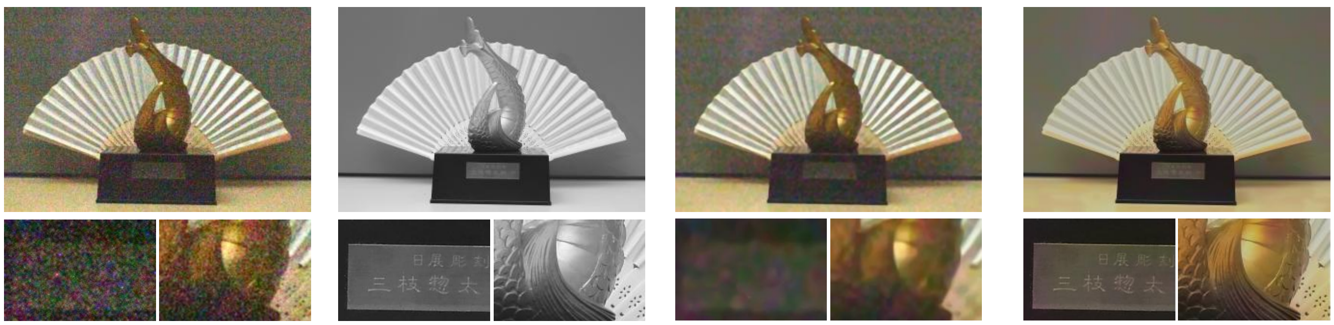
IEEE International Workshop on Computational Cameras and Displays
(formerly PROCAMS) (CCD/PROCAMS2013), June, 2013

Low-Light Scene Color Imaging based on Luminance Estimation from Near-Infrared Flash Image

Koichi Takeuchi, Masayuki Tanaka, Masatoshi Okutomi
Tokyo Institute of Technology
Tokyo, Japan
k.takeuchi@ok.ctrl.titech.ac.jp, mtanaka@ctrl.titech.ac.jp, mxo@ctrl.titech.ac.jp



(a) Noisy-color image (b) NIR-flash image (c) BM3D [2] (d) Proposed method

Figure 1: Examples of input and output images, where (a) is the noisy-color image, (b) is the NIR-flash image, (c) is the denoised image of the noisy-color image by BM3D, and (d) is the result image of the proposed method using the noisy-color and NIR-flash image pair.

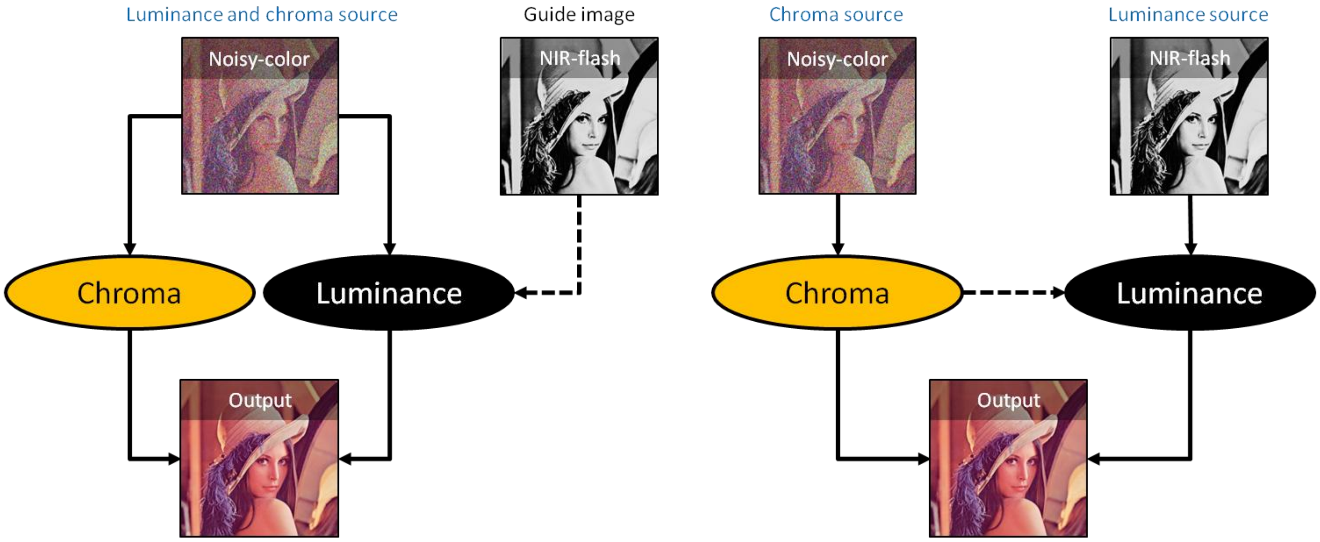
Abstract

In a low-light scene, we need a high-gain setting or a long-exposure setting to avoid a visible flash when we take pictures. However, these settings lead to severe degradations such as high-level noise and/or motion blur. Another approach for low-light scene capture is to use an invisible flash. In this paper, we propose a novel approach for color imaging using a noisy-color image and an invisible flash image. The basic idea of the proposed method is that the luminance component and the chroma component of a color image are estimated from different image sources. The luminance component is estimated mainly from the invisible flash image via a spectral estimation. The chroma component is estimated from the noisy-color image by denoising. Experimental comparisons demonstrate that the proposed method outperforms both a state-of-the-art single image denoising method and an existing method using a noisy-color image and an invisible flash image.

1. Introduction

A visible flash is required when we take pictures of a dark or a low-light scene. However, the visible flash is unacceptable when we try to take pictures of a nocturnal animal or a sleeping baby, because of the strong emission of the visible flash. In order to avoid the visible flash, we need a high-gain or a long-exposure setting to take pictures for the low-light scene. However, these settings meet severe degradations, i.e. significant noise in a high-gain setting and motion blur in a long-exposure setting. Although there is a long history of researches on denoising [1-6] and deblurring [7, 8], denoising and deblurring are still very challenging problems and these single-image-based processing are limited. Then, several color imaging methods using a visible-flash and no-flash image pair are proposed [9, 10]. These color imaging methods perform much better than the single-image-based processing. However, as mentioned above, in many situations, a strong visible flash is unacceptable.

One of other approaches to avoid the visible flash is to use an invisible flash instead. The most promising invisible flash is a near-infrared (NIR) flash, because image sensors equipped in commercial color cameras usually have some sensitivity to the NIR, while humans and many kinds of animals do not have any sensitivity to the NIR. Although commercial color cameras have a “hot mirror” which reflects infrared light, we can easily capture NIR-flash images by just removing the hot mirror from the camera. However, the NIR-flash image does not have color information. Color imaging methods using a noisy-color and a NIR-flash image pair have been proposed [11-14]. In these methods, the NIR-flash image was used as a “guide” image which helps to denoise and enhance the noisy-color image.



(a) Existing methods

(b) Proposed method

Figure 2: Conceptual difference between existing methods and our proposed method, where the broken lines represent guide data. Existing methods extract both the luminance and the chroma components from the single noisy-color image, while the proposed method extracts the chroma component from the noisy-color image, and the luminance component mainly from the NIR-flash image.

It is known that the NIR-flash image and the color image for the same scene are strongly correlated to each other, especially in high-frequency components. Therefore, the edge preserving denoising methods using the NIR-flash image have been proposed. A bilateral filter [5] and a WLS filter [6] are well known powerful non-linear filters. The bilateral filter was extended to a dual bilateral filter by using the NIR-flash image as the guide image [11]. The WLS filter is also extended to a dual WLS filter [12]. In addition, the NIR-flash image was used to enhance the details of the denoised color image [13]. In [14], the NIR-flash and a near-ultraviolet (UV) flash images were used for color image reconstruction. The common concept of these methods using the NIR and/or the UV flash image is to denoise and enhance the noisy-color image. However, even if the NIR-flash image can help denoising, this denoising approach has limitation. The denoising approach cannot extract fine details from very noisy color images perfectly.

In this paper, we propose a novel color imaging method using the noisy-color and the NIR-flash image pair. A color image consists of a luminance component and a chroma component. The luminance component represents the intensity or total power of the light at each pixel. The chroma component represents the color or the shape of the spectral distribution at each pixel. The basic idea of the proposed method is to extract the luminance component and the chroma component from different image sources; the NIR-flash image and the noisy-color image. In contrast, existing methods [11-14] try to extract both the luminance and the chroma components from a single image source that is the noisy-color image. We can extract the chroma component by applying strong denoising to the noisy-color image because it is known that the chroma component is dominated by low frequencies. However, we cannot extract the luminance component with fine details because it is very difficult to distinguish the fine details from the noise. Then, we extract the luminance component mainly from the NIR-flash image assuming the strong correlation between the luminance component and the NIR-flash image.

Figure 1 shows an example of the resultant color image by the single image denoising, BM3D [2], and the proposed method. We can find that the proposed method can reconstruct the fine details with reducing noise. We experimentally demonstrate that the proposed method can generate sharp and noise free image from a noisy-color and NIR-flash image pair.

2. Color imaging from different image sources

The color image consists of the luminance component and the chroma component. As shown in Fig. 2, existing denoising methods using the noisy-color and NIR-flash image pair extract both the luminance and the chroma components from the single noisy-color image, although those methods effectively use the NIR-flash image as the guide image. In contrast, out proposed method generates the color image based on the luminance and the chroma components extracted from the NIR-flash and the noisy-color images, respectively. For the luminance estimation in the proposed method, the estimated chroma is used to improve the accuracy of the luminance estimation.

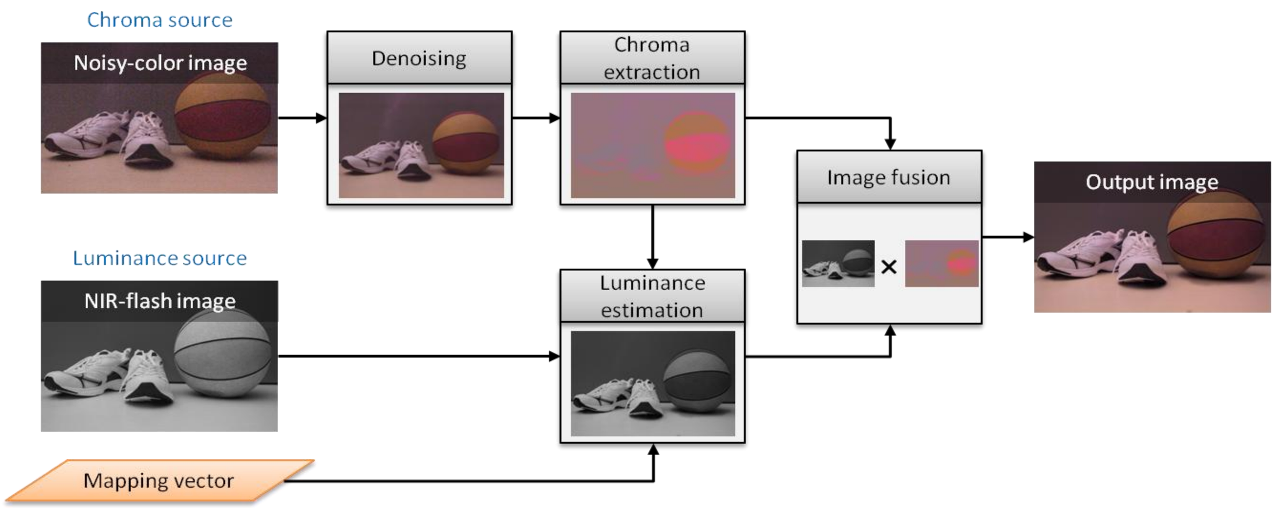


Figure 3: The processing flow of our proposed method.

Figure 3 shows the image processing flow of the proposed method.

Before explaining the proposed method, we clarify the luminance and chroma representation. There are various color space models which represent the luminance component and the chroma component [15]. The most popular color representation is the RGB representation by three primary color components; R, G, and B. The data of the RGB representation can be considered as the three-dimensional Euclidian data, $\boldsymbol{v}$. In this paper, we consider that the luminance is Y channel of YCbCr color space. The RGB data normalized by the luminance provides the chroma information. The RGB data can be decomposed into the luminance component and the chroma component or the luminance-normalized RGB data as

$$\boldsymbol{v} = y\boldsymbol{u}, \tag{1-a}$$
$$y = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{v},, \tag{1-b}$$

$$\boldsymbol{u} = \boldsymbol{v}/(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{v})\,, \tag{2}$$

where $\boldsymbol{a}$ is the vector which represents the coefficients of transform from RGB to Y, T represents the transpose operator, y represents the luminance component of the RGB data, and $\boldsymbol{u}$ is the luminance-normalized RGB data which represents the chroma. In the proposed method, the luminance component, y, is extracted from the NIR-flash image and the chroma component, $\boldsymbol{u}$, is extracted from the noisy-color image. Then, the extracted luminance components and the chroma components are fused into the output color image.

2.1. Chroma extraction from the noisy-color image

We apply a denoising method to the noisy-color image with strong denoising parameters. The strong denoising produces over-smoothed color image without fine details. However, the over-smoothed color image is sufficient to extract the chroma component, because the chroma component is dominated by low frequencies. In this paper, we use the BM3D [2] for this denoising. The chroma component at each pixel is estimated by

$$\hat{\boldsymbol{u}} = \tilde{\boldsymbol{v}}_{\mathrm{n}}/(\boldsymbol{a}^{T}\tilde{\boldsymbol{v}}_{n}), \tag{2}$$

where $\tilde{\boldsymbol{v}}_{\mathrm{n}}$ represents the RGB data of the denoised noisy-color image. We can easily extract the chroma component from the noisy-color image by just applying the denoising and the luminance normalization. However, it is very challenging to extract the luminance component with fine details from the noisy-color image.

2.2. Color image fusion

The NIR-flash image has strong correlation to the luminance component especially in the high-frequency component. The most naive luminance extraction from the NIR-flash image is to directly use the NIR-flash image as the luminance [16]. Obviously, the NIR-flash image is not exactly the same as the luminance component. However, this naive estimation is sufficient for a certain purpose. We also discuss the luminance estimation method to improve image quality in the next section.

The final output of the proposed method can be expressed as

$$\hat{\boldsymbol{v}} = \hat{y}\hat{\boldsymbol{u}} = \hat{y}\frac{\tilde{\boldsymbol{v}}_{n}}{\boldsymbol{a}^{T}\tilde{\boldsymbol{v}}_{n}}, \tag{3}$$

where $\hat{y}$ is the estimated luminance value. We can obtain

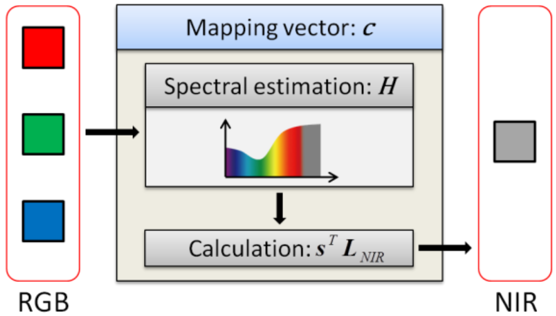


Figure 5: Schematic of transform from the RGB data to the NIR-flash image value via spectral estimation.

the output image by applying this process for each pixel.

3. Luminance estimation

In the previous section, we assume that the NIR-flash image can be used as the luminance component. This simple assumption is good enough for some applications. However, it is also true that this simple assumption produces wrong luminance. Therefore, we propose a method to estimate the luminance based on the NIR-flash image value.

3.1. Luminance estimation via spectral estimation

If we know the spectral reflectance of the target object, we can easily calculate the NIR-flash image value by assuming the spectral density of the NIR-flash and the sensor sensitivity. Although the spectral reflectance is usually estimated with multi-spectral images, this reflectance estimation method can be applied to the RGB data. We follow the same manner of [17] to estimate the spectral reflectance, assuming that an ambient light is white. The resultant estimated spectral reflectance can be expressed by the linear transformation as

$$\boldsymbol{r} = \boldsymbol{H}\,\boldsymbol{v}, \tag{4}$$

where $\boldsymbol{r}$ is the discretized spectral reflectance, and $\boldsymbol{H}$ is the spectral estimation matrix which transforms the RGB data to the discretized spectral reflectance. Once we can estimate the spectral reflectance from the RGB data, we can easily estimate the NIR-flash image value from the estimated spectral reflectance as shown in Fig. 4. The NIR-flash image value, n, can be estimated by

$$\begin{aligned} n &= \boldsymbol{s}^{\mathrm{T}}\,\boldsymbol{L_{NIR}}\,\boldsymbol{r} = \boldsymbol{s}^{\mathrm{T}}\,\boldsymbol{L_{NIR}}\boldsymbol{H}\,\boldsymbol{v} \\ &= \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{c}^{\mathrm{T}}(y\boldsymbol{u})\,, \end{aligned} \tag{5}$$

where $\boldsymbol{s}$ is the vector which represents the sensitivity of the sensor, $\boldsymbol{L_{NIR}}$ is the diagonal matrix which represents the spectral distribution of the NIR-flash, and $\boldsymbol{c}$ is the vector

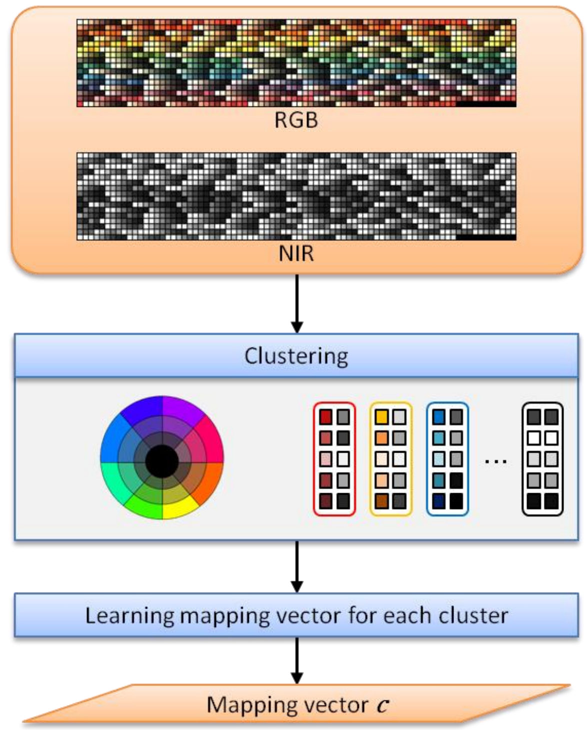


Figure 4: The processing flow for learning the mapping vector. The clustering is based on hue and saturation of the RGB data.

which maps the RGB data to the NIR-flash image value.

In the proposed method, we only observe the NIR-flash image value, n, and the luminance-normalized RGB data, $\boldsymbol{u}$, which is extracted from the noisy-color image. For given these data, we can estimate the luminance by minimizing the cost function:

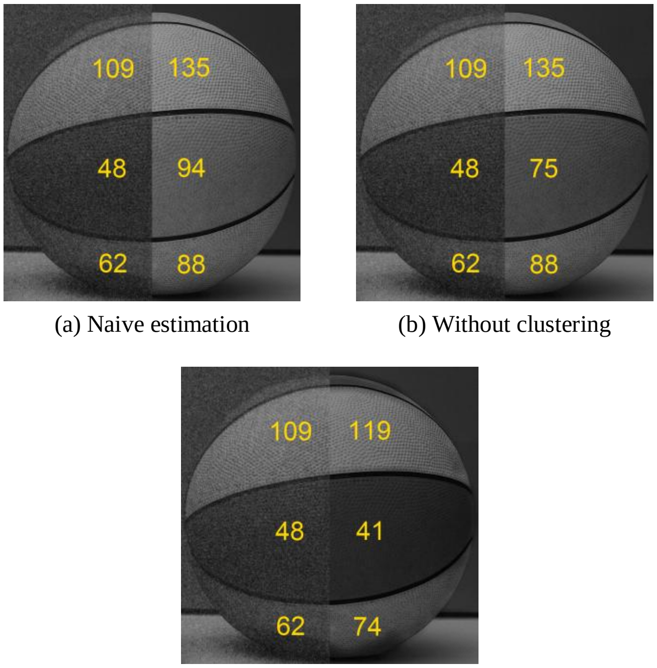
$$\hat{y} = \arg\min_{y} \|n - \boldsymbol{c}^T(y\hat{\boldsymbol{u}})\|_2^2 = \frac{n}{\boldsymbol{c}^T\hat{\boldsymbol{u}}}. \tag{6}$$

3.2. Mapping vector learning

As described in the previous section, we can estimate the luminance with the NIR-flash image value, the luminance-normalized RGB data, and the mapping vector which maps the RGB data to the NIR-flash image value. The mapping vector can be learned with learning data pairs of the RGB and the NIR-flash image values.

In this paper, we first generate the learning data pairs from the Munsell colors matt reflectance database [18] assuming that the ambient light is white. For the learning data pair generation, we use the typical spectral sensitivities of the color camera and the spectral distribution of the NIR light. The mapping vector can be learned by

$$\hat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}} \sum_{k=1}^{N} \|n_k - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v}_k\|_2^2\,, \tag{7}$$



(a) Naive estimation

(b) Without clustering

(c) With clustering

Figure 6: Comparison of three luminance estimation methods, where the left-side of each image is the input noisy color image, the right-side of each image is the estimated luminance. The numbers in each image represent average luminances around them.

where N is the number of the learning data pairs, n_k is k-th NIR-flash image value, and $\boldsymbol{v}_k$ is k-th RGB data.

As mentioned above, the mapping vector is derived via the spectral estimation. However, the spectral estimation from the RGB data is very challenging, because the RGB data only has three primary spectral components. Three spectral components are too few to estimate the spectral distribution. It is reported that the clustering-and-mapping approach can improve the performance of the spectral estimation [19]. Then, we apply the same approach to improve the performance of the luminance estimation.

First, the RGB data and the NIR-flash image value pairs are clustered based on the hue and the saturation of the RGB data as shown in Fig. 5. Then, for each cluster, the mapping vector is learned by Eq. (7).

In the luminance estimation phase, the luminance-normalized RGB data is first assigned to the nearest cluster. The mapping vectors are assigned to all pixel locations. We apply spatial smoothing to the mapping vectors to reduce discontinuities at borders of clusters. Then, the luminance is estimated based on the smoothed mapping vector.

4. Experiments

In [11, 12, 13], they used a multi-camera system to simultaneously capture the color image and the NIR-flash image. In this paper, we adopt a single camera system with a NIR-flash [14], because the single camera system is much simpler than the multi-camera system. We use the Canon EOS 7D digital still camera with removing the hot mirror to capture the NIR-flash image. We assume the noisy-color image and the NIR-flash image are perfectly aligned. The image alignment for a dynamic scene is one of our feature works.

4.1. Validation of the proposed luminance estimation

In this paper, we have described three types of luminance estimation methods; (1) the naive estimation which directly uses the NIR-flash image value as the luminance, (2) the luminance estimation from the luminance-normalized RGB data and the NIR-flash image value without clustering, and (3) the luminance estimation from the luminance-normalized RGB data and the NIR-flash image value with clustering. The differences appear only in the luminance components. Figure 6 shows the comparison of three luminance estimation methods. The left-side numbers in each image represent the average luminances of the input noisy-color image and the right-side numbers represent the average estimated luminances. From these comparisons, we can find that the luminance estimation with the clustering provides the closest estimation among three methods. Then, we use the luminance estimation with the clustering for later experiments.

4.2. Comparisons with existing methods

In this section, we compare the proposed method with the state-of-the-art denoising algorithm, BM3D [2], and the existing method using a noisy-color and NIR-flash image pair [11]. Figure 7 shows the visual comparisons for two scenes.

In general, denoising using a weak denoising parameter yields finely-detailed but still noisy results, and denoising using a strong denoising parameter yields less-noise but over-smoothed results. It is a trade-off problem. We can see that the denoising results in Fig. 7 (c) and (j) are over-smoothed and still disturbed by the noise.

As shown in Fig. 7 (d) and (k), Bennett's method [11] using the noisy-color and the NIR-flash image pair provides better results than single-image-based denoising since the NIR-flash image helps to enhance the image quality. However, the details of the result by Bennett's method are blurred compared to that of the results of the proposed method. For example, the proposed method can recover the fine texture of the ball in (f) and the sharp characters in (m).

(a) Noisy-color image

(b) NIR-flash image

(h) Noisy-color image

(i) NIR-flash image

(c) BM3D [2]

(d) Bennett [11]

(j) BM3D [2]

(k) Bennett [11]

(e) Proposed method
(naïve estimation)

(f) Proposed method
(estimation with clustering)

(l) Proposed method
(naïve estimation)

(m) Proposed method
(estimation with clustering)

(g) Long exposure

(n) Long exposure

Figure 7: Comparisons of the proposed method with the BM3D denoising and Bennett's method.

(a) NIR-Flash image (b) Noisy-color image (c) Proposed method (d) Long-exposure

Figure 8: Limitation of proposed method.

4.3. Limitation

We assume there is a strong correlation between the luminance component and the NIR-flash image. However, the assumption is not always true in some cases. We show an example of limitation in Fig. 8. We can see the result of our proposed method in Fig. 8 (c) significantly differs from the image taken by long-exposure setting in Fig. 8 (d). The reason of these differences is that the paint of the T-shirt does not reflect NIR. Then the NIR-flash image has no correlation to the luminance component. In such cases, our assumption fails and our method yields different result from that of the long-exposure setting.

5. Conclusions

We have proposed the novel color imaging method for low-light scene photography. Our proposed method extracts the luminance and the chroma components from different image sources; the NIR-flash image and the noisy-color image. The luminance component is estimated from the NIR-flash image via spectral estimation. We apply denoising and luminance normalization to the noisy-color image to extract the chroma component. Then, the estimated luminance and chroma components are fused to the color image. The experimental comparisons demonstrate that the proposed method provides sharp results without noise and color artifacts compared to the state-of-the-art single image denoising and the existing method using the noisy-color and NIR-flash image pair.

Our future study includes the image alignment between the color image and the NIR-flash image for a dynamic scene.

References

[1] S. Roth and M. Black. Fields of experts: a framework for learning image priors. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 860-867, 2005.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Bm3d image denoising with shape-adaptive principal component analysis. In *Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, 2009.

[3] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 7, pp. 629-639, 1990.

[4] A. Buades, B. Coll, and J.-M.Morel. A non-local algorithm for image denoising. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* Vol. 2, pp. 60-65, 2005.

[5] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *International Conference on Computer Vision (ICCV)*, pp. 839-846, 1998.

[6] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Trans. Graph*., Vol. 27, No. 3, pp. 67:1-67:10, 2008.

[7] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In *ACM Trans. Graph*., Vol. 25, No. 3, pp. 787-794, 2006.

[8] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. In *ACM Trans. Graph*., Vol. 27, No. 3, pp. 73:1-73:10, 2008.

[9] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. In *ACM SIGGRAPH*, pp. 664-672, 2004.

[10] E. Eisemann and F. Durand. Flash photography enhancement via intrinsic relighting. In *ACM Trans. Graph.*, Vol. 23, No. 3, pp. 673-678, 2004.

[11] E. Bennett, J. Mason, and L. McMillan. Multispectral bilateral video fusion. In *IEEE Transactions on Image Processing*, Vol. 16, No. 5, pp. 1185-1194, 2007.

[12] S. Zhuo, X. Zhang, X. Miao, and T. Sim. Enhancing low light images using near infrared flash images. In *IEEE International Conference on Image Processing (ICIP), pp.* 2537-2540, 2010.

[13] S. Matsui, T. Okabe, M. Shimano, and Y. Sato. Image enhancement of low-light scenes with near-infrared flash images. In *Asian Conference on Computer Vision (ACCV)*, pp. 213-223, 2010.

[14] D. Krishnan and R. Fergus. Dark Flash photography. In *ACM Transactions on Graphics, SIGGRAPH*, 28, 2009.

[15] H. Kang. *Computational Color Technology*. *Society of Photo*

Optical, 2006.

[16] C. Fredembach and S. Süsstrunk. Colouring the near infrared. In *Proceedings of the IS&T/SID Color Imaging Conference*, pp. 176-182, 2008.

[17] J. Park, M. Lee, M. Grossberg, and S. Nayar. Multispectral imaging using multiplexed illumination. In *IEEE International Conference on Computer Vision (ICCV), pp*. 1-8, 2007.

[18] R. Lenz, M. Osterberg, J. Hiltunen, T. Jaaskelainen, and J. Parkkinen. Unsupervised Filtering of color spectra. *JOSA A*, Vol. 13, pp. 1315-1324, 1996.

[19] Y. Murakami, M. Yamaguchi, and N. Ohyama. Class-based spectral reconstruction based on unmixing of low-resolution spectral information. *JOSA A*, Vol. 28, pp. 1470-1481, 2011.