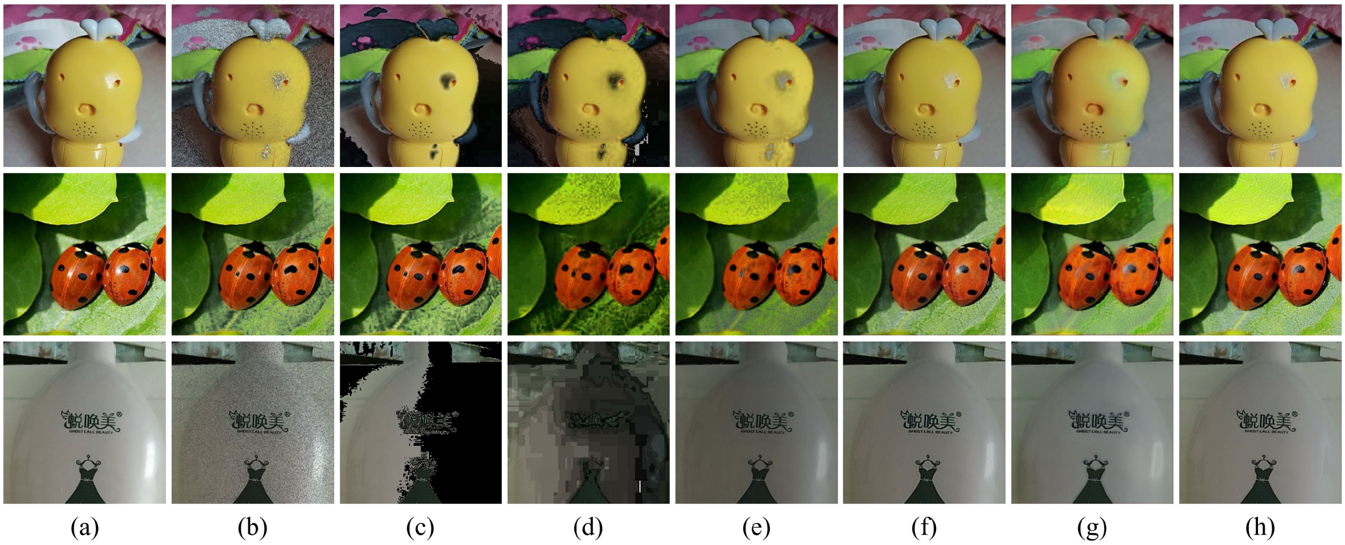

FIGURE 9. Evaluation with WHU-Specular dataset [40]. (a) input image, (b-h) the diffuse reflection of (b) Akashi and Okatani [41], (c) Yamamoto and Nakazawa [42], (d) Yang et al. [43], (e) Fu et al. [44], (f) JSHDR [39], (g) JSHDR-Trans [45], (h) the proposed method.

the conventional methods fail to remove highlights due to their poor intrinsic image decomposition.

Fig. 8 compares the proposed method with Yoo et al. [17] which is a closely related work. In this paper, we assume that the input image has been already white-balanced. For fair comparison, the estimated illuminant of [17] is used for white-balancing in the proposed method. Since [17] exploits the diffuse chromaticity dictionary for modeling the diffuse reflection, it has limitation in representing various diffuse colors and artifacts occur in saturated regions. As shown in the first row of Fig. 8, the colors of the three reddish objects in the proposed method are closely restored to (b), while they are distorted in [17]. Also, the proposed method performs better in separation of diffuse and specular reflection, compared with [17] as shown in the second row.

Since there is no ground truth reflection of the proposed dataset, performance comparison is made with several no-reference image quality metrics. Table 1 shows quantitative comparison with BRISQUE [46], NIQE [47], and PIQUE [48]. The proposed method achieved the best performance for BRISQUE and PIQUE. Although JSHDR-Trans [45] achieved lower NIQE than the proposed method, it suffers from blurred and color distortions as in Fig.7.

The performance comparison is made with the real image datasets, WHU-Specular dataset [40] and SHIQ [39]. It is shown in Fig. 9, Fig. 10 and Table 2. Since WHU-Specular dataset [40] and SHIQ are a single image dataset, the multi-image based methods, Tsuji [33] and Yoo et al. [17] cannot be evaluated. Although the proposed method is trained with multiple frames, the network can be evaluated with

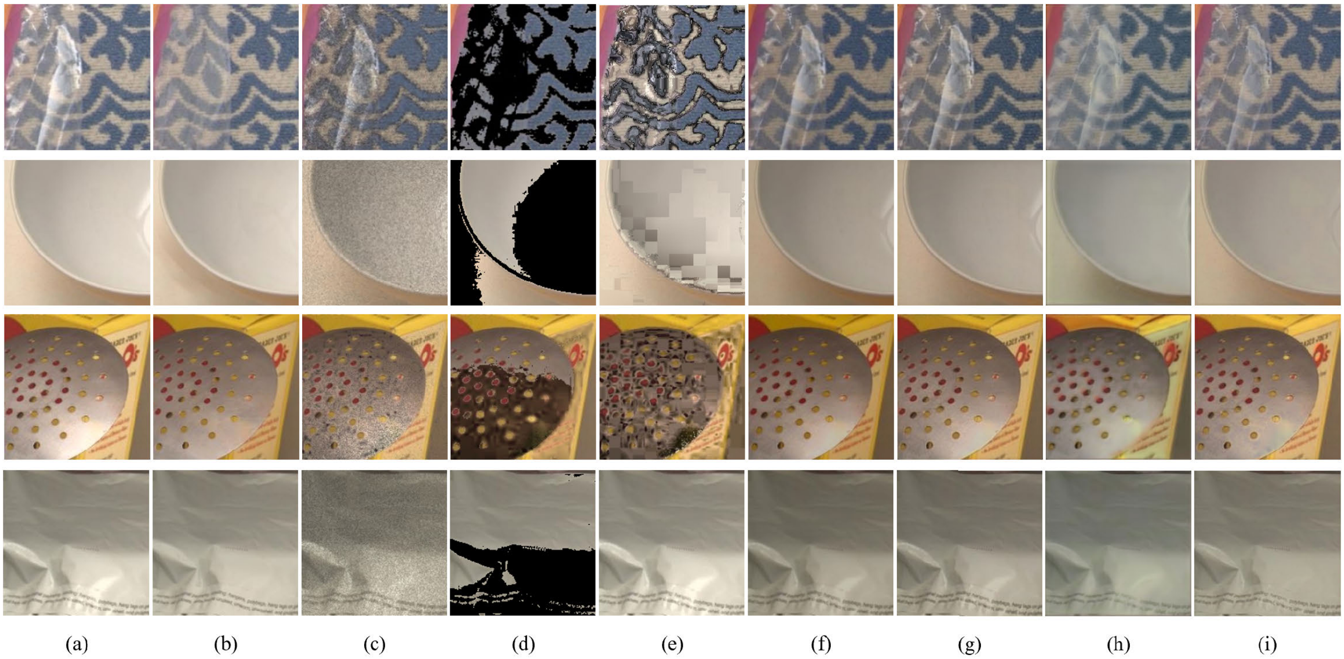

FIGURE 10. Evaluation with SHIQ [39]. (a) input image, (b) ground truth, (c-i) the diffuse reflection of (c) Akashi and Okatani [41], (d) Yamamoto and Nakazawa [42], (e) Yang et al. [43], (f) Fu et al. [44], (g) JSHDR [39], (h) JSHDR-Trans [45], (i) the proposed method.

TABLE 2. PSNR and SSIM comparison for the real dataset, SHIQ [39].

Methods	Akashi *et al.* [41]	Yamamoto *et al.* [42]	Yang *et al.* [43]	Fu *et al.* [44]	JSHDR [39]	JSHDR-Trans [45]	Proposed
PSNR	22.17	12.27	23.20	19.89	22.58	21.56	**24.69**
SSIM	0.80	0.53	0.87	0.94	0.96	0.90	**0.97**

a single frame. However, TDP cannot be generated with a single image dataset, and input image is used instead of TDP. Although TDP is not available for the evaluation, the performance of the proposed method exceeds the conventional methods in both qualitative and quantitative aspects. The proposed method achieved the highest PSNR and SSIM.

C. ABLATION STUDY

Ablation studies were performed to demonstrate the effectiveness of the local modules in the proposed network. Note that the proposed dataset is used for training and test in the ablation studies. First, the TDP and Blur encoders in the lower subnet of Fig. 3 were removed to evaluate the effect of the TDP and its blurred version. As shown in Fig. 11, we can observe the color distortion and the larger hole artifacts around the saturated regions in (b) and (c). In the blue boxed region in Fig. 11, the specularity is significantly reduced by the addition of TDP encoder. It means that TDP plays an important role for performance improvement, and it is the primary contribution of the proposed method. However, the red boxed region, which indicates highly saturated regions, still shows hole artifacts. Since fully saturated regions do not show AC variation, they result in hole artifacts and poor specularity removal. To alleviate these hole artifacts, Blur encoder is also proposed. The green boxed regions that have small hole artifacts caused by saturation are reduced by adding the Blur encoder. The results show the role and importance of the TDP and Blur encoders. The proposed method reconstructs the diffuse chromaticity to some extent in the region, as expected, while the others still suffer from hole artifacts. As shown in the zoomed-in portion in the first row of Fig. 11, the colors of the saturated regions are filled closer to the ground truth in the proposed method, while the others are incorrectly restored, leading to color distortion. It is confirmed that the TDP and its blurred image help reconstruct the diffuse chromaticity on highlight regions.

Second, our proposed method exploits sinusoidal temporal variations as a constraint on the network outputs (diffuse and specular generations). This is reflected in the AC fitting loss, and its usefulness is evaluated. If the AC fitting loss is removed in the loss function, the diffuse component is inaccurately separated and color distortions occur in some regions, as shown in Fig. 12 (c). It is thought that the sinusoidal property plays a strong role in constraining the diffuse component. The smooth losses (eq (13) and eq (17)) that assume the smooth surface and uniform chromaticity are commonly used in highlight removal. To confirm the effect of this assumption, the proposed method is trained without smooth losses (Fig. 12 (e)). The proposed method shows better color recovery on saturated regions.

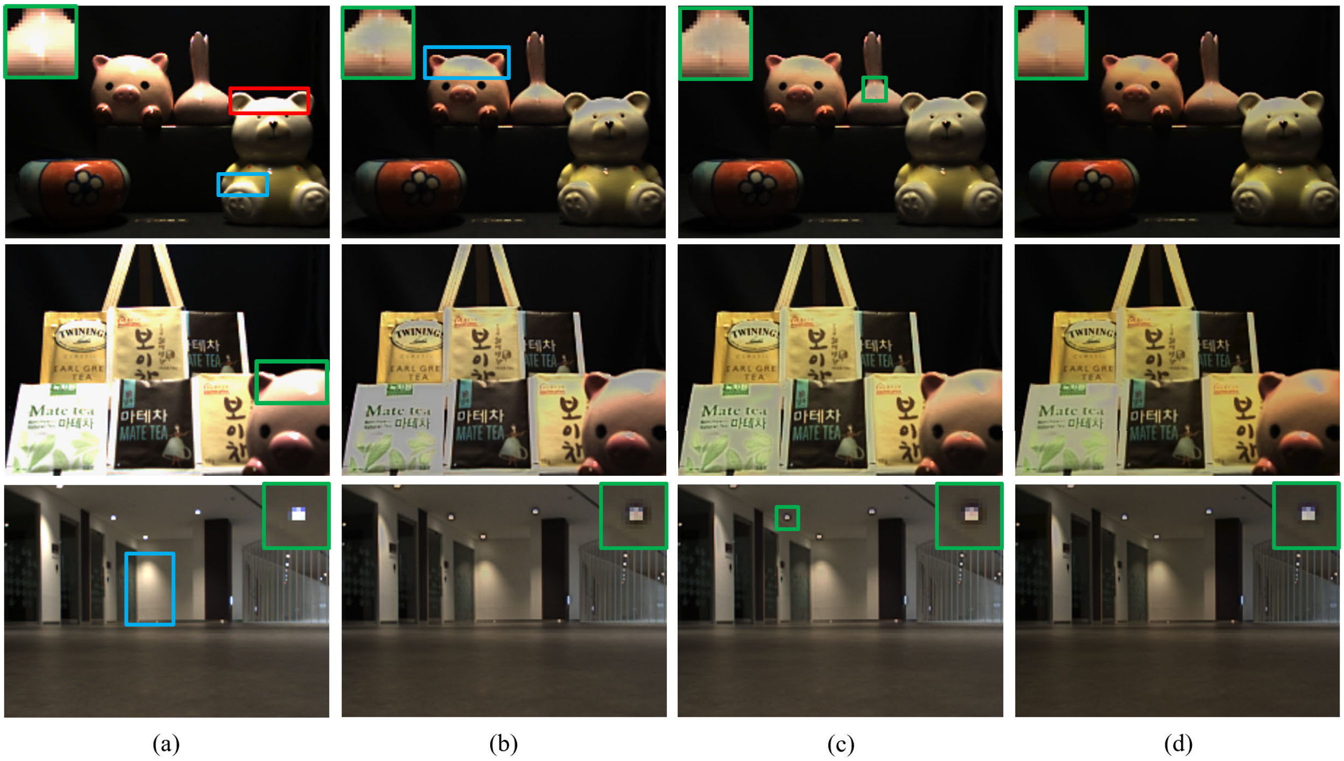

FIGURE 11. (a) input image, (b-d) the estimated diffuse component by using (b) Target encoder, (c) Target and TDP encoder, (d) Target, TDP, and Blur encoder (the proposed method).

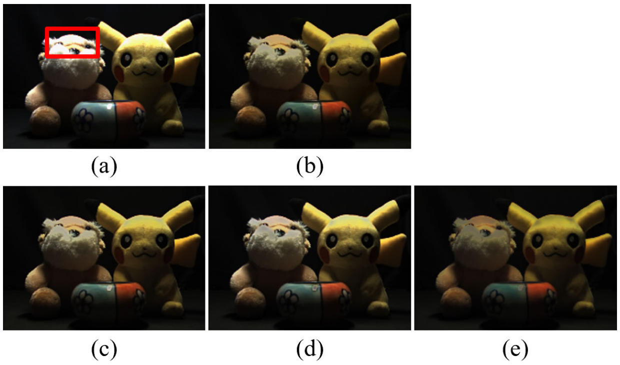

FIGURE 12. (a) input image, (b-d) the diffuse reflection of (b) the proposed method, (c) without AC fitting loss, (d) with three input frames (e) without smooth losses.

Third, to evaluate the effect of the number of the input frames, we conducted an experiment with three frames, and the results are shown in Fig. 12 (d). Decreasing the input frames probably affects the temporal features directly, which can specifically degrade the performance of the separation. This hypothesis was confirmed experimentally. The color distortion of the highlight region appears in the diffuse component, as shown in the red box in Fig. 12 (d). In another analysis, it is assumed that the TDP would be more similar to the true diffuse chromaticity with more input frames. As shown in Fig. 4, the TDP obtained from three frames (d) is less accurate than that obtained from five frames (c). This result demonstrates the importance of an accurate TDP in dichromatic image decomposition.

To confirm the superiority of the TDP, other priors were used for comparison. First, conventional priors in Kim et al. [15] and Shan et al. [12] are evaluated. In [15] and [16], the dark channel prior was used as a pseudo specular component and a pseudo diffuse component was obtained by subtracting the dark channel prior from the input image. Shan et al. [12] proposed highlight removal using a pseudo diffuse component based on [19]. Also, mean, median and max frames among the input video are exploited instead of TDP. They are generated with mean, median, and max operators instead of min operator in eq (3). Instead of the proposed prior, the diffuse priors in Kim et al. [15], Shan et al. [12], mean, median, and max frame and their blurred versions are fed into the proposed network. Fig. 13 shows a performance comparison between the TDP and the other diffuse priors in the proposed network. The conventional diffuse priors [12], [15] have large color differences in strong specular regions and white objects. This leads to severe color distortion of the diffuse component. This shows that the proposed TDP is relatively more useful for recovering the diffuse component. Compared with mean, median and max frames, the proposed TDP showed better recovery performance especially on strong specular region, where the estimation of intrinsic

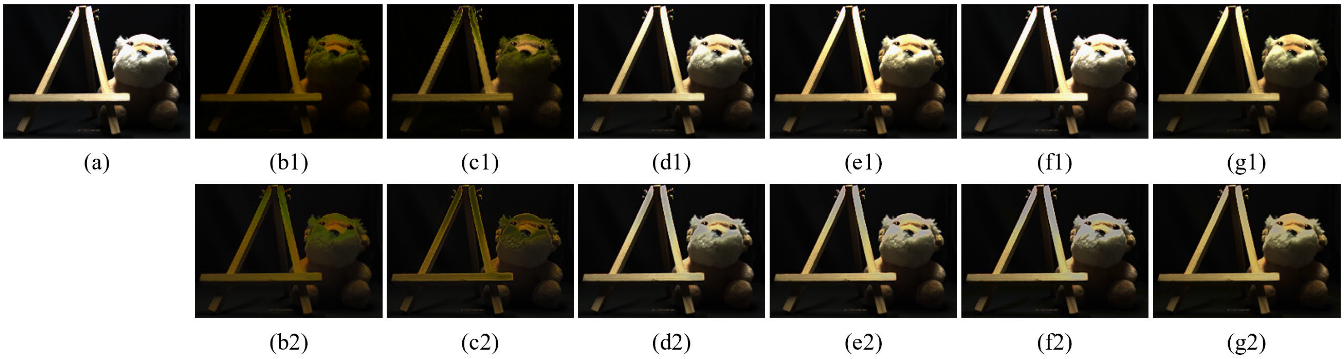

FIGURE 13. (a) input image, (b1-g1) the diffuse prior and (b2-g2) the estimated diffuse component using prior. (b1, b2) Kim et al. [15], (c1, c2) Shan et al. [12], (d1, d2) mean prior, (e1, e2) median prior, (f1, f2) max prior, (g1, g2) TDP.

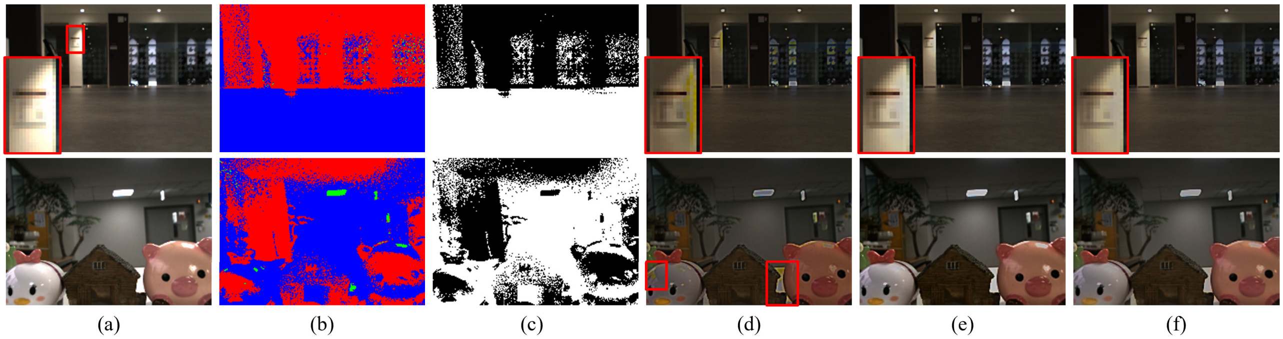

FIGURE 14. (a) input image, (b) phase map, (c) dominant phase (blue) map, and highlight removal result with (d) three phases model, (e) dominant phase model, (f) single phase model (proposed).

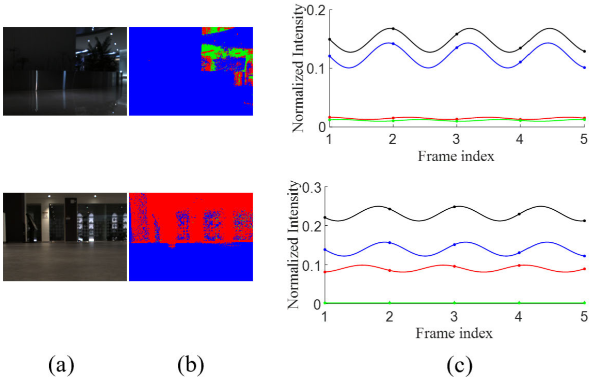

FIGURE 15. (a) input image, (b) phase map, (c) sine curve fit for each phase. Red: $\phi = 0$, Green: $\phi = 2\pi/3$, Blue: $\phi = 4\pi/3$, Black: proposed (single phase model).

chromaticity is very challenging. It indicates that TDP contains more information about inherent chromaticity rather than the other priors.

Since the open scenes contain multiple bulbs whose fluctuations are different, we evaluate the accuracy of the proposed single phase model. Each pixel is fitted with a sine curve, and the phase is modelled by the optimal one among the group $\{0, 2\pi/3 \text{ and } 4\pi/3\}$ as in [9]. This is visualized as the phase map in Fig. 15. The color indicates the optimal phase, and the sine curves fit with average intensity for each phase are shown. As shown in Fig. 15, the optimal phase is spatially varying. The dominant phase (which has the largest amplitude among the sine curves fit with three phases) is chosen, and its AC model is compared with the other two phases and the proposed single phase model. The other two phases models (red and green in Fig. 15) have relatively small sinusoidal variations than the dominant. Thus, the dominant phase is representative in the scene. Also, there exists very marginal phase difference between the proposed and the dominant phase model. This means that the proposed single phase model is quite reasonable. To consider the variations of multiple bulbs, we trained the proposed network by calculating AC fitting loss for both the optimal phases on the phase map and the dominant phases only (no AC loss for the other two phases), and the result are shown in Fig. 14. Considering all three phases (three phases model) produces some artifacts around highlight regions. For weak AC variation with small amplitude, the determined optimal phase tends to be incorrect due to the difficulty of modelling the AC variation. This causes color distortion as shown in the red boxes. Thus, dominant phase model is superior to three phases model. In case of using the only dominant phase (dominant phase model), the variations of the other two phases are not reflected to the AC loss.

V. CONCLUSION

We proposed a deep highlight removal method based on the dichromatic model under AC light sources. It is an ill-posed problem to decompose an image into diffuse and specular components in the dichromatic model. To overcome this challenge, we propose to utilize the TDP, which corresponds to a pseudo diffuse component. Unlike the existing methods with specular priors, the proposed diffuse prior contributes to reducing the hole artifacts in highlight regions. In addition, the AC variations of incident light can be used as a strong constraint on both the diffuse and specular outputs of the network, resulting in more accurate image decomposition. The experimental results show that incorporating the temporal feature (the TDP and the AC fitting loss) can improve the separation capability of the dichromatic components and alleviate hole artifacts, outperforming state-of-the-art methods in terms of dichromatic decomposition and highlight removal.

The proposed method exploits the temporal variation of the AC illuminants. Thus, its performance is fundamentally limited, depending on the available temporal variation. For instance, signal variation is hardly observed in strongly saturated regions, leading to poor decomposition of specular and diffuse components. In addition, small illuminant regions are errorneously treated as strong specularities and the network tries to fill them. They are filled with chromaticity similar to surrounding pixels like real saturation region, as confirmed in the yellow boxed region of Fig. 7 and Fig. 8. These limitations have been also quite challenging to the existing methods, and further study is required.

Our previous work [17] is successful in estimating the dichromatic model which is very challenging. It, however, still shows some limitations such as color distortion and highlight removal performance in saturated regions as illustrated in Fig. 8. In this paper, the model estimation problem is alleviated by image decomposition and the assumption of white illuminant, primarily concentrating on highlight removal. Thanks to these two points, we could obtain more accurate diffuse reflection. It would be interesting to perform both highlight removal and color constancy via image decomposition without the assumption of white illuminant.

REFERENCES

[1] S. A. Shafer, "Using color to separate reflection components," *Color Res. Appl.*, vol. 10, no. 4, pp. 210–218, 1985.

[2] Y. Wang, F. Fu, F. Lai, W. Xu, J. Shi, and J. Wang, "Efficient road specular reflection removal based on gradient properties," *Multimedia Tools Appl.*, vol. 77, no. 23, pp. 30615–30631, Dec. 2018.

[3] S. Hou, C. Wang, W. Quan, J. Jiang, and D.-M. Yan, "Text-aware single image specular highlight removal," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*. Cham, Switzerland: Springer, 2021, pp. 115–127.

[4] S. Alsaleh, "Specular reflection removal for endoscopic data with applications to medical robotics," Ph.D. dissertation, School Eng. Appl. Sci., George Washington Univ., Washington, DC, USA, 2020.

[5] G. L. Oliveira, W. Burgard, and T. Brox, "Efficient deep models for monocular road segmentation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 4885–4891.

[6] V. Christlein, C. Riess, E. Angelopoulou, G. Evangelopoulos, and I. Kakadiaris, "The impact of specular highlights on 3D-2D face recognition," *Proc. SPIE*, vol. 8712, May 2013, Art. no. 87120T.

[7] I. Ahmed B. K., G. Ahmed, A. Saleem, and S. Ahmed, "Enhancement of the iris-texture by removal of specular reflections for an accurate iris segmentation," in *Proc. 2nd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2020, pp. 586–589.

[8] J.-S. Yoo and J.-O. Kim, "Dichromatic model based temporal color constancy for AC light sources," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12329–12338.

[9] M. Sheinin, Y. Y. Schechner, and K. N. Kutulakos, "Computational imaging on the electric grid," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6437–6446.

[10] J.-S. Yoo, K.-K. Lee, C.-H. Lee, J.-M. Seo, and J.-O. Kim, "Deep spatio-temporal illuminant estimation under time-varying AC lights," *IEEE Access*, vol. 10, pp. 15528–15538, 2022.

[11] Q. Xu and L. Zhou, "A specular removal algorithm based on improved specular-free image and chromaticity analysis," in *Proc. 13th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2020, pp. 104–109.

[12] W. Shan, C. Xu, and B. Feng, "Image highlight removal based on double edge-preserving filter," in *Proc. IEEE 5th Int. Conf. Signal Image Process. (ICSIP)*, Oct. 2020, pp. 263–268.

[13] W. Xia, E. C. S. Chen, S. E. Pautler, and T. M. Peters, "A global optimization method for specular highlight removal from a single image," *IEEE Access*, vol. 7, pp. 125976–125990, 2019.

[14] M. Arnold, A. Ghosh, S. Ameling, and G. Lacey, "Automatic segmentation and inpainting of specular highlights for endoscopic imaging," *EURASIP J. Image Video Process.*, vol. 2010, no. 9, pp. 1–12, 2010.

[15] H. Kim, H. Jin, S. Hadap, and I. Kweon, "Specular reflection separation using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1460–1467.

[16] V. S. Ramos, L. G. D. Q. Silveira Junior, and L. F. D. Q. Silveira, "Single image highlight removal for real-time image processing pipelines," *IEEE Access*, vol. 8, pp. 3240–3254, 2020.

[17] J.-S. Yoo, C.-H. Lee, and J.-O. Kim, "Deep dichromatic model estimation under AC light sources," *IEEE Trans. Image Process.*, vol. 30, pp. 7064–7073, 2021.

[18] R. T. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse-intensity chromaticity space," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 21, no. 3, pp. 321–334, 2004.

[19] R. T. Tan and K. Ikeuchi, "Separating reflection components of textured surfaces using a single image," in *Digitally Archiving Cultural Objects*. Cham, Switzerland: Springer, 2008, pp. 353–384.

[20] D. An, J. Suo, X. Ji, H. Wang, and Q. Dai, "Fast and high quality highlight removal from a single image," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5441–5454, Nov. 2016.

[21] J. Yang, L. Liu, and S. Z. Li, "Separating specular and diffuse reflection components in the HSI color space," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 891–898.

[22] H.-L. Shen and Z.-H. Zheng, "Real-time highlight removal using intensity ratio," *Appl. Opt.*, vol. 52, no. 19, pp. 4483–4493, Jul. 2013.

[23] A. Rodriguez-Sanchez, D. Chea, G. Azzopardi, and S. Stabinger, "A deep learning approach for detecting and correcting highlights in endoscopic images," in *Proc. 7th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Nov. 2017, pp. 1–6.

[24] Y.-C. Chang, C.-N. Lu, C.-C. Cheng, and W.-C. Chiu, "Single image reflection removal with edge guidance, reflection classifier, and recurrent decomposition," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2033–2042.

[25] S. K. Nayar, X.-S. Fang, and T. Boult, "Separation of reflection components using color and polarization," *Int. J. Comput. Vis.*, vol. 21, no. 3, pp. 163–186, Feb. 1997.

[26] S. Umeyama and G. Godin, "Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 639–647, May 2004.

[27] M. O'Toole, S. Achar, S. G. Narasimhan, and K. N. Kutulakos, "Homogeneous codes for energy-efficient illumination and imaging," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–13, Jul. 2015.

[28] R. Feris, R. Raskar, K.-H. Tan, and M. Turk, "Specular reflection reduction with multi-flash imaging," in *Proc. 17th Brazilian Symp. Comput. Graph. Image Process.*, Oct. 2004, pp. 316–321.

[29] R. Nakao, Y. Iwahori, Y. Adachi, A. Wang, M. K. Bhuyan, and B. Kijsirikul, "Detecting and removing specular reflectance components based on image linearization," *Proc. Comput. Sci.*, vol. 159, pp. 1576–1583, Jan. 2019.

[30] S. M. Z. A. Shah, S. Marshall, and P. Murray, "Removal of specular reflections from image sequences using feature correspondences," *Mach. Vis. Appl.*, vol. 28, nos. 3–4, pp. 409–420, 2017.

[31] S. Lin, Y. Li, S. B. Kang, X. Tong, and H.-Y. Shum, "Diffuse-specular separation and depth recovery from image sequences," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2002, pp. 210–224.

[32] S. W. Lee and R. Bajcsy, "Detection of specularity using color and multiple views," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 1992, pp. 99–114.

[33] T. Tsuji, "Specular reflection removal on high-speed camera for robot vision," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 1542–1547.

[34] V. Prinet, M. Werman, and D. Lischinski, "Specular highlight enhancement from video sequences," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 558–562.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[36] G. Eilertsen, J. Kronander, G. Denes, R. Mantiuk, and J. Urger, "HDR image reconstruction from a single exposure using deep CNNs," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–15, 2017.

[37] J. Guo, Z. Zhou, and L. Wang, "Single image highlight removal with a sparse and low-rank reflection model," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 268–283.

[38] Y. Liu, Z. Yuan, N. Zheng, and Y. Wu, "Saturation-preserving specular reflection separation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3725–3733.

[39] G. Fu, Q. Zhang, L. Zhu, P. Li, and C. Xiao, "A multi-task network for joint specular highlight detection and removal," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7752–7761.

[40] G. Fu, Q. Zhang, Q. Lin, L. Zhu, and C. Xiao, "Learning to detect specular highlights from real-world images," in *Proc. ACM Multimedia*, 2020, pp. 1873–1881.

[41] Y. Akashi and T. Okatani, "Separation of reflection components by sparse non-negative matrix factorization," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 611–625.

[42] T. Yamamoto and A. Nakazawa, "General improvement method of specular component separation using high-emphasis filter and similarity function," *ITE Trans. Media Technol. Appl.*, vol. 7, no. 2, pp. 92–102, 2019.

[43] Q. Yang, J. Tang, and N. Ahuja, "Efficient and robust specular highlight removal," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1304–1311, Jun. 2015.

[44] G. Fu, Q. Zhang, C. Song, Q. Lin, and C. Xiao, "Specular highlight removal for real-world images," *Comput. Graph. Forum*, vol. 38, no. 7, pp. 253–263, 2019.

[45] Z. Wu, J. Guo, C. Zhuang, J. Xiao, D.-M. Yan, and X. Zhang, "Joint specular highlight detection and removal in single images via unet-transformer," *Comput. Vis. Media*, vol. 9, no. 1, pp. 141–154, Mar. 2023.

[46] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[47] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[48] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *Proc. 21st Nat. Conf. Commun. (NCC)*, Feb. 2015, pp. 1–6.

JEONG-WON HA received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 2021, where she is currently pursuing the M.S. degree in electrical engineering. Her research interests include color constancy, dichromatic model, and intrinsic image decomposition.

KANG-KYU LEE received the B.S. degree in electronic engineering and the Ph.D. degree in electrical engineering from Korea University, Seoul, South Korea, in 2015 and 2022, respectively. His current research interests include the area of intrinsic image decomposition, multi-exposure fusion, and color constancy.

JUN-SANG YOO received the B.S. degree and the Ph.D. degree in electrical engineering from the School of Electrical Engineering, Korea University, Seoul, South Korea, in 2015 and 2021, respectively. He joined the Samsung Advanced Institute of Technology, Gyeonggi-do, South Korea, in 2021, where he is currently a Staff Researcher. His current research interests include sparse representations, super-resolution, and color constancy.

JONG-OK KIM (Member, IEEE) received the B.S. and M.S. degrees in electronic engineering from Korea University, Seoul, South Korea, in 1994 and 2000, respectively, and the Ph.D. degree in information networking from Osaka University, Osaka, Japan, in 2006. From 1995 to 1998, he was an Officer with Korea Air Force. From 2000 to 2003, he was with the SK Telecom Research and Development Center and Mcubeworks Inc., South Korea, where he was involved in research and development on mobile multimedia systems. From 2006 to 2009, he was a Researcher with the Advanced Telecommunication Research Institute International (ATR), Kyoto, Japan. He joined Korea University, in 2009, where he is currently a Professor. His current research interests include image processing, computer vision, and intelligent media systems. He was a recipient of the Japanese Government Scholarship, from 2003 to 2006.