IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 April 2023, accepted 23 April 2023, date of publication 27 April 2023, date of current version 3 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3271114

RESEARCH ARTICLE

Joint Deep Estimation of Intrinsic and Dichromatic Image Decomposition

JEONG-WON HA, KANG-KYU LEE, AND JONG-OK KIM, (Member, IEEE)
School of Electrical Engineering, Korea University, Seoul 02841, South Korea

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

This work was supported by the National Research Foundation of Korea (NRF) funded by the Korean Government [Ministry of Science and ICT (MSIT)] under Grant 2023R1A2C2003554.

ABSTRACT This paper proposes an image formation model that jointly combines dichromatic and intrinsic image decomposition models. The two decomposition models analyze image formation process from a different perspective, and they can be combined synergistically. It is confirmed that the proposed method performs better than the individual decomposition. The joint estimation and the study of the decomposition order ('intrinsic + dichromatic' or 'dichromatic + intrinsic') are the first attempt to the best of our knowledge. It was confirmed that the proposed 'intrinsic + dichromatic' is more optimal through experimental evaluations. We also exploit the temporal property of AC light sources, which can further improve the decomposition performance. The experimental results show that the proposed model can make an accurate image decomposition and achieve a remarkable color constancy performance.

INDEX TERMS Intrinsic image decomposition, dichromatic model, color constancy, AC light, high-speed video.

I. INTRODUCTION

There are two image decomposition models to inversely find out the process of color image formation from the observed image. They are the intrinsic image decomposition and dichromatic models which describe the properties of surface reflection. The former assumes that the image can be expressed as the product of reflectance (R) and illumination (L):

$$I = R \otimes L, \tag{1}$$

where $\otimes$ is pixel wise multiplication [1]. The latter assumes that the reflected light is the sum of diffuse (D) and specular (S) reflection [2]:

$$I = D + S. \tag{2}$$

The image formation of two model is described in Fig. 1. They are one of the most fundamental tasks in computer vision and graphics communities. The two models are closely related to each other in that they commonly deal with surface reflection in the image formation process. The reflectance component of the intrinsic model and the diffuse component of the dichromatic model represent the inherent color of an object in common. The other illumination and specular components are dependent on illumination environment.

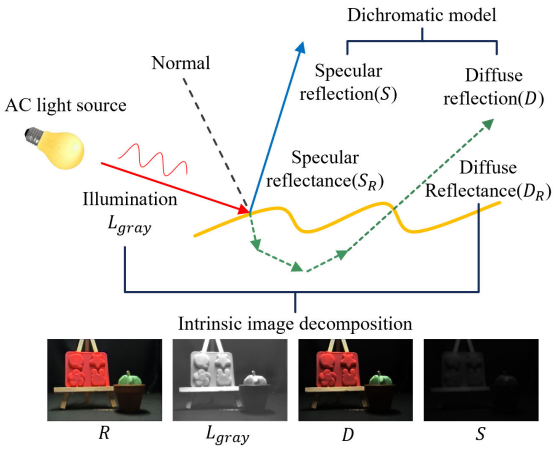

FIGURE 1. Image formation scenario of the proposed model.

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

Because the reflected properties between objects and illumination are described well with these models, they have been popularly exploited for image quality enhancement. The dichromatic model is useful for color constancy [3], [4], [5] and highlight removal [6], while the intrinsic model is for low-light enhancement [7] and relighting [8], [9]. However, the two decomposition models have fundamental limitations for unveiling image formation process thoroughly. The existing intrinsic model assumes Lambertian surface, and thus, it works poorly for real scenes with highlight or saturation. The dichromatic model focuses on surface reflection only. It has difficulty in obtaining object intrinsic characteristics. For shading regions, it is hard to recover chromaticity unlike the intrinsic model as shown in Fig. 1. In other words, they look into image formation process from a different perspective, and can be combined synergistically for understanding its details.

Therefore, we propose to jointly learn the dichromatic and intrinsic models in order to accurately separate the reflection components from the observed image. This enables us to deeply understand the details of color image formation process. The proposed network learns the two models together, and it decomposes an input image in two ways simultaneously. The simultaneous learning can further improve the accuracy of the model estimation rather than the individual learning because the two inverse problems are highly ill-posed.

The original intrinsic model often approximates the reflection component to diffuse reflection, and neglects specular reflection. Recently, it is extended by considering the specular component as an additive residue term [10], [11]. The extended model first removes highlight, and is followed by the conventional intrinsic decomposition. However, in the proposed method, intrinsic decomposition is first made, and the separated reflectance is further decomposed into the diffuse and specular components. This logically follows the imaging process in sequence where incident light is reflected on surface in two ways (diffuse and specular) [3], [12]. Our work thoroughly studies the order of the decompositions (i.e., 'intrinsic + dichromatic' or 'dichromatic + intrinsic'). It was confirmed that the proposed 'intrinsic + dichromatic' is more optimal through experimental evaluations. Also, it was found that the gain of the joint decomposition is superior to individual decomposition. The joint estimation and the study of the decomposition order are the first attempt to the best of our knowledge.

Estimating the two models from a single image is a highly ill-posed problem. Conventional methods [7], [13], [14], [15], [16], [17], [18], [19] assume white-light environment for simplicity, but the proposed method also attempts to estimate illuminant chromaticity which is more general and practical. As reported in the previous works, the sinusoidal variation of AC (alternative current) powered light sources can be an important clue for illumination chromaticity estimation and image decomposition [5], [6], [20]. However, the previous studies simply exploit this prior as the cost of the deep network. In the proposed method, to exploit the temporal feature more efficiently, knowledge distillation [21], [22], [23], [24], [25] is used. The feature of a teacher network that learns temporal feature is transferred to the student network that learns image decomposition. By leveraging the AC variation, the proposed network showed better image decomposition performance.

II. RELATED WORKS

A. INTRINSIC IMAGE DECOMPOSITION

Although the intrinsic image decomposition has been extended to $I = R \otimes L + S$ (Lambertian shading L, reflectance R, and specularity S) recently [10], [11], [26], it decomposes an image into two components (R and L) by ignoring specularity for simplicity in many previous works [27], [28], [29], [30]. Therefore, intrinsic image decomposition in this paper, means separation into reflectance and illumination. Intrinsic image decomposition is an ill-posed problem, and some priors have been studied in conventional methods. One of the priors is the Retinex model [31], which has been widely used. Recently, deep learning based intrinsic decomposition has been popularly studied, and it achieves a superior performance. However, these models are trained in a supervised manner and require ground truth of intrinsic decomposition for training [32], [33], [34], [35], [36]. Because they use synthetic images which are far from real scenes, there exists a fundamental limitation from a perspective of practical applications. Although the human-labeled real-world dataset (IIW [37] and SAW [38]) were created, they have sparse annotation and it is difficult to collect annotation in a large scale [19]. On the other hand, there are several studies that utilize time-lapse sequences [18], [39]. They assume varying illumination and constant reflectance in a scene. However, they require a large number of images and often fail in indoor scenes, because the assumption works primarily for outdoors [39]. Also, the conventional intrinsic model is commonly inadequate for highlight, strong shadow regions, and colored illumination [39]. No decomposition between specular and diffuse reflection makes the model work poorly for strongly-illuminated objects, leading to distorted visual quality (Fig. 8 (b - g)). The proposed method attempts to overcome these challenges by jointly combining intrinsic decomposition, dichromatic decomposition and color constancy tasks in a cooperative way.

B. DICHROMATIC MODEL BASED DECOMPOSITION

Dichromatic model based decomposition separates diffuse and specular reflection from an input image, and due to its ill-posed property, various priors have been investigated. The thresholded Value in the HSV color space [40], [41] and minimum intensity among the RGB channels for each pixel [42], [43] were explored as the prior of specular reflection. Several methods use color dictionary to recover the chromaticity of

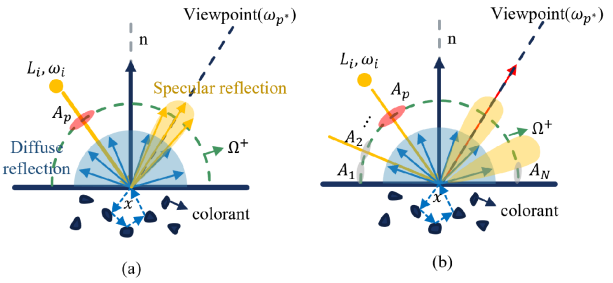

FIGURE 2. Reflection modeling of (a) single ray and (b) *N* multiple segments.

diffuse reflection, based on the assumption that diffuse color can be expressed as a linear combination of some representative colors [44], [45], [46]. There are previous studies that use multiple images, which are captured in different viewpoints or directions of a light source [47], [48], [49], [50], [51], while our proposed method has no constraint on the position of a camera and a light source. In these multiple images based methods including the proposed, constant diffuse chromaticity becomes a useful prior. While many conventional methods rely on only spatial features of images, the proposed method utilizes both temporal and spatial features which are obtained from natural high-speed video frames.

There are few studies that have exploited temporal features for dichromatic decomposition. The works in [6], [20], and [52] use the intensity fluctuation of AC lights captured in high speed video. Tsuji [20] assumes the linear relation between minimum and maximum luminance of a high-speed video. Yoo et al. [6] proposed a deep network that estimates all parameters of the dichromatic model including illuminant color. Ha et al. [52] introduced a new temporal dark prior for dichromatic model based decomposition. However, Tsuji [20] still showed color distortion on strong highlight regions which is a common problem of highlight removal. Also, Yoo et al. [6] have limitation on diffuse color recovery as other dictionary based methods do.

C. MODEL IMPROVEMENT

Because intrinsic decomposition commonly assumes Lambertian surface, specularity is hardly considered [10], [11], [26]. Cheng et al. [26] simply assumes that diffuse reflection is dominant over specular reflection as follows:

$$I \approx D = R \otimes L \tag{3}$$

Then, a couple of previous works [10], [11] extended the intrinsic decomposition to accommodate specularity for highlight removal, and (1) is extended by

$$I = R \otimes L + S. \tag{4}$$

Modeling as (4) is similar to the proposed method in that it combines the intrinsic model with the dichromatic one. However, it just treats the separation of the specular component as the preprocess of intrinsic decomposition for Lambertian surface input, while we model the image formation process by closely combining both models.

III. THE PROPOSED METHOD

A. THE PROPOSED IMAGE FORMATION MODEL

The total reflected light on a surface point x, observed at viewpoint ω_{p^*} under incident light $L_i(x, \omega_i)$ whose incident direction is ω_i can be expressed as follows:

$$L(x, \omega_{p^*}) = \int_{\Omega^+} f_r(x, \omega_i, \omega_{p^*}) L_i(x, \omega_i)(\omega_i \cdot n) d\omega_i \tag{5}$$

where Ω^+ means positive hemisphere to sample the whole incident light and n is a surface normal. In (5), $f_r(x, \omega_i, \omega_{p^*})$ is bidirectional reflectance distribution function (BRDF), which is the fraction of reflected radiance observed from a direction ω_{p^*} for each incident direction ω_i. If non-Lambertian is considered more generally, BRDF in (5) is extended to the sum of a diffuse isotropic lobe (f_d) and a specular lobe (f_s) [53]:

$$f_{NL}(x, \omega_i, \omega_{p^*}) = f_d(x, \omega_i, \omega_{p^*}) + f_s(x, \omega_i, \omega_{p^*}). \tag{6}$$

To derive the proposed image decomposition model, assume for a single ray environment as Fig. 2 (a), first. Under non-Lambertian assumption, the reflected radiance caused by a single ray that comes through the i^{th} segment of Ω^+ (denoted by A_i) can be expressed as:

$$\begin{aligned} &L(x, \omega_{p^*}, A_i) \\ &= \int_{A_i} \left\{ f_d(x, \omega_i, \omega_{p^*}) + f_s(x, \omega_i, \omega_{p^*}) \right\} L_i(x, \omega_i)(\omega_i \cdot n) d\omega_i \end{aligned} \tag{7}$$

Diffuse reflection does not depend on the incident direction and f_d has a constant value α_d. Also, the intensity of specular observed at ω_{p^*} can be assumed as constant reflectance ($\alpha_{s,i}$) in a single ray environment. Then, the reflected radiance in (7) can be re-expressed as:

$$\begin{aligned} &L(x, \omega_{p^*}, A_i) \\ &= \alpha_d \int_{A_i} L_i(x, \omega_i)(\omega_i \cdot n) d\omega_i + \alpha_{s,i} \int_{A_i} L_i(x, \omega_i)(\omega_i \cdot n) d\omega_i \\ &= (\alpha_d + \alpha_{s,i}) L_{A_i} \end{aligned} \tag{8}$$

where L_{A_i} is total amount of light incident on A_i. Because of the directional property of the specular reflection, the specular reflection observed at ω_{p^*} is generated by incident light that comes through the small area A_p. Therefore, the specular reflectance by incident direction in A_p is assumed as constant (α_s) with respect to ω_i, and 0 for the other directions:

$$\alpha_{s,i} = \begin{cases} \alpha_s, & \omega_i \in A_p \\ 0, & \omega_i \notin A_p \end{cases} \tag{9}$$

For N multiple incident rays (in Fig. 2 (b)), each ray generates diffuse and specular reflection. The total reflected light is

expressed as follows:

$$L(x, \omega_{p^*}) = \sum_{i=1}^{N} L(x, \omega_{p^*}, A_i) \tag{10}$$

$$= \sum_{i=1}^{N} (\alpha_d + \alpha_{s,i}) L_{A_i} \tag{11}$$

Then, by the relation of $\alpha_{s,i}$ and ω_i in (9):

$$L(x, \omega_{p^*}) = (\alpha_d + k\alpha_s) L_t \tag{12}$$

where k is the ratio of incident light between A_p and positive hemisphere (Ω^+), and L_t is total illumination. From (12), we derive the proposed joint decomposition model as follows:

$$I = (D_R + S_R) \otimes L. \tag{13}$$

With the proposed model, the image (I) is decomposed as diffuse reflectance (D_R), specular reflectance (S_R), and illumination (L). The specular reflectance is given by $k\alpha_s$ that means the ratio between specular reflection and whole incident light. So, the specular reflectance is highly affected by the direction of viewpoint, while the diffuse reflectance does not depend on the direction. Reflectance in the conventional intrinsic image decomposition model under Lambertian assumption corresponds to the diffuse reflectance of the proposed model. Since the intrinsic model assumes for the Lambertian reflection, the specular term is ignored. In the proposed decomposition model, the specular term is considered as specular reflectance which means the ratio of illumination and specular reflection.

So many conventional works have dealt with the decomposition of imaging formation process, which is a crucial part for high-quality imaging. The intrinsic model mainly describes the reflective phenomenon of incident light, while the dichromatic model further analyzes the reflectance into diffuse and specular reflection. The previous improved model in (4) primarily concentrates on intrinsic decomposition, and specular reflection is just added to intrinsic decomposition. However, we attempt to estimate both models simultaneously in a single deep network, targeting at improving the decomposition accuracy better than the individual model estimation. Following the order of imaging process, intrinsic image decomposition is first made, and the resulting reflectance component is further separated into the diffuse and specular components based on the dichromatic model. This sequence of the image decompositions is actually equal to the reflection flow of the incident light for image formation. The integrated model is cooperatively learned within a single deep network for more accurate decomposition.

B. NETWORK STRUCTURE

Fig. 3 shows the overall network structure of the proposed method. The proposed network consists of a Temporal Feature Network (Teacher, TF-Net) and Image Decomposition Network (Student, ID-Net). The ID-Net consists of the Illumination and Reflectance subnets that learn the features of illumination and reflectance, respectively. The subnets adopt a convolutional auto-encoder structure based on VGG16 as [54]. In the most conventional teacher-student learning studies [21], [22], [23], [24] for the network compression, the student network is a lightweight version of the teacher network. However, the proposed method treats the Teacher network as a temporal feature extractor, and VGG16 based auto-encoder is used identically to the ID-Net.

TF-Net learns temporal feature by estimating AC fitting map, M_{AC}, generated with high-speed frames. The intensity variation under AC light source can be modeled as a sine curve [5], [6], [55]. Fig. 4 shows the estimated sine curves of highly illuminated region (red) and low illuminated region (blue). The more the regions are affected by the AC light sources, the larger intensity variation is observed. Therefore, we generated AC fitting map with amplitude of each pixel variation, and it reflects the effect of illumination. Fig. 5 shows the examples of input video and its M_{AC}. By training TF-Net to estimate the AC fitting map with N frames of high-speed video, it can learn the temporal variation of the incident light. By transferring these features to the ID-Net, it is expected that the temporal feature can be extracted more efficiently than just reflecting it to the cost only.

The proposed method exploits high speed video as an input, and estimates illuminant chromaticity, the achromatic illumination component, and the specular and diffuse components (corresponding to reflectance in (1)). The input image I_t is t^{th} frame of input video, and every frame of input video is sequentially fed into the Illumination-encoder and Reflectance-encoder. N is the number of frames of the input video. For t^{th} input frame, illumination $L_{gray,t}$ and its chromaticity Γ_t, the specular component of reflectance $S_{R,t}$, and the diffuse component of reflectance $D_{R,t}$ are generated through the proposed network.

With the Illumination subnet, the illumination $L_{gray,t}$ and its chromaticity are estimated. Motivated from FC4 [56], the proposed method estimates illuminant as a weighted sum of local illuminants and its confidence map. The Illuminant estimation decoder generates 4 channels of output that represent local illuminant and confidence map with 1/8 resolution of the input. Note that the truncated decoder is used for the Illuminant estimation decoder in order to efficiently generate a single illuminant RGB. Recall that the proposed method considers chromatic illumination. Since the chromaticity of illumination probably leads to inaccurate prediction of reflectance color [39], the input image is white-balanced with the predicted illuminant, I_{wb}, and then, it is put into the Reflectance subnet. The Reflectance subnet generates diffuse and specular components separated from the reflectance. As illustrated in Fig. 3, the Reflectance decoder outputs 32 channels of features (F), which go to the two different convolutional blocks. The concatenation of F and the white-balanced input goes through convolutional blocks, leading to the estimation of the specular component, $S_{R,t}$. Then, the concatenation of F, $S_{R,t}$, and I_{wb} are fed into another convolutional blocks, which generates the diffuse component, $D_{R,t}$. The prediction of $S_{R,t}$ is followed

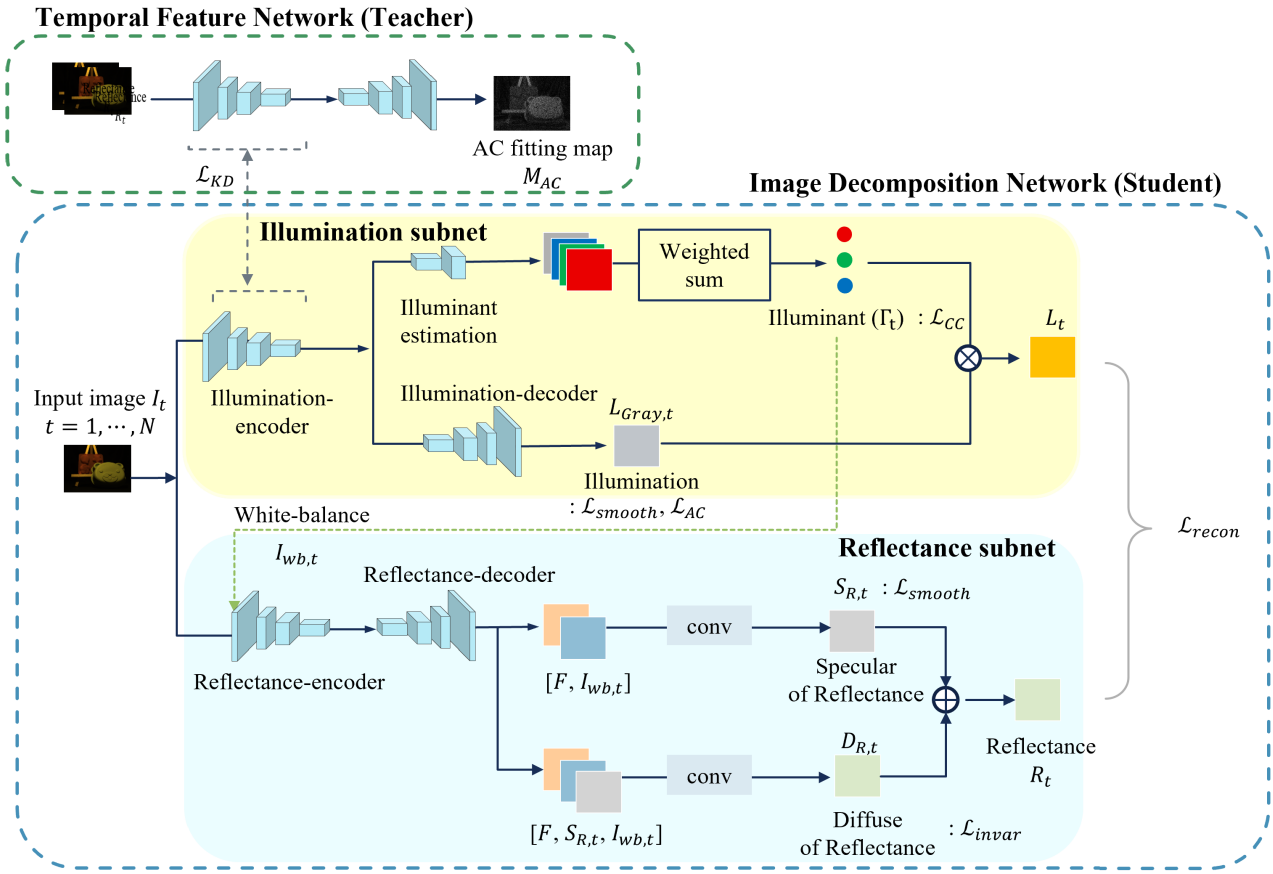

FIGURE 3. Overall architecture of the proposed network. The proposed network consists of Temporal Feature Network (Teacher, TF-Net) and Image Decomposition Network (Student, ID-Net) that estimates illumination, specular and diffuse component of reflectance.

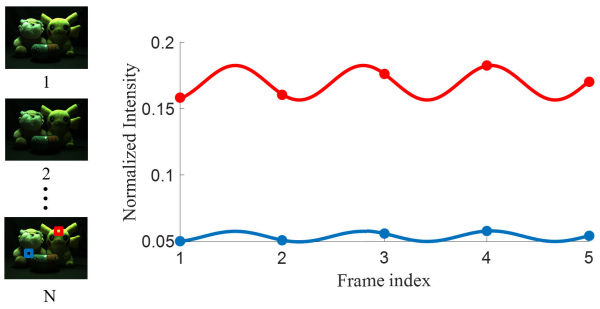

FIGURE 4. AC variations of highly (red) and low (blue) illuminated regions.

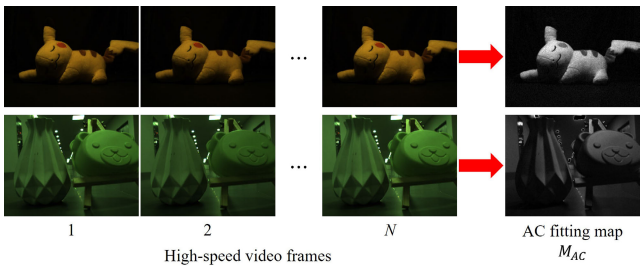

FIGURE 5. Input frames of high-speed video and their AC fitting map generated with amplitude of sinusoidal variation.

by that of $D_{R,t}$ sequentially, and that is originally inspired from [57].

As explained above, the proposed network finally estimates the diffuse and specular reflection of the dichromatic model, and the reflectance and illumination of intrinsic decomposition. These are clearly confirmed by deriving the dichromatic model from (13) straightforwardly as follows:

$$I = D_R \otimes L + S_R \otimes L = D + S \tag{14}$$

where D and S indicate the diffuse and specular components of the dichromatic model.

C. LOSS FUNCTIONS

To train the network, several losses that reflect the characteristics of the two models are exploited. The network is trained with the weighted sum of losses as follows:

$$\mathcal{L}_{tot} = \mathcal{L}_{recon} + w_1\mathcal{L}_{CC} + w_2\mathcal{L}_{invar} + w_3\mathcal{L}_{smooth} + w_4\mathcal{L}_{AC} + w_5\mathcal{L}_{KD} \tag{15}$$

The sub-losses $\mathcal{L}_{recon}$, $\mathcal{L}_{CC}$, $\mathcal{L}_{invar}$, $\mathcal{L}_{smooth}$, $\mathcal{L}_{AC}$ and $\mathcal{L}_{KD}$ mean the reconstruction, color constancy, invariant, smooth, AC fitting, and knowledge distillation losses, respectively.

1) RECONSTRUCTION LOSS

Based on our proposed decomposition model, a target frame I_t should be equal to the reconstructed frame with the network output. The reconstructed frame of the input frame I_t can be represented as (13). For saturated pixels, the reconstructed value is larger than 255. This may lead the network to be trained with inaccurate reconstruction loss. So, the reconstruction loss is calculated on non-saturated regions as follows:

$$\mathcal{L}_{recon} = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_{ij} \parallel M_{sat} \cdot \{(D_{R,j} + S_{R,j}) \otimes L_i \cdot \Gamma_i - I_i\} \parallel_1 \tag{16}$$

where α_{ij} is 1 for $i = j$, and otherwise it is smaller than 1. M_{sat} is a saturated region mask. Objects in a scene and camera are assumed to be static in the input video, and this is a quite reasonable assumption because the time interval of the high-speed video frames is very short. Thus, the reflectance of all input frames should be constant. So, the reconstructed frame with illumination L_i and reflectance components($S_{R,j}$ and $D_{R,j}$) should be the same as the input frame I_i.

To find saturated pixels, previous studies [6], [40], [41] depend on only pixel intensity, while our proposed method leverages temporal constraint additionally. Under AC light sources, the intensity of a saturated pixel is constant, while the non-saturated pixel varies sinusoidally. So, the saturated pixels have zero temporal gradients, and it is used for determining saturated regions. The pixels with small temporal gradient $TG(i)$ and high intensity are determined as saturated and it is expressed as follows:

$$M_{sat}(i) = \begin{cases} 0, & I(i) > Th_1,\ TG(i) < Th_2 \\ 1, & otherwise. \end{cases} \tag{17}$$

where Th_1 and Th_2 are threshold values of intensity and temporal gradient and i is a pixel index.

2) COLOR CONSTANCY LOSS

Unlike other dichromatic and intrinsic decomposition researches, our proposed model does not assume gray illumination and estimating illumination chromaticity is crucial. As a loss for illumination color estimation, angular error which is the common quality measure of color constancy is exploited. The angular error between the estimated illuminant Γ_i and ground truth illuminant Γ_{gt} is expressed as:

$$\mathcal{L}_{CC} = arccos\left(\frac{\Gamma_i \cdot \Gamma_{gt}}{\parallel \Gamma_i \parallel \parallel \Gamma_{gt} \parallel}\right). \tag{18}$$

3) INVARIANT LOSS

As described in 'Reconstruction loss', the reflectance for all frames should be constant. The invariance of diffuse reflectance is expressed with L1 loss as follows:

$$\mathcal{L}_{invar} = \sum_{t=1}^{N-1}\sum_{t'=t+1}^{N} \parallel D_{R,t} - D_{R,t'} \parallel_1 . \tag{19}$$

4) SMOOTH LOSS

By the Retinex model, the illumination should be smooth [59]. The large gradients of an image come from reflectance variations and small gradients are relatively related to illumination information. To reflect this property, TV-L2 loss is applied to illumination [59]. Also, the specular reflection is spatially smooth on surfaces [60], and it is reflected to TV-L2 loss. These smoothness losses can contribute to extract the reflection components closer to ground truth, and $\mathcal{L}_{smooth}$ is represented as:

$$\mathcal{L}_{smooth} = \sum_{t=1}^{N}(\lambda_1 \parallel \nabla L_t \parallel_2 + \lambda_2 \parallel \nabla S_{R,t} \parallel_2). \tag{20}$$

5) AC FITTING LOSS

As the input high-speed video is captured under AC light source environments, the intensity of incident light varies sinusoidally by double the AC standard frequency, and the reflected light also fluctuates accordingly. This periodic variation is fit with the Gauss-Newton method [61], and the regression error is measured as AC fitting loss. The mean values of all the illumination frames are fit with a sinusoidal function as in [6].

6) KNOWLEDGE DISTILLATION LOSS

The teacher network is pretrained with MSE loss between the estimated AC fitting map and its groundtruth, and it is not updated while training the student network. The down-sampling layers of the network are used as the breakpoints. Since the tasks of teacher and student are different, the every feature channel of TF-Net might not be equally beneficial. So, we use meta-network to decide which channel of the teacher network is useful for the ID-Net. The features after meta-network are transferred to the student with MSE loss.

7) NETWORK TRAINING

The proposed network is trained with two types of losses: temporal and spatial loss. To calculate the temporal loss ($\mathcal{L}_{AC}$ and $\mathcal{L}_{invar}$), the outputs of all input frames are required, while the spatial loss ($\mathcal{L}_{smooth}$ and $\mathcal{L}_{CC}$) is calculated for each frame. Therefore, the network is updated by every video sequence (N frames), not by a single frame.

IV. EXPERIMENTAL RESULTS

The proposed network was trained with a high-speed video dataset proposed in [6]. The Adam optimizer was used for training with a batch size of 16. The initial learning rate was 1×10^{-4}, and learning rate is decayed with epochs. The number of frames (N) used for training was 5.

FIGURE 6. The dichromatic model results. (a) input image, the diffuse reflection of (b) Akashi et al. [13], (c) Yamamoto et al. [14], (d) Yang et al. [58] (e) Tsuji [20], (f) Fu et al. [16], (g) JSHDR [57], (h) DDME [6] and (i) the proposed method.

A. COMPARISONS WITH CONVENTIONAL METHODS

The performance was compared with several conventional methods that conduct dichromatic model, intrinsic image decomposition and color constancy. Since the dataset has no ground truth for dichromatic and intrinsic decomposition, a qualitative comparison is made for high-speed video dataset. A quantitative evaluation for highlight removal was conducted with SHIQ [57].

1) DICHROMATIC MODEL RESULT

The proposed method is compared with the dichromatic based methods such as Akashi et al. [13], Yamamoto et al. [14], Yang et al. [58], Fu et al. [16], and JSHDR [57] which are the single-image approaches, and Tsuji [20] and DDME [6] which are the multiple-image approach that exploits high-speed video captured under AC light source (close to our method). Since conventional methods except DDME [6] assume gray illumination condition, the white-balanced image with ground truth illuminant is used for the input. Since JSHDR [57] is a supervised method and there is no ground-truth in the high-speed video dataset, the model is trained with the same loss as the proposed method in an unsupervised manner. Network structure in [57] was not changed. Note that the learning-based models (JSHDR, DDME, and the proposed method) are trained with

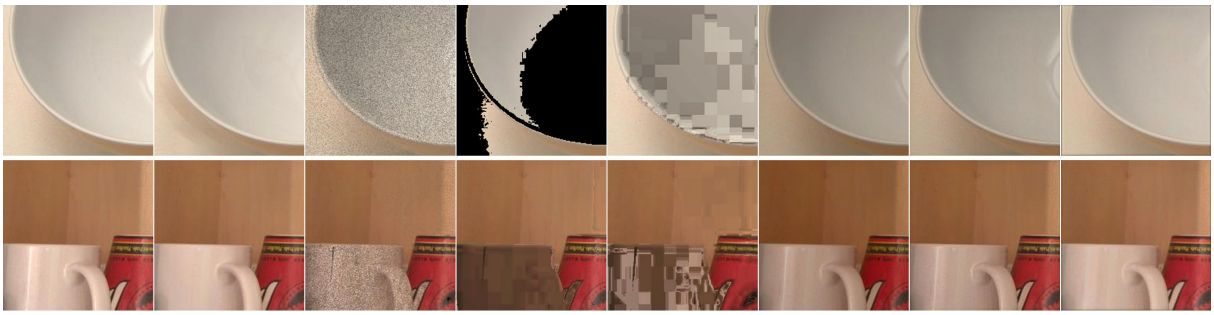

FIGURE 7. SHIQ result comparison. (a) input image, (b) ground truth, (c-h) the diffuse reflection of (c) Akashi et al. [13], (d) Yamamoto et al. [14], (e) Yang et al. [58], (f) Fu et al. [16], (g) JSHDR [57], and (h) the proposed method.

TABLE 1. PSNR and SSIM comparison for the real dataset, SHIQ [57].

Methods	Akashi *et al.* [13]	Yamamoto *et al.* [14]	Yang *et al.* [58]	Fu *et al.* [16]	JSHDR [57]	**Proposed**
PSNR	22.17	12.27	23.20	19.89	22.58	**25.80**
SSIM	0.80	0.53	0.87	0.94	**0.96**	**0.96**

the high-speed video dataset. Fig. 6 compares the diffuse reflection component. As shown in the red boxed regions which have strong specularity, conventional methods suffer from color distortion or fail to remove highlight properly, while the proposed method successfully reconstructs the inherent color. The methods that exploit both temporal and spatial features have better performance than single image methods with only spatial feature. As mentioned in the section of 'Introduction', the dichromatic model has a fundamental limitation in reconstructing chromaticity of shadow and dark regions. The proposed method further alleviates this problem by jointly learning dichromatic and intrinsic image decomposition, and this can be observed in the results of (a4) in Fig. 6.

The quantitative comparison is made with the real image dataset (SHIQ) proposed in [57], and it is shown in Table 1 and Fig. 7. Since SHIQ is a single image dataset, the multi-image based methods, Tsuji [20] and DDME [6] cannot be evaluated. Although the proposed method is trained with multiple frames, the network can be evaluated with a single frame. The proposed network was fine-tuned for the gray illumination input. The performance of the proposed method exceeds the conventional methods in both qualitative and quantitative aspects by achieving the highest PSNR and SSIM.

2) INTRINSIC IMAGE DECOMPOSITION MODEL RESULT

Li et al. [17], Lettry et al. [18], Wei et al. [7], JiePP [62], STAR [63], and UIDNet [64] are evaluated to compare their performances with the proposed method. Wei et al. [7] and Lettry et al. [18] take the multiple-image approach that exploits low/normal light image pairs and time-lapse image dataset. Since these conventional methods except STAR [63] and JiePP [62] require illumination chromaticity as a prior, the white-balanced image with groundtruth illuminant is used as an input. The reflectance of JiePP [62] and STAR [63] is white-balanced with its estimated illuminant calculated as the global average of the illumination. The learning-based methods (Lettry et al. [18], Wei et al. [7], UIDNet [64], and the proposed method) are trained with high-speed video dataset.

The experimental results are shown in Fig. 8. Our proposed network generates the reflectance component (which contains specularity) and its separated diffuse component, and they are shown in (h) and (i). It is shown that the intrinsic chromaticity is accurately recovered by removing specularity. Since other intrinsic models do not consider the specularity of real scenes, they often fail to recover the chromaticity of strong specularity regions, as shown in red boxed regions of Fig. 8. Also, the conventional methods cause severe artifacts around saturated regions, while the proposed method accurately separates illumination and reflectance. One of the weak points for intrinsic image decomposition is the failure on strong shadow regions. As shown in the blue boxed region of Fig. 8, the strong shadow caused color distortion and artifacts in previous studies, while the proposed method successfully removes shadow and reconstructs the intrinsic chromaticity.

3) COLOR CONSTANCY COMPARISON

The result of the proposed decomposition can usefully contribute to color constancy, and its performance is compared with the SOTA methods in Table 2. As shown in Table 2, the proposed method achieved a remarkable performance. Although the task of the proposed method is primarily on image decomposition, its performance is better than the color constancy methods thanks to its accurate decomposition capability. DDME [6], JiePP [62] and STAR [63] estimate illuminant by decomposing image based on dichromatic and intrinsic model.

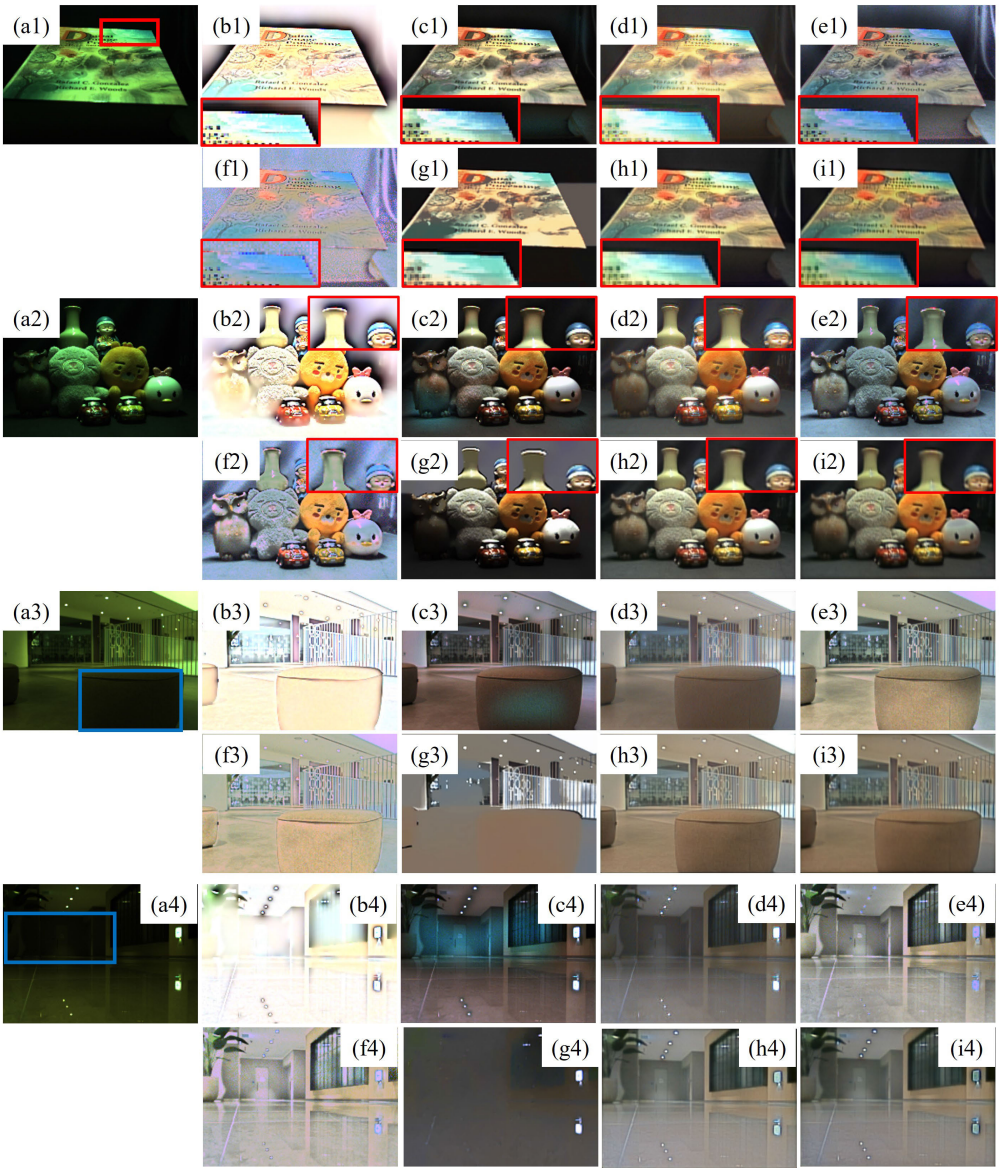

FIGURE 8. Intrinsic image decomposition results. (a) input image, the reflectance of (b) Li et al. [17] (c) Lettry et al. [18], (d) Wei et al. [7], (e) JieP [62], (f) STAR [63], (g) UIDNet [19] and (h, i) the proposed method. Note that two reflectance components w/wo specular are shown in the proposed method.

TABLE 2. Angular error comparisons with conventional color constancy methods.

Method		Mean	Best-25%	Worst-25%	Closed	Ambient
Image decomposition	DDME [6]	1.16	0.26	2.75	0.90	**1.28**
	JieP [62]	3.99	1.42	8.05	5.25	3.36
	STAR [63]	4.00	1.35	8.16	5.46	3.27
	The proposed method	**0.98**	**0.16**	**2.49**	**0.40**	**1.28**
Learning	Bianco *et al.* [65]	1.79	0.36	4.42	1.44	1.97
	FFCC [66]	1.42	**0.12**	4.18	**0.19**	2.04
	FC4 [56]	2.26	0.76	4.17	2.30	2.25
	DDME [6]	1.16	0.26	2.75	0.90	1.28
	ReWU [67]	1.26	0.30	3.30	0.87	1.46
	ColorTiger [68]	7.79	4.91	11.45	7.97	7.70
	Spatio-Temporal [69]	1.00	0.26	2.57	0.56	**1.22**
	The proposed method	**0.98**	0.16	**2.49**	0.40	1.28

B. ABLATION STUDY

To confirm the effectiveness of the proposed decomposition model, three ablation studies were conducted. As shown in Fig. 10, we evaluated several image decomposition models, which are intrinsic decomposition, dichromatic decomposition, and the extended intrinsic decomposition. The decomposition results are shown in Fig. 9 (b-d) and their color constancy performances are compared at Table 3. The results

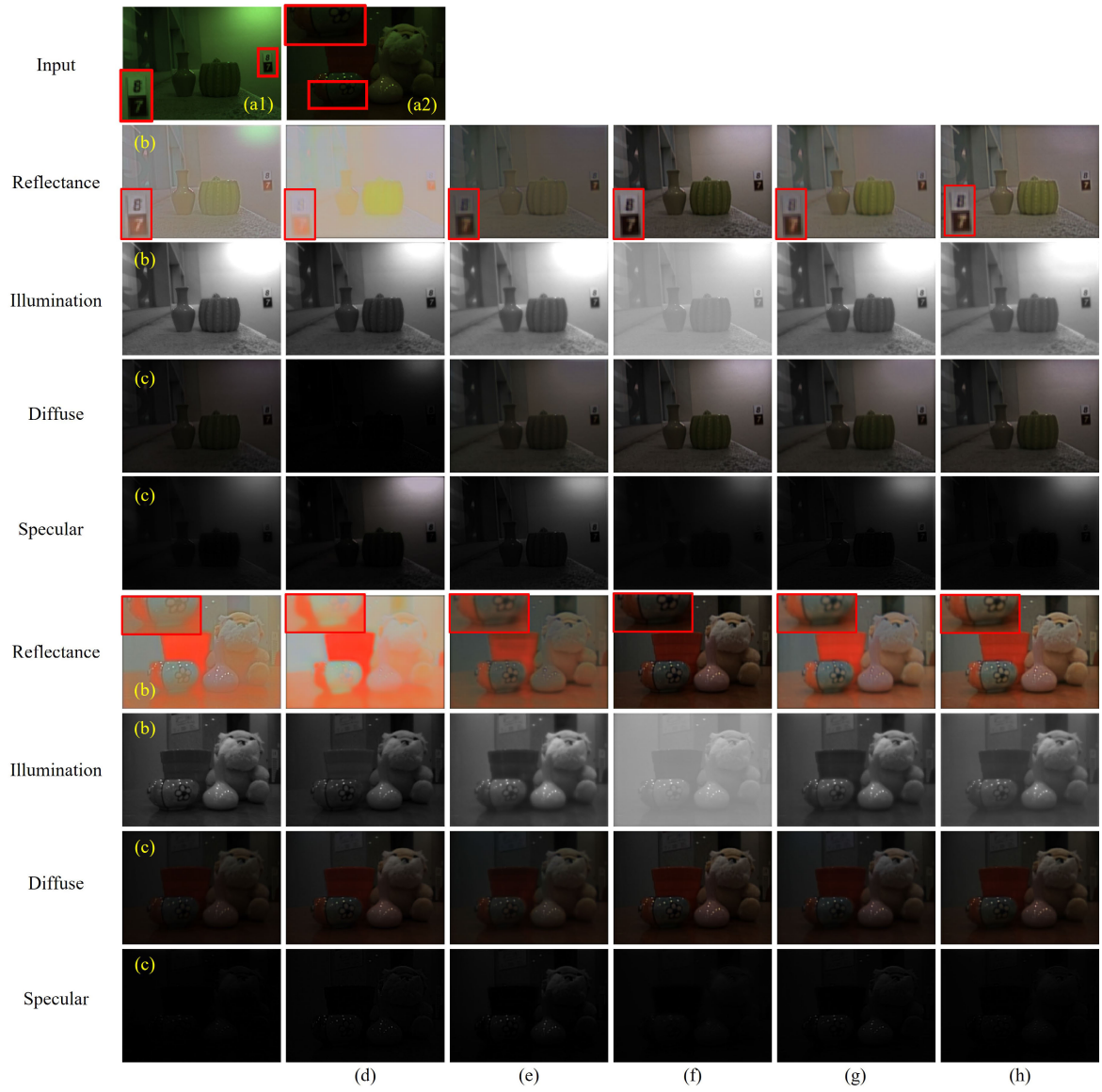

FIGURE 9. (a) Input image and decomposition results for (b) the model A, (c) the model B, (d) the model C in Fig. 10, (e) without white-balance in the Reflectance subnet, (f) a single frame input (no temporal feature), (g) without knowledge distillation, and (h) the proposed method. Note that all methods except for (f) accept multiple video frames as an input.

TABLE 3. Average angular error comparisons of ablation studies.

Method	proposed	A in Fig. 10	B in Fig. 10	C in Fig. 10	w/o WB	single input	w/o KD
average angular error	**0.98**	1.14	1.98	1.24	1.09	1.21	1.52

show that our proposed method is superior to the other possible decomposition models for both color constancy and image decomposition. Specularity is removed in the reflectance of the model C and the proposed. Unlike other models, the proposed method accurately separates specular components and accurate reflectance is obtained. Since the original intrinsic decomposition (A in Fig. 10) does not consider specularity, color distortion happens in high specular regions. It demonstrates the importance of considering specularity in intrinsic decomposition. The result of dichromatic decomposition (B in Fig. 10) is also degraded by severe color distortion and fails to reconstruct inherent chromaticity on saturated regions. Although the extended intrinsic decomposition model (C in Fig. 10), which is the same assumption as [10] and [11] accomplishes better performance than the model A and B in Fig. 10, it showed over-smoothed result as in the red boxes. The image details are not reconstructed in reflectance, while the proposed method correctly classifies the pattern to reflectance.

To examine the effect of illuminant color in the input image of the reflectance subnet, the original input image is used without white-balance. As shown in Fig. 9 (e), the

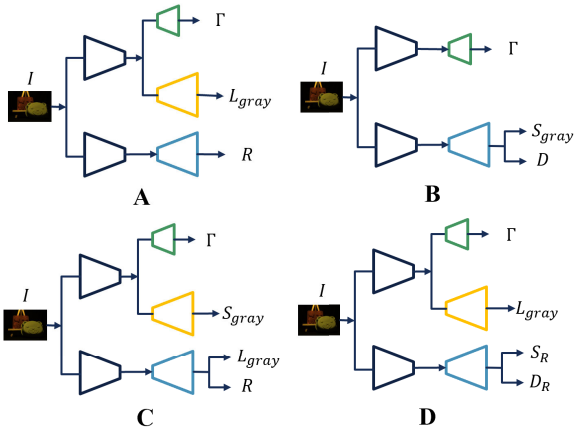

FIGURE 10. A: intrinsic decomposition in (1), B: dichromatic decomposition in (2), C: extended intrinsic decomposition in (4), and D: the proposed decomposition in (13).

result without white-balance is suffered from severe color distortion. By transferring illumination chromaticity from the Illumination subnet to the Reflectance subnet, reflectance chromaticity recovery is improved and the color distortion is alleviated.

To confirm the importance of temporal features, we conducted a couple of experiments. First, a single frame input is used instead of N frames, and accordingly, the temporal losses ($\mathcal{L}_{AC}$ and $\mathcal{L}_{invar}$) and $\mathcal{L}_{KD}$ are removed. Second, with N frames of input, only temporal feature distillation is removed. Fig. 9 (f) and (g) are the results of a single input and without distillation loss. It is confirmed that temporal features are helpful for both color constancy and image decomposition. The single image method is poor in separating the reflection components, and the intrinsic chromaticity the shadow region is not reconstructed successfully. Also, Fig. 9 (g) shows more blurred reflectance than the proposed method. Unlike previous study [6] that reflects temporal variation in training cost, the proposed method further improves performance by exploiting the temporal feature more efficiently with knowledge distillation.

C. LIMITATION

Although our proposed method performs superior to the conventional methods, it still has limitation in some cases, because of the ill-posed property of image decomposition. The conventional methods reported that some regions (weak texture, strong shadow and saturation) are incorrectly decomposed. Although this mis-classification has been further improved in our method, there are still limitations for perfectly reconstructing intrinsic property for saturation and strong shadow as shown in Fig. 6 and Fig. 8. This is because saturated regions and weakly-illuminated shadow are lack of AC variation. Note that the stronger the AC variation is, the better temporal features are [70].

V. CONCLUSION

In this paper, we proposed a new image formation model that conducts dichromatic and intrinsic decomposition jointly. The experimental result shows that the proposed model performs better than each single decomposition. Also, it was experimentally found that the decomposition performance depends on the order of the two decompositions. Namely, the proposed method ('intrinsic + dichromatic') performs better than the conventional model ('dichromatic + intrinsic'). Specular reflection is generally very weak and sparse, and thus, its separation is more difficult than the illumination of the intrinsic model. The fundamental limitation of intrinsic image decomposition is Lambertian assumption and poor working for real scenes, which is easily solved with the proposed method. Unlike conventional methods, the proposed model is trained in an semi-supervised manner with real scenes by leveraging the temporal property of AC light sources. The performance was further improved by temporal features. Illumination chromaticity is estimated in the Illumination subnet, and is used for white-balance in the Reflectance subnet, leading to the significant reduction of color distortion. Although our main task is decomposition, color constancy performance is better than SOTA methods.

REFERENCES

[1] H. Barrow, J. Tenenbaum, A. Hanson, and E. Riseman, "Recovering intrinsic scene characteristics," *Comput. Vis. Syst.*, vol. 2, nos. 3–26, p. 2, 1978.
[2] S. A. Shafer, "Using color to separate reflection components," *Color Res. Appl.*, vol. 10, no. 4, pp. 210–218, 1985.
[3] R. T. Tan, K. Ikeuchi, and K. Nishino, "Color constancy through inverse-intensity chromaticity space," in *Digitally Archiving Cultural Objects*. Berlin, Germany: Springer, 2008, pp. 323–351.
[4] S.-M. Woo, S.-H. Lee, J.-S. Yoo, and J.-O. Kim, "Improving color constancy in an ambient light environment using the Phong reflection model," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1862–1877, Apr. 2018.
[5] 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.*, Jun. 2019, pp. 12329–12338.
[6] 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.
[7] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex decomposition for low-light enhancement," 2018, *arXiv:1808.04560*.
[8] S. Beigpour and J. Van De Weijer, "Object recoloring based on intrinsic image estimation," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 327–334.
[9] S. Duchene, C. Riant, G. Chaurasia, J. L. Moreno, P.-Y. Laffont, S. Popov, A. Bousseau, and G. Drettakis, "Multiview intrinsic images of outdoors scenes with an application to relighting," *ACM Trans. Graph.*, vol. 34, no. 5, pp. 1–16, Nov. 2015.
[10] J. Shi, Y. Dong, H. Su, and S. X. Yu, "Learning non-Lambertian object intrinsics across ShapeNet categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1685–1694.
[11] R. Yi, P. Tan, and S. Lin, "Leveraging multi-view image sets for unsupervised intrinsic image decomposition and highlight separation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 12685–12692.
[12] H.-C. Lee, E. J. Breneman, and C. P. Schulte, "Modeling light reflection for computer color vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 4, pp. 402–409, Apr. 1990.
[13] 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.

[14] 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.

[15] K.-F. Yang, S.-B. Gao, and Y.-J. Li, "Efficient illuminant estimation for color constancy using grey pixels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2254–2263.

[16] 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.

[17] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust Retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.

[18] L. Lettry, K. Vanhoey, and L. Van Gool, "Unsupervised deep single-image intrinsic decomposition using illumination-varying image sequences," *Comput. Graph. Forum*, vol. 37, pp. 409–419, Oct. 2018.

[19] Y. Liu, Y. Li, S. You, and F. Lu, "Unsupervised learning for intrinsic image decomposition from a single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3248–3257.

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

[21] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.

[22] X. Cheng, Z. Rao, Y. Chen, and Q. Zhang, "Explaining knowledge distillation by quantifying the knowledge," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12925–12935.

[23] M. Phuong and C. Lampert, "Towards understanding knowledge distillation," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 5142–5151.

[24] J. H. Cho and B. Hariharan, "On the efficacy of knowledge distillation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4794–4802.

[25] M. Suin, K. Purohit, and A. N. Rajagopalan, "Degradation aware approach to image restoration using knowledge distillation," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 2, pp. 162–173, Feb. 2021.

[26] Z. Cheng, Y. Zheng, S. You, and I. Sato, "Non-local intrinsic decomposition with near-infrared priors," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2521–2530.

[27] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2335–2342.

[28] C. Rother, M. Kiefel, L. Zhang, B. Scholkopf, and P. Gehler, "Recovering intrinsic images with a global sparsity prior on reflectance," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 24, 2011, pp. 1–9.

[29] J. T. Barron and J. Malik, "Color constancy, intrinsic images, and shape estimation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2012, pp. 57–70.

[30] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[31] E. H. Land and J. J. McCann, "Lightness and Retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.

[32] A. S. Baslamisli, H.-A. Le, and T. Gevers, "CNN based learning using reflection and Retinex models for intrinsic image decomposition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6674–6683.

[33] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2012, pp. 611–625.

[34] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "ShapeNet: An information-rich 3D model repository," 2015, *arXiv:1512.03012*.

[35] Q. Fan, J. Yang, G. Hua, B. Chen, and D. Wipf, "Revisiting deep intrinsic image decompositions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8944–8952.

[36] S. Kim, K. Park, K. Sohn, and S. Lin, "Unified depth prediction and intrinsic image decomposition from a single image via joint convolutional neural fields," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 143–159.

[37] S. Bell, K. Bala, and N. Snavely, "Intrinsic images in the wild," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–12, 2014.

[38] B. Kovacs, S. Bell, N. Snavely, and K. Bala, "Shading annotations in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6998–7007.

[39] Z. Li and N. Snavely, "Learning intrinsic image decomposition from watching the world," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9039–9048.

[40] 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.

[41] 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.

[42] 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.

[43] V. S. Ramos, L. G. Q. S. 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.*, 2004, pp. 316–321.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] J.-W. Ha, K.-K. Lee, J.-S. Yoo, and J.-O. Kim, "Deep highlight removal using temporal dark prior in high-speed domain," *IEEE Access*, vol. 11, pp. 20136–20149, 2023.

[53] E. Garces, C. Rodriguez-Pardo, D. Casas, and J. Lopez-Moreno, "A survey on intrinsic images: Delving deep into Lambert and beyond," *Int. J. Comput. Vis.*, vol. 130, no. 3, pp. 836–868, Mar. 2022.

[54] 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.

[55] 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.

[56] Y. Hu, B. Wang, and S. Lin, "FC4: Fully convolutional color constancy with confidence-weighted pooling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4085–4094.

[57] 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.

[58] 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.

[59] Z. Jiang, H. Li, L. Liu, A. Men, and H. Wang, "A switched view of Retinex: Deep self-regularized low-light image enhancement," *Neurocomputing*, vol. 454, pp. 361–372, May 2021.

[60] 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.

[61] Å. Bjorck, *Numerical Methods in Matrix Computations*, vol. 59. Berlin, Germany: Springer, 2015.

[62] B. Cai, X. Xu, K. Guo, K. Jia, B. Hu, and D. Tao, "A joint intrinsic-extrinsic prior model for Retinex," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4000–4009.
[63] J. Xu, Y. Hou, D. Ren, L. Liu, F. Zhu, M. Yu, H. Wang, and L. Shao, "STAR: A structure and texture aware Retinex model," *IEEE Trans. Image Process.*, vol. 29, pp. 5022–5037, 2020.
[64] Q. Zhang, J. Zhou, L. Zhu, W. Sun, C. Xiao, and W.-S. Zheng, "Unsupervised intrinsic image decomposition using internal self-similarity cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9669–9686, Dec. 2022.
[65] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 81–89.
[66] J. T. Barron and Y.-T. Tsai, "Fast Fourier color constancy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 886–894.
[67] J. Qiu, H. Xu, and Z. Ye, "Color constancy by reweighting image feature maps," *IEEE Trans. Image Process.*, vol. 29, pp. 5711–5721, 2020.
[68] N. Banic, K. Koscevic, and S. Loncaric, "Unsupervised learning for color constancy," 2017, *arXiv:1712.00436*.
[69] 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.
[70] J.-W. Ha, J.-S. Yoo, and J.-O. Kim, "Deep color constancy using temporal gradient under AC light sources," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 2355–2359.

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 intrinsic image decomposition, multi-exposure fusion, 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.