

Image Enhancement of Low-Light Scenes with Near-Infrared Flash Images

Sosuke Matsui¹, Takahiro Okabe¹, Mihoko Shimano^{1,2}, and Yoichi Sato¹

¹ Institute of Industrial Science, The University of Tokyo

² PRESTO, Japan Science and Technology Agency

{matsui,takahiro,miho,ysato}@iis.u-tokyo.ac.jp

Abstract. We present a novel technique for enhancing an image captured in low light by using near-infrared flash images. The main idea is to combine a color image with near-infrared flash images captured at the same time without causing any interference with the color image. In this work, near-infrared flash images are effectively used for removing annoying effects that are commonly observed in images of dimly lit environments, namely, image noise and motion blur. Our denoising method uses a pair of color and near-infrared flash images captured simultaneously. Therefore it is applicable to dynamic scenes, whereas existing methods assume stationary scenes and require a pair of flash and no-flash color images captured sequentially. Our deblurring method utilizes a set of near-infrared flash images captured during the exposure time of a single color image and directly acquires a motion blur kernel based on optical flow. We implemented a multispectral imaging system and confirmed the effectiveness of our technique through experiments using real images.

1 Introduction

When taking a picture in low light, photographers usually face the dilemma of using flash or not. The quality of an image captured without flash is often degraded by noise and motion blur. On the other hand, noise and motion blur in an image captured with flash are significantly reduced. However, flash causes undesired artifacts such as flat shading and harsh shadows. As a result, the atmosphere of the original scene evoked by dim light is destroyed. Thus, there are positive and negative points with using flash.

We propose two methods for enhancing an image captured in low light according to the following two scenarios. In the first scenario, we reduce the noise of a color image captured in low light with a short exposure time, since the image is not blurry, but contains a significant amount of noise due to large gain or high ISO. In the second scenario, we remove motion blur of a color image captured in low light with a long exposure time, since the image is not noisy, but is blurry due to camera shake or scene motion.

The main idea of our methods is to combine a color image captured without flash and additional near-infrared (NIR) images captured with NIR flash for reducing noise and motion blur in the color image. Because the spectrum of NIR

light is different from that of visible light, we can capture both a color image and NIR images at the same time without causing any interference by using a multispectral imaging system composed of a color camera and an NIR camera. In addition, NIR flash provides sufficient amount of light in the NIR spectrum, thus suppressing noise and motion blur in the NIR images.

Our denoising method uses a pair of color and NIR flash images captured simultaneously, which is applicable to dynamic scenes, whereas existing methods [1,2,3] assume stationary scenes and require a pair of flash and no-flash color images captured sequentially. More specifically, we first decompose the color image into a large-scale image (low-frequency components) and a detail image (high-frequency components); the former mainly includes global textures and shading caused by lighting, and the latter mainly includes subtle textures, edges, and noise. Then, taking the difference in spectrum into consideration, we carefully denoise the detail image by using a novel algorithm termed *joint non-local mean algorithm* which is a multispectral extension of a non-local mean algorithm [4]. Finally, we combine the large-scale and the revised detail images and obtain a denoised color image. We experimentally show that our method works better than Bennett’s method [5] which also uses NIR images to reduce noise in a video shot in a dimly lit environment.

Our deblurring method uses a set of NIR flash images captured during the exposure time of a single color image and directly acquires a motion blur kernel based on optical flow in a similar manner to Ben-Ezra’s method [6], which combines videos with different temporal and spatial resolutions. Then, the Richardson-Lucy deconvolution algorithm [7,8] is used for deblurring the color image. We demonstrate that combining images with different temporal resolutions is effective also for deblurring an image captured in low light by incorporating a multispectral imaging system.

The rest of this paper is organized as follows. We briefly summarize related work in Section 2. We introduce our denoising and deblurring methods in Section 3. We present experimental results in Section 4 and concluding remarks in Section 5.

2 Related Work

We briefly summarize previous studies related to our technique from two distinct points of view: denoising and deblurring.

Denoising

Petschnigg *et al.* [1] and Eisemann and Durand [2] independently proposed methods for denoising an image taken in low light by using a pair of flash and no-flash images captured using a single color camera. They combine the strengths of flash and no-flash images; flash captures details of a scene, and no-flash captures ambient illumination. More specifically, they decompose the no-flash image into a large-scale image and a detail image by using a bilateral filter [9]. Then, they revise the noisy detail image by transferring the details of the scene from the flash image. They recombine the large-scale and the revised detail images and

finally obtain a denoised image. Agrawal *et al.* [3] made use of the fact that the orientation of image gradient is insensitive to illumination conditions and proposed a method for removing artifacts, such as highlights, caused by flash.

However, these methods share common limitations. That is, they assume stationary scenes and require a pair of flash and no-flash color images captured sequentially. On the other hand, our method uses a color image as well as an NIR flash image, which can be captured at the same time without causing any interference with the color image. The use of an NIR flash image enables us to apply our method to dynamic scenes, which is the advantage of our method.

NIR images are used for denoising a video shot in a dimly lit environment [5] and for enhancing contrast and textures of an image of a high-dynamic range scene [10]. Particularly, our method is similar to the former method proposed by Bennett [5] in that it uses a pair of color and NIR images for noise reduction. However, our method differs from it with respect to the manner in which we revise the detail image. Bennett's method revises the detail image in the visible spectrum by transferring the details from the NIR image. Since it combines intensities observed in different spectra, it causes artifacts such as color shifts. On the other hand, we revise the detail image by non-locally averaging the color image with the weights computed based on the NIR flash image. We experimentally show that our method works better for denoising an image of a low-light scene.

Recently, Krishnan and Fergus [11] used dark flash consisting of IR and UV light for denoising an image taken in low light. They achieve dazzle-free flash photography by hiding the flash in invisible spectrum. However, their method also requires a pair of flash and no-flash images captured sequentially, and therefore assumes stationary scenes.

Deblurring

Yuan *et al.* [12] proposed an image enhancement method using a pair of images captured in low light using a single color camera successively with long and short exposure times. Their basic idea is denoising the image with the short exposure time and estimating the motion blur kernel of the image with the long exposure time based on the denoised image. They proposed an iterative deconvolution scheme focusing on the residuals of denoising so that ringing artifacts inherent in image deconvolution are reduced.

On the other hand, our method using a multispectral imaging system is considered to be a hardware approach to image enhancement. This system captures a set of NIR flash images during the exposure time of a single color image and directly acquires the motion blur kernel based on optical flow. Thus, the use of multispectral images makes deconvolution more tractable.

Ben-Ezra and Nayar [6] proposed a hybrid imaging system which captures images of a scene with high spatial resolution at a low frame rate and with low spatial resolution at a high frame rate. They directly measure the motion blur kernel of the image with the low temporal resolution by using the images with high temporal resolution. Recently, Tai *et al.* [13] extended their method assuming a spatially-uniform blur kernel to deal with spatially-varying blur kernels.

One of the main contributions of our study is to demonstrate that combining images with different temporal resolutions is effective also for deblurring an image of a low-light scene by incorporating a multispectral imaging system.

3 Proposed Methods

We explain our methods for removing noise and motion blur in images of dimly lit environments with the help of NIR flash images. We describe our denoising method in Section 3.1 and our deblurring method in Section 3.2.

3.1 Noise Reduction by Using NIR Flash Image

We explain how noise in a color image captured in low light with a short exposure time is reduced by using an NIR flash image, as shown in Fig.1. First, we decompose the color image into a large-scale image and a detail images by using a dual bilateral filter [5]. The former mainly includes global textures and shading caused by lighting, and the latter mainly includes subtle textures, edges, and noise. We preserve the large-scale image as is so that the shading caused by lighting of a scene is preserved. Second, we denoise the detail image by using our joint non-local mean algorithm so that the details are recovered and noise is reduced. Finally, we recombine the large-scale and the revised detail images and obtain a denoised color image.

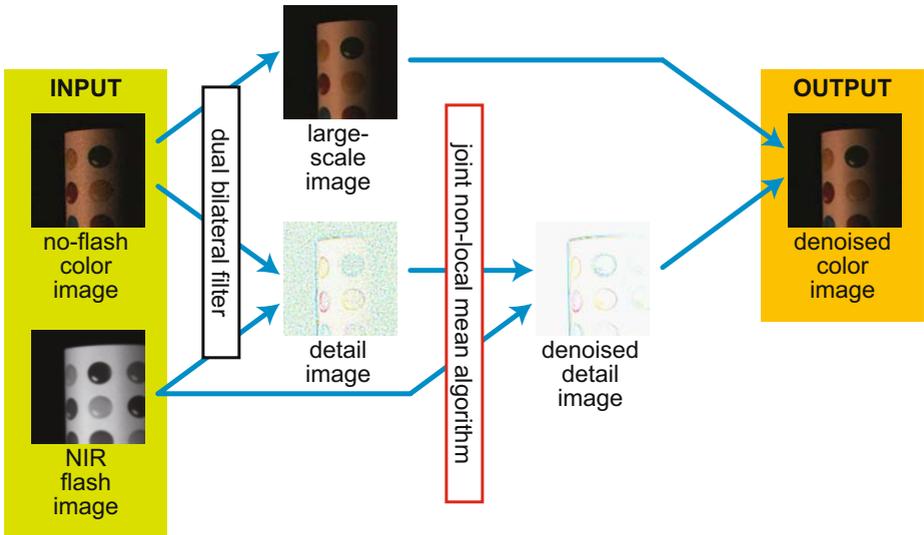


Fig. 1. Flow of our denoising method with help of NIR flash image

Decomposing color image into large-scale and detail images

First, we decompose a color image into a large-scale image and a detail image by using the dual bilateral filter [5]. The dual bilateral filter incorporates the weights calculated based on the NIR channel into a conventional bilateral filter [9]. Since the NIR flash image is captured under sufficient amount of light and is not noisy, the dual bilateral filter significantly alleviates the effects of noise contained in the color image.

More specifically, we convert the color space of an input image I_c from RGB ($c = R, G, B$) to YUV ($c = Y, U, V$). Then, we obtain the large-scale image of the Y component as

$$I_Y^{\text{large_scale}}(\mathbf{p}) = \frac{1}{Z_B(\mathbf{p})} \sum_{\mathbf{q} \in \Omega_B(\mathbf{p})} G_D(\mathbf{p} - \mathbf{q}) \times G_{\text{NIR}}(I_{\text{NIR}}(\mathbf{p}) - I_{\text{NIR}}(\mathbf{q})) G_Y(I_Y(\mathbf{p}) - I_Y(\mathbf{q})) I_Y(\mathbf{q}). \quad (1)$$

Here, $Z_B(\mathbf{p})$ is a normalization constant, and $\Omega_B(\mathbf{p})$ is a certain area around a pixel \mathbf{p} . $I_{\text{NIR}}(\mathbf{p})$, $I_Y(\mathbf{p})$, and $I_Y^{\text{large_scale}}(\mathbf{p})$ are intensities at the pixel \mathbf{p} in the IR, Y, and large-scale images. G_D , G_{NIR} , and G_Y are the weights calculated with the Gaussian functions whose means are zero and variances are σ_D^2 , σ_{NIR}^2 , and σ_Y^2 respectively.

As for the U and V channels, we use the bilateral filter and obtain $I_U^{\text{large_scale}}$ and $I_V^{\text{large_scale}}$. Then, we combine the filtered YUV images and obtain a large-scale image $I_c^{\text{large_scale}}$. By dividing the original color image by the large-scale image, we acquire a noisy detail image $I_c^{\text{noisy_detail}}$ as

$$I_c^{\text{noisy_detail}}(\mathbf{p}) = \frac{I_c(\mathbf{p}) + \epsilon}{I_c^{\text{large_scale}}(\mathbf{p}) + \epsilon}, \quad (2)$$

where ϵ is a small constant for avoiding division by zero.

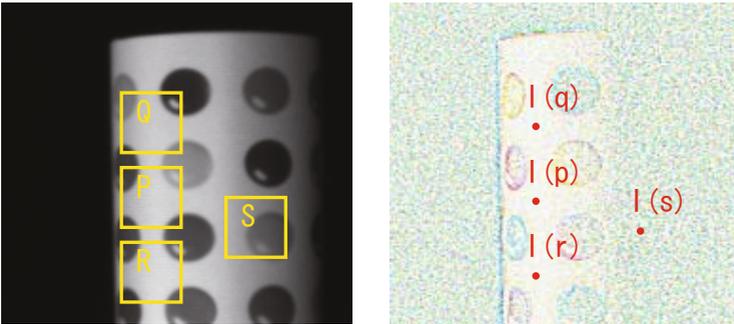


Fig. 2. Basic idea of our joint non-local mean algorithm. Pixel value $I(p)$ is replaced by weighted average of $I(q)$, $I(r)$, $I(s)$, and so on. Larger weights are assigned to $I(q)$ and $I(r)$ with similar local textures Q and R to P , whereas smaller weight is assigned to $I(s)$ with dissimilar local texture S to P .

Denoising detail image using joint non-local mean algorithm

Second, we carefully denoise the noisy detail image by taking the difference in spectra into consideration. In contrast to the existing methods which transfer the details of a scene from a flash image [1,2] or an NIR image [5], we denoise the noisy detail image by non-locally averaging it with the weights computed based on the NIR flash image.

More specifically, we assume that the intensity of a certain pixel is similar to the intensity of another pixel if the appearance of patches around the pixels resemble each other (see Fig.2). Then, a detail image I_c^{detail} is acquired as

$$I_c^{\text{detail}}(\mathbf{p}) = \frac{1}{Z_N(\mathbf{p})} \sum_{\mathbf{q} \in \Omega_N(\mathbf{p})} G(\mathbf{v}(\mathbf{p}) - \mathbf{v}(\mathbf{q})) I_c^{\text{noisy_detail}}(\mathbf{q}). \quad (3)$$

Here, $Z_N(\mathbf{p})$ is a normalization constant, and $\Omega_N(\mathbf{p})$ is a search area around the pixel \mathbf{p} . We represent the appearance of the patch by concatenating the $k \times k$ pixel intensities around the pixel \mathbf{p} into a vector $\mathbf{v}(\mathbf{p})$.

This joint non-local mean algorithm is a multispectral extension of the non-local mean algorithm [4], which uses the appearance of patches in the visible spectrum for determining the weights. Since the NIR flash image is captured under sufficient amount of light and is not noisy, our joint non-local mean algorithm works well even when the color image is captured under dim lighting, and as a result, is significantly contaminated by noise.

We understand that the joint non-local mean algorithm may degrade because the NIR flash image captures the radiance of the scene in a different spectrum from the visible spectrum. However, as far as we know from our experiments, our

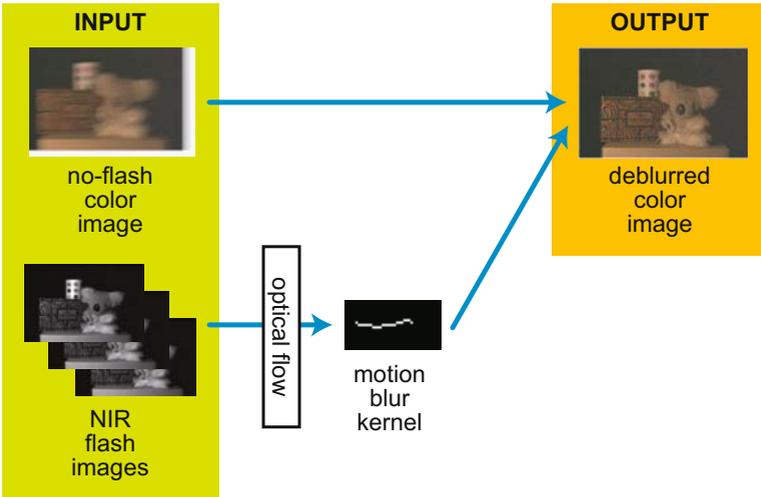


Fig. 3. Flow of our deblurring method with help of successive NIR flash images

algorithm is insensitive to the difference in spectra and outperforms the most closely related method [5].

Combining large-scale and revised detail images

Finally, we recombine the large-scale and the revised detail images and obtain a denoised color image $I_c^{\text{denoised}}(\mathbf{p})$ as

$$I_c^{\text{denoised}}(\mathbf{p}) = I_c^{\text{large-scale}}(\mathbf{p}) \times I_c^{\text{detail}}(\mathbf{p}). \quad (4)$$

3.2 Blur Removal by Using Sequence of NIR Flash Images

We explain how blur in a color image captured in low light with a long exposure time is removed by using NIR flash images, as shown in Fig.3. First, we take a sequence of NIR flash images during the exposure time of a single color image. Then, we directly acquire a motion blur kernel based on optical flow in a similar manner to Ben-Ezra’s method [6]. Finally, we use the Richardson-Lucy deconvolution algorithm [7,8] for deblurring the blurry color image.

Estimating blur kernel from NIR flash images

We assume a spatially-uniform motion and estimate the blur kernel from NIR flash images as follows. First, we compute the motion between successive frames of NIR images by using optical flow. Then, we join the successive motion and obtain the path of the motion during the exposure time of a single color image. Finally, we convert the motion path into the blur kernel by taking the energy conservation constraint into consideration. As for the implementation details, see Ben-Ezra and Nayar [6].

4 Experiments

We implemented a multispectral imaging system composed of a 3CCD color camera and an NIR camera, as shown in Fig.4. The image of a scene is split by a half mirror. We used SONY XC-003 as the color camera, XC-EI50 as the NIR

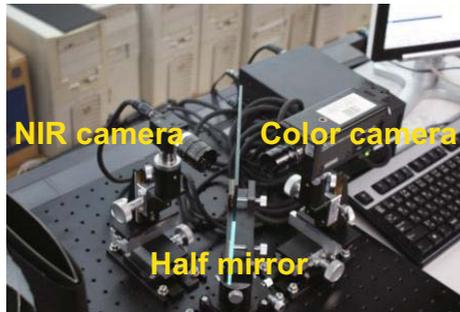


Fig. 4. Prototype of our multispectral imaging system

camera, and a white light source covered with an NIR pass filter. The image coordinates of the two cameras are calibrated based on homography [14].

In the current implementation, we empirically set the parameters as follows. In eq.(1), the variances for the dual bilateral filter are $\sigma_D^2 = 100$, $\sigma_{\text{NIR}}^2 = 87.6$, and $\sigma_Y^2 = 22.5$. $\Omega_B(\mathbf{p})$ is an area with 7×7 pixels around the pixel \mathbf{p} . In eq.(2), ϵ is set to 0.02. In eq.(3), $\Omega_N(\mathbf{p})$ is an area with 21×21 pixels around the pixel \mathbf{p} and $k = 3$. The variance of the Gaussian is set to 1.5.

4.1 Denoising Results

First, we demonstrate that our denoising method is applicable to dynamic scenes. The images shown in the first row in Fig.5 are the input no-flash color images. The dynamic range of the images is linearly expanded for display purpose only. The images in the second row are the images simultaneously captured with the NIR camera. The images in the third row are the close-ups of the bounding

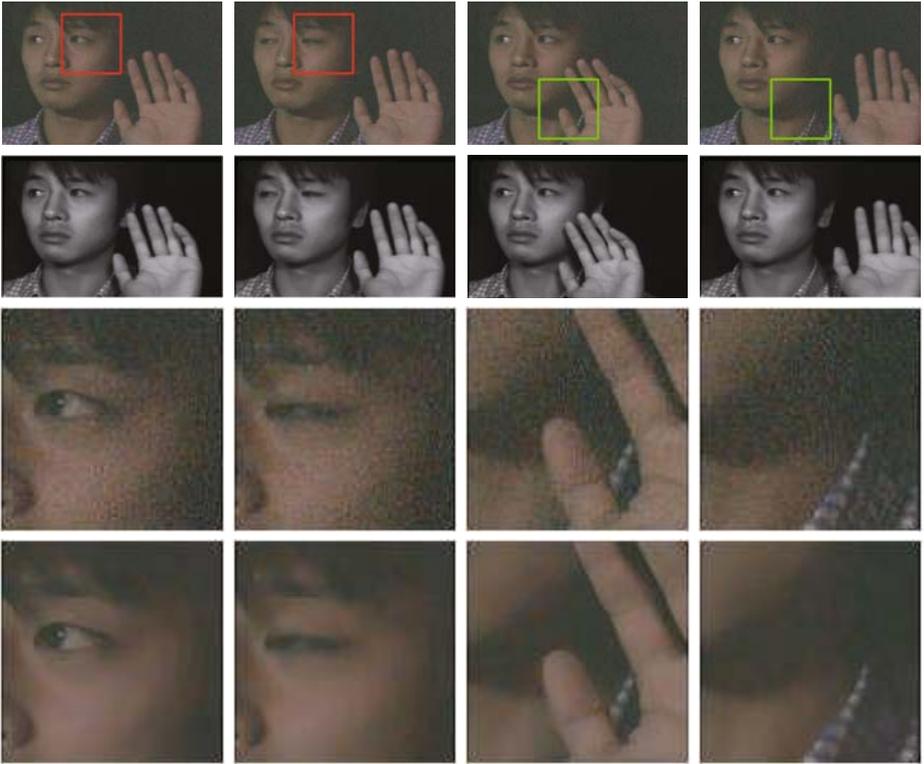


Fig. 5. Results for dynamic scene. Images in first and second rows are color and NIR flash images captured at same time. Images in third row are close-ups of bounding boxes in images in first row. Corresponding results of our method are shown in fourth row.

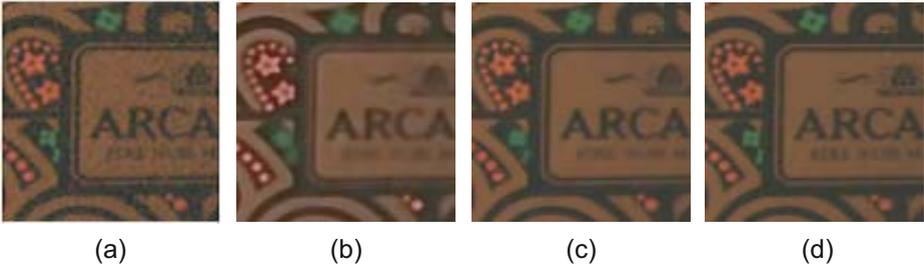


Fig. 6. Results for static scene. (a) input no-flash image, (b) result obtained using Bennett's method, (c) result obtained using our method, and (d) temporal average.

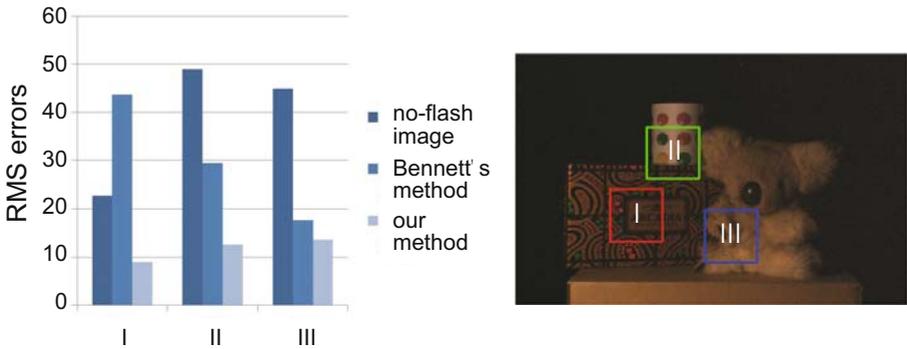


Fig. 7. Quantitative comparison between our method and Bennet's method

boxes in the images in the first row, and the images in the fourth row are the corresponding results. One can see that our method significantly reduces noise in images even for a dynamic scene by using a pair of color and NIR flash images.

Second, we applied our method to a static scene where the temporal average of no-flash color images is considered to be the ground truth of the denoised image if we assume zero-mean image noise. Fig.6 shows (a) an input no-flash color image, (b) the result obtained using the Bennett's method, (c) the result obtained using our method, and (d) the temporal average. One can see that the result obtained from our method resembles the temporal average. On the other hand, one can see that the Bennett's method causes artifacts such as color shifts and blurs. These results demonstrate that our method, which carefully revises the detail image by taking the difference in spectra into consideration, outperforms the Bennett's method.

Next, we quantitatively evaluated the performance of our method. Fig.7 compares the root-mean-square (RMS) errors of the pixel values in the corresponding bounding boxes in the color image. We consider the temporal average image as the ground truth of the denoised image. One can see that our method decreases



Fig. 8. When NIR image is saturated due to highlights caused by NIR flash, our denoising method does not work well

the RMS errors compared with the input color image and outperforms the Bennett’s method.

Finally, Fig.8 demonstrates an example where our method does not work well. One can see that a portion of (b) an NIR flash image is saturated due to highlights caused by NIR flash. In this case, the weights in eq.(3) are large for patches in the highlights because the textures disappeared due to saturations. Thus, (d) the resulting image is blurry.

4.2 Deblurring Results

As shown in Fig.9, we captured (a) nine NIR flash images of a scene during the exposure time of (c) a single no-flash color image. We estimated (b) the spatially-uniform motion blur kernel from the sequence of NIR flash images and obtained (d) the deblurred image. One can see that the motion blur decreases although some artifacts are still visible.

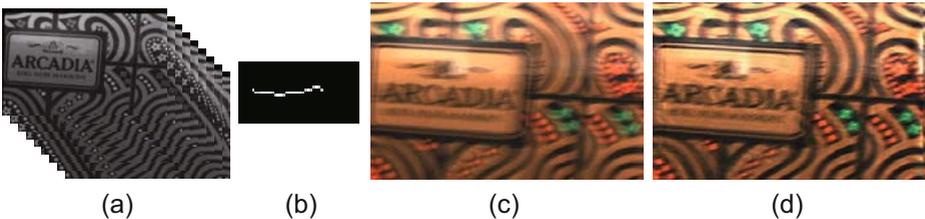


Fig. 9. (a) NIR flash images for estimating motion blur kernel, (b) estimated blur kernel, (c) blurry input image, and (d) deblurred image

5 Conclusions and Future Work

We presented a novel technique for enhancing an image captured in low light by using a multispectral imaging system, which captures a color image and NIR flash images without causing any interference. The experimental results demonstrate that our denoising method using a pair of color and NIR flash

images is applicable to dynamic scenes and outperforms the existing method that is most closely related to ours. We demonstrated that combining images with different temporal resolutions is effective also for deblurring an image of a low-light scene.

The directions of our future work include the enhancement of a noisy and blurry image since our methods are used for a noisy image without blur or a blurry image without noise. In addition, we plan to remove the unpleasant effects of highlights caused by NIR flash.

References

1. Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., Toyama, K.: Digital photography with flash and no-flash image pairs. In: Proc. SIGGRAPH 2004, pp. 664–672 (2004)
2. Eisemann, E., Durand, F.: Flash photography enhancement via intrinsic relighting. In: Proc. SIGGRAPH 2004, pp. 673–678 (2004)
3. Agrawal, A., Raskar, R., Nayar, S., Li, Y.: Removing photography artifacts using gradient projection and flash-exposure sampling. In: Proc. SIGGRAPH 2005, pp. 828–835 (2005)
4. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Proc. CVPR 2005, pp. II-60–65 (2005)
5. Bennett, E.: Computational video enhancement. PhD thesis, The University of North Carolina at Chapel Hill (2007)
6. Ben-Ezra, M., Nayar, S.: Motion deblurring using hybrid imaging. In: Proc. CVPR 2003, pp. I-657–664 (2003)
7. Richardson, W.: Bayesian-based iterative method of image restoration. *JOSA* 62(1), 55–59 (1972)
8. Lucy, L.: An iterative technique for the rectification of observed distributions. *Astronomical Journal* 79(6), 745–754 (1974)
9. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proc. ICCV 1998, pp. 839–846 (1998)
10. Zhang, X., Sim, T., Miao, X.: Enhancing photographs with near infra-red images. In: Proc. CVPR 2008, pp. 1–8 (2008)
11. Krishnan, D., Fergus, R.: Dark flash photography. In: Proc. SIGGRAPH 2009 (2009)
12. Yuan, L., Sun, J., Quan, L., Shum, H.-Y.: Image deblurring with blurred/noisy image pairs. In: Proc. SIGGRAPH 2007 (2007)
13. Tai, Y.W., Du, H., Brown, M., Lin, S.: Image/video deblurring using a hybrid camera. In: Proc. CVPR 2008, pp. 1–8 (2008)
14. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press, Cambridge (2004)