Robust Reflection Removal with Reflection-free Flash-only Cues

Chenyang Lei Qifeng Chen
The Hong Kong University of Science and Technology

Abstract

We propose a simple yet effective reflection-free cue for robust reflection removal from a pair of flash and ambient (no-flash) images. The reflection-free cue exploits a flash-only image obtained by subtracting the ambient image from the corresponding flash image in raw data space. The flash-only image is equivalent to an image taken in a dark environment with only a flash on. We observe that this flash-only image is visually reflection-free, and thus it can provide robust cues to infer the reflection in the ambient image. Since the flash-only image usually has artifacts, we further propose a dedicated model that not only utilizes the reflection-free cue but also avoids introducing artifacts, which helps accurately estimate reflection and transmission. Our experiments on real-world images with various types of reflection demonstrate the effectiveness of our model with reflection-free flash-only cues: our model outperforms state-of-the-art reflection removal approaches by more than 5.23dB in PSNR, 0.04 in SSIM, and 0.068 in LPIPS. Our source code and dataset are publicly available at github.com/ChenyangLEI/flash-reflection-removal.

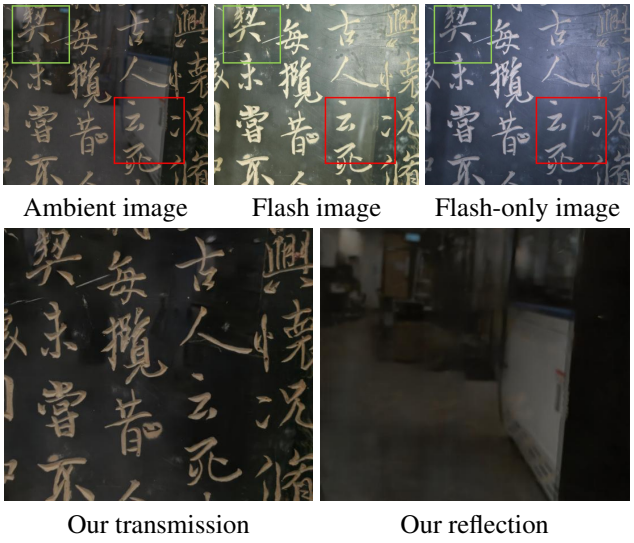


Figure 1. A reflection-free flash-only image is computed from a pair of ambient/flash images to help remove reflection. Our transmission image does not absorb the artifacts in the flash-only image.

1. Introduction

An image taken by a camera in front of a glass often contains undesirable reflection. In the process of image formation with reflection, the irradiance received by a camera can be approximately modeled as the sum of transmission and reflection. In this paper, we are interested in recovering a clear transmission image by removing reflection from the ambient image (captured image). Reflection removal is an important application in computational photography, which can highly improve image quality and pleasantness. Furthermore, computer vision algorithms can be more robust to images with reflection, as the reflection can be largely erased by a reflection removal method.

Reflection removal is challenging because the reflection component is usually unknown. Since both reflection and transmission are natural images, it is hard to distinguish between reflection and transmission from an input image. Therefore, many methods adopt various assumptions on the appearance of reflection for reflection removal. For example, some single image-based methods [3, 42] assume that the reflection is not in-focus and blurry. The ghosting cue [32] is another assumption that holds when the glass is thick. However, reflection in real-world images is diverse, and these assumptions do not necessarily hold [19, 35]. As a result, existing methods cannot perfectly remove reflection with diverse appearance from real-world images [19].

We propose a novel *reflection-free flash-only cue* that facilitates inferring the reflection in an ambient image. This cue is robust since it is independent of the appearance and strength of reflection, unlike assumptions adopted in previous single image reflection removal methods [32, 42] or flash-based methods [5]. The reflection-free cue is based on a physics-based phenomenon of an image obtained by subtracting an ambient image from the corresponding flash image (in raw data space). This flash-only image is equivalent to an image captured under the flash-only illumination: the environment is completely dark, and a single flash is the sole light source. A key observation is that the reflection is

invisible in the flash-only image.

While flash-only images provide reflection-free cues to distinguish reflection, they also have weaknesses. For instance, in Fig. 1, we can observe artifacts (e.g., color distortion, illuminated dust) due to uneven flash illumination, occlusions, and other reasons. These artifacts prevent us from obtaining a high-quality transmission easily.

To utilize the reflection-free cue and avoid introducing the flash-only image artifacts, we design a dedicated architecture for obtaining high-quality transmission. Specifically, we first estimate a reflection image instead of a transmission image. Then, to further avoid introducing artifacts in the flash-only image, only the input ambient image and the estimated reflection are given to the second network that estimates the transmission.

Combining our dedicated architecture with the reflection-free cue, we can robustly and accurately remove various kinds of reflection to recover the underlying transmission image. Although we need an extra flash image compared with single image methods, a flash/no-flash image pair can be captured with a single shutter-press using customized software, as shown in Fig. 6. Hence, general users can easily apply our method for robust reflection removal. In summary, our contributions are as follows:

- We propose a novel cue - the reflection-free flash-only cue that makes distinguishing reflections simpler for reflection removal. This cue is robust since it is independent of the appearance and strength of reflection.
- We propose a dedicated framework that can avoid introducing artifacts of flash-only images while utilizing reflection-free cues. We improve more than 5.23dB in PSNR, 0.04 in SSIM, and 0.068 in LPIPS on a real-world dataset compared with state-of-the-art methods.
- We construct the first dataset that contains both raw data and RGB data for flash-based reflection removal.

2. Related Work

2.1. Reflection Removal

Single image reflection removal. In single image reflection removal, the defocused reflection assumption and ghosting cue are commonly used. The defocused reflection assumption means that reflections are not in focus. Hence, prior work can assume they are more blurry compared with the transmission. Following this assumption, learning-based methods [9, 46] can synthesize abundant data for training, and non-learning based methods can suppress the reflection based on image gradient [3, 42]. The ghosting cue means multiple reflections are visible on the glass [32]. However, the ghosting cue only exists when the glass is thick. Hence, algorithms that are based on ghosting cue might fail on the thin glass.

There are many attempts to relax assumptions of reflection. Wei et al. [38] and Ma et al. [26] use generative adversarial networks [11] to synthesize realistic reflection under the guidance of real-world reflections. Kim et al. [16] propose a physics-based method to render the reflection and mixed image, which improves the quality of training data a lot. Also, Zhang et al. [46], Wei et al. [37], and Li et al. [21] collected real-world data for improving the quality of training data. However, as reported by Lei et al. [19], these methods [3, 9, 42] are still far from perfectly removing reflections for diverse real-world data.

Multiple images reflection removal. Some reflection removal methods utilize the motion cue of reflection and transmission in multiple images for reflection removal [12, 13, 22, 23, 33, 40]. In these motion-based methods, SIFT-flow [22], homography [12] and optical flow [23, 40] are used to find correspondences among multiple images to distinguish reflection and transmission. However, taking images with different motion cost more effort, and some assumptions are required (e.g., all pixels in transmission must appear in at least one image [40]). Polarization is also used in reflection removal to achieve great performance [10, 18, 25, 27, 29, 43]. The inputs are usually images through various polarizers, which contain polarization information of light. Since polarization of reflection and transmission is usually different, it can be used to distinguish them. However, a polarizer is usually required to be shifted to take images, which is complicated. Recently, a camera [21, 31] that can take several polarization images appears but this kind of camera is yet to be widely used.

Flash-based reflection removal. Various properties of a pair of flash/ambient images are adopted in previous work [1, 5]. Agrawal et al. [1] claim that gradient orientations are consistent in the image pair, assuming that depth edges, shadows, and highlights are few. However, they cannot generate reasonable results for undesirable regions (e.g., shadows, specular reflection), and their results tend to be over-smooth. SDN [5] utilize the assumption that reflection can be obviously suppressed by flash, but the suppression effect is sensitive to the strength of reflection: when reflection is strong, the suppression is no longer effective.

2.2. Flash Photography

Flash images are used in various tasks. Petschnigg et al. [30] use the flash image for denoising, detail transfer, etc. Drew et al. [8] use the flash-only image for shadow removal. Sun et al. [34] observe that the change of intensity is different for near objects and background in the flash-only image and apply it to image matting. Cao et al. [4] use the flash-only image for shape and albedo recovery under the assumption of the Lambertian model.

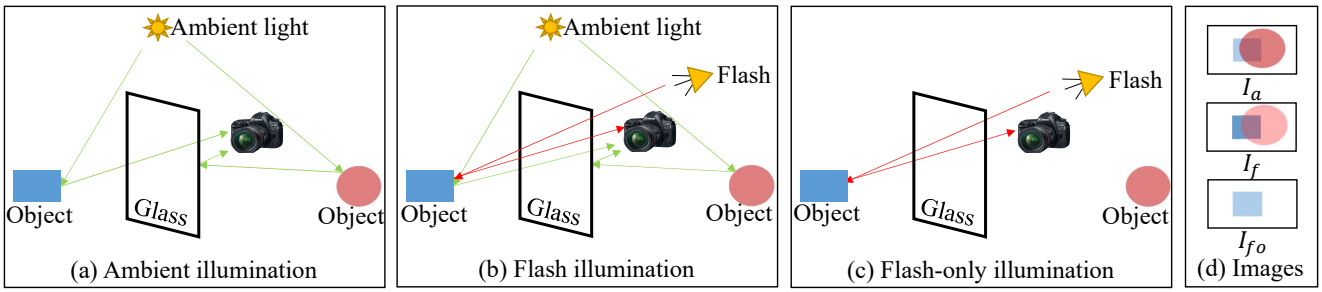


Figure 2. An illustration model of the reflection-free cue. Since objects in reflection cannot *directly* receive flash and reflected flash from glass is often weak, flash-only images are visually reflection-free. Note that the flash-only image I_{fo} is obtained from I_a and I_f.

Figure 3. Examples of reflection-free flash-only images. Flash-only images are visually reflection-free but have many artifacts.

3. Reflection-free Flash-only Cues

Flash-only images. Let $\{I_a, I_f\}/\{I_a^{raw}, I_f^{raw}\}$ be the RGB/raw images under ambient and flash illuminations. A flash-only image I_{fo}^{raw} can be computed from I_a^{raw} and I_f^{raw}. Since the flash image is the sum of ambient image and flash-only image for a linear response camera in linear space [30], we can obtain the flash-only image through:

$$I_a^{raw} = R_a^{raw} + T_a^{raw}, \tag{1}$$

$$I_f^{raw} = R_a^{raw} + T_a^{raw} + R_{fo}^{raw} + T_{fo}^{raw}, \tag{2}$$

$$I_{fo}^{raw} = I_f^{raw} - I_a^{raw} = R_{fo}^{raw} + T_{fo}^{raw}, \tag{3}$$

where R and T are the reflection and transmission. For simplicity, we also use I^{raw} to denote the image after linearization. Reflection-free cues exist in flash-only images. The flash-only image is equivalent to an image captured in a completely dark environment, and the flash is the sole light source, as shown in Fig. 2(b). Note that I_{fo}^{raw} is invariant to different ambient illuminations as long as I_a^{raw} and I_f^{raw} do not have saturated pixels.

Reflection-free cues. The reflection-free cue denotes a physics-based phenomenon: reflections of ambient image R_a^{raw} are invisible in the flash-only image. Besides, we have $R_{fo}^{raw} \approx \mathbf{0}$ in most cases. Fig. 2 shows an illustration of this phenomenon. Reflections exist in ambient images because objects in the reflection receive ambient light and then reflect it to the camera through the glass. In Fig. 2(b), objects in reflection do not directly receive light from the flash. Besides, since reflectance of glass is mostly much weaker than transmittance, the reflected flash is almost negligible (please check the supplement for detailed analysis). Hence, objects in reflection are barely illuminated and reflections do not appear in flash-only images.

To verify reflection-free cues, we capture pairs of ambient and flash images under different illumination and scenes, and we compute I_{fo}^{raw} following Eq. 1. As shown in Fig. 3, reflections disappear in flash-only images, even when reflections are strong. Also, the 2nd example shows this cue is valid not only for semi-reflecting surfaces.

Undesirable artifacts. Although flash-only images are visually reflection-free, they usually have undesirable artifacts, as shown in Fig. 3. We can analyze reasons of degradation formally from flash-only radiance by Eq. 4 [14]:

$$L_o^{fo} = \int_\Omega f_r(\omega_i, \omega_o) L_i(\omega_i)(\omega_i \cdot \mathbf{n}) d\omega_i, \tag{4}$$

$$L_{i,d}^{fo}(w_i) = \frac{L^{fo}(w_i)}{d^2}, \tag{5}$$

where L_o and L_i are the radiance of outgoing and incident light, ω_i and ω_o are the light direction of outgoing and incident light, f_r is the bidirectional reflectance distribution function (BRDF), $\mathbf{n}$ is the surface normal and Ω is the hemisphere. Flash-only images can contain the following artifacts that require to be resolved:

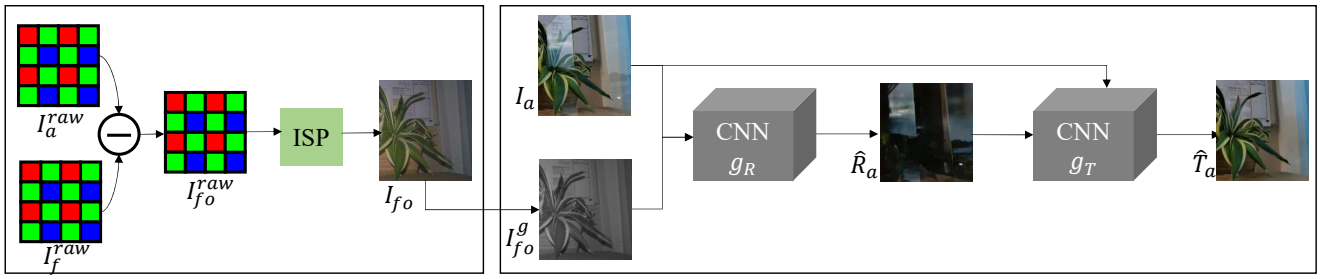
(1) Color distortion usually appears since the flash light L^{fo} is different from ambient illumination. Similarly, the shading also changes since light direction w_i is different.

(2) Uneven illumination is a common problem due to irradiance falloff in Eq. 5, irradiance (and thus radiance $L_{i,d}^{fo}$) is different due to different distance d to the flash.

(3) New shadows are brought by occlusion.

(4) Highlights caused by flash might appear on the glass.

(5) If the glass is dirty, dust can be illuminated on the glass, as shown in 1st example in Fig. 3.



(a) Pre-processing (b) Architecture

Figure 4. The overall architecture of our approach. We compute the I_{fo} from I_f^{raw} and I_a^{raw}. Then, our dedicated architecture estimates the reflection first to avoid absorbing artifacts of flash-only images. Finally, the transmission is estimated with the guidance of reflection.

4. Method

Given an ambient image I_a^{raw} and a flash image I_f^{raw} in raw space, our approach aims to estimate the transmission T_a under ambient illumination. In Fig. 4(a), we first take raw images I_a^{raw}, I_f^{raw} and implement pre-processing to obtain the RGB flash-only image I_{fo}, as introduced Sec. 4.1. Then, our dedicated architecture in Sec. 4.2 takes RGB images I_a, I_{fo} as input to remove the reflection.

4.1. Pre-processing

We first capture two raw images I_a^{raw} and I_f^{raw} through pipeline in Sec. 5. Given I_a^{raw} and I_f^{raw}, we implement the following pipeline to obtain RGB images:

1) Subtraction. We first implement linearization to convert images to linear space using the black-level and white-level information from the metadata. After this step, the range of each pixel is transferred to $[0, 1]$. Then the flash-only image is obtained through Eq. 1 since the linearity between pixel values and physical light is preserved well. At last, the flash-only image is converted back to raw space using the black-level and white-level information.

2) Image signal processing (ISP). We implement a regular ISP [15] that includes linearization, demosaiced, white balance, color correction, and gamma correction to convert raw images to RGB images using the original metadata of images. We adopt the metadata of I_a^{raw} to process I_{fo}^{raw} since it is obtained by I_a^{raw}, I_f^{raw}, and no metadata is available. Note that the white balance of I_{fo} is usually not as good as I_a since it does not have its own metadata. For I_a at test time, we can use our ISP to obtain the sRGB image or use the original sRGB image processed by the camera's ISP. A learning-based ISP [6, 39, 44] can also be used here.

4.2. Architecture

As shown in Fig. 3, the reflection in I_a does not exist in I_{fo}. Except for the artifacts and color distortion, the I_{fo} is quite similar to our target transmission T_a. Hence, we first try to use a network to directly estimate transmission from I_a and I_{fo}, which we denote as base model g_B (note that this model g_B is *not our final model*):

$$\hat{T}_B = g_B(I_a, I_{fo}; \theta_B), \tag{6}$$

where θ_B is the parameters of network g_B. However, we observe that although this model can correctly remove various types of reflectionsR_a, the estimated $\hat{T}_B$ has undesirable artifacts, especially for the area that contains shadows, highlight in flash-only images. Also, the color might be closer to I_{fo} in some area (i.e., color distortion), as shown in Fig. 5(d).

We argue that: since the transmission component is the intersection of I_{fo} and I_a, the network tends to fuse I_{fo} and I_a to obtain the estimated transmission, and artifacts of I_{fo} are inevitably fused too.

4.2.1 Reflection-pass network

To solve the drawback of g_B, we only estimate the reflection first instead of directly estimating T_a. Stated in another way, only the reflection passes the first network. As reflections R_a only exist in I_a, it must be extracted from I_a and avoid introducing artifacts of I_{fo}. On the other hand, since there is no reflection in the flash-only image I_{fo}, it can provide strong guidance for reflection estimation. Specifically, we first convert the flash-only image to grayscale image I_{fo}^g to avoid the influence of color distortion. In practice, we find that the grayscale flash-only image can provide enough structure information for estimating the reflection. Then, I_a, I_{fo}^g are concatenated as input to the network g_R:

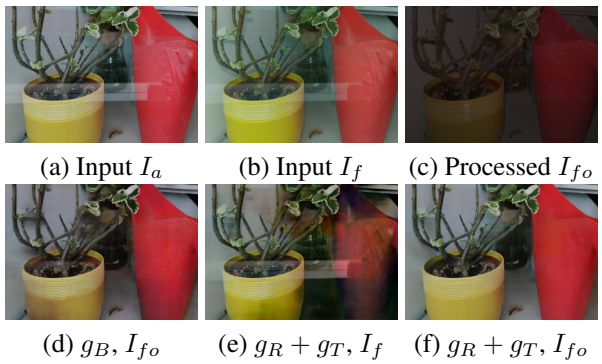
$$\hat{R}_a = g_R(I_a, I_{fo}^g; \theta_R), \tag{7}$$

$$L_R(R_a, \hat{R}_a) = ||R_a - \hat{R}_a||_2^2, \tag{8}$$

where θ_R is the parameters of network g_R. We adopt the L2 loss for training g_R.

4.2.2 Reflection-guided transmission estimation

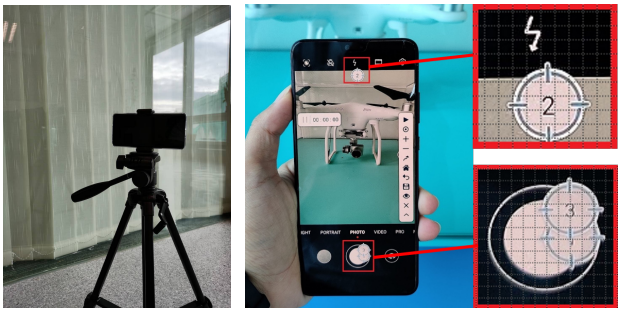
The effectiveness of adopting reflection as guidance has been proven in previous work [19, 21, 41]. Hence, we directly use the estimated reflection and the ambient image to



(a) Input I_a (b) Input I_f (c) Processed I_{fo}

(d) g_B, I_{fo} (e) $g_R + g_T, I_f$ (f) $g_R + g_T, I_{fo}$

Figure 5. Qualitative comparison among multiple implementations. Combining the reflection-free I_{fo} with our dedicated architecture achieves the best performance.



(a) (b)

Figure 6. A picture of (a) our data acquisition setup for constructing the dataset and (b) a real application. In (b), a user only needs to click a button to capture a pair of flash/ambient images.

estimate the transmission. Note that the flash-only image I_{fo} is *not input* to g_T to avoid introducing artifacts. The transmission $\hat{T}_a$ is then estimated:

$$\hat{T}_a = g_T(I_a, \hat{R}_a; \theta_T), \tag{9}$$

where θ_T is the parameters of g_T. We also adopt a L2 loss for training g_T. As shown in Fig. 5(f), the result of $g_R + g_T$ does not contain obvious artifact (e.g., color distortion), which is much better than the result of g_B in Fig. 5(d).

Discussion One might argue that reflection-free flash-only images (Eq. 1) can be learned implicitly using a large amount of data. However, note that the linearity does not exist in RGB images after non-linear ISP operation. Using the same training setting, replacing I_{fo} with I_f can lead to artifacts on strong reflection, as shown in Fig. 5(e).

4.2.3 Implementation details

We train for 150 epochs with batch size 1 on an Nvidia RTX 2080 Ti GPU. We use the Adam optimizer [17] to update the weights with an initial learning rate of 10^{-4}. The plain U-Net [28] is used for the two networks (with trivial modification [20]). The two networks g_R and g_T are trained simultaneously. We implement random cropping for images with more than 640,000 pixels. For the other images, the network is trained on complete images rather than patches.

4.3. Limitation

The reflection-free cue is based on the quality of flash-only image. If all objects in transmission are too far and are not illuminated by the flash, there would be no difference between the ambient image and the flash image (i.e., the flash-only image will be totally black except for reflected flash) due to the irradiance falloff problem in Eq. 5. In this case, our model will be degraded to single image reflection removal. Also, if the objects in transmission move rapidly (larger motion within the exposure time), the flash-only image can have a serious misalignment problem. We believe other methods should be proposed to solve these cases well.

5. Flash-only Reflection Removal Dataset

Real-world data. Our method requires a pair of raw flash/ambient images. Since there is no existing dataset, we construct the first real-world dataset that contains raw data for flash-based reflection removal. This dataset is collected by Nikon Z6 and a smartphone camera Huawei Mate30. We control the camera setting (e.g., exposure) to make sure that Eq. 1 holds. The collection procedure is as follows:

1) Fix the focal length, aperture, exposure time, and ISO.
2) Take the ambient image I_a (I_a^{raw}).
3) Turn on the flash and take the flash image I_f (I_f^{raw}).
4) To get the ground truth T_a, we turn off the flash and take an extra reflection image R_a (R_a^{raw}). Note that this step is unnecessary at test time.

To collect high-quality data with perfect alignment, we use a tripod to fix the camera, as shown in Fig. 6(a). In practice, steps (1)-(3) can be programmed to be implemented automatically with a single shutter-press on mobile phones, as shown in Fig. 6(b). By doing so, the extra cost is mainly longer exposure time compared with single image methods. In the next section, we demonstrate that this extra flash image can robustly and effectively improve performance.

We collect ground truth ambient transmission T_a for training and evaluation. Specifically, we obtain $T_a^{raw} = I_a^{raw} - R_a^{raw}$ in raw (linear) space [19]. Thus, an extra reflection image under ambient illumination is captured. Then, ISP is implemented for each raw image similar to processing pipeline in Sec. 4.1. We adopt the metadata of I_a^{raw} to process the raw data T_a^{raw}. At last, we crop the area where the transmission is valid following Lei et al. [19]. Briefly speaking, we capture a set $\{I_a^{raw}, I_f^{raw}, R_a^{raw}\}$ and process these three images to get the set $\{I_a, I_f, I_{fo}, T_a, R_a\}$. In total, we collect 157 sets of

	Input I_a	Zhang et al. [46]	BDN [41]	Wei et al. [37]	Kim et al. [16]	Li et al. [21]	Agrawal et al. [1]	SDN [5]	Ours
#Input images	1	1	1	1	1	1	2	2	2
PSNR↑	22.72	23.76	21.41	23.89	21.67	<u>24.53</u>	23.13	22.63	**29.76**
SSIM↑	0.874	0.873	0.802	0.864	0.821	<u>0.890</u>	0.853	0.827	**0.930**
LPIPS↓	<u>0.205</u>	0.242	0.410	0.238	0.298	0.224	0.251	0.269	**0.156**

Table 1. Quantitative comparison results among our method and previous methods on a real-world dataset.

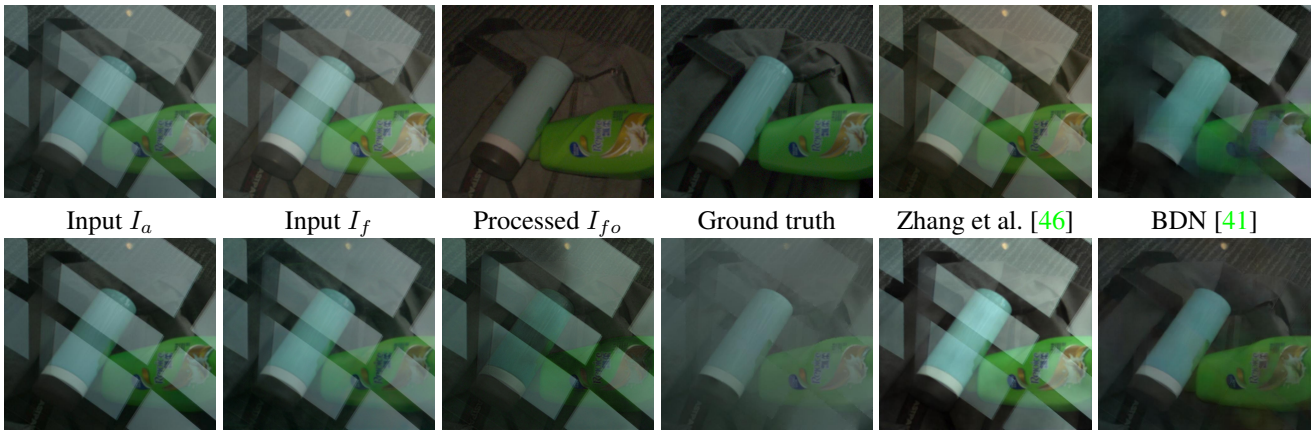


Figure 7. Qualitative comparison to baselines on a real-world image that contains strong reflection.

real-world images.

Synthetic data. Since the real-world dataset cannot provide enough data for training, we construct an extra synthetic dataset. We use 1964 ambient transmission images T_a and flash-only transmission images I_{fo} from a flash dataset [2]. Two kinds of reflections R_a are provided for each T_a to synthesize the ambient image I_a. The first type of reflection is real-world reflections collected by Wan et al. [36]. Then, since there are many blurry reflections and few sharp reflections in their dataset [36], we use an arbitrary ambient image that is quite sharp as the second type of reflection. We reverse gamma correction to mimic the raw data and synthesize I_a^{raw} by $I_a^{raw} = R_a^{raw} + T_a^{raw}$.

Dataset split. For the real-world dataset, we use 77, 30, 50 sets of images for training, validation, and evaluation. There is no overlapping reflection or transmission between the training and test sets. The synthetic data is only used as a supplement for training since the real-world reflection images in CoRRN [36] are in a chaotic order, and no dataset split is available.

6. Experiments

6.1. Comparison to Baselines

We first select two flash-based reflection removal methods: Agrawal et al. [1] and SDN [5]. Then we select several single image methods for comparison, including Zhang et al. [46], Wei et al. [37], BDN [41], Li et al. [21], and Kim et al. [16]. For Agrawal et al. [1], we observe that it is wrongly used in the comparison of SDN [5]: they use the flash image instead of the flash-only image as guidance; in our comparison, we adopt the flash-only image as the input to Agrawal et al. [1]. For SDN [5], we use predicted ambient transmission for quantitative comparison. We retrain the models whose training codes are available on our constructed training set and choose the better results between pretrained models and retrained models.

In Table 1, we adopt PSNR, SSIM, and LPIPS [45] as quantitative evaluation metrics, and our model obtains the best scores on all metrics. Specifically, our method outperforms state-of-the-art reflection removal approaches by more than 5.23dB in PSNR, 0.04 in SSIM, and 0.068 in LPIPS on the real-world dataset.

In Fig. 7, we compare our approach with all mentioned baselines. Both single image baselines [16, 21, 37, 41, 46] and flash-based baselines [1, 5] cannot correctly remove reflection. As can be seen, our approach can remove very strong reflection and recover underlying transmission. It is because processed flash-only image I_{fo} is still reflection-free for strong reflection, and thus provides strong guidance.

In Fig. 8, we further compare our method with single image methods [21, 37] that obtains quantitative scores. In the

Input ambient image | Processed I_{fo} | Li et al. [21] | Wei et al. [37] | Ours

Figure 8. Qualitative comparison to single image based baselines [21, 37] on real-world images.

Input I_a | Li et al. [21] | Ours

Figure 9. Comparison to Li et al. [21] on a real-world image that contains blurry transmission.

Architecture	Input	PSNR↑	SSIM↑	LPIPS↓
g_B	I_a+I_f	26.99	0.911	0.204
g_B	I_a+I_{fo}	27.55	0.917	0.187
$g_R + g_T$	I_a	25.13	0.888	0.258
$g_R + g_T$	I_a+I_f	27.21	0.917	0.196
$g_R + g_T$	I_a+I_{fo}	**29.76**	**0.930**	**0.156**

Table 2. Quantitative comparison among our complete model and multiple ablated models of our methods on a real-world dataset.

first row, the edge of reflection is sharp. The second row is a picture of calligraphy writing, in which both reflection and transmission have rare semantic information. As can be seen, two single-image methods [21, 37] cannot remove the reflections. Our method removes reflections well since the reflection-cue is independent of the appearance (e.g., smoothness and semantic information) of reflection.

In Fig. 9, we compare with Li et al. [21] on an image that contains blurry transmission. The result of Li et al. [21] remove transmission wrongly since their method cannot distinguish the reflection correctly. As a comparison, our approach can easily distinguish the reflection and avoid removing transmission wrongly because the reflection-free cue is independent to smoothness.

In Fig. 7 and Fig. 10, we compare our method with two flash-based methods [1, 5]. The results of Agrawal et al. [1] usually remove too many details and cannot completely remove reflection. For SDN [5], they can remove weak reflection but cannot remove strong reflection. It is because they require the reflection is well suppressed, but the strong reflection cannot be suppressed by flash. Our method removes both weak and strong reflection. Also, the details and color are consistent with ambient images in our results.

6.2. Ablation Study

Reflection-free cues. Although it is quite simple to compute the reflection-free flash-only image, it can improves quantitative and qualitative results a lot. To demonstrate the importance of I_{fo}, we modify the input of the first network: (1) Replace I_{fo} with I_f. (2) Use a single I_a as input. Table 2 shows the quantitative results. Under the same training setting, using a single I_a gets the worst scores, and replacing I_{fo} with I_f also degrades the performance.

We find that the weakness of using the flash image I_f instead of the flash-only image I_{fo} is similar to SDN [5] in the qualitative comparison. In Fig. 5, the reflection is not well suppressed by flash, but the flash-only image is still reflection-free. In this case, replacing I_{fo} with I_f performs poorly. Another example is also shown in Fig. 11(d), replacing I_{fo} by I_f leads to obvious artifacts when reflection cannot be suppressed by the flash. Moreover, it cannot handle novel shadows brought by the flash.

Dedicated architecture. As introduced in Sec. 4.2, the dedicated architecture is vital to avoid absorbing artifacts of flash-only images. In Fig. 5, the base model can remove the reflection well, but artifacts (e.g., color distortion) appear in the result. As a comparison, the result of complete model

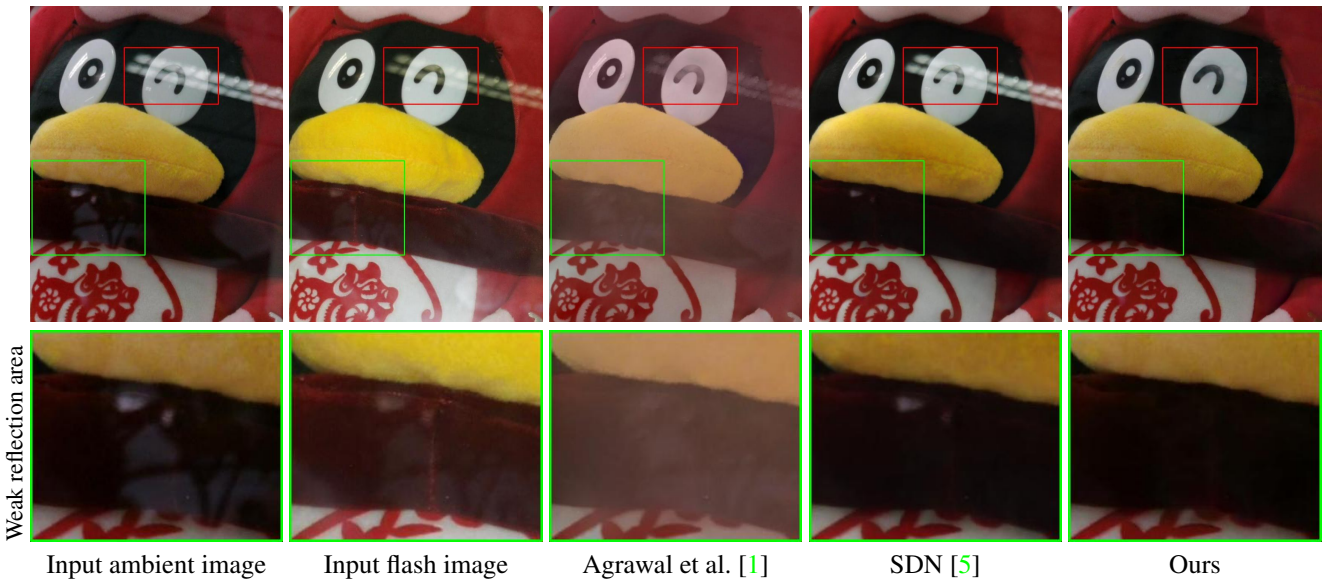


Figure 10. Qualitative comparison to flash-based reflection removal baselines on real-world images. Results of Agrawal et al. [1] contain reflection residuals and are over-smooth. For SDN [5], they can remove the weak reflection but cannot remove the strong reflection.

(a) Input I_a (b) Input I_f (c) Processed I_{fo} (d) Our $\hat{T}$: using I_f (e) Our $\hat{T}$: using I_{fo}

Figure 11. Qualitative comparison between using $I_a + I_f$ and $I_a + I_{fo}$ as input. Note that the edge suppressed in flash image (b) is removed in (d). However, most reflections are not suppressed, and thus the perceptual quality of (d) is not good.

does not contain obvious artifacts. In addition to, using our dedicated architecture also improves the quantitative performance a lot, as shown in Table 2.

Note that although 'g_B, $I_a + I_{fo}$' has similar quantitative scores with '$g_R + g_T, I_a + I_f$', the reasons for degradation are different: we observe the former can remove most reflection but usually has artifacts of flash-only images; the latter generally cannot remove strong reflection correctly.

6.3. Discussion

Flash-only images can be applied to various tasks. In addition to our reflection-free flash-only cues, we also notice other attractive cues exist in the flash-only images. For example, the shadow of the ambient image is invisible in the corresponding flash-only image. Namely, shadow-free flash-only cues also exist and they can be used for shadow removal [24]. Besides, there is only a single light source in the flash-only image, which is an important property to many tasks (such as photometric stereo [7]). We believe more applications of flash-only images are yet to be studied.

7. Conclusion

We propose a very simple yet effective cue called *reflection-free cue* for reflection removal, which is independent of the appearance and strength of reflection. The reflection-free cue is based on the fact that objects in reflection do not directly receive light from the flash and the reflected flash is weak. With a reflection-free flash-only image as guidance, estimating the reflection becomes much easier. Since the flash-only image has obvious artifacts, we propose a dedicated architecture to avoid absorbing artifacts of flash-only images and utilize the cue better. As a result, our model outperforms state-of-the-art methods significantly on a real-world dataset. Also, the qualitative results show that our method can robustly remove various kinds of reflections. We also analyze the flash-based method's feasibility and find it simple to continuously take two images, making it practical in real-world applications.

Acknowledgements

We thank Xuaner Zhang, Changlin Li, Yazhou Xing, and anonymous reviewers for helpful discussions on the paper.

References

[1] Amit Agrawal, Ramesh Raskar, Shree K Nayar, and Yuanzhen Li. Removing photography artifacts using gradient projection and flash-exposure sampling. In *SIGGRAPH*. 2005. 2, 6, 7, 8

[2] Yagiz Aksoy, Changil Kim, Petr Kellnhofer, Sylvain Paris, Mohamed Elgharib, Marc Pollefeys, and Wojciech Matusik. A dataset of flash and ambient illumination pairs from the crowd. In *ECCV*, 2018. 6

[3] Nikolaos Arvanitopoulos, Radhakrishna Achanta, and Sabine Susstrunk. Single image reflection suppression. In *CVPR*, 2017. 1, 2

[4] Xu Cao, Michael Waechter, Boxin Shi, Ye Gao, Bo Zheng, and Yasuyuki Matsushita. Stereoscopic flash and no-flash photography for shape and albedo recovery. In *CVPR*, 2020. 2

[5] Yakun Chang, Cheolkon Jung, Jun Sun, and Fengqiao Wang. Siamese dense network for reflection removal with flash and no-flash image pairs. *Int. J. Comput. Vis.*, 128(6):1673–1698, 2020. 1, 2, 6, 7, 8

[6] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, 2018. 4

[7] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K. Wong. Self-calibrating deep photometric stereo networks. In *CVPR*, 2019. 8

[8] Mark S Drew, Cheng Lu, and Graham D Finlayson. Removing shadows using flash/noflash image edges. In *ICME*, 2006. 2

[9] Qingnan Fan, Jiaolong Yang, Gang Hua, Baoquan Chen, and David Wipf. A generic deep architecture for single image reflection removal and image smoothing. In *ICCV*, 2017. 2

[10] H. Farid and E. H. Adelson. Separating reflections and lighting using independent components analysis. In *CVPR*, 1999. 2

[11] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, 2014. 2

[12] Xiaojie Guo, Xiaochun Cao, and Yi Ma. Robust separation of reflection from multiple images. In *CVPR*, 2014. 2

[13] Byeong-Ju Han and Jae-Young Sim. Reflection removal using low-rank matrix completion. In *CVPR*, 2017. 2

[14] James T Kajiya. The rendering equation. In *SIGGRAPH*, 1986. 3

[15] Hakki Can Karaimer and Michael S Brown. A software platform for manipulating the camera imaging pipeline. In *ECCV*, 2016. 4

[16] Soomin Kim, Yuchi Huo, and Sung-Eui Yoon. Single image reflection removal with physically-based training images. In *CVPR*, 2020. 2, 6

[17] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *ICLR*, 2015. 5

[18] Naejin Kong, Yu-Wing Tai, and Joseph S. Shin. A physically-based approach to reflection separation: From physical modeling to constrained optimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(2):209–221, 2014. 2

[19] Chenyang Lei, Xuhua Huang, Mengdi Zhang, Qiong Yan, Wenxiu Sun, and Qifeng Chen. Polarized reflection removal with perfect alignment in the wild. In *CVPR*, 2020. 1, 2, 4, 5

[20] Chenyang Lei, Yazhou Xing, and Qifeng Chen. Blind video temporal consistency via deep video prior. In *NeurIPS*, 2020. 5

[21] Chao Li, Yixiao Yang, Kun He, Stephen Lin, and John E Hopcroft. Single image reflection removal through cascaded refinement. In *CVPR*, 2020. 2, 4, 6, 7

[22] Yu Li and Michael S Brown. Exploiting reflection change for automatic reflection removal. In *ICCV*, 2013. 2

[23] Yu-Lun Liu, Wei-Sheng Lai, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Learning to see through obstructions. In *CVPR*, 2020. 2

[24] Cheng Lu, Mark S Drew, and Graham D Finlayson. Shadow removal via flash/noflash illumination. In *MMSP*, 2006. 8

[25] Youwei Lyu, Zhaopeng Cui, Si Li, Marc Pollefeys, and Boxin Shi. Reflection separation using a pair of unpolarized and polarized images. In *NeurIPS*, 2019. 2

[26] Daiqian Ma, Renjie Wan, Boxin Shi, Alex C. Kot, and Ling-Yu Duan. Learning to jointly generate and separate reflections. In *ICCV*, 2019. 2

[27] Shree K Nayar, Xi-Sheng Fang, and Terrance Boult. Separation of reflection components using color and polarization. *Int. J. Comput. Vis.*, 21(3):163–186, 1997. 2

[28] Ronneberger Olaf, Fischer Philipp, and Brox Thomas. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015. 5

[29] Wieschollek Patrick, Gallo Orazio, Gu Jinwei, and Kautz Jan. Separating reflection and transmission images in the wild. In *ECCV*, 2018. 2

[30] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael F. Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM Trans. Graph.*, 23(3):664–672, 2004. 2, 3

[31] Li Rui, Qiu Simeng, Zang Guangming, and Heidrich Wolfgang. Reflection separation via multi-bounce polarization state tracing. In *ECCV*, 2020. 2

[32] YiChang Shih, Dilip Krishnan, Fredo Durand, and William T Freeman. Reflection removal using ghosting cues. In *CVPR*, 2015. 1, 2

[33] Chao Sun, Shuaicheng Liu, Taotao Yang, Bing Zeng, Zhengning Wang, and Guanghui Liu. Automatic reflection removal using gradient intensity and motion cues. In *ACM MM*, 2016. 2

[34] Jian Sun, Yin Li, Sing Bing Kang, and Heung-Yeung Shum. Flash matting. In *SIGGRAPH*, 2006. 2

[35] Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot. Benchmarking single-image reflection removal algorithms. In *ICCV*, 2017. 1

[36] Renjie Wan, Boxin Shi, Haoliang Li, Ling-Yu Duan, Ah-Hwee Tan, and Alex Kot Chichung. Corrn: Cooperative reflection removal network. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(12):2672–2680, 2019. 6

[37] Kaixuan Wei, Jiaolong Yang, Ying Fu, David Wipf, and Hua Huang. Single image reflection removal exploiting misaligned training data and network enhancements. In *CVPR*, 2019. 2, 6, 7

[38] Qiang Wen, Yinjie Tan, Jing Qin, Wenxi Liu, Guoqiang Han, and Shengfeng He. Single image reflection removal beyond linearity. In *CVPR*, 2019. 2

[39] Yazhou Xing, Zian Qian, and Qifeng Chen. Invertible image signal processing. In *CVPR*, 2021. 4

[40] Tianfan Xue, Michael Rubinstein, Ce Liu, and William T. Freeman. A computational approach for obstruction-free photography. *ACM Trans. Graph.*, 34(4):79:1–79:11, 2015. 2

[41] Jie Yang, Dong Gong, Lingqiao Liu, and Qinfeng Shi. Seeing deeply and bidirectionally: a deep learning approach for single image reflection removal. In *ECCV*, 2018. 4, 6

[42] Yang Yang, Wenye Ma, Yin Zheng, Jian-Feng Cai, and Weiyu Xu. Fast single image reflection suppression via convex optimization. In *CVPR*, 2019. 1, 2

[43] Y. Schechner Yoav, Shamir Joseph, and Kiryati Nahum. Polarization-based decorrelation of transparent layers: The inclination angle of an invisible surface. *ICCV*, 1999. 2

[44] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *CVPR*, 2020. 4

[45] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 6

[46] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *CVPR*, 2018. 2, 6