Image Fusion for Single-Shot High Dynamic Range Imaging with Spatially Varying Exposures

Chihiro Go*, Yuma Kinoshita*, Sayaka Shiota* and Hitoshi Kiya*
* Tokyo Metropolitan University, Tokyo, Japan
E-mail: go-chihiro@ed.tmu.ac.jp, kinoshita-yuma@ed.tmu.ac.jp, sayaka@tmu.ac.jp, kiya@tmu.ac.jp

***Abstract*—This paper proposes a novel multi-exposure image fusion scheme for single-shot high dynamic range imaging with spatially varying exposures (SVE). Single-shot imaging with SVE enables us not only to produce images without color saturation regions from a single-shot image, but also to avoid ghost artifacts in the producing ones. However, the number of exposures is generally limited to two, and moreover it is difficult to decide the optimum exposure values before the photographing. In the proposed scheme, a scene segmentation method is applied to input multi-exposure images, and then the luminance of the input images is adjusted according to both of the number of scenes and the relationship between exposure values and pixel values. The proposed method with the luminance adjustment allows us to improve the above two issues. In this paper, we focus on dual-ISO imaging as one of single-shot imaging. Experiments are demonstrated to confirm the effectiveness of the proposed scheme under both of dual-ISO imaging and single-ISO one. In an experiment, the effectiveness of the proposed scheme is demonstrated by comparing the previous dual-ISO imaging and with single-ISO one.**

I. Introduction

The low dynamic range (LDR) imaging sensors used in modern digital cameras cannot express the dynamic range of a real scene, due to a limited dynamic range which imaging sensors have. The limit results in the low contrast of images taken by digital cameras.

A method for capturing high dynamic range (HDR) images is to use a wide dynamic image sensor [1]. However, such devices are very expensive, so they are not widespread yet. Therefore, the most common approach for HDR imaging is to fuse multi-exposure images which are to merge a set of LDR images taken with different exposure times. This approach requires to capture multi-exposure images by taking at the different time, so there are ghost artifact issues, due to the movement of the camera and the subject. Therefore, various research works on motion correction and ghost reduction have been conducted for the multi-exposure image fusions [2]–[5].

A ghost-free tecqunic for HDR imaging is to employ spatially varying exposures (SVE) [6]. In the SVE-based imaging, a scene is captured with varying exposures for each pixel in a single image, and multiple sub-images with each exposure are obtained. The dual-ISO imaging has been proposed as one of single-shot imaging with SVE, in which the ISO speeds alternates every two lines in a single raw Bayer image [7]–[9]. In [10]–[12], the exposure time alternates row-wise varying exposures in a single raw Bayer image with two exposure times. In Quad Bayer pixel structure [13], integration can be divided into long-time integration and short-time integration for every two pixels in the Quad array. However, in these methods, the number of exposures is generally limited, and moreover, it is difficult to decide the optimum exposure values before the photographing.

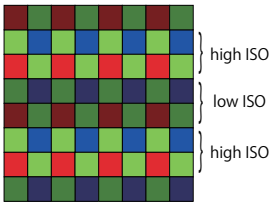


Fig. 1. Raw Bayer image sensed with dual-ISO sensor

Because of such a situation, in this paper, we propose a new image fusion method for single-shot imaging with SVE. In the proposed method, a new scene segmentation method is applied to input multi-exposure images, and then the luminance of the input images is automatically adjusted by analysing segmented images [14], [15]. This paper focuses on dual-ISO imaging as one of single-shot imaging with SVE to conform the effectiveness of the proposed fusion method. Some experiments demonstrate that the proposed method outperforms conventional methods in terms of two objective quality metrics, HIGRADE [16] and discrete entropy [17].

II. Preparation

In this paper, we focus on dual-ISO imaging as one of single-shot imaging. Here we summarize dual-ISO imaging and conventional fusion schemes for the imaging.

A. Dual-ISO imaging

A raw Bayer image sensed with a dual-ISO sensor is illustrated in Fig.1, where the ISO speed alternates every two lines in the Bayer image [7]. By using the dual-ISO sensor, raw images with two exposures are produced. The Bayer image captured with two ISO speeds are fused as shown in Fig.2.

1) Separation and interpolation: A raw image $\mathbf{X}$ with the size of $M \times N$ is first divided into two raw images with the size of $M/2 \times N$, according to the difference of ISO speed. Next, interpolation processing is applied to each raw image for producing two raw images with the size of $M \times N$: $\mathbf{X}_{low}$ and $\mathbf{X}_{high}$ as in Fig.2. In [7], AMaZE [18] was used as the interpolation method.

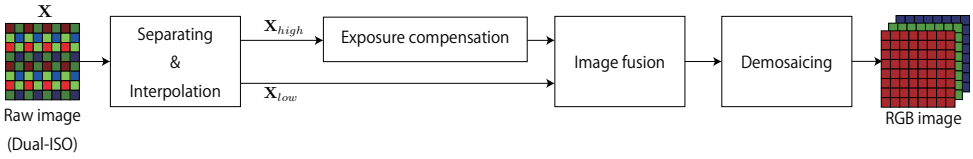


Fig. 2. Conventional method for dual-ISO imaging

2) Exposure compensation: The relationship between a pixel value $x(i,j)$ and the exposure value $e(i,j)$ at the pixel (i,j) is given by

$$e(i,j) = \log_2(\max(x(i,j),1)). \tag{1}$$

Before carrying out image fusion, in the previous work [7], the exposure value of $\mathbf{X}_{high}$ is adjusted as

$$\hat{e}_{high}(i,j) = e_{high}(i,j) - \log_2\left(\frac{\mathrm{iso}_{high}}{\mathrm{iso}_{low}}\right), \tag{2}$$

where iso_{high} and iso_{low} are high ISO speed and low ISO speed respectively, and $e_{high}(i,j)$ is the exposure value of $\mathbf{X}_{high}$ at the pixel (i,j).

3) Image fusion: $\mathbf{X}_{low}$ has generally detailed information in bright areas. On the other hand, $\mathbf{X}_{high}$ has detailed information in dark areas. Considering these properties, an equation for image fusion was proposed to fuse two exposure values, e_{low} and $\hat{e}_{high}$ as

$$e_{mix}(i,j) = \hat{e}_{high}(i,j)\cdot K + e_{low}(i,j)\cdot(1-K), \tag{3}$$

where K, $0 \le K \le 1$ is a mix factor and $e_{low}(i,j)$ is the exposure value of $\mathbf{X}_{low}$. From eq.(1), the adjusted pixel value $\hat{x}(i,j)$ can be calculated from the exposure value $e_{mix}(i,j)$ by

$$\hat{x}(i,j) = 2^{e_{mix}(i,j)}. \tag{4}$$

4) Demosaicing: To obtain a RGB image, an image demosaicing algorithm is finally applied to the fused raw image.

In the conventional signle-shot imaging, the number of exposures is generally limited to two, and moreover it is difficult to decide the optimum exposure values before the photographing as described above.

III. Proposed method

In order to improve the two issues that the conventional signle-shot imaging has, we propose a new image fusion method for the imaging. The outline of the proposed method is shown in Fig.3, where the main contribution of this work is in scene-segmentation based exposure competition. The exposure competition consists of the following five steps (See Fig.4).

A. Local contrast enhancement

Since the number of exposures is generally limited to two, $\mathbf{X}$ can not always represent the scene clearly. A local contrast enhancement algorithm is used to enhance detailed information in $\mathbf{X}$. Therefore, in this paper, the local contrast enhancement using the dodging and burning algorithm [19] is performed as

$$L'_k(i,j) = \frac{L_k^2(i,j)}{L_{ak}(i,j)},\ k \in \{low,\ high\}, \tag{5}$$

where $L_k(i,j)$ is the luminance value of $\mathbf{X}_k$ at the place (i,j), and $L_{ak}(i,j)$ is the local average of luminance $L_k(i,j)$ around pixel (i,j). Here, a bilateral filter is performed to obtain $L_{ak}(i,j)$ as in [19].

The luminance $\mathbf{L}_k = \{L_k(i,j)\}$ is required to calculate eq.(5), but $X_k(i,j)$ has only a R, G or B value. In this paper, in order to obtain luminance $\mathbf{L}_k$ from $\mathbf{X}_k$, $\mathbf{X}_k$ is divided into non-overlapping blocks with 2×2 pixels as shown in Fig.5. Since $X_k(i,j)$ is a R value when both i and j are odd, the luminance $L_k(i,j)$ is calculated in each block according to ITU-R BT.601 [20], as

$$\begin{aligned} L_k(i,j) &= 0.257 X_k(i,j) \\ &\quad + 0.504\left(\frac{X_k(i+1,j)+X_k(i,j+1)}{2}\right) \\ &\quad + 0.098 X_k(i+1,j+1) + 16, \end{aligned} \tag{6}$$

$$L_k(i,j) = L_k(i+1,j) = L_k(i,j+1) = L_k(i+1,j+1). \tag{7}$$

B. Scene segmentation

Using $\mathbf{L}'_{high} = \{L'_{high}(i,j)\}$ and $\mathbf{L}'_{low} = \{L'_{low}(i,j)\}$, the set R of all pixels is divided into S subsets $\{\mathrm{R}_1, \mathrm{R}_2, \ldots, \mathrm{R}_S\}$, where each area R_s, $s \in \{1,2,\ldots,S\}$ has specific brightness and $\mathrm{R}_1 \cup \mathrm{R}_2 \cup \cdots \cup \mathrm{R}_S = \mathrm{R}$. For this segmentation, a Gaussian mixture distribution is utilized to model the luminance distribution of the input images in this paper. Pixels are classified by a clustering algorithm based on a Gaussian mixture model (GMM).

To obtain a model considering the luminance values $\mathbf{L}'_{high}$ and $\mathbf{L}'_{low}$, we regard luminance values at a pixel (i,j) as a 2-dimensional vector $\mathbf{l}(i,j) = \{L'_{low}(i,j), L'_{high}(i,j)\}^{\mathrm{T}}$, where T denotes the transpose of a vector. Then the distribution of the vector $\mathbf{l}(i,j)$ is modeled by a GMM with D areas as

$$p\left(\mathbf{l}(i,j)\right) = \sum_{d=1}^{D} \pi_d G\left(\mathbf{l}(i,j)|\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d\right), \tag{8}$$

where π_d is a mixing coefficient and $G\left(\mathbf{l}(i,j)|\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d\right)$ is a 2-dimensional Gaussian distribution with mean $\boldsymbol{\mu}_d$ and variance covariance matrix $\boldsymbol{\Sigma}_d$.

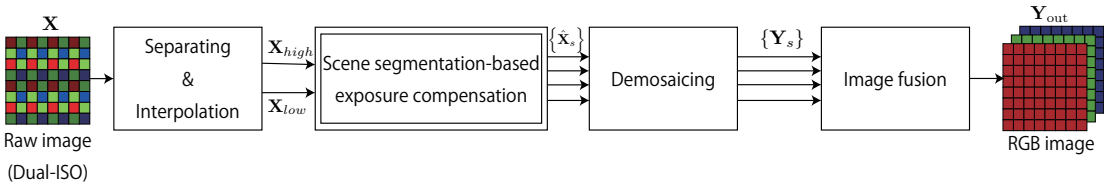


Fig. 3. Outline of proposed method

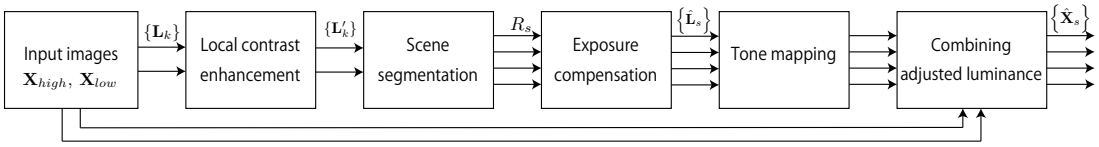


Fig. 4. Details of scene segmentation-based exposure compensation

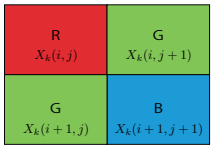


Fig. 5. Block for luminance calculation

To fit the GMM into given $\mathbf{l}(i,j)$, a variational Bayesian algorithm [21] is utilized. Compared to the traditional maximum likelihood approach, the variational Bayesian algorithm allows us to avoid overfitting even when we choose a large D. Therfore, unnecessary mixture components are automatically removed by using the approach together with a large D. $D = 10$ is used in this paper, as the maximum of the partition number S.

A cluster for an observation $\mathbf{l}(i,j)$ is determined by the responsibility $\gamma\left(z_d|\mathbf{l}(i,j)\right)$ which is given as the following conditional probability:

$$\gamma\left(z_d|\mathbf{l}(i,j)\right) = p(z_d = 1|\mathbf{l}(i,j)) = \frac{\pi_d G(\mathbf{l}(i,j)|\boldsymbol{\mu}_d, \Sigma_d)}{\sum_{j=1}^{D} \pi_j G(\mathbf{l}(i,j)|\boldsymbol{\mu}_j, \Sigma_j)}, \tag{9}$$

where z_d is a particular element equal to 1 and all other elements are equal to 0. When an pixel $(i,j) \in \mathrm{R}$ is given and s satisfies the following equation

$$s = \underset{d}{\operatorname{argmax}}\, \gamma(z_d|\mathbf{l}(i,j)), \tag{10}$$

the pixel (i,j) is assigined to a subset R_s of R.

By using the above approach, the number of segmented areas, $S \leq D$ is generally larger than two, that is the number of exposure values used in the conventional single-shot image.

C. Exposure compensation

The scaled luminance $\hat{\mathbf{L}}_s$ which clearly represents an area R_s is obtained by

$$\hat{L}_s(i,j) = \alpha_s L'_k(i,j), \tag{11}$$

where parameter $\alpha_s > 0$ indicates the degree of adjustment.

The approximate brightness of an area R_s is calculated as the geometric mean of luminance values on R_s. We thus estimate an adjusted multi-exposure image $\hat{L}_s(i,j)$, so that the geometric mean of its luminance equals to middle-gray of the displayed image, or 0.18 on a scale from zero to one, as in [22]. The geometric mean $g(L'_k|\mathrm{R}_s)$ of luminance L'_k on pixel set R_s is calculated using

$$g(L'_k|\mathrm{R}_s) = \exp\left(\frac{1}{|R_s|}\sum_{(i,j)\in \mathrm{R}_s} \ln(\max(L'_k(i,j), \epsilon))\right), \tag{12}$$

where ϵ is set to a small value to avoid singularities at $L'_k(i,j) = 0$. By using eq.(12), parameter α_s is calculated as

$$\alpha_s = \frac{0.18}{g(L'_k|\mathrm{R}_s)}. \tag{13}$$

Since a smaller value for parameter α_s is better, k is chosen as

$$k = \underset{h\in\{low,\, high\}}{\operatorname{argmin}} (0.18 - g(L'_h|\mathrm{R}_s))^2. \tag{14}$$

By using the exposure compensation, exposure values are automatically adjusted, even when the values have no approximate brightness.

D. Tone Mapping

Because the scaled luminance value often exceeds the maximum value of the common image format, pixel values might be lost due to truncation of the values. This problem is overcome by using a tone mapping operation to fit the adjusted luminance value into the interval $[0, 1]$. In this paper,

Reinhard's global operator [22] is used as a tone mapping operator.

E. Combining adjusted luminance and input images

A set $\{\hat{\mathbf{L}}_s\}$ of luminance adjusted by the scene segmentation-based exposure compensation is combined with an input image $\{\mathbf{X}_k\}$ to obtain adjusted images $\{\hat{\mathbf{X}}_s\}$. To associate each $\hat{\mathbf{L}}_s = \{\hat{L}_s(i,j)\}$ with an input image $\mathbf{X}_k = \{X_k(i,j)\}$, eq.(14) is utilized. As a result, combining $\hat{\mathbf{L}}_s$, the selected input image $\mathbf{X}_{high}$ or $\mathbf{X}_{low}$ and its luminance $\mathbf{L}_{high}$ or $\mathbf{L}_{low}$, adjusted multi-exposure images $\{\hat{\mathbf{X}}_s\}$ are obtained.

Therefore, the adjusted pixel value $\hat{X}_s(i,j)$ is computed by

$$\hat{X}_s(i,j) = \frac{\hat{L}_s(i,j)}{L_k(i,j)} X_k(i,j), \quad k \in \{low,\ high\}. \tag{15}$$

Similarly, from eq.(7), other three pixels in the block are computed with $\hat{L}_s(i,j)$, as

$$\begin{aligned} \hat{X}_s(i+1,j) &= \frac{\hat{L}_s(i,j)}{L_k(i,j)} X_k(i+1,j), \\ \hat{X}_s(i,j+1) &= \frac{\hat{L}_s(i,j)}{L_k(i,j)} X_k(i,j+1), \\ \hat{X}_s(i+1,j+1) &= \frac{\hat{L}_s(i,j)}{L_k(i,j)} X_k(i+1,j+1). \end{aligned} \tag{16}$$

F. Demosaicing

Since $\{\hat{\mathbf{X}}_s\}$ are Raw images, a demosaicing algorithm is carried out to obtain the RGB images $\{\mathbf{Y}_s\}$. In this paper, we apply a simple image demosaicing algorithm [23] to $\{\hat{\mathbf{X}}_s\}$.

G. Image fusion

A final image $\mathbf{Y}_{\rm out}$ is produced as

$$\mathbf{Y}_{\rm out} = \mathcal{F}(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_S), \tag{17}$$

where $\mathcal{F}(\cdot)$ indicates a function to fuse S images $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_S$ into a single image. Any existing multi-exposure image fusion methods are applicable for the proposed scheme. The fusion method proposed by Mertens et al. [24] is used in this paper as $\mathcal{F}(\cdot)$.

IV. Simulation

Some simulations were carried out to demonstrate the effectiveness of the proposed scheme for single-shot HDR imaging.

A. Simulation conditions

Photographs taken by Canon EOS 5D Mark II camera were directly used as input image $\mathbf{X}$. We also used Magic Lantern [25], which is a firmware to perform dual-ISO sensing. The shutter speed and the aperture were set by auto exposure of the camera at ISO 100. This condition means that the exposure value is 0 EV at ISO 100. For the dual-ISO imaging, ISO 100 and ISO 1600 correspond to 0 EV and +4 EV respectively. Gamma compensation with $\gamma = 2.2$ was also applied to all images.

B. Comparison with conventional methods

The performance of the proposed scheme was compared with the conventional dual-ISO imaging [7] and single-ISO one. The quality of images produced by each method was evaluated according to HIGRADE [16] and discrete entropy [17] which are objective quality assessments without any reference images. HIGRADE represents the quality of SDR images obtained by tone mapping, multi-exposure fusion methods. Discrete entropy represents the amount of information in an image. For each score, a larger value means higher quality.

C. Simulation results

Figure 6 shows examples of output images produced by each method. Figures 6(b) and 6(c) do not clearly represent dark areas, while Fig. 6(d) does not clearly represent bright areas. From these results, the conventional dual-ISO imaging in Fig.6(b) does not successfully capture high-quality images, although two exposure values are utilized. Compared with the conventional one, the proposed imaging successfully captures images as shown in Fig.6(a), where the number of segmented areas in Fig.6(a) was $S = 5$ as shown in Fig.6(e). The objective quality assessments, HIGRADE and discrete entropy, also show that the proposed imaging outperforms other methods.

Figures 7 summarizes the average score and variance range of the objective assessments for eight input images, where the middle line indicates the average scores of eight images under the use of each method, and the upper and lower line denote the variance range. From Fig. 7, the proposed imaging provides higher quality images than the conventional imaging.

V. Conclusion

This paper proposes a novel multi-exposure image fusion scheme for single-shot high dynamic range imaging with spatially varying exposures. The main contribution of this work is in scene-segmentation based exposure competition. In this paper, we have focused on dual-ISO imaging as one of single-shot imaging with SVE. Experimental results have shown that the proposed scheme outperforms the conventional algorithm with two ISO speeds and the direct capture with single ISO speed in terms of two no reference objective metrics, HIGRADE and discrete entropy.

In the future, the proposed method will be applied to other single-shot imaging methods with SVE such as Quad Bayer pixel structure.

References

[1] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile hdr video production system," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4, p. 41, 2011.

[2] E. A. Khan, A. O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Image Processing, 2006 IEEE International Conference on*. IEEE, 2006, pp. 2005–2008.

[3] C. Lee and E. Y. Lam, "Computationally Efficient Truncated Nuclear Norm Minimization for High Dynamic Range Imaging," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4145–4157, September 2016.

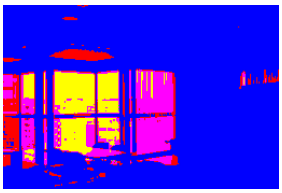
[4] Y. Kinoshita, T. Yoshida, S. Shiota, and H. Kiya, "Pseudo multi-exposure fusion using a single image," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017*. IEEE, 2017, pp. 263–269.

(a) Proposed with ISO 100 and ISO 1600
HIGRADE $= 0.5817$
discrete entropy $= 7.0609$

(b) Conventional [7] with ISO 100 and ISO 1600
HIGRADE $= -0.1829$
discrete entropy $= 3.0898$

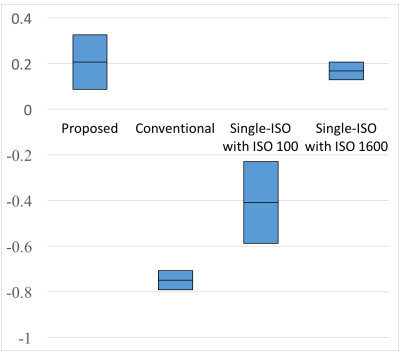
(c) Single-ISO with ISO 100
HIGRADE $= 0.0138$
discrete entropy $= 5.6019$

(d) Single-ISO with ISO 1600
HIGRADE $= 0.3390$
discrete entropy $= 6.6461$

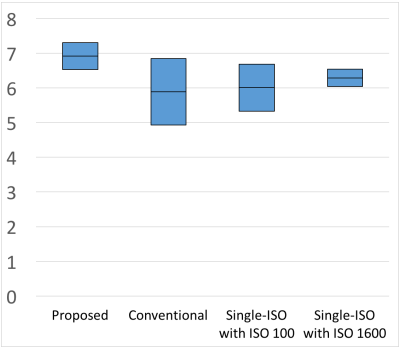


(e) Scene segmentation result

Fig. 6. Examples of fused Images $\mathbf{Y}_{out}$



(a) HIGRADE



(b) Discrete Entropy

Fig. 7. Average and variance for eight images

[5] Y. Kinoshita, S. Shiota, and H. Kiya, "A pseudo multi-exposure fusion method using single image," *IEICE Trans. Fundamentals*, vol. 101, no. 11, November 2018.

[6] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 1. IEEE, 2000, pp. 472–479.

[7] A1EX, "Dynamic range improvement for some canon dslrs by alternating iso during sensor readout," http://acoutts.com/a1ex/dual_iso.pdf, 2013.

[8] S. Hajsharif, J. Kronander, and J. Unger, "Hdr reconstruction for alternating gain (iso) sensor readout," in *Eurographics, Strasbourg, France, April 7-11, 2014*, 2014.

[9] R. Gil Rodríguez and M. Bertalmío, "High quality video in high dynamic range scenes from interlaced dual-iso footage," in *IS&T International Symposium on Electronic Imaging Science and Technology; 2016 Febr. 14-18; San Francisco (CA, USA). Springfield: IST 2016. Digital Photography and Mobile Imaging XII, p. DPMI-245.1 [7 p.].* The Society for Imaging Science and Technology (IS&T), 2016.

[10] H. Cho, S. J. Kim, and S. Lee, "Single-shot high dynamic range imaging using coded electronic shutter," in *Computer Graphics Forum*, 2014.

[11] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar, "Coded rolling shutter photography: Flexible space-time sampling," in *Computational Photography (ICCP), 2010 IEEE International Conference on*, 2010, pp. 1–8.

[12] V. G. An and C. Lee, "Single-shot high dynamic range imaging via deep convolutional neural network," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC 2017, Kuala Lumpur, Malaysia, December 12-15, 2017*, 2017, pp. 1768–1772.

[13] SONY, "IMX294CJK — Sony Semiconductor Solutions," https://www.sony-semicon.co.jp/products_en/new_pro/may_2017/imx294cjk_e.html.

[14] Y. Kinoshita, S. Shiota, and H. Kiya, "Automatic exposure compensation for multi-exposure image fusion," in *IEEE International Conference on Image Processing (ICIP).*, 2018.

[15] Y. Kinoshita, T. Yoshida, S. Shiota, and H. Kiya, "Multi-exposure image fusion based on exposure compensation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 1388–1392.

[16] D. Kundu, D. Ghadiyaram, A. C. Bovik, and B. L. Evans, "No-reference image quality assessment for high dynamic range images," in *Signals, Systems and Computers, 2016 50th Asilomar Conference on*. IEEE, 2016, pp. 1847–1852.

[17] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[18] E. Martinec, "Amaze demosaic algorithm (aliasing minimization and zipper elimination)," http://code.google.com/p/rawtherapee/source/browse/rtengine/amaze_demosaic_RT.cc.

[19] Y. Huo, F. Yang, and V. Brost, "Dodging and burning inspired inverse tone mapping algorithm," *Journal of Computational Information Systems*, vol. 9, no. 9, pp. 3461–3468, 2013.

[20] "ITU-R BT.601-5," http://www.itu.int.

[21] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[22] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM transactions on graphics (TOG)*, vol. 21, no. 3, pp. 267–276, 2002.

[23] H. S. Malvar, L.-w. He, and R. Cutler, "High-quality linear interpolation for demosaicing of bayer-patterned color images," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 3. IEEE, 5 2004, pp. 485–488.

[24] *Exposure fusion: A simple and practical alternative to high dynamic range photography*. Wiley Online Library, 2009.

[25] "Magic Lantern," http://www.magiclantern.fm.