IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 December 2022, accepted 23 December 2022, date of publication 30 December 2022,
date of current version 5 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3233546

RESEARCH ARTICLE

Fast and Effective Scene Segmentation Method for Luminance Adjustment of Multi-Exposure Image

SEIICHI KOJIMA AND NORIAKI SUETAKE, (Member, IEEE)
Graduate School of Science and Technology for Innovation, Yamaguchi University, Yamaguchi 753-8512, Japan
Corresponding author: Seiichi Kojima (c002wbw@yamaguchi-u.ac.jp)

This work was supported by JSPS KAKENHI under Grant JP22K12097.

ABSTRACT Multi-exposure image fusion is an important task for high dynamic range imaging. The performance of a fusion method is highly dependent on the quality of the input multi-exposure image. However, it is often difficult to obtain a multi-exposure image thoroughly covering the dynamic range of a scene. In such a situation, some areas are unclear in all the images contained by the multi-exposure image, leading to a low-visibility fusion result. To overcome this problem, Kinoshita and Kiya proposed scene segmentation-based luminance adjustment (SSLA). In SSLA, the scene is segmented into regions and the luminance of each region is adjusted so that the region has mid-level brightness. A fusion result with better visibility can be obtained by fusing the luminance adjusted multi-exposure image. Kinoshita and Kiya proposed two scene segmentation approaches for SSLA. One of them is fast but sometimes fails to segment dark areas and bright areas into different regions. The other can generate a better scene segmentation result but it requires a time-consuming iterative process. In this paper, we propose a new scene segmentation method for SSLA. The proposed scene segmentation method uses information obtained from the nonlinear luminance distribution and area occupancy of the segmented regions. A fast implementation for the proposed scene segmentation method is also described. Experimental results show that the proposed scene segmentation method is fast and stably generates good scene segmentation results.

INDEX TERMS High dynamic range imaging, multi-exposure image fusion, scene segmentation-based luminance adjustment.

I. INTRODUCTION

The dynamic range of luminance in natural scenery is very wide, whereas the dynamic range of image sensors used for general cameras and smart phones is relatively narrow. When an image sensor with a narrow dynamic range is used, the pixels whose luminance exceeds the dynamic range of the sensor cause white-out or black-out.

Many attempts have been made to obtain high-quality images in wide dynamic range scenes. One such attempt is to use a high-spec image sensor whose dynamic range is wider than those of general image sensors [1], [2]. Since such image sensors are more expensive than general image sensors, their applications are limited to special uses such as security cameras and in-vehicle cameras. Another common attempt to obtain high-quality images of wide dynamic range scenes is to use a multi-exposure image, which is a set of multiple images taken under different exposure conditions (in most cases, the exposure time is varied). A multi-exposure image can be easily obtained with a general image sensor. Many methods have been proposed to create a single high dynamic range (HDR) image by integrating a multi-exposure image into a single image [3], [4], [5]. An HDR image differs from a low dynamic range (LDR) image such as a jpeg image in that its pixel values are linear to the luminance of the actual scene and its dynamic range is wider than that of an LDR image. To display an HDR image on a device with a standard dynamic range, the range of pixel values must be compressed in accordance with the dynamic range of the device. Such dynamic range compression is called tone mapping [6], [7], [8]. Many methods of obtaining an LDR image directly from a multi-exposure image without creating HDR images have been proposed. Such methods are called exposure fusion (EF) [9], [10], [11], [12], [14], [15], [16], [17]. EF has

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Sharif.

been widely studied in recent years because it is a simple and convenient approach that does not require the creation of an HDR image.

The quality of an image obtained by EF highly depends on the quality of the input multi-exposure image. The following two problems often occur when acquiring a multi-exposure image.

Problem 1 There may be some misalignments between images due to moving objects in the scene and camera motion.

Problem 2 The dynamic range of a multi-exposure image may not cover the entire luminance dynamic range in a scene.

Problem 1 causes an artifact called a ''ghost'' in the fused image. A multi-exposure image can be obtained without spanning different points in time by using multiple sensors [18], [19] or spatially varying exposures [20], [21], [22] to avoid the occurrence of Problem 1. However, in these methods, the number of images that can be acquired at one time is limited due to the constraint of the hardware architecture. The effect of a ghost can be alleviated by using a fusion method that is robust to dynamic scenes [14], [17]. These methods have the advantage that they do not require a special hardware design. Fig. 1 shows a multi-exposure image in which Problem 2 occurs. In Fig. 1, the right side of the scene is dark and unclear in all the images. This indicates that the dynamic range of the scene is not fully covered by the multi-exposure image shown in Fig. 1. In areas whose luminance is not covered by the dynamic range of the multi-exposure image, the fusion result will be unclear and its visibility will be low. Problem 2 also degrades the performance of the fusion algorithm for ghost removal. Problem 2 tends to occur when the shooting time is limited. For example, in hand-held photography, setting a long exposure time is discouraged to avoid image blurring, making it difficult to obtain long-exposure images. A high shutter speed is required to take a photography of rapidly moving objects, which makes it difficult to obtain long-exposure images. In these situations, it is extremely difficult to capture a large number of images under various exposure conditions. In real-life shooting situations, it is often difficult to obtain a multi-exposure image that covers the luminance dynamic range of the scene clearly. Nevertheless, most fusion algorithms are designed under the assumption that the input multi-exposure image covers the entire luminance dynamic range of the scene. Recently, Kinoshita and Kiya proposed an effective pre-processing method of EF to suppress the adverse effect caused by Problem 2 [23]. They named it scene segmentation-based luminance adjustment (SSLA). SSLA can be used in various applications including single image enhancement [24].

The main part of SSLA consists of three steps: (1) local contrast enhancement of the input multi-exposure image based on dodging and burning [25], (2) scene segmentation, and (3) luminance scaling based on the result of scene segmentation. In [23], two approaches are proposed for scene segmentation. In the first approach, the scene is segmented by selecting one image with intermediate brightness and dividing its luminance dynamic range equally. In the second approach, the luminance distribution of the input multi-exposure image is modeled by the Gaussian mixture model (GMM) and the scene is segmented by clustering pixels based on the probability density function obtained through GMM fitting. The first approach is fast, but it often fails to generate good scene segmentation results because it makes little use of the luminance distribution information. The second approach generates better scene segmentation results. However, fitting a GMM requires a long computation time. In addition, the second approach sometimes generates regions with very few pixels. Such regions have little effect on the final fusion result while increasing the computational time for the EF process.

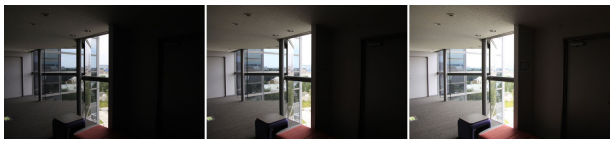


FIGURE 1. Multi-exposure image that fails to cover the luminance dynamic range of the scene clearly.

In this paper, we propose a new scene segmentation method for SSLA. The proposed method repeatedly chooses one region and divides it into two regions until the scene segmentation result satisfies the stopping condition. In the proposed method, the nonlinear luminance distribution and the area occupancy of each segmented region are exploited. The proposed method stably generates a better scene segmentation result than Approach A by considering the information of the nonlinear luminance distribution. It also reduces the number of images in the output multi-exposure image, thus reducing the overall computational cost for the following processes such as EF by suppressing the generation of regions with very few pixels. Furthermore, the proposed scene segmentation method is much faster than Approach B and its processing speed is comparable to that of Approach A. Through various experiments, it is revealed that the proposed scene segmentation method outperforms not only Approach A but also Approach B.

The rest of this paper is organized as follows. Chapter II describes the notations used in this paper. In Chapter III, the algorithm of SSLA is described. In Chapter IV, our new scene segmentation method for SSLA is proposed. Chapter V describes the experimental results. Chapter VI concludes this paper.

II. NOTATIONS IN THIS PAPER

In this section, the notations used in this paper are described.

The set of all pixels of an image is denoted as $\mathbb{P}$. Each element of $\mathbb{P}$ is a two-dimensional vector $\boldsymbol{p} = (u, v)$. A partition of $\mathbb{P}$, which is obtained by grouping elements of $\mathbb{P}$ into M parts, is denoted as $\{\mathbb{P}_m | m \in \{1, 2, \cdots, M\}\}$. For a set with indexed elements such as i, the simplified notation $\{\mathbb{P}_m\}$ is used to denote the same set. The height and width of an image

are denoted as U and V, respectively. The number of elements in a set is represented by $|\cdot|$. For example, the number of pixels of an image is denoted as $|\mathbb{P}|$.

A color image whose pixel value is the linear-RGB value is represented by a boldface character as in $\boldsymbol{x}$. The term "linear-RGB value" is used here to mean a value represented by a coordinate in a color space that has a linear relationship with the CIE-XYZ color space. The pixel value of $\boldsymbol{x}$ at position $\boldsymbol{p}$ is denoted as $\boldsymbol{x}(\boldsymbol{p})$. $\boldsymbol{x}(\boldsymbol{p})$ can be transformed to a value in the CIE-XYZ color space by applying a linear transformation. In this paper, the luminance of a color image is defined as the Y component of the CIE-XYZ color space.

The pixel value of a monochrome image l at position $\boldsymbol{p}$ is denoted as $l(\boldsymbol{p})$. The geometric mean for $\mathbb{P}_m$ of l is defined as follows:

$$G\left(l|\mathbb{P}_m\right) = \exp\left(\frac{1}{|\mathbb{P}_m|}\sum_{\boldsymbol{p}\in\mathbb{P}_m}\log\left(\max(l(\boldsymbol{p}),\epsilon)\right)\right), \quad (1)$$

where ϵ is a small positive constant. The geometric mean of a whole image is simply denoted as $G(l)$. $G(l)$ is obtained by replacing $\mathbb{P}_m$ by $\mathbb{P}$ in the right-hand side of (1).

III. SSLA

In this chapter, each process of SSLA is described. Note that the local contrast enhancement process of SSLA is referred to "CE" in this paper.

A. CE

Let $\{\boldsymbol{x}_n\} = \{\boldsymbol{x}_n | n \in \{1, 2, \cdots, N\}\}$ be the input multi-exposure image, which contains N color images. In SSLA, the luminance components of $\{\boldsymbol{x}_n\}$ are first calculated, which are denoted as $\{l_n\} = \{l_n | n \in \{1, 2, \cdots, N\}\}$. The contrast-enhanced luminance images are denoted as $\{l'_n\} = \{l'_n | n \in \{1, 2, \cdots, N\}\}$. Let $\bar{l}_n$ be the image obtained by applying a bilateral filter [26], [27] to l_n. In accordance with [25], $l'_n(\boldsymbol{p})$ is calculated as follows:

$$l'_n(\boldsymbol{p}) = \frac{l_n(\boldsymbol{p})}{\bar{l}_n(\boldsymbol{p})} l_n(\boldsymbol{p}). \quad (2)$$

If a user prefers, CE can be skipped. In such a case, $\{l'_n\}$ is obtained by the simple operation $\{l'_n\} \leftarrow \{l_n\}$, where $\leftarrow$ denotes substitution. Note that a bilateral filter can be replaced with other edge-preserving smoothing filters such as guided image filter [28] and weighted guided image filter [29].

B. SCENE SEGMENTATION

In [23], Kinoshita and Kiya proposed two approaches for scene segmentation. In this paper, Approach 1 and Approach 2 in [23] are referred to Approach A and Approach B, respectively.

Approach A: In Approach A, the geometric mean of l'_n, namely $G(l'_n)$, is first calculated for each n. Secondly, n^{med} that satisfies $G(l'_{n^{\mathrm{med}}}) = \mathrm{median}_n G(l'_n)$ is calculated. Then, θ_ω is calculated as follows:

$$\theta_\omega = \frac{\omega}{M}\left(\max_{\boldsymbol{p}} l'_{n^{\mathrm{med}}}(\boldsymbol{p}) - \min_{\boldsymbol{p}} l'_{n^{\mathrm{med}}}(\boldsymbol{p})\right) + \min_{\boldsymbol{p}} l'_{n^{\mathrm{med}}}(\boldsymbol{p}), \quad (3)$$

where $\omega \in \{0, 1, 2, \cdots, M\}$. In Approach A, $\mathbb{P}_m$ is determined as follows:

$$\mathbb{P}_m = \begin{cases} \left\{\boldsymbol{p} \middle| \theta_{m-1} \le l'_{\mathrm{med}}(\boldsymbol{p}) < \theta_m\right\} (m < M) \\ \left\{\boldsymbol{p} \middle| \theta_{m-1} \le l'_{\mathrm{med}}(\boldsymbol{p}) \le \theta_m\right\} (m = M) \end{cases}, \quad (4)$$

where $m \in \{1, 2, \cdots, M\}$. In Approach A, M is manually set by a user.

Approach B: In Approach B, the N dimensional vector $\boldsymbol{l}'(\boldsymbol{p}) = (l'_1(\boldsymbol{p}), l'_2(\boldsymbol{p}), \cdots, l'_N(\boldsymbol{p}))$ is first generated for each pixel $\boldsymbol{p}$. Secondly, the distribution of $\boldsymbol{l}'(\boldsymbol{p})$ is modeled by using the GMM and its parameters are estimated. The GMM model for $\boldsymbol{l}'(\boldsymbol{p})$ is represented as follows:

$$p(\boldsymbol{l}'(\boldsymbol{p})) \propto \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{l}'(\boldsymbol{p}) | \boldsymbol{\mu}_k, \Sigma_k\right), \quad (5)$$

where $p(\boldsymbol{l}'(\boldsymbol{p}))$ is the probability of generating $\boldsymbol{l}'(\boldsymbol{p})$. $\mathcal{N}\left(\cdot|\boldsymbol{\mu}_k, \Sigma_k\right)$ is an N-dimensional Gaussian distribution with mean $\boldsymbol{\mu}_k$ and variance-covariance matrix Σ_k. π_k and K are a mixing coefficient and the number of components, respectively. In [23], variational Bayesian inference is utilized to estimate model parameters. Variational Bayesian inference can avoid overfitting by automatically removing some components from all K components. In Approach B, M, namely the number of regions, is determined by the number of components remaining after variational Bayesian inference. The set of indexes of the remaining components is defined by $\{k_m | m \in \{1, 2, \cdots, M\}\}$, where $k_m \in \{1, 2, \cdots, K\}$. Finally, $\mathbb{P}_m$ is determined as follows:

$$\mathbb{P}_m = \left\{\boldsymbol{p} \middle| \arg\max_j \pi_{k_j} \mathcal{N}\left(\boldsymbol{l}'(\boldsymbol{p}) \middle| \boldsymbol{\mu}_{k_j}, \Sigma_{k_j}\right) = m\right\}. \quad (6)$$

C. LUMINANCE SCALING

Fig. 2 shows the explanatory diagram for luminance scaling in SSLA. In the luminance scaling process, for each region $\mathbb{P}_m$, one image of $\{l'_n\}$ is selected and its luminance is scaled so that the luminance-scaled image has mid-level brightness in $\mathbb{P}_m$. The M obtained luminance-scaled images are denoted

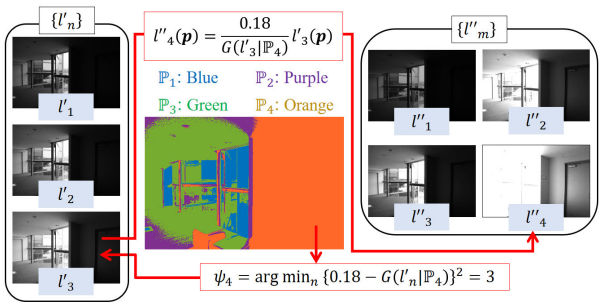


FIGURE 2. Explanatory diagram for luminance scaling in SSLA.

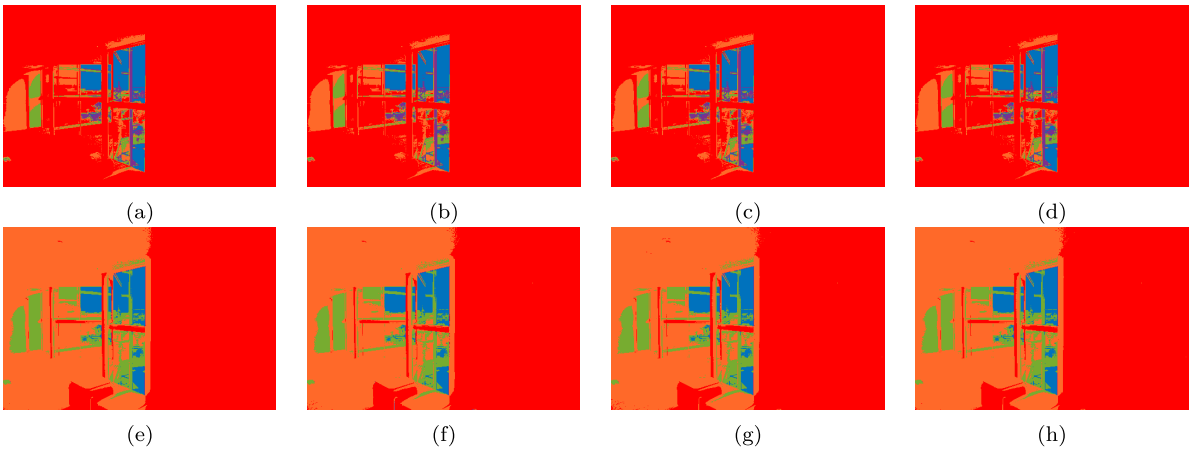


FIGURE 3. Scene segmentation results of Approach A and Approach B for the multi-exposure image shown in Fig. 1. (a) Approach A without CE, (b) Approach A with CE (bilateral filter), (c) Approach A with CE (guided image filter), (d) Approach A with CE (weighted guided image filter), (e) Approach B without CE, (f) Approach B with CE (bilateral filter), (g) Approach B with CE (guided image filter), (h) Approach B with CE (weighted guided image filter).

as $\{l''_m\} = \{l''_m | m \in \{1, 2, \cdots, M\}\}$. To obtain $\{l''_m\}$, ψ_m and α_m are calculated as follows:

$$\psi_m = \arg\min_n \left(0.18 - G(l'_n|\mathbb{P}_m)\right)^2, \tag{7}$$

$$\alpha_m = \frac{0.18}{G(l'_{\psi_m}|\mathbb{P}_m)}. \tag{8}$$

Then, $l''_m(\boldsymbol{p})$ is calculated as follows:

$$l''_m(\boldsymbol{p}) = \alpha_m l'_{\psi_m}(\boldsymbol{p}). \tag{9}$$

From $\{l'_n\}$, (7) is used to select the image with highest quality in region $\mathbb{P}_m$. This image is the most suitable for generating the luminance-scaled image corresponding to $\mathbb{P}_m$. Note that the luminance scaling in (9) is applied not to only the pixels in $\mathbb{P}_m$ but to all the pixels of the image. As the same luminance scaling operation is applied to all the pixels, l''_m and l'_{ψ_m} have the same image structure.

D. TONE MAPPING AND COLOR IMAGE GENERATION

After $\{l''_m\}$ is obtained, Reinhard's tone mapping operator [6] is applied to each image of $\{l''_m\}$. Let $\hat{l}_m$ be the tone mapping result of l''_m. $\hat{l}_m(\boldsymbol{p})$ is calculated as follows:

$$\hat{l}_m(\boldsymbol{p}) = \frac{l''_m(\boldsymbol{p})}{1 + l''_m(\boldsymbol{p})}\left(1 + \frac{l''_m(\boldsymbol{p})}{{L_m}^2}\right), \tag{10}$$

where L_m is the parameter that controls the value of $l''_m(\boldsymbol{p})$ satisfying $\hat{l}_m(\boldsymbol{p}) = 1$. With $l''_m(\boldsymbol{p}) = L_m$, $\hat{l}_m(\boldsymbol{p}) = 1$ holds.

The output multi-exposure image of SSLA is denoted as $\{\hat{\boldsymbol{x}}_m\} = \{\hat{\boldsymbol{x}}_m | m \in \{1, 2, \cdots, M\}\}$. $\hat{\boldsymbol{x}}_m(\boldsymbol{p})$ is calculated as follows:

$$\hat{\boldsymbol{x}}_m(\boldsymbol{p}) = \frac{\hat{l}_m(\boldsymbol{p})}{l_{\psi_m}(\boldsymbol{p})}\boldsymbol{x}_{\psi_m}(\boldsymbol{p}). \tag{11}$$

The hue of $\hat{\boldsymbol{x}}_m(\boldsymbol{p})$ is made the same as that of $\boldsymbol{x}_{\psi_m}(\boldsymbol{p})$ [23] by using (11).

E. PROBLEMS OF SCENE SEGMENTATION METHOD PROPOSED BY KINOSHITA AND KIYA

Figs. 3(a), 3(b), 3(c), and 3(d) show the scene segmentation results of Approach A ($M = 5$) for the multi-exposure image shown in Fig. 1. In Fig. 3(a), CE was not applied before scene segmentation. In Figs. 3(b), 3(c), and 3(d), CE was applied before scene segmentation with bilateral filter, guided image filter [28], and weighted guided image filter [29], respectively. The parameter setting of the bilateral filter in SSLA is the same as that in [23]. For the bilateral filter, the fast implementation described in [27] is utilized. The filter radius and regularization parameter in the guided image filter and weighted guided image filter are set to 15 and 0.0003, respectively. For the guided image filter and weighted guided image filter, the linear time implementation of mean filtering [30] is utilized. From Figs. 3(a), 3(b), 3(c), and 3(d), it can be seen that Approach A was not able to separate the right dark areas and upper-left bright areas regardless of whether CE is applied or not applied. Fig. 4(a) shows the output multi-exposure image generated by Approach A without CE. From Fig. 4(a), we observe that Approach A failed to generate an image that is sufficiently bright in the right area. Approach A may be able to separate the right dark areas and upper-left bright areas by increasing M. However, increasing M means increasing the memory consumption and the computational cost for the fusion process following SSLA.

Figs. 3(e), 3(f), 3(g), and 3(h) show the scene segmentation results of Approach B for the multi-exposure image shown in Fig. 1. In Fig. 3(e), CE was not applied before scene segmentation. In Figs. 3(f), 3(g), and 3(h), CE was applied before scene segmentation with bilateral filter, guided image filter, and weighted guided image filter, respectively. The parameter settings for the bilateral filter, guided image filter, and weighted guided image filter are the same as those used to obtain the results in Figs. 3(b), 3(c), and 3(d). In this example, K was set to 10 and five images were generated as

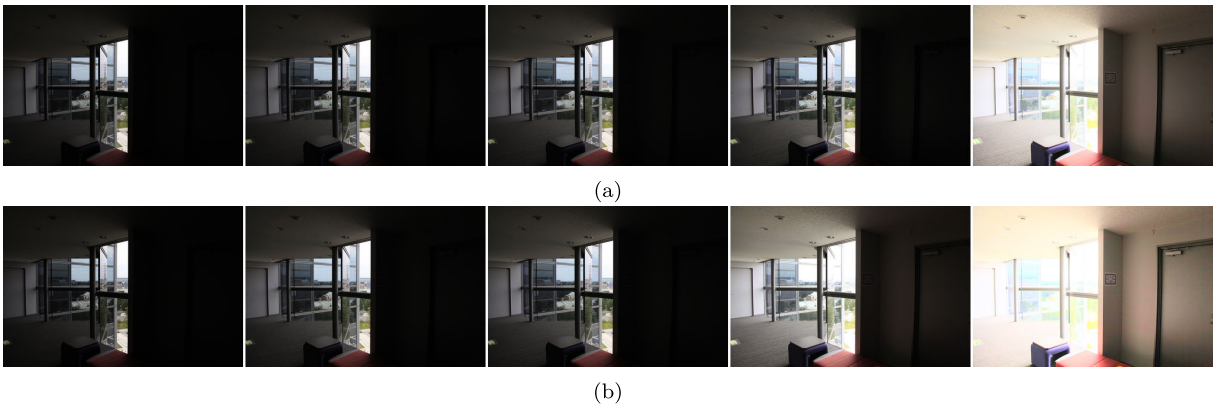


FIGURE 4. Output multi-exposure image generated by Approach A and Approach B for the multi-exposure image shown in Fig. 1. (a) Approach A without CE, (b) Approach B without CE.

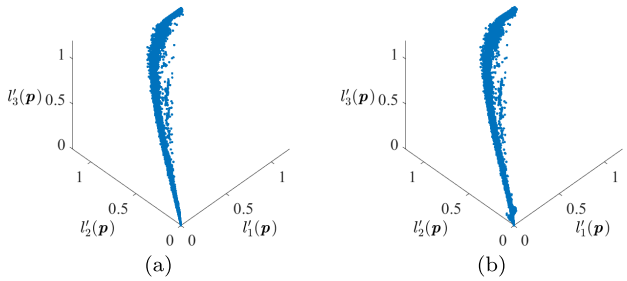


FIGURE 5. Distribution of $\boldsymbol{l}'(\boldsymbol{p})$ for the multi-exposure image shown in Fig. 1. (a) Without CE, (b) with CE (bilateral filter).

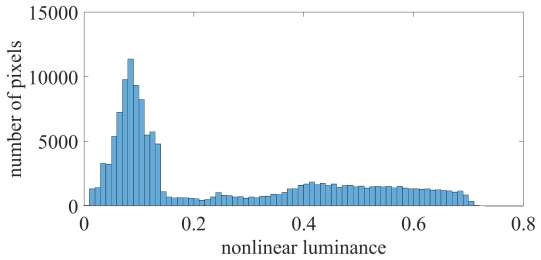


FIGURE 6. Histogram of $\left\{\left(l'_{\psi_m}(p)\right)^{\frac{1}{2.2}} \middle| p \in \mathbb{P}_m\right\}$ for the widest region in Fig. 3(a), which is colored red.

a result (i.e., $M = 5$). From Figs. 3(e), 3(f), 3(g), and 3(h), it can be seen that Approach B was able to separate the right dark areas and upper-left bright areas. Fig. 4(b) shows the output multi-exposure image generated by Approach B without CE. From Fig. 4(b), we see that Approach B was able to generate an image that is sufficiently bright in the right area.

Fig. 5 shows the distribution of $\boldsymbol{l}'(\boldsymbol{p}) = (l'_1(\boldsymbol{p}), l'_2(\boldsymbol{p}), l'_3(\boldsymbol{p}))$ for the multi-exposure image shown in Fig. 1. From Fig. 5, it is difficult to conclude that the GMM model is suitable for modeling the distribution. In addition, one of the regions generated by Approach B contains only 0.2% of all the pixels in the image. Such a region has only a small effect on the final fusion result, whereas it increases M. The second and third images from the left in Fig. 4(b) have almost the same brightness, which indicates that one of them can be removed. The other problem of Approach B is its long computational time, variational Bayesian inference takes a long time even if the maximum number of iterations is limited as described in [23].

IV. PROPOSED SCENE SEGMENTATION METHOD FOR SSLA

In this chapter, our new scene segmentation method for SSLA is proposed.

A. IMPORTANT INFORMATION FOR SCENE SEGMENTATION IN SSLA

Fig. 6 shows the histogram of $\left\{\left(l'_{\psi_m}(\boldsymbol{p})\right)^{\frac{1}{2.2}} \middle| \boldsymbol{p} \in \mathbb{P}_m\right\}$ for the widest region in Fig. 3(a), which is colored red. The histogram indicates that l'_{ψ_m} has not only dark pixels but also pixels with mid-level brightness in $\mathbb{P}_m$. Therefore, the geometric mean of $\left\{l'_{\psi_m}(\boldsymbol{p}) \middle| \boldsymbol{p} \in \mathbb{P}_m\right\}$ (i.e., $G(l'_{\psi_m}|\mathbb{P}_m)$) becomes closer to 0.18. In such a case, the correction coefficient α_m calculated using (8) becomes close to 1. This means that the dark pixels in $\mathbb{P}_m$ of l'_{ψ_m} are not brightened sufficiently. To avoid such insufficient luminance correction, the distribution information obtained from $\left\{\left(l'_{\psi_m}(\boldsymbol{p})\right)^{\frac{1}{2.2}} \middle| \boldsymbol{p} \in \mathbb{P}_m\right\}$ is important for scene segmentation in SSLA. As explained in Section III-E, the generation of a region containing too few pixels should be avoided in SSLA. Hence $|\mathbb{P}_m|$ is also important information for helping detect whether the number of pixels in $\mathbb{P}_m$ is too small.

B. PROPOSED SCENE SEGMENTATION METHOD FOR SSLA

The proposed scene segmentation method is also applied to $\{l'_n\}$ similarly to the methods described in Section III-B. In the proposed scene segmentation method, $\mathbb{P}$ is iteratively

segmented. Let $\{\mathbb{P}_m^{(i)}\}$ and $M^{(i)}$ be the scene segmentation results and the number of regions in the ith iteration, respectively. In the initial state ($i = 0$), $\mathbb{P}_1^{(0)} = \mathbb{P}$ and $M^{(0)} = 1$ hold. In the $(i+1)$ th iteration, $\psi_m^{(i)}$ and $\alpha_m^{(i)}$ are first calculated with (7) and (8), respectively. Then, the set $S_m^{(i)}$ is obtained as follows:

$$S_m^{(i)} = \left\{ \left(l'_{\psi_m^{(i)}}(\boldsymbol{p}) \right)^{\frac{1}{2.2}} \middle| \boldsymbol{p} \in \mathbb{P}_m^{(i)} \right\}. \tag{12}$$

After $S_m^{(i)}$ is obtained, $\kappa^{(i)}$ is calculated as follows:

$$\kappa^{(i)} = \arg\max_m \frac{|\mathbb{P}_m^{(i)}|}{|\mathbb{P}|} \lambda_m^{(i)}, \tag{13}$$

$$\lambda_m^{(i)} = \sqrt{\frac{1}{|S_m^{(i)}|} \sum_{s \in S_m^{(i)}} \left(s - \bar{S}_m^{(i)} \right)^2}, \tag{14}$$

$$\bar{S}_m^{(i)} = \frac{1}{|S_m^{(i)}|} \sum_{s \in S_m^{(i)}} s. \tag{15}$$

Note that $|S_m^{(i)}| = |\mathbb{P}_m^{(i)}|$ holds. If $\lambda_m^{(i)}$ is large, $l'_{\psi_m^{(i)}}$ has pixels with the widely varying brightness similarly to the example in the previous section. $\mathbb{P}_m^{(i)}$ with a large value of $\lambda_m^{(i)}$ should be divided on the basis of the discussion in the previous section. On the other hand, if $\frac{|\mathbb{P}_m^{(i)}|}{|\mathbb{P}|}$ is too small, $\mathbb{P}_m^{(i)}$ contains too few pixels. To avoid generating a region containing too few pixels, it should be checked whether $\frac{|\mathbb{P}_m^{(i)}|}{|\mathbb{P}|}$ is sufficiently large before segmenting $\mathbb{P}_m^{(i)}$. Equation (13) is used to determine $\kappa^{(i)}$ so that both $\lambda_{\kappa^{(i)}}^{(i)}$ and $\frac{|\mathbb{P}_{\kappa^{(i)}}^{(i)}|}{|\mathbb{P}|}$ are large. $\mathbb{P}_{\kappa^{(i)}}^{(i)}$ is divided into two regions by the Otsu method [31]. In this method, the threshold t_{otsu} is calculated as follows:

$$t_{\text{otsu}} = \arg\max_{t \in S_\kappa} |Q_1(t)||Q_2(t)| \left(\bar{Q}_1(t) - \bar{Q}_2(t) \right)^2, \tag{16}$$

$$Q_1(t) = \{ s | s \geq t, s \in S_\kappa \}, \tag{17}$$

$$Q_2(t) = \{ s | s < t, s \in S_\kappa \}, \tag{18}$$

$$\bar{Q}_1(t) = \frac{1}{|Q_1(t)|} \sum_{q \in Q_1(t)} q, \tag{19}$$

$$\bar{Q}_2(t) = \frac{1}{|Q_2(t)|} \sum_{q \in Q_2(t)} q, \tag{20}$$

where the superscript (i), which means the ith iteration, is omitted for notational convenience. $\mathbb{P}_{\kappa^{(i)}}^{(i)}$ is divided into two regions, $\mathbb{P}_{\kappa_1}^{(i)}$ and $\mathbb{P}_{\kappa_2}^{(i)}$, by using $t_{\text{otsu}}^{(i)}$ as follows:

$$\mathbb{P}_{\kappa_1}^{(i)} = \left\{ \boldsymbol{p} \in \mathbb{P}_\kappa^{(i)} \middle| \left(l'_{\psi_m^{(i)}}(\boldsymbol{p}) \right)^{\frac{1}{2.2}} \geq t_{\text{otsu}}^{(i)} \right\}, \tag{21}$$

$$\mathbb{P}_{\kappa_2}^{(i)} = \left\{ \boldsymbol{p} \in \mathbb{P}_\kappa^{(i)} \middle| \left(l'_{\psi_m^{(i)}}(\boldsymbol{p}) \right)^{\frac{1}{2.2}} < t_{\text{otsu}}^{(i)} \right\}. \tag{22}$$

The scene segmentation result in the $(i+1)$ th iteration, namely $\{\mathbb{P}_m^{(i+1)}\}$, is obtained as follows.

$$\mathbb{P}_m^{(i+1)} = \begin{cases} \mathbb{P}_m^{(i)} & (m \neq \kappa) \\ \mathbb{P}_{\kappa_1}^{(i)} & (m = \kappa) \\ \mathbb{P}_{\kappa_2}^{(i)} & (m = i+1 = M^{(i)}). \end{cases} \tag{23}$$

In the proposed method, the above segmentation process is repeated until the stopping condition is satisfied. The stopping condition is defined as follows:

$$\max_{m \in \{1,2,\cdots,M^{(i)}\}} \frac{|\mathbb{P}_m^{(i)}|}{|\mathbb{P}|} \lambda_m^{(i)} \leq t_{\text{stop}}, \tag{24}$$

where t_{stop} is a parameter. To avoid the generation of too few regions, the minimal number of regions is set to 3 in the proposed method (i.e., M must be greater than 2).

Fig. 7 shows the segmentation result of the proposed scene segmentation method for the multi-exposure image shown in Fig. 1. In Fig. 7(a), CE was not applied before scene segmentation. In Figs. 7(b), 7(c), and 7(d), CE was applied before scene segmentation with bilateral filter, guided image filter [28], and weighted guided image filter [29], respectively. The parameter setting for scene segmentation is the same as those described in Subsection V-A4. The parameter settings for the bilateral filter, guided image filter, and weighted guided image filter are the same as those used to obtain the results in Figs. 3(b), 3(c), and 3(d). In Fig. 7, it can be observed that the right under-exposed region and the upper-left medium-exposed region are properly segmented as different regions. It also can be observed that the choice of the edge-preserving smoothing filter for CE has no significant effect on the result of scene segmentation. In the rest of this paper, the bilateral filter is utilized for CE. Fig. 8 shows the output of SSLA based on the scene segmentation result shown in Fig. 7(a). Similarly to in Fig. 4(a) and 4(b), CE was not applied. In Fig. 8, it can be seen that an image that is sufficiently bright in the right region is generated (see the rightmost image).

C. FAST IMPLEMENTATION OF PROPOSED SCENE SEGMENTATION METHOD

$\psi_m^{(i)}$, $\alpha_m^{(i)}$, and $t_{\text{otsu}}^{(i)}$ are calculated on the basis of simple statistics such as the mean and standard deviation. The mean and standard deviation can be approximately estimated by using some samples instead of all the data. Hence, the computational cost of calculating $\psi_m^{(i)}$, $\alpha_m^{(i)}$, and $t_{\text{otsu}}^{(i)}$ can be easily reduced by limiting the maximum number of samples to be used.

V. EXPERIMENT

A. SIMULATION CONDITIONS

1) METRICS USED FOR QUANTITATIVE EVALUATIONS

In this experiment, the following four metrics were used for quantitative evaluation.

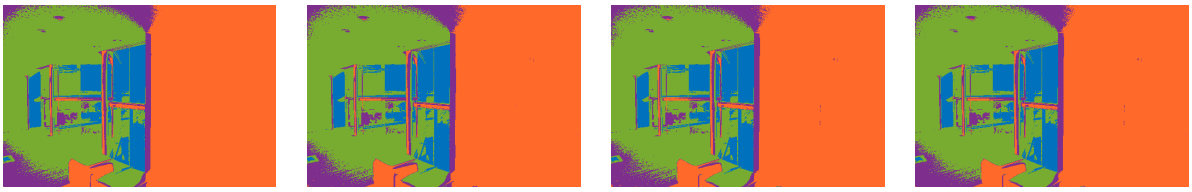


FIGURE 7. Scene segmentation results of the proposed method for the multi-exposure image shown in Fig. 1. (a) Without CE, (b) with CE (bilateral filter), (c) with CE (guided image filter), (d) with CE (weighted guided image filter).

FIGURE 8. Output multi-exposure image generated by the proposed method for the multi-exposure image shown in Fig. 1. Note that CE was not applied.

a: MEF-SSIM

MEF-SSIM [32] is the metric for measuring the quality of the fusion result for a multi-exposure image. To calculate MEF-SSIM, the structural similarity is calculated between the fusion result and the multi-exposure image in a multi-scale manner. MEF-SSIM can evaluate both the global and local structure preservation. MEF-SSIM assumes that the multi-exposure image contains a sufficient number of images and that the overall quality of the images is high.

b: TMQI

TMQI [33] is a metric for measuring the quality of the tone mapping result of an HDR image. TMQI consists of structural fidelity and statistical naturalness (SN) measurements. Structural fidelity measures the structural similarity between the HDR image and the tone mapping result. SN measures the naturalness of the luminance distribution of the image obtained by tone mapping.

c: STATISTICAL NATURALNESS (SN)

SN, which is calculated in TMQI [33], can be calculated for an arbitrary 24-bit color image. SN is defined as $\mathcal{N}\left(\bar{L}|115.94, 27.99^2\right)\mathcal{B}\left(\bar{\sigma}_l/64.29|4.4, 10.1\right)/O$, where $\mathcal{B}\left(\cdot|\alpha, \beta\right)$ is a beta probability density function with parameters α and β. $\bar{L}$ and $\bar{\sigma}_l$ are the average luminance of an image and the average of the local standard deviation of luminance, respectively. O is a normalization factor.

d: DISCRETE ENTROPY (DE)

Discrete entropy (DE) is a metric for measuring the global contrast of a monochrome image. To calculate the DE, an image histogram with 256 bins is first created. The DE is calculated as the entropy of the histogram. In this paper, the DE is calculated for the luminance of the final fusion result.

2) DATASET USED FOR EXPERIMENTS

In this experiment, we use five different datasets, which are denoted as **Dataset 1**, **Dataset 2**, **Dataset 3**, **Dataset 4**, and **Dataset 5**. For each set of multi-exposure images, a few of the darker images are selected as the input for SSLA. **Dataset 1** comprises multi-exposure images created by applying the tone mapping operator to 104 HDR images selected from an online database [34]. In accordance with [23], the tone mapping operator is applied with a linear response function. For each HDR image, a multi-exposure image with EV values of $\{-7, -6, \cdots, 6, 7\}$ is created. From the 15 created images, the images with EV values of -7, -5, and -3 are selected as the input of SSLA. MEF-SSIM is calculated from the final fusion result and the 15 images with all EV values. TMQI is calculated from the final fusion result and the HDR image used to create the multi-exposure image. In this experiment, each HDR image is downsized so that it has 1024 pixels on the longer side. **Dataset 2** is the dataset used in [3] and contains eight sets of multi-exposure images. Each set in **Dataset 2** is a multi-exposure image of a dynamic scene and contains some misalignments. Feeding Time consists of three images and the other sets consist of five or more images. For Feeding Time, the two darkest images are selected as the input for SSLA. For the other seven sets, the three darkest images are used as the input for SSLA. **Dataset 3** is taken from an online database [35] and contains 12 sets of multi-exposure images of a static scene. For Cave, even the second brightest image is so dark that more than 15% of the pixels are zero due to clipping. Therefore, all the images are used as the input for SSLA. For the sets except for Cave, the two darkest images are used as the input for SSLA. **Dataset 4** and **Dataset 5** are the dataset extracted from the dataset used in [36]. The indexes of the sets extracted from the dataset in [36] are summarized in Table 1. **Dataset 4** consists of 45 sets of multi-exposure

(a) (b) (c)

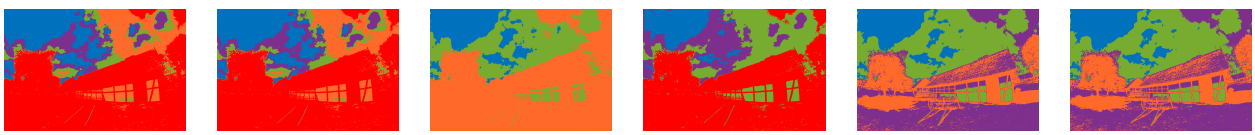
FIGURE 9. Three sets of multi-exposure images used for the explanations in Section V-C. (a) Bar Harbor Sunrise in **Dataset 1** (-7EV, -5EV, -3EV), (b) Baby at Window in **Dataset 2** (three darker images), (c) Kluki in **Dataset 2** (two darker images).

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

FIGURE 10. Scene segmentation results for Bar Harbor Sunrise. (a)-(f) are the results of A, A (CE), B, B (CE), Pr., and Pr. (CE), respectively.

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

FIGURE 11. Scene segmentation results for Baby at Window. (a)-(f) are the results of A, A (CE), B, B (CE), Pr., and Pr. (CE), respectively.



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

FIGURE 12. Scene segmentation results for Kluki. (a)-(f) are the results of A, A (CE), B, B (CE), Pr., and Pr. (CE), respectively.

TABLE 1. Indexes of sets extracted from the dataset used in [36]. Note that the dataset used in [36] is divided into Part 1 and Part 2.

Dataset 4	2, 3, 16, 23, 42, 49, 52, 57, 61, 71, 84, 102, 111, 112, 113, 116, 117, 119, 120, 129, 131, 223, 251, 255, 256, 257, 259, 260 from Part1 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125 from Part2
Dataset 5	54, 66, 67, 70, 72, 87, 90, 104, 122, 123, 222, 261, 264, 269, 270 from Part1

images of a static scene and **Dataset 5** consists of 15 sets of multi-exposure images of a dynamic scene. For each set of **Dataset 4** and **Dataset 5**, the two darkest images are used as the input for SSLA. For **Dataset 2**, **Dataset 3**, **Dataset 4**, and **Dataset 5**, MEF-SSIM and TMQI are not used for quantitative evaluation.

3) FUSION METHODS USED IN EXPERIMENTS

The fusion methods used in this experiment are those of Mertens et al. [9], Nejati et al. [12], Kou et al. [13], and Li et al. [17]. For each fusion method, the codes provided by the authors and the default parameter settings in the codes are used. For **Dataset 2** and **Dataset 5**, only Li et al.'s method is applied since only this method can be applied to the multi-exposure image of a dynamic scene.

4) PARAMETER SETTING FOR SSLA

The parameter setting of the bilateral filter in SSLA is the same as that in [23]. For the bilateral filter, the fast implementation described in [27] is utilized. For the GMM fitting in Approach B, l'_n is downsized in accordance with [23] so that it has 256 pixels on the longer side to reduce the computational cost. The maximum number of iterations in the GMM fitting is set to 100 in accordance with [23]. For the scene segmentation of the proposed method, the maximum number of samples used for calculating $\psi_m^{(i)}$, $\alpha_m^{(i)}$, and $t_{\mathrm{otsu}}^{(i)}$ is limited to 10,000. t_{stop} is set to 0.05. L_m in (10) is determined as $\max_{\boldsymbol{p}} l''_m(\boldsymbol{p})$.

FIGURE 13. Comparison of three segmentation methods (Approach A, Approach B, and proposed scene segmentation method) for Bar Harbor Sunrise. (a)-(d) are the results with Mertens et al.'s fusion method [9] for w/o SSLA, A (CE), B (CE), and Pr. (CE), respectively. (e)-(h) are the results with Nejati et al.'s method [12] for w/o SSLA, A (CE), B (CE), and Pr. (CE), respectively. (i)-(l) are the results with Kou et al.'s method [13] for w/o SSLA, A (CE), B (CE), and Pr. (CE), respectively. (m)-(p) are the results with Li et al.'s method [17] for w/o SSLA, A (CE), B (CE), and Pr. (CE), respectively.

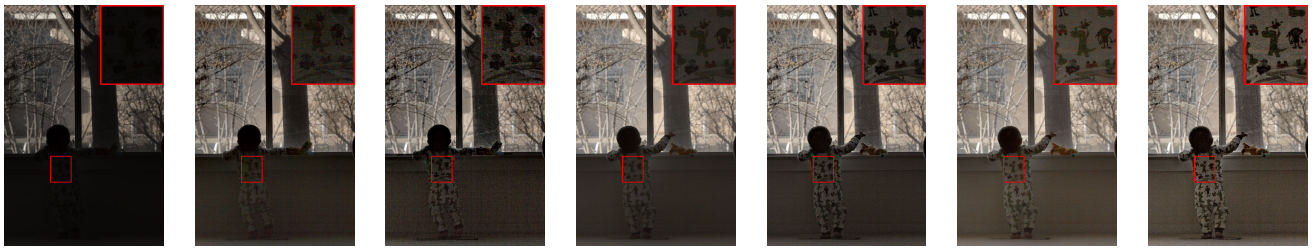


FIGURE 14. Comparison of three segmentation methods (Approach A, Approach B, and proposed scene segmentation method) for Baby at Window. Li et al.'s method [17] was used for EF. (a)-(g) are the results for w/o SSLA, A, A (CE), B, B (CE), Pr., and Pr. (CE), respectively.

B. NOTATIONS IN THIS CHAPTER

In this chapter, the following notations are used. **w/o SSLA** means that SSLA was not applied. **A**, **B**, and **Pr.** mean that SSLA was applied without CE and the scene segmentation in the SSLA was implemented by Approach A, Approach B, and the proposed scene segmentation method, respectively. **(CE)** after **A**, **B**, and **Pr.** means that CE was applied before scene segmentation.

C. VISUAL EVALUATION

Fig. 9 shows three sets of multi-exposure images used for the explanations in this chapter. Figs. 9(a), 9(b), and 9(c) show Bar Harbor Sunrise in **Dataset 1**, Baby at Window in **Dataset 2**, and Kluki in **Dataset 3**, respectively. Note that Fig. 9 shows only the images that were input to SSLA. Figs. 10, 11, and 12 show the scene segmentation results for Bar Harbor Sunrise, Baby at Window, and Kluki, respectively. We see that the proposed method was able to separate the dark regions and bright regions, whereas Approach A and Approach B were not able to do so.

Fig. 13 shows the fusion results for Bar Harbor Sunrise. It can be seen that **Pr.** and **Pr. (CE)** were able to generate fusion results with high visibility, whereas **A**, **A (CE)**, **B**, and **B (CE)** failed to generate fusion results with high visibility

FIGURE 15. Comparison of three segmentation methods (Approach A, Approach B, and proposed scene segmentation method) for Kluki. (a)-(d) are the results with Mertens et al.'s fusion method [9] for w/o SSLA, A (CE), B (CE), and Pr. (CE), respectively. (e)-(h) are the results with Nejati et al.'s method [12] for w/o SSLA, A (CE), B (CE), and Pr. (CE), respectively. (i)-(l) are the results with Kou et al.'s method [13] for w/o SSLA, A (CE), B (CE), and Pr. (CE), respectively. (m)-(p) are the results with Li et al.'s method [17] for w/o SSLA, A (CE), B (CE), and Pr. (CE), respectively.

in the bottom of the scene. This is because Approach A and Approach B were not able to separate the dark region at the bottom of the scene, as seen in Fig. 10. Fig. 14 shows the fusion results for Baby at Window. We see that **Pr.** and **Pr. (CE)** generated fusion results with higher visibility than those of **A**, **A (CE)**, **B**, and **B (CE)**. From the boxed regions in Fig. 14, we also see that **A (CE)**, **B (CE)**, and **Pr. (CE)** amplified noise in the lower dark region. This indicates that CE sometimes amplifies noise as pointed out in [23]. From Fig. 15, which shows the fusion results for Kluki, we see that the proposed scene segmentation method generated a better effect on the fusion result than Approach A and Approach B, as already observed from Figs. 13 and 14.

D. QUANTITATIVE EVALUATION

Fig. 16 summarizes the results of the quantitative evaluation for **Dataset 1** in terms of MEF-SSIM [32], TMQI [33], SN, and DE. In each box plot, the box represents the range from the first quartile q_1 to the third quartile q_3. The whiskers represent the maximum and minimum scores in the range $[q_1-1.5(q_3-q_1), q_1+1.5(q_3-q_1)]$. The horizontal line in the box represents the median and the cross represents the mean. We see that **Pr.** generated better scores for MEF-SSIM than **A** and **B**. This indicates that the proposed method performed better than Approach A and Approach B. From Fig. 16, we also see that **Pr.** stably generated favorable scores for DE, SN, and TMQI when any of the fusion methods were applied. Table 2 shows the results of the quantitative evaluation for **Dataset 2** in terms of SN and DE, where the boldface numbers are the maximum value in each row. It can be seen that **Pr. (CE)** and **B (CE)** generated good scores. Fig. 17 summarizes the results of the quantitative evaluation for **Dataset 3** in terms of SN and DE. We see that **Pr.** and **Pr. (CE)** stably generated favorable scores when any of the fusion methods were applied. The same thing can be observed from Fig. 18, which summarizes the results of the quantitative evaluation for **Dataset 4** in terms of SN and DE. Fig. 19 summarizes the results of the quantitative evaluation for **Dataset 5** in terms of SN and DE. We see that **Pr.** stably generated favorable scores compared with **A** and **B**.

The numbers of regions generated (i.e., M) for all 184 sets in **Dataset 1**, **Dataset 2**, **Dataset 3**, **Dataset 4**, and **Dataset 5**

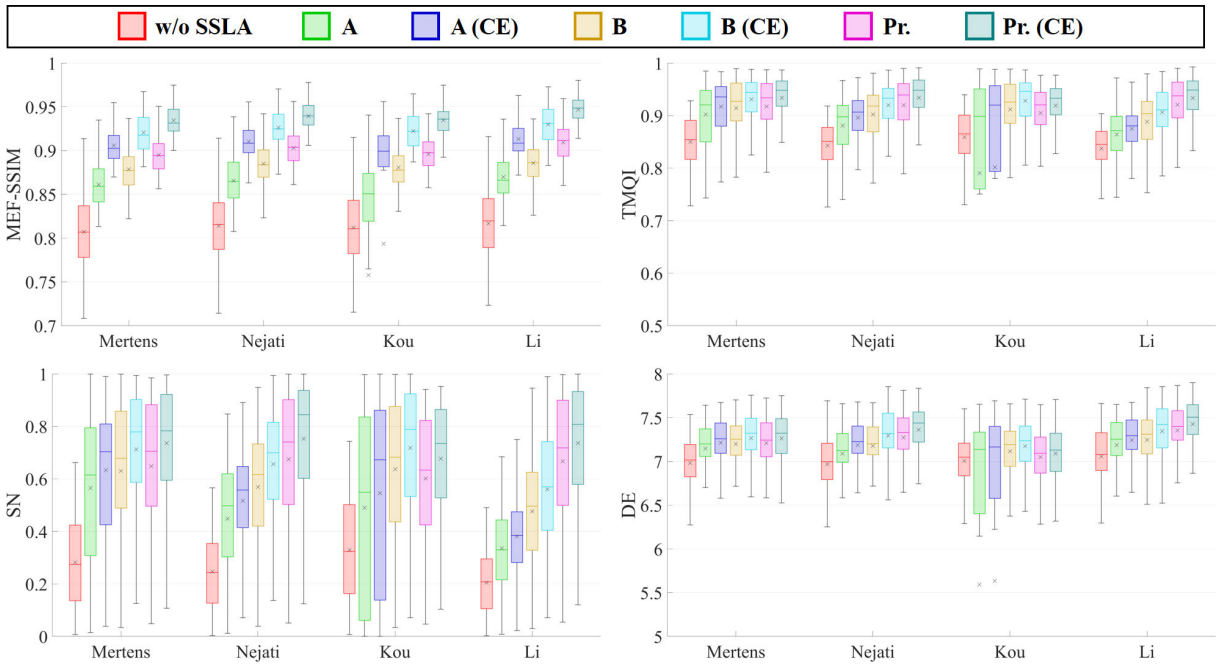


FIGURE 16. Quantitative evaluation scores for **Dataset 1.**

TABLE 2. Quantitative evaluation scores for **Dataset 2.**

Metric	Scene	w/o SSLA	A	A (CE)	B	B (CE)	Pr.	Pr. (CE)
SN	Baby at Window	0.007	0.205	0.297	0.444	0.623	0.630	**0.860**
	Baby on Grass	0.100	0.210	0.298	0.319	0.470	0.578	**0.740**
	Christmas Rider	0.004	0.486	0.378	0.658	0.805	0.795	**0.979**
	Feeding Time	0.095	0.239	0.418	0.288	**0.511**	0.339	0.485
	High Chair	0.049	0.233	0.429	0.297	0.452	0.444	**0.632**
	Lady Eating	0.003	0.120	0.250	0.394	**0.708**	0.335	0.617
	Piano Man	0.001	0.480	0.381	0.654	0.551	**0.764**	0.760
	Santas Little Helper	0.001	0.549	0.246	0.867	0.844	0.801	**0.904**
DE	Baby at Window	6.229	7.138	7.312	7.258	7.435	7.328	**7.511**
	Baby on Grass	6.925	7.065	7.168	7.154	7.263	7.088	**7.272**
	Christmas Rider	6.276	7.340	7.536	7.425	7.798	7.649	**7.800**
	Feeding Time	7.233	7.408	7.449	7.429	**7.483**	7.393	7.439
	High Chair	7.166	7.325	7.460	7.500	7.566	7.417	**7.676**
	Lady Eating	6.565	7.363	7.472	7.531	7.657	7.563	**7.692**
	Piano Man	5.812	7.446	7.648	7.565	7.727	7.651	**7.825**
	Santas Little Helper	5.910	7.303	7.567	7.416	7.708	7.622	**7.817**

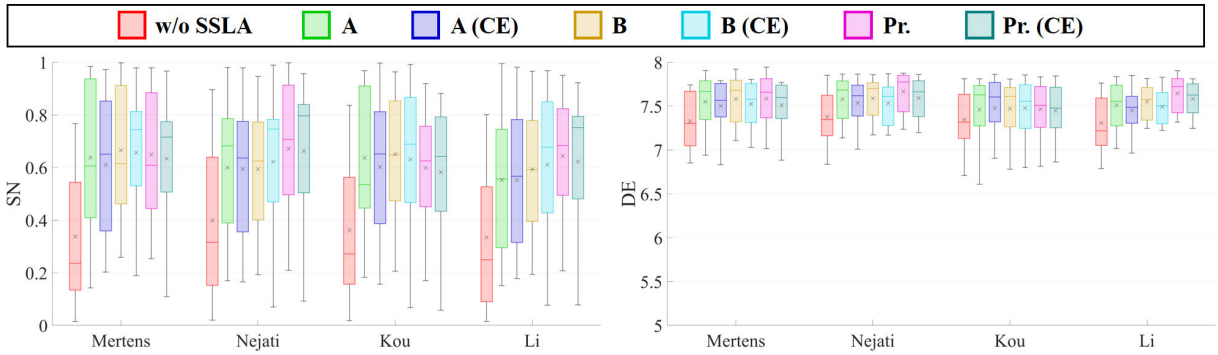


FIGURE 17. Quantitative evaluation scores for **Dataset 3.**

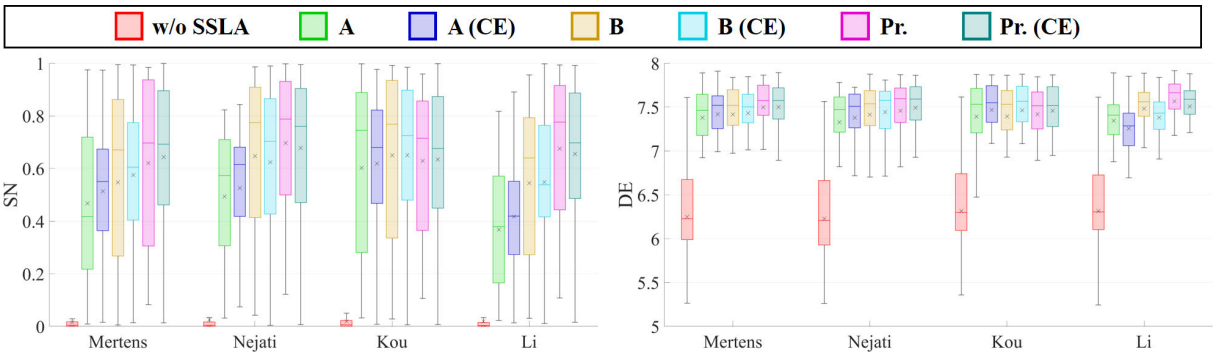


FIGURE 18. Quantitative evaluation scores for **Dataset 4.**

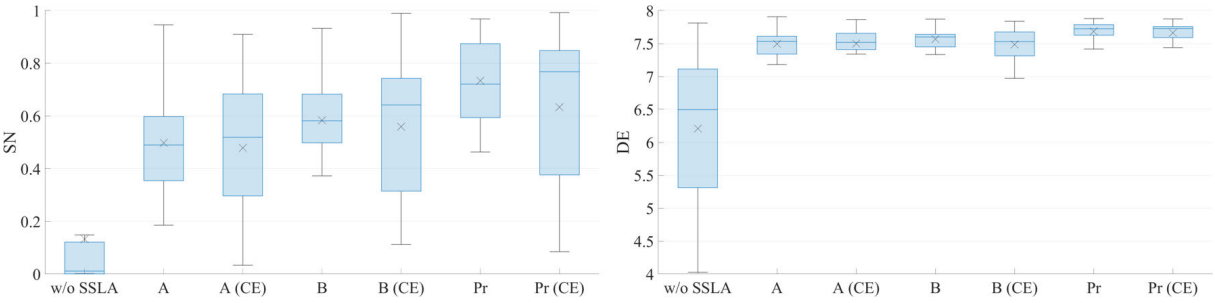


FIGURE 19. Quantitative evaluation scores for **Dataset 5.**

TABLE 3. Numbers of generated regions for all sets in the dataset. Note that **A** and **A (CE)** are not included since the numbers of generated regions are manually set for them.

	$M \leq 2$	$M = 3$	$M = 4$	$M = 5$	$M \geq 6$
B	22	43	56	35	28
B (CE)	14	43	63	34	30
Pr.	0	156	27	1	0
Pr. (CE)	0	156	27	1	0

are summarized in Table 3. Note that **A** and **A (CE)** are not included because $M = 5$ was manually set for Approach A. We see that **Pr.** and **Pr. (CE)** generated fewer regions than **B** and **B (CE)** in most cases. On the other hand, **B** and **B (CE)** often generated six or more regions. This has a non-negligible impact on the memory consumption and the computational time required for the fusion process. From Table 3, we also see that applying CE has no significant effect on M.

E. COMPARISON OF COMPUTATIONAL EFFICIENCY OF EACH SCENE SEGMENTATION METHOD

The average executing time for **Dataset 3** is shown in Fig. 20. An Intel®Core™ i9-11900 2.50GHz CPU with 16.0 GB of memory and a Windows 11 Pro operating system were used for the experiments. MATLAB R2021a was used as the programming language. The executing time for the fusion process was the average of those for the methods of Mertens et al. [9], Wang et al. [16], Kou et al. [13], and Li et al. [17]. The maximum, minimum, and average numbers of pixels in **Dataset 3** are 425430, 1306200, and 634425.8, respectively.

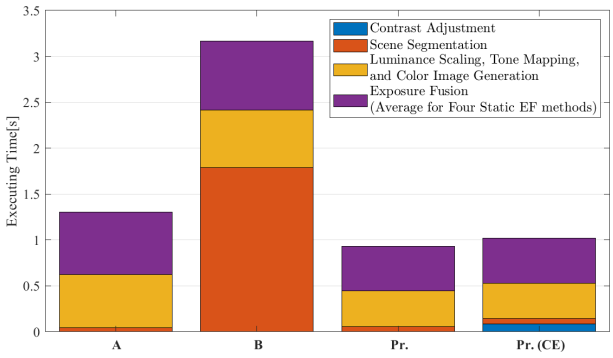


FIGURE 20. Average executing time for **Dataset 3.**

From Fig. 20, it can be seen that **A** and **Pr.** are much faster than **B**. It can also be seen that **Pr.** required less computational time than **A** and **B** for the fusion process. The same trend can be seen for luminance scaling, tone mapping, and color image generation. This is because the number of regions generated by **Pr.** was smaller than those generated by **A** and **B** in most cases.

VI. CONCLUSION

In this paper, we proposed a new scene segmentation method for SSLA. The proposed method exploits information of the nonlinear luminance distribution of each segmented region, which enables us to avoid the situation where the effect of luminance scaling in SSLA is weakened. In addition to this merit, the proposed scene segmentation method can avoid the

generation of regions with very few pixels, which reduces the computational cost for the processes following scene segmentation. The computational cost of the proposed method itself can also be easily reduced. In the experiment, the proposed method generated good scene segmentation results with a small number of regions. Overall, the experimental results showed that the proposed scene segmentation method is very effective for generating high-visibility fusion results and is faster than the conventional scene segmentation approaches for SSLA.

REFERENCES

[1] I. Takayanagi, N. Yoshimura, K. Mori, S. Matsuo, S. Tanaka, H. Abe, N. Yasuda, K. Ishikawa, S. Okura, S. Ohsawa, and T. Otaka, "An over 90 dB intra-scene single-exposure dynamic range CMOS image sensor using a 3.0 μm triple-gain pixel fabricated in a standard BSI process," *Sensors*, vol. 18, no. 2, p. 203, Jan. 2018.

[2] I. Takayanagi and R. Kuroda, "HDR CMOS image sensors for automotive applications," *IEEE Trans. Electron Devices*, vol. 69, no. 6, pp. 2815–2823, Jun. 2022.

[3] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, B. G. Dan, and E. Shechtman, "Robust patch-based HDR reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–11, Nov. 2012.

[4] T.-H. Oh, J.-Y. Lee, Y.-W. Tai, and I. S. Kweon, "Robust high dynamic range imaging by rank minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1219–1232, Jun. 2015.

[5] Q. Yan, D. Gong, Q. Shi, A. Van Den Hengel, C. Shen, I. Reid, and Y. Zhang, "Attention-guided network for ghost-free high dynamic range imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1751–1760.

[6] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, Jul. 2002.

[7] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Comput. Graph. Forum*, vol. 22, no. 3, pp. 419–426, Sep. 2003.

[8] Z. Liang, J. Xu, D. Zhang, Z. Cao, and L. Zhang, "A hybrid L_1-L_0 layer decomposition model for tone mapping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4758–4766.

[9] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, Sep. 2009.

[10] Z. G. Li, J. H. Zheng, and S. Rahardja, "Detail-enhanced exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4672–4676, Nov. 2012.

[11] T. Sakai, D. Kimura, T. Yoshida, and M. Iwahashi, "Hybrid method for multi-exposure image fusion based on weighted mean and sparse representation," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 809–813.

[12] M. Nejati, M. Karimi, S. M. R. Soroushmehr, N. Karimi, S. Samavi, and K. Najarian, "Fast exposure fusion using exposedness function," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2234–2238.

[13] F. Kou, Z. Li, C. Wen, and W. Chen, "Multi-scale exposure fusion via gradient domain guided image filtering," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 1105–1110.

[14] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, "Robust multi-exposure image fusion: A structural patch decomposition approach," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2519–2532, May 2017.

[15] Z. Li, Z. Wei, C. Wen, and J. Zheng, "Detail-enhanced multi-scale exposure fusion," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1243–1252, Mar. 2017.

[16] Q. Wang, W. Chen, X. Wu, and Z. Li, "Detail-enhanced multi-scale exposure fusion in YUV color space," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2418–2429, Aug. 2020.

[17] H. Li, K. Ma, H. Yong, and L. Zhang, "Fast multi-scale structural patch decomposition for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 5805–5816, 2020.

[18] M. Aggarwal and N. Ahuja, "Split aperture imaging for high dynamic range," *Int. J. Comput. Vis.*, vol. 58, no. 1, pp. 7–17, Jun. 2004.

[19] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video production system," *ACM Trans. Graph. (TOG)*, vol. 30, no. 4, pp. 1–10, Jul. 2011.

[20] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2000, pp. 472–479.

[21] V. G. An and C. Lee, "Single-shot high dynamic range imaging via deep convolutional neural network," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Dec. 2017, pp. 1768–1772.

[22] A. G. Vien and C. Lee, "Single-shot high dynamic range imaging via multiscale convolutional neural network," *IEEE Access*, vol. 9, pp. 70369–70381, 2021.

[23] Y. Kinoshita and H. Kiya, "Scene segmentation-based luminance adjustment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4101–4116, Aug. 2019.

[24] Y. Kinoshita and H. Kiya, "Automatic exposure compensation using an image segmentation method for single-image-based multi-exposure fusion," *APSIPA Trans. Signal Inf. Process.*, vol. 7, no. 1, p. e22, 2018.

[25] Y. Huo, F. Yang, and V. Brost, "Dodging and burning inspired inverse tone mapping algorithm," *J. Comput. Inf. Syst.*, vol. 9, no. 9, pp. 3461–3468, May 2013.

[26] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, 1998, pp. 839–846.

[27] J. Chen, S. Paris, and F. Durand, "Real-time edge-aware image processing with the bilateral grid," *ACM Trans. Graph.*, vol. 26, no. 3, p. 103, Jul. 2007.

[28] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[29] Z. Li, J. Zheng, Z. Zhu, W. Yao, and S. Wu, "Weighted guided image filtering," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 120–129, Jan. 2015.

[30] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[31] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

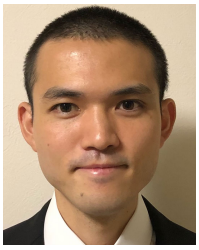
[32] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.

[33] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.

[34] (Nov. 2022). *The HDR Photographic Survey*. [Online]. Available: http://markfairchild.org/HDRPS/HDRthumbs.html

[35] (Nov. 2022). *HDR Photography Gallery*. [Online]. Available: https://www.easyhdr.com/examples/

[36] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.



SEIICHI KOJIMA received the B.Ec. degree from Kyoto University, Japan, in 2016, and the M.S. degree from Yamaguchi University, Japan, in 2020, where he is currently pursuing the Ph.D. degree in science. His research interests include low-light image enhancement and high dynamic range imaging. He is a Student Member of IEICE.

NORIAKI SUETAKE (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in control engineering and science from the Kyushu Institute of Technology, Japan, in 1992, 1994, and 2000, respectively.

He is currently with the Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Japan, where he is a Professor. His research interests include digital signal processing, image processing, and intelligent systems. He is a member of the OPTICA and IEICE.