FLASH/NO-FLASH IMAGE INTEGRATION USING CONVEX OPTIMIZATION

Tatsuya Baba, Ryo Matsuoka*, Shunsuke Ono§, Keiichiro Shirai†, and Masahiro Okuda**

*Univ. of Kitakyushu, §Tokyo Institute of Technology, †Shinshu Univ.

ABSTRACT

When high ISO sensitivity is used to acquire images of dark scenes, their detail textures are often deteriorated by sensor noise. On the other hand, using flash photography with artificial light, one can shorten the exposure time, and obtain a sharp image under the low ISO sensitivity. However, the use of flash light changes the color tone and often generates unnatural images due to a specific color temperature of the additional light. This paper presents a new efficient method for flash/no-flash image integration. In contrast to conventional integration methods assuming that the flash image has a sharp texture without any noise, our method can successfully remove noise. Specifically, our method separately handles regions within the reach of the flash light and other regions out of range of the flash light, because the two regions have much different characteristics. As for the former well-exposed regions, we transferred the detail of the flash image to no-flash image by optimization and component separation. As for the latter under-exposed regions, an optimization based joint bilateral filtering that uses information of a flash image is performed to remove noise. Experimental results show the effectiveness of our method compared to the conventional methods.

***Index Terms*—** flash/no-flash image integration, image denoising, convex minimization

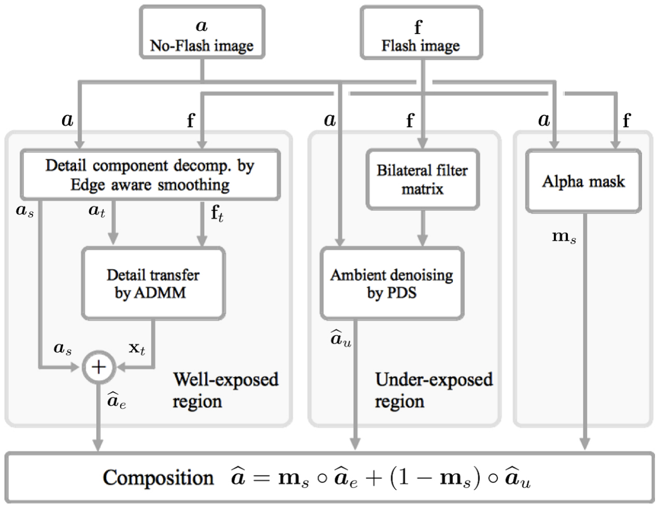

Fig. 1. Flowchart of flash/no-flash image integration.

1. INTRODUCTION

In the photographing of dark scenes, one often faces deterioration of images due to motion blur or sensor noise. One solution to the problem is to utilize flash photography. The use of flash, however, may change the color and atmosphere of a scene, which results in an unnatural image, since the color temperature of the artificial light is often unsuitable for the scene. On the other hand, photography with high ISO sensitivity can shorten its exposure time and reduces motion blur artifacts, but it also gains up the influence of various noise.

Petschnigg et al. [1] addresses the problem by recombining the structure and texture of the flash and no-flash images. To perform this, joint-bilateral filtering (joint-BLF) [2] is introduced, i.e., the weights of a bilateral filter for the unexposed regions of a no-flash image are obtained from the corresponding well-exposed regions of a flash image. Then sharp edges of the flash image (detail image), and vivid colors of the no-flash image (ambient image) are simultaneously restored. However, some weak edges that are not preserved in joint-BLF yield discoloration (color unevenness) and halo artifacts along the edges.

The guided filter [3, 4] can also deal with these problems. They consider a local linear model between a guidance (flash) image and an input (no-flash) image, and yield a noiseless image by color transform. In particular, [3] is able to reduce the effect of shadows and specular lights, and suitable for flash/no-flash integration. HaCohen et al.'s [5] propose, a non-rigid color mapping method, that is applicable to the flash/no-flash problem. As a great advantage, the method dues not require image alignment unlike the aforementioned methods. Unfortunately the performance becomes unstable when the intensities of two images are quite different.

The problem is that all the aforementioned methods assume flash images have sharp textures without noise. However when taking a dark scene in practice, they often have underexposed noisy regions. This paper presents a new effective method that can handle more practical situations. In our framework, for the regions where a flash image has more detailed texture than a no-flash image, we transfer the detail of a flash image to the corresponding no-flash image by minimizing an appropriate cost function. For the regions where the flash image is unclear, we perform edge preserving filtering by using a reference image that is achieved by solving a convex optimization problem. Finally, the two results are integrated using an alpha map. As a result, our method is better than the conventional methods.

2. PROPOSED METHOD

2.1. Outline

In our method, a pair of flash and no-flash image is used to restore the deteriorated no-flash image that has the atmosphere of a scene. Since the effect of flash light fades with increasing distance from a camera, distant regions appear dark due to the underexposure. We call underexposed regions "no-flash regions". On the other hand, regions where the flash light adequately reaches is called "flash re-

Thanks to JSPS and KDDI foundation for funding.

gions". Fig. 1 illustrates the outline of our method. Specifically, first, each region is processed by different methods that are elaborated on in Sec. 2.2 and 2.3. Then the integration of images is performed by a blending method described in Sec. 2.4.

2.2. Edge Transfer

First we describe a denoising method for the flash region, that is, the area within an effective range of flash. We assume that the no-flash image is taken with a high ISO sensitivity and an appropriate setting of the white-balance, while the flash image is taken with a lower ISO sensitivity and shorter exposure time. Our aim of this section is to denoise the noisy no-flash image by transferring the sharp edges of the flash image while preserving the color balance.

Let $\mathbf{f}$ and $\boldsymbol{a} \in \mathbb{R}^N$ be a vectorized flash image and a no-flash image, where N denote the number of pixels. The flash and no-flash images are decomposed into the structure and the texture component as shown in Fig. 2. The structure component consists of almost flat areas with uniform values or smooth gradation. The texture component consists of lines and contours with strong directionality. The structure components $\mathbf{f}_s$ and $\boldsymbol{a}_s$ are obtained by edge-preserving filter for $\mathbf{f}$ and $\boldsymbol{a}$. As the filter, we use domain transform filter [6] because of its fast execution time and the comparable performance to the bilateral filter [7]. The texture component of the two images are simply given by

$$\boldsymbol{a}_t = \boldsymbol{a} - \boldsymbol{a}_s, \qquad \mathbf{f}_t = \mathbf{f} - \mathbf{f}_s.$$

Although the edge-preserving smoothing methods possess a denoising capability, they often fail to keep edges. Thus the texture component $\boldsymbol{a}_t$ may contain not only noise but also edges of a scene. To extract the edges from $\boldsymbol{a}_t$, We apply an optimization-based smoothing operation only to the texture component. Our method is based on the assumption that the noise is only included in the texture component $\boldsymbol{a}_t$ of the no-flash images. The color tone of the scene is kept in the structure component $\boldsymbol{a}_s$. The restored image is given by solving the following convex optimization problem:

$$\widehat{\boldsymbol{a}}_e = \boldsymbol{a}_s + \arg\min_{\mathbf{x}_t} \left\{ \gamma \|\mathbf{D}\mathbf{x}_t - \mathbf{D}\mathbf{f}_t\|_1 + \frac{1}{2}\|\mathbf{x}_t - \boldsymbol{a}_t\|_2^2 \right\}, \quad (1)$$

where $\|\cdot\|_p$ denotes l_p norm, $\mathbf{D} = [\mathbf{D}_x; \mathbf{D}_y]^T \in \mathbb{R}^{2N\times N}$ is the vertically concatenated first-order difference operators $\mathbf{D}_x$ and $\mathbf{D}_y \in R^{N\times N}$ for the horizontal and vertical directions. The term $\|\mathbf{D}\mathbf{x}_t - \mathbf{D}\mathbf{f}_t\|_1$ can be regarded as the regularization term with anisotropic total variation (TV-term) acting on the high frequency components of $\mathbf{x}_t - \mathbf{f}_t$. This TV-term plays the role of transferring the edges of the flash image to $\widehat{\boldsymbol{a}}_e$, and the l_2-term is introduced to correct the color tone around the edges.

The optimization problem in (1) is solved by the alternating direction method of multipliers (ADMM) [8, 9]. By choosing $\mathbf{z}^0, \mathbf{u}^0 \in R^{2N}$ and $\rho \in (0, \infty)$, an ADMM specialized for the problem (1) is

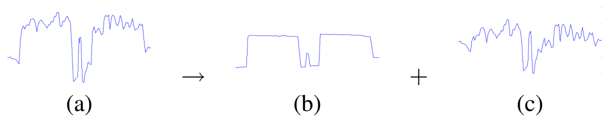

Fig. 2. Structure/Texture decomposition: (a) original 1D signal, (b) structure component, (c) texture component.

given by

$$\left| \begin{array}{l} \mathbf{x}_t^{k+1} := \arg\min_{\mathbf{x}_t} \frac{1}{2}\|\mathbf{x}_t - \boldsymbol{a}_t\|_2^2 + \frac{\rho}{2}\|\mathbf{D}\mathbf{x}_t - \mathbf{D}\mathbf{f}_t - \mathbf{z}^k + \mathbf{u}^k\|_2^2 \\ \mathbf{z}^{k+1} := \arg\min_{\mathbf{z}} \gamma\|\mathbf{z}\|_1 + \frac{\rho}{2}\|\mathbf{D}\mathbf{x}_t^{k+1} - \mathbf{D}\mathbf{f}_t - \mathbf{z} + \mathbf{u}^k\|_2^2 \\ \mathbf{u}^{k+1} := \mathbf{u}^k - (\mathbf{D}\mathbf{x}_t^{k+1} - \mathbf{D}\mathbf{f}_t - \mathbf{z}^{k+1}). \end{array} \right. \quad (2)$$

Since the first step of the algorithm is a quadratic optimization problem w.r.t. $\mathbf{x}_t$, the solution is obtained by solving a matrix inversion, where the matrix can be diagonalized using the fast Fourier transform (FFT) thanks to its block circulant with circulant blocks (BCCB) structure. Thus, we can efficiently solve the equation as follows:

$$\mathbf{x}_t^{k+1} = \mathscr{F}^{-1}\left\{ \frac{\mathscr{F}(\boldsymbol{a}_t) + \rho(\overline{\mathscr{F}(\partial_x)}\mathscr{F}(\dot{\mathbf{z}}_x) + \overline{\mathscr{F}(\partial_y)}\mathscr{F}(\dot{\mathbf{z}}_y))}{\mathscr{F}(\mathbf{I}) + \rho(\overline{\mathscr{F}(\partial_x)}\mathscr{F}(\partial_x) + \overline{\mathscr{F}(\partial_y)}\mathscr{F}(\partial_y))} \right\}, \quad (3)$$

where $\dot{\mathbf{z}}_x = \mathbf{z}_x^k + \mathbf{D}_x\mathbf{f}_t - \mathbf{u}_x^k$ and $\dot{\mathbf{z}}_y = \mathbf{z}_y^k + \mathbf{D}_y\mathbf{f}_t - \mathbf{u}_y^k$. Furthermore $\mathbf{z}_{x,y}$ and $\mathbf{u}_{x,y}$ are horizontal and vertical direction components, respectively. $\mathscr{F}(\cdot)$ and $\overline{\mathscr{F}(\cdot)}$ denote FFT and its complex conjugate. $\mathscr{F}(\mathbf{I})$ is a FFTed delta function. ∂_x and ∂_y are difference operators corresponding to $\mathbf{D}_x$ and $\mathbf{D}_y$, respectively. Then, the solution of the second step in the algorithm is simply given by a soft-thresholding operation For details, refer to [10].

2.3. Under-exposed region

As for the no-flash regions in a flash and a no-flash image, the no-flash image has shape edges but suffers from noise, the flash image has less noise but suffers from low contrast. Since the no-flash regions of a flash image are often under-exposed, applying the method in Sec. 2.2 to the no-flash domain tends to excessively smooth the edges. To denoise the image $\boldsymbol{a}$ using $\mathbf{f}$ as a guidance image, we introduce the following convex optimization problem for obtaining a denoised image $\boldsymbol{a}_u$:

$$\widehat{\boldsymbol{a}}_u = \arg\min_{\mathbf{x}_a} \left\{ \lambda\|(\mathbf{I} - \mathbf{B})\mathbf{x}_a\|_1 + \frac{1}{2}\|\mathbf{x}_a - \boldsymbol{a}\|_2^2 \right\}, \quad (4)$$

where $\mathbf{I} \in \mathbb{R}^{N\times N}$ is an identity matrix, and $\mathbf{B} \in \mathbb{R}^{N\times N}$ is a bilateral filter matrix for joint-BLF. The coefficient $b_{i,j}$ is calculated from the flash image:

$$b_{i,j} = \frac{1}{C_i} g_s(i-j) g_r(f_i - f_j), \quad (5)$$

where $C_i = \sum_j g_s(i-j)g_r(f_i - f_j)$ is the normalization term, $j \in \mathcal{N}(i)$ is neighboring pixels centered at a pixel i, and $g_s(\cdot)$ and $g_r(\cdot)$ are weight functions of the bilateral filter for the spatial domain and range domain. As for the function g, we use a Gaussian function $\{g_i(x) = \exp(-\|x-\mu\|^2/\sigma_i^2) |\, i \in \{s, r\}\}$, where $\mu = 0$ is the mean of x and σ is the standard deviation of x. The detail of parameters σ_s, σ_r are described in Sec. 3.

In this case, $\mathbf{B}^T\mathbf{B} \in \mathbb{R}^{N\times N}$ is not a BCCB matrix, and it cannot be diagonalized by FFT. Therefore applying ADMM to (4) requires some iterative schemes, such as the conjugate gradient method [11, 12], in the first step. Hence, instead of the ADMM, we use the primal-dual splitting algorithm [13], which has also been recently

(a) No-flash (b) Flash (c) Alpha map

Fig. 3. Example of alpha map.

applied to color image restoration [14]. A primal-dual splitting algorithm specialized for the problem in (4) is given by

$$\left\lfloor \begin{array}{l} \mathbf{x}_a^{k+1} := \mathbf{x}_a^k - \tau_1(\mathbf{x}_a^k - \boldsymbol{a}) - \tau_1 \mathbf{L}^T \mathbf{z}^k \\ \widetilde{\mathbf{z}}^{k+1} := \mathbf{z}^k + \tau_2 \mathbf{L}(2\mathbf{x}_a^{k+1} - \mathbf{x}_a^k) \\ \mathbf{z}^{k+1} := \widetilde{\mathbf{z}}^{k+1} - \tau_2 \Psi_{\lambda/\tau_2}(\tau_2^{-1} \widetilde{z}^{k+1}), \end{array} \right. \tag{6}$$

where $\mathbf{L} = \mathbf{I} - \mathbf{B}$ and Ψ is the soft thresholding (shrinkage) operator defined as $\Psi_\kappa(\mathbf{z}) = (\mathbf{z} - \kappa)_+ - (-\mathbf{z} - \kappa)_+$ where $(\mathbf{z})_+ = \max(\mathbf{z}, 0)$ takes nonnegative part of each element of $\mathbf{z}$.

2.4. Alpha Blending

We finally integrate the two images $\widehat{\boldsymbol{a}}_e$ and $\widehat{\boldsymbol{a}}_u$ obtained from (1) and (4) using alpha blending. Assuming the final image consists of flash regions and no-flash regions, we aim to assign the regions of $\boldsymbol{a}_e$ and $\boldsymbol{a}_u$ to the adequate regions of the final image. The alpha map $\mathbf{m}$ is created by comparing the luminance of original flash and no-flash image as

$$\mathbf{m}_i = \begin{cases} 1 & \text{if} \quad \mathbf{f}_i \geq \boldsymbol{a}_i \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

The discrete separation of blending coefficients by $\{0, 1\}$, however causes artifacts around the edges such as double contour and noise. Therefore we use Gaussian filtering to smooth the alpha map, and denote it as $\mathbf{m}_s$. The standard deviation of the Gaussian filter is set as 2.0.

Finally, the images $\widehat{\boldsymbol{a}}_e$ and $\widehat{\boldsymbol{a}_u}$ are integrated to produce the final image $\widehat{\boldsymbol{a}}$ by

$$\widehat{\boldsymbol{a}} = \mathbf{m}_s \circ \widehat{\boldsymbol{a}}_e + (1 - \mathbf{m}_s) \circ \widehat{\boldsymbol{a}}_u, \tag{8}$$

where $\circ$ denotes element-wise multiplication. An example of a flash map is illustrated in Fig.3.

3. EXPERIMENTAL RESULTS

Here we show two experimental results. In the first experiment, the original flash and no-flash images are taken by Canon EOS 20D, and f-number is set to 4.5. ISO sensitivity of the flash and no-flash images are set to 100 and 3200, respectively. The pixel values of each image are normalized to the range [0,1]. The images (no-flash images) used in the experiment are shown in Fig.4(a)-(c), while the images in Fig.4(d)-(g) are used for the numerical evaluation (described later). Fig.5 shows the results for some parts of the three images, in which we show the flash images, the noisy inputs, the results obtained using the conventional methods [1, 3], and our method. The parameters of the conventional methods are carefully adjusted so as to give similar denoising effect. The standard deviations of bilateral filter used in the our method are $\sigma_s = 2$ and $0.01 \leq \sigma_r \leq 0.1$ in this paper (σ_s and σ_r are standard deviations of g_s and g_r in (5)).

Table 1. Comparison in PSNR for images (d)-(g).

Image	noisy	[1]	[3]	Our method
(d)	26.71	32.48	30.82	32.88
(e)	26.88	31.88	29.29	31.85
(f)	26.68	31.53	28.81	31.82
(g)	26.14	28.49	28.22	29.94

One can see from the figure that our method yields images with less noise and sharper edges. The method [1] has some drawbacks such as blurring artifact, especially for regions with some color gradation. In Fig.5(a), the method [1] results in more blurred images than the others. The results of the method [3] performs well, but denoising capability and sharpness are insufficient in some areas. In the flash region, our method restore sharper texture than the others as shown in Fig.5(b). As for the no-flash region, the methods [1, 3] (Fig.5(c-3) and (c-4)) have unclear contrast in some parts, while our method (Fig.5(c-5)) achieves efficient noise reduction without reducing contrast.

Next we show quantitative comparison to show the validity of our method. We prepare clear flash/no-flash image pairs, artificially add white Gaussian noise ($\sigma = 0.05$, $\mu = 0$), and then apply the methods to them. The no-flash image and long exposure image (clear no-flash image) have slightly different color due to the ISO sensitivity. Thus the colors of the resulting images are different from those of long exposures images. To avoid this issue, we do not use the no-flash image but a image that noise add to long exposure image as an image to be used in PSNR. Table 1 shows the results (PSNR) of four images Fig.5(d)-(g). The results show that our method outperforms the others quantitatively as well. Additionally, Fig.5(f) and (g) show close-up images, which illustrate the results of the flash and the no-flash regions, respectively. In Fig.5(g), even though the flash images have only a little texture, our method gains clearer contrast and better scores in PSNR than the other methods.

4. CONCLUSION

Our method deals with the no-flash regions as well, and two methods are applied for the two regions. Meanwhile, the conventional methods for flash/no-flash image integration are focused only on the flash region. Our method solves the two convex optimization problems to the flash/no-flash regions to yield two images for the regions, and then integrate them using the alpha map. Our method significantly improves the denoising capabilities, compared with the conventional image integration methods. Alpha map to integrate a flash/no-flash image is not required strictly. Performance should be improved, if alpha map that can distinguish flash region and no-flash region is obtained. In the end, this paper presented a new approach using convex optimization for integration of a pair image that has pixel of corresponding. However, in real images the flash/no-flash image pair is not registered, and we would like to consider it for our future work.

5. REFERENCES

[1] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama, "Digital photography with flash and no-flash image pairs," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 23, no. 3, pp. 664–672, 2004.

(a) (b) (c) (d) (e) (f) (g)

Fig. 4. Full frames of samples in first experiment (a,b,c) and second experiment (d,e,f,g).

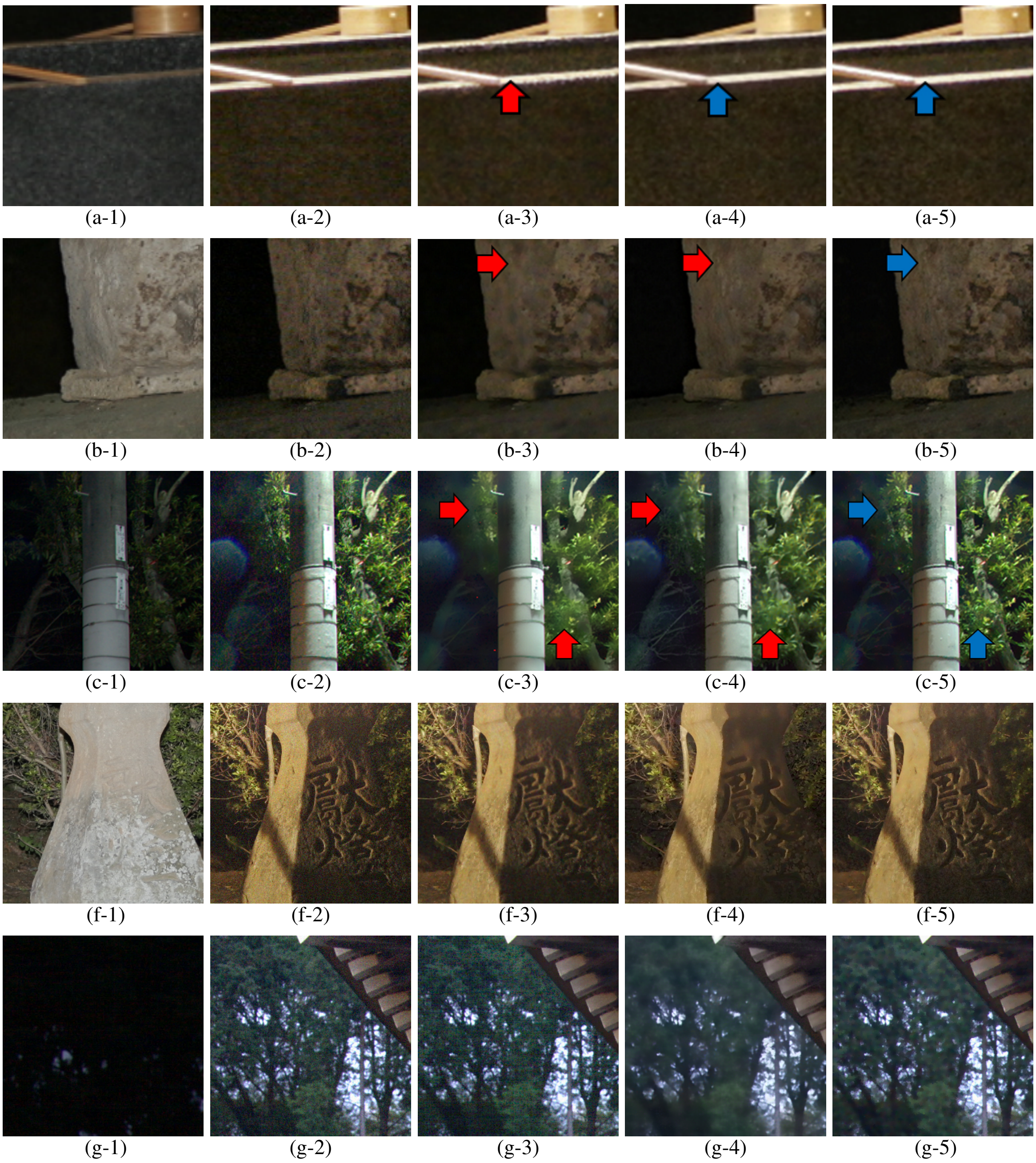

Fig. 5. Results of the first experiment and the second experiment: (∗-1) Flash image, (∗-2) Noisy input (no-flash), (∗-3) Petschnigg et al. [1], (∗-4) Shirai et al. [3], (∗-5) Our method.

[2] Elmar Eisemann and Fro Durand, "Flash photography enhancement via intrinsic relighting," *ACM T. Graph. (Proc. SIGGRAPH)*, vol. 23, pp. 673–678, 2004.

[3] Keiichiro Shirai, Masaaki Ikehara, and Masayuki Okamoto, "Noiseless no-flash photo creation by color transform of flash image," in *Proc. IEEE ICIP*, 2011, pp. 3437–3440.

[4] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," in *Proc. ECCV*, 2010, pp. 1–14.

[5] Yoav HaCohen, Eli Shechtman, Dan B. Goldman, and Dani Lischinski, "Non-rigid dense correspondence with applications for image enhancement," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 30, no. 4, pp. 70:1–70:10, 2011.

[6] Eduardo S. L. Gastal and Manuel M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 30, no. 4, pp. 69:1–69:12, 2011.

[7] Carlo Tomasi and Roberto Manduchi, "Bilateral filtering for gray and color imagegs," in *Proc. IEEE ICCV*, 1998, pp. 839–846.

[8] Jonathan Eckstein and Dimitri P. Bertsekas, "On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, pp. 293–318, 1992.

[9] Manya V. Afonso and M'ario A. T. Figueiredo Jos'e M. Bioucas-Dias, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.

[10] David L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[11] R. M. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Research National Bureau Standards*, vol. 49, no. 6, pp. 409–436, 1952.

[12] R. Barrett, T. F. Chan M. W. Berry, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Vorst, *Templates for the solution of linear systems: building blocks for iterative methods*, SIAM, 1994.

[13] Laurent Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.

[14] Shunsuke Ono and Isao Yamada, "A convex regularizer for reducing color artifact in color image recovery," in *Proc. CVPR*, 2013, pp. 1775–1781.