# Dual Exposure Fusion with Entropy-based Residual Filtering

**Yong Seok Heo[1], Soochahn Lee[2], and Ho Yub Jung[3*]**

[1] Dept. of Electrical and Computer Engineering, Ajou University, Suwon, Korea
[e-mail: ysheo@ajou.ac.kr]
[2] Dept. of Electrical Engineering, Soonchunhyang University, Asan, Korea
[e-mail: sclsch@sch.ac.kr]
[3] Dept. of Computer Engineering, Chosun University, Gwangju, Korea
[e-mail: jung.ho.yub@gmail.com]
*Corresponding author: Ho Yub Jung

---

## *Abstract*

This paper presents a dual exposure fusion method for image enhancement. Images taken with a short exposure time usually contain a sharp structure, but they are dark and are prone to be contaminated by noise. In contrast, long-exposure images are bright and noise-free, but usually suffer from blurring artifacts. Thus, we fuse the dual exposures to generate an enhanced image that is well-exposed, noise-free, and blur-free. To this end, we present a new scale-space patch-match method to find correspondences between the short and long exposures so that proper color components can be combined within a proposed dual non-local (DNL) means framework. We also present a residual filtering method that eliminates the structure component in the estimated noise image in order to obtain a sharper and further enhanced image. To this end, the entropy is utilized to determine the proper size of the filtering window. Experimental results show that our method generates ghost-free, noise-free, and blur-free enhanced images from the short and long exposure pairs for various dynamic scenes.

---

---

# 1. Introduction

There are many factors of camera parameters that can change the appearance of its images, such as exposure time, aperture size, ISO, camera gain, focal length, and lens properties. By controlling these parameters, images taken from the same scene can have a variety of different appearances and properties [1][2][3].

Among these factors, the exposure time is an important one that determines the level of brightness and quality of an image. In general, images taken with a short exposure time tend to contain sharp properties, but they are dark and often suffer from noise due to an insufficient amount of photons incident to the camera sensor. On the other hand, images taken with a long exposure time are bright and usually free from noise. Its disadvantages, however, include blurry properties due to camera shake or object movement while accumulating photons for the camera sensors. **Fig. 1 (a)** and **Fig. 1 (b)** show an example of a long- and short-exposure image pair for the same scene that are taken from consecutive time periods. From this example, it is not difficult to see the complementary properties of different exposures. That is, **Fig. 1 (a)** is bright and has good color distribution, but it is blurry and some of the regions such as the sky region are over-saturated. In contrast, **Fig. 1 (b)** is sharp, and the sky regions show clear detail. It is, however, noisy and dark for most regions.

Thus, it is difficult to select a single exposure time for arbitrary scenes that contain dynamic ranges, because the range of radiance incident to the camera is far larger than the intensity range that can be expressed in the images. In this situation, it is a natural idea to combine multiple exposures to generate more enhanced image quality.

Static scenes can be easily fused to generate an enhanced image that contains good complementary properties that can be borrowed from each of the exposures. In general, however, it is not an easy task to combine images from moving scenes because of their different properties, such as darkness and brightness, as well as noise and blur, though this occurs more often in practical scenarios.

In this paper, a practical method is proposed to combine different exposures for dynamic scenes that contain moving objects. We assume that the short exposure and long exposure contain a certain amount of noise and blur, respectively. In this situation, most methods fail to generate satisfactory fusion results, because they usually suffer from ghost artifacts, where the same object appears multiple times, and suffer from color artifacts by neglecting the noise and blur problems that can easily occur for short- and long-exposure images, respectively.

**Fig. 1 (c)** shows the result of a state-of-the-art high dynamic range (HDR) imaging method [4], whereas **Fig. 1 (d)** shows the result of our method. By comparing **Fig. 1 (c)** and **Fig. 1 (d)**, the HDR method [4] generates noise as well as artifact results for over-saturated regions such as the sky region. In addition, some regions in the result of [4] contain blurry properties due to inaccurate estimation of correspondences between the dual exposures. In contrast, our method generates a better result that contains smaller noise and blur.

In Section 2, we review previous works for various image fusion methods using multiple exposures. Section 3 describes the proposed method. Section 4 demonstrates the experimental evaluation and comparative results for various datasets. Finally, we conclude in Section 5.
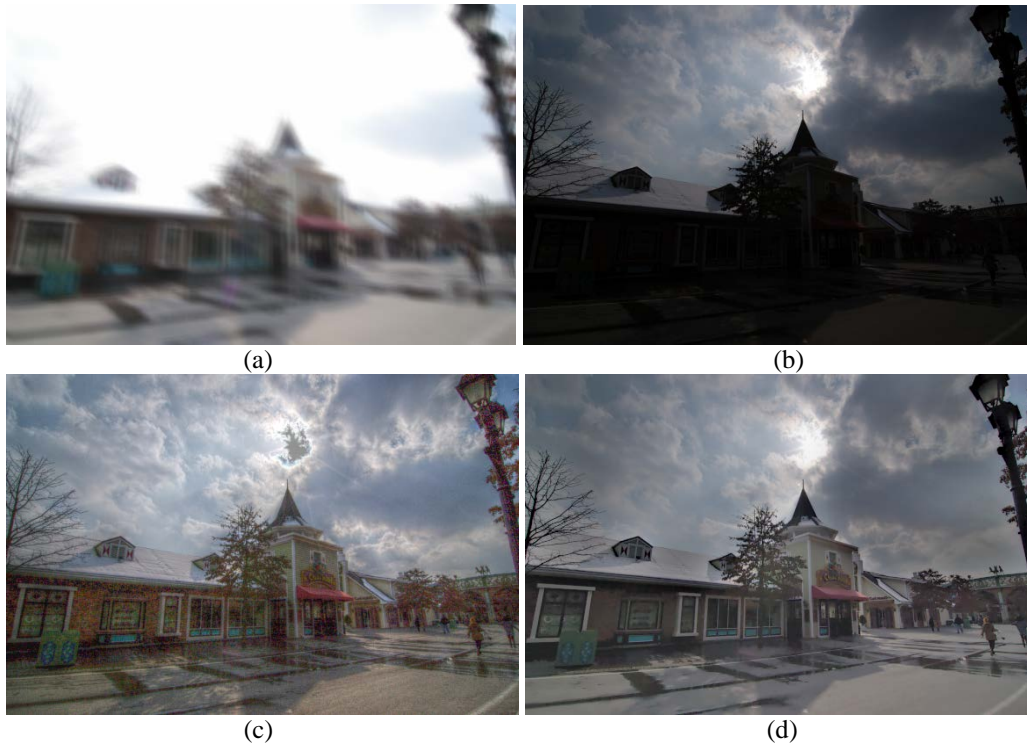
**Fig. 1.** Comparison of different methods. (a) Long exposure image $I^L$. (b) Short exposure image $I^S$. (c) Result of [4] using (a) and (b). (d) Result of our method using (a) and (b).

## 2. Related Work

The topic of image enhancement using multiple exposure images has earned growing attention due to its complementary properties such as blur and noise as well as brightness and darkness. To address these problems, various image fusion methods have been developed to generate more visually pleasing and enhanced images.

In general, the HDR imaging approach [5] is a common method that aims to estimate the radiance value for each pixel by estimating an inverse camera response function (CRF) so that a single well-exposed image can be generated. This inverse CRF is usually estimated using multiple exposures [6]. Most HDR works mainly focus on addressing the ghost problems that usually occur with moving scenes, where the same object can appear multiple times in the resulting images. To this end, Jacobs et al. [7] used two measures, variance and uncertainty, to detect regions of the ghost so that the ghost regions can be excluded when fusing the input exposures. Grosch [8] also detected regions of the ghost using pixel colors that were estimated from the inverse CRF. They determined the ghost regions by thresholding the absolute difference between the estimated color value using the inverse CRF and the input color value. However, the ghost detection results of these methods [7] [8] are easily affected by the threshold values of those measures. Gallo et al. [9] estimated the ghost pixels using a block-wise window comparison. Then, the block boundaries are blended to reduce the difference in the color values between neighboring pixels. Raman et al. [10] also presented a ghost detector using a block-wise comparison of different exposures similar to [9]. However,

these methods [9][10] are not free from color artifacts around the block boundaries due to inaccurate blending.

Meanwhile, there are other methods that address this ghost problem by defining a proper weighting function when fusing images. Khan et al. [11] proposed a pioneering ghost removal method where the weighting function is composed of two weighting terms, the weight of being correctly exposed and the weight of belonging to the background. Heo et al. [12] proposed a generalized weighted filtering method for effective ghost detection and elimination for robust HDR imaging. Park and Park [13] proposed a new HDR imaging method combining multiple aperture images. Recently, Sen et al. [4] proposed an image synthesis method that generates an HDR image that contains information from all the input exposures. For each exposure level, they synthesize a new exposure image changed from the reference exposure by finding optimal patches using the PatchMatch (PM) [14] between the reference and new exposure images.

On the other hand, there are methods that directly combine multiple images that have different properties. Mertens et al. [15] presented an exposure fusion (EF) method by fusing a Laplacian pyramid of multiple exposures. This method is fast and simple, but it assumes that the scenes in the multiple exposures are stationary. Tico et al. [16] proposed a method to combine multiple exposures where there is motion blur in one of them. They combined two types of EF results, such as EF using original input exposures and EF using images that are color-calibrated from the shortest exposure as a reference. The limitation of this approach is an ambiguity problem when a moving object and its background have similar brightness, because its analysis is focused only on the intensity comparison. Agarwala et al. [17] presented a method to combine images where the objects are severely moving. Their method is based on an interactive stroke to denote important regions that can be combined using gradient fusion. Petschnigg et al. [18] presented a method that synthesizes a new image from a pair of flash and no-flash images by merging the ambient component of the no-flash image with the high-frequency detail of the flash image. Similarly, Yuan et al. [19] proposed a deblurring method for a noisy and blurry image pair by iteratively estimating the blur kernel and performing deconvolution. The limitation of [18] and [19] is that the input image pair should be stationary. Thus, for images that contain moving objects, [18] and [19] cannot generate a stable result. Zhang et al. [20] presented an analysis comparing deblurring using a single image and denoising using multiple images for HDR imaging from high-resolution cameras. They concluded that denoising is a more reliable solution than deblurring for obtaining a better image. Joshi and Cohen [21] presented a method that generates an enhanced image from a sequence of images that suffer from shot and quantization noise, sensor dust, and haze of the scene by using a novel local weighted averaging method. HaCohen et al. [22] proposed a method to generate a new image from a target and source image pair. The method finds corresponding patches, and computes a global color mapping function. However, they do not deal with different exposure settings and the noise problems that usually occur in the dark images.

Most of these works do not address the noise and blur problem together when combining the different exposures. Thus, our method assumes the most practical cases where there are noise and blur in the input exposures as well as dynamic motion between exposures. A detailed explanation of the proposed method follows in the next section.
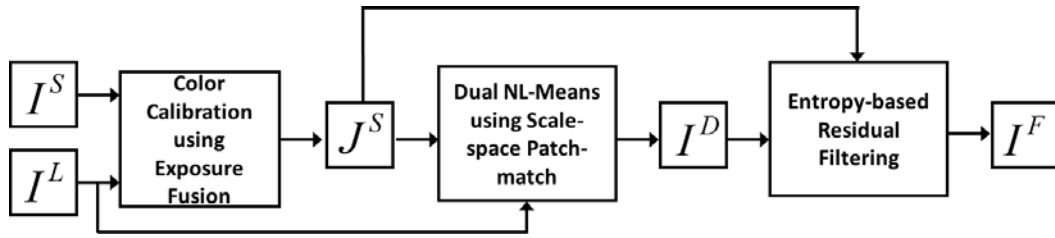
**Fig. 2.** Block diagram of the proposed method.

## 3. Proposed Method

**Fig. 2** shows a block diagram of the proposed method. First, we calibrate the color of the short exposure to that of the long exposure for the following procedures. To this end, we propose to exploit histogram matching [23] with EF [15]. The histogram matching [23] is used to transform the short exposure to the long exposure so that the brightness levels of them can be similar. However, if there are over-exposure properties such as over-saturated regions in the long exposure, those unwanted properties are also transferred to the resultant image $I_h^S$. Thus, we also adopt the EF [15] which selects and combines well-exposed regions between the short exposure $I^S$ and the transformed short exposure $I_h^S$ in order to generate more enhanced image $J^S$.

The problem of $J^S$ is that it contains the amplified noise that is contained in the short exposure as shown in **Fig. 1** (d). To deal with this problem, we propose a dual non-local (DNL) means method which extended the idea of [24] in order to generate an initial enhanced image. Buades et al. [24] generated a denoised image using non-local average by finding patches that have similar structure for all pixels in the search window. In general, it is not a trivial task to naively extend the framework of [24] to the multiple exposures which include noisy and blurry image pair, because it is difficult to find correspondences between them. Thus, we also propose a scale-space patch-match method to find correspondences between the noisy and blurry image pair.

Finally, a residual filtering method based on entropy is presented to further enhance the image. Our method is modification of the previous method [25] that proposed a method to filter the noise layer after separating it from the original noisy image based on three weight terms including color, spatial, and structure weights. They aim to eliminate the image structure in the noise layer, and generate a sharper image from an initial denoised image. However, [25] fixed the size of the filtering window that can results in blurring for some regions. Thus, we propose an adaptive filtering method by determining the size of window based on the entropy of the window. Detailed explanations of each process are given below:

### 3.1 Color Calibration Using Exposure Fusion

In order to find correspondences between different exposures, we calibrate the brightness of the short-exposure image $I^S$ to that of the long-exposure image $I^L$. First, the histogram of $I^S$ is matched to that of $I^L$ as follows:

(a)                                                        (b)

**Fig. 3.** Color calibration result using our method. (a) Histogram matching result $I_h^S$ using **Fig. 1** (a) and **Fig. 1** (b) as the input image pair. (b) Exposure fusion result $J^S$ using **Fig. 3** (a) and **Fig. 1** (b).

$$I_h^S = h_S^L(I^S), \tag{1}$$

where $h_S^L$ denotes the histogram matching operation [23] that changes the color of $I^S$ to that of $I^L$, and $I_h^S$ is the result of histogram matching. **Fig. 3** (a) shows an example of $I_h^S$ using $I^L$ and $I^S$ in **Fig. 1** (a) and **Fig. 1** (b), respectively. Although the brightness of $I_h^S$ is similar to $I^L$, it still contains saturated regions where $I^S$ has clearer details such as cloudy sky regions. Thus, we combine $I_h^S$ and $I^S$ using the EF technique [15] as follows:

$$J^S = I_h^S \oplus I^S, \tag{2}$$

where $J^S$ is the result of EF, and the $\oplus$ operation represents the EF process. **Fig. 3** (b) shows an example of $J^S$. Note that although $J^S$ shows better details compared to the original input images in **Fig. 1** (a) and **Fig. 1** (b), it still contains notably amplified noise and color artifacts. Thus, we need to decrease the noise and enhance it with the long exposure image using the following proposed method.

## 3.2 Dual NL-Means using Patch-Match with a Coherent Position Constraint

Short exposure is sharp, dark and noisy, whereas long exposure is blurry, bright and noiseless. These complementary properties motivate us to combine the two images to generate a sharper, noise-free, and visually pleasing image that has good exposure. To this end, it is required to find good correspondences between two exposures, because we assume there are movements between them. Let us denote the correspondence function using our patch-match method with a coherent position constraint (PMC) as $f(\cdot)$ such that $f(p) = p'$, where $p$ and $p'$ are the corresponding pixel pair in images $J^S$ and $I^L$, respectively. In general, it is not an easy task to find accurate correspondences between noisy and blurry images because of their different properties. The resized small image, however, tends to have less noise and blur compared to the original full-size images. Thus, it enables us to find more reliable matches between $J^S$
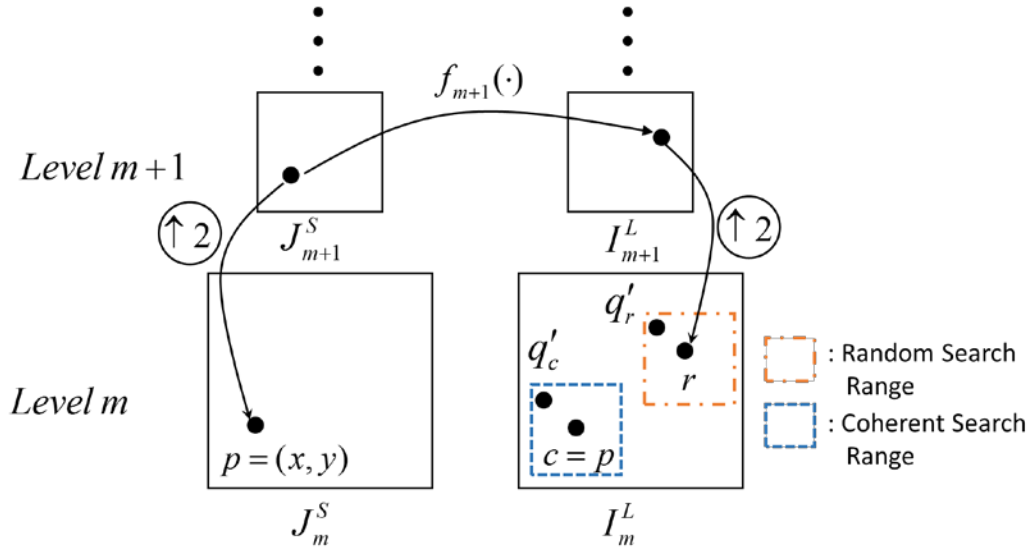
**Fig. 4.** Search selection scheme for our PMC method.

and $I^L$. This is one of the reasons that we use a scale-space analysis on both $J^S$ and $I^L$ for applying patch-match [14]. Hence, we build image pyramids for both images, where $J_m^S$ and $I_m^L$ represent the $m^{th}$ level of the pyramid of $J^S$ and $I^L$, respectively, where the coarsest level is $m = M - 1$ and the finest level is $m = 0$. The number of levels of the pyramid is defined by $M = \log\left(\dfrac{W_{\min}}{\min(W, H)}\right) / \log\left(\dfrac{1}{\sqrt{2}}\right)$, where $W$ and $H$ respectively represent width and height of the image, and we set $W_{\min}$ as $W_{\min} = 29$. Here, we add a coherent position constraint in our scale-space patch-match method, because we assume that dual exposures are taken in consecutive time periods and most of the pixels move slightly, except for the moving objects between the short and long exposures.

  Our method in each pyramid level is composed of four steps, including 1) initialization, 2) propagation and random search, 3) coherent position search, and 4) search selection. Explanations of each step follow.

**Matching Measure.** Matching measure is an important criterion for determining the quality of match between two patches. In our method, it is required to measure the accuracy of matches between noisy and blurry patches. This circumstance hinders conventional matching measures such as sum of squared difference (SSD) or sum of absolute difference (SAD) from finding proper matches. Thus, our method uses the Gaussian weighted window matching as a matching measure for two patches in order to give more weight to the center pixel. The definition of our measure $D(\cdot, \cdot)$ is as follows:

**Fig. 5.** Patch-match result for input images in **Fig. 1** (a) and **Fig. 3** (b).
(a) Result using [14]. (b) Result using our PMC method.

$$D(p, p') = \frac{1}{Z} \sum_{\substack{q \in N(p), \\ q' \in N(p'),}} \exp\left(\frac{-\|p - q\|^2}{2\sigma_d^2}\right) \left\|J^S(q) - I^L(q')\right\|^2, \tag{3}$$

where $Z$ represents the normalization factor, and $N(p)$ and $N(p')$ are the set of pixels in the $5 \times 5$ window centered at pixel $p \in J^S$, and $p' \in I^L$, respectively, and $\sigma_d$ is empirically set as $\sigma_d = 0.67$.

**Initialization.** Our algorithm starts from the coarsest $M - 1^{th}$ level. Here, we randomly assign the correspondences between $J_{M-1}^S$ and $I_{M-1}^L$ as the original PM method [14]. Except at the $M - 1^{th}$ level, we use the result of the previous $m + 1^{th}$ scale for the current $m^{th}$ scale as initial correspondences by simply scaling up the result from the previous scale. This initialization makes our method more accurate than that of starting from a random guess.

**Propagation and Random Search.** In our propagation stage, we use four neighborhoods to find more appropriate patches around the center pixel, unlike the original PM method [14], which uses only two neighborhood pixels. The random search phase is the same as [14], which randomly samples and selects one of the pixels in the search range by testing the improvement of the current match.

**Coherent Position Search.** Unlike the original PM method [14], we put an additional constraint in our framework. The coherent position search finds an optimal patch around the current position. This is because we assume that, for most regions, there are no significant movements between two images because of the short period of time between the two images. Thus, most pixels have a high probability that a proper patch can be found around the current position.

**Search Selection.** Finally, we select a more appropriate match as a good match between the result of the random search and that of the coherent position search, based on the measure $D(\cdot, \cdot)$ in Eq. (3), that is defined by

$$f(p) = \arg\min_{\substack{q'_c \in N(c), \\ q'_r \in N(r)}} \{D(p,q'_c), D(p,q'_r)\}, \tag{4}$$

where $c$ is the same position as $p$, $r$ is the position that is determined from scaling up the correspondence result $f_{m+1}(\cdot)$ of the previous $m+1^{th}$ scale, and $q'_c$ and $q'_r$ are the position within the search range of the coherent search and random search, respectively. **Fig. 4** depicts our search selection scheme.

**Fig. 5** compares the results of [14] and our PMC method for the input pair $J^S$ and $I^L$. Note that our method generates a more plausible result, whereas the result using the original method [14] is blurry because of incorrect correspondences.

On the other hand, although the generated image using our PMC method contains less noise and more detail than [14], it still contains pixels that suffer from noisy and blurry artifacts due to inaccurate correspondences. Here, using the correspondence function $f(\cdot)$ obtained from Eq. (4), we present a DNL means framework using both $J^S$ and $I^L$. By using both the short and long exposures, we have advantages such as enhanced color distribution, which is borrowed from the long exposure, as well as enhanced denoising performance due to the increased number of support pixels. Thus, our enhanced value $I^D(p)$ for pixel $p$ is defined by

$$I^D(p) = \frac{1}{Z_p} \sum_{\substack{q \in W_D(p), \\ q' \in W_D(p'),}} \left[ w_S(p,q)J^S(q) + w_L(p,q')I^L(q') \right], \tag{5}$$

where $p' = f(p)$, and $W_D(p)$ and $W_D(p')$ are the set of pixels in the window size of $21 \times 21$ around pixel $p \in J^S$ and $p' \in I^L$, respectively, and $Z_p$ is the normalization factor. The weight function $w_S(p,q)$ is defined by

$$w_S(p,q) = \exp\left( \frac{-D(p,q)}{\kappa} \right), \tag{6}$$

where $p$ and $q$ are pixel positions in $J^S$, the distance function $D(p,q)$ is defined similarly to Eq. (3), and $\kappa$ is empirically set as $\kappa = 100.0$. Note that it is necessary to carefully select support pixels from $I^L$ with blurry properties in order to prevent quality degradation in our DNL framework. Thus, we define $w_L(p,q')$ as follows:

$$w_L^0(p,q') = \exp\left( \frac{-D(p,q')}{\kappa} \right),$$

$$w_L(p,q') = \begin{cases} w_L^0(p,q') & \text{if } w_L^0(p,q') > \eta \cdot w_S^{\min}(p) \\ 0 & \text{if } w_L^0(p,q') \leq \eta \cdot w_S^{\min}(p) \end{cases}, \tag{7}$$

where $q'$ are pixel positions in $I^L$, $\eta$ is empirically set as $\eta = 10$, and $w_S^{\min}(p)$ is defined by
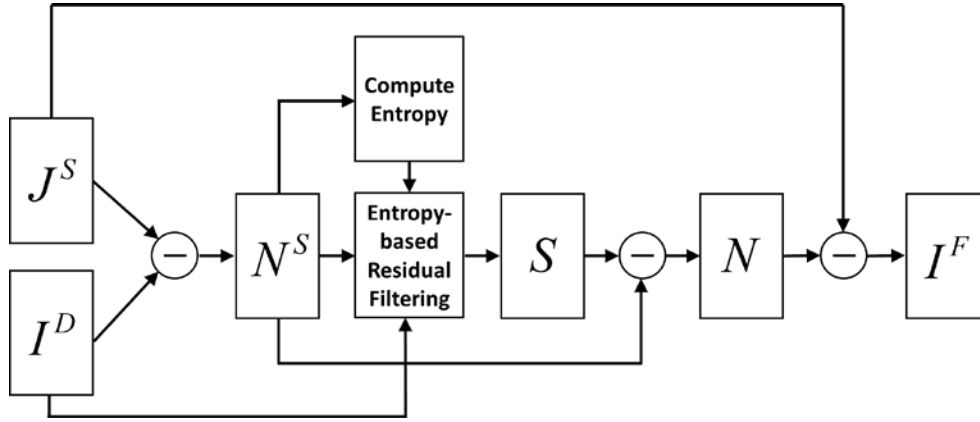
**Fig. 6.** Flow diagram of our entropy-based residual filtering.

$$w_S^{\min}(p) = \min_q w_S(p, q).$$  (8)

**Fig. 9** (e) shows an example of the $I^D$ result from Eq. (5). Although this initial enhanced image $I^D$ is clearly better than that of the result using Eq. (4), it tends to suffer from blur due to over-smoothing in the DNL-means framework. Thus, we still need to sharpen and further enhance the result using our entropy-based residual filtering as described in the following subsection.

## 3.3 Entropy-Based Residual Filtering

**Fig. 6** depicts the procedure of our entropy-based residual filtering method. The blurring property in $I^D$ is due to the existence of the structure component in the residual image $N^S$, which is defined [25] as

$$N^S \equiv J^S - I^D = N + S.$$  (9)

Note that $N^S$ contains the noise component $N$ as well as the structure component $S$, where $S$ is the unwanted component we need to eliminate by our filtering method. To this end, it is important to determine a proper window size for filtering. For structured regions of the residual image, the filtering size should be larger than that of pure noisy regions. Thus, we adaptively determine the size of the filtering window using entropy as a measure of the randomness of the patch. Hence, the region of larger entropy would have a smaller-size filtering window, where the entropy $\varepsilon(p)$ for pixel $p$ is defined by

$$\varepsilon(p) = \sum_{z=-255}^{255} -P(z) \log P(z),$$

$$P(z) = \frac{1}{Z_\varepsilon} \sum_{q \in W_\varepsilon(p)} T[N^S(q) = z],$$  (10)

where $T[\cdot]=1$ if the argument is true and 0 otherwise, and $Z_\varepsilon$ is the normalization factor. The probability $P(z)$ is constructed by counting the number of pixels that have intensity gradient value $z \in [-255, 255]$, and $W_\varepsilon(p)$ is fixed as a $7 \times 7$ size of window for pixel $p$.

Thus, the structure component $S$ in $N^S$ can be extracted similar to [25] using the guide image $I^D$ that is defined by

$$S(p, I^D) = \frac{1}{Z_F} \sum_{q \in W_F(p)} V_{tot}(q, I^D) N^S(q), \tag{11}$$

where $W_F(p)$ is the set of pixels in the filtering window centered at pixel $p$, and $Z_F$ is the normalization factor. The radius $R_F(p)$ of this filtering window $W_F(p)$ is adaptively determined by

$$R_F(p) = 3\sigma_D(p), \tag{12}$$

where $\sigma_D(p)$ is defined using the entropy $\varepsilon(p)$ in Eq. (10) as follows:

$$\sigma_D(p) = \sigma_D^0 \exp\left(\frac{-\varepsilon(p)}{\varepsilon_0}\right), \tag{13}$$

where $\sigma_D^0$ and $\varepsilon_0$ are empirically set as $\sigma_D^0 = 7$ and $\varepsilon_0 = 1.0$, respectively, and $\varepsilon(p)$ plays the role of attenuating the radius when the entropy of the residual component is large, which indicates there is pure noise rather than some structures. **Fig. 7** (d) shows an example of the entropy for each region, where brighter pixels indicate the higher entropy values.

Meanwhile, the weight $V_{tot}$ in Eq. (11) is composed of three terms similar to [25] as follows:

$$V_{tot}(q, I^D) = V_G(q) \cdot V_C(q, I^D) \cdot V_S(q, I^D), \tag{14}$$

where each weight term is defined by

$$V_G(q) = \exp\left(\frac{-\|p - q\|^2}{\sigma_D^2(p)}\right),$$

$$V_C(q, I^D) = \exp\left(\frac{-\|I^D(p) - I^D(q)\|^2}{\sigma_C^2}\right), \tag{15}$$

$$V_S(q, I^D) = \frac{V_S(q; p, I^D) + V_S(p; q, I^D)}{2}.$$

Here, $V_G(q)$, $V_C(q, I^D)$, and $V_S(q, I^D)$, respectively, represent the geometric weight, color weight, and structure weight. The value of $\sigma_C$ is empirically set as $\sigma_C = 25$, and $V_S(q; p, I^D)$ is defined by

$$V_S(q; p, I^D) = \exp\left(\frac{-(p-q)^T D_p^{-1}(p-q)}{\sigma_D^2(p)}\right),$$

$$D_p = \frac{1}{|W(p)|} \sum_{q \in W(p)} \nabla I^D(q) \nabla I^D(q)^T, \tag{16}$$

where $\nabla I$ is the gradient vector, and $D_p$ is the structure tensor that measures the structure around pixel $p$.

From Eq. (9) and Eq. (11), the noise component $N$ from the residual component $N^S$ is computed by

$$N = N^S - S. \tag{17}$$

Now, the final enhanced image $I^F$ can be obtained by subtracting the input noisy image from the estimated noise component as follows

$$I^F = J^S - N. \tag{18}$$

**Fig. 7** shows an example of each component $N^S$, $S$, $N$, entropy values, and our final result $I^F$. **Fig. 8** shows magnified views for cropped regions of the images in **Fig. 7**. Note that the structure component $S$ is eliminated in our final noise component $N$, which shows a more random property than that of $N^S$. **Fig. 9** compares the intermediate results for each process in our method. Our final result $I^F$ shows clearer and noise-free results compared to the intermediate results.



(a)

(b)

(c)

(d)
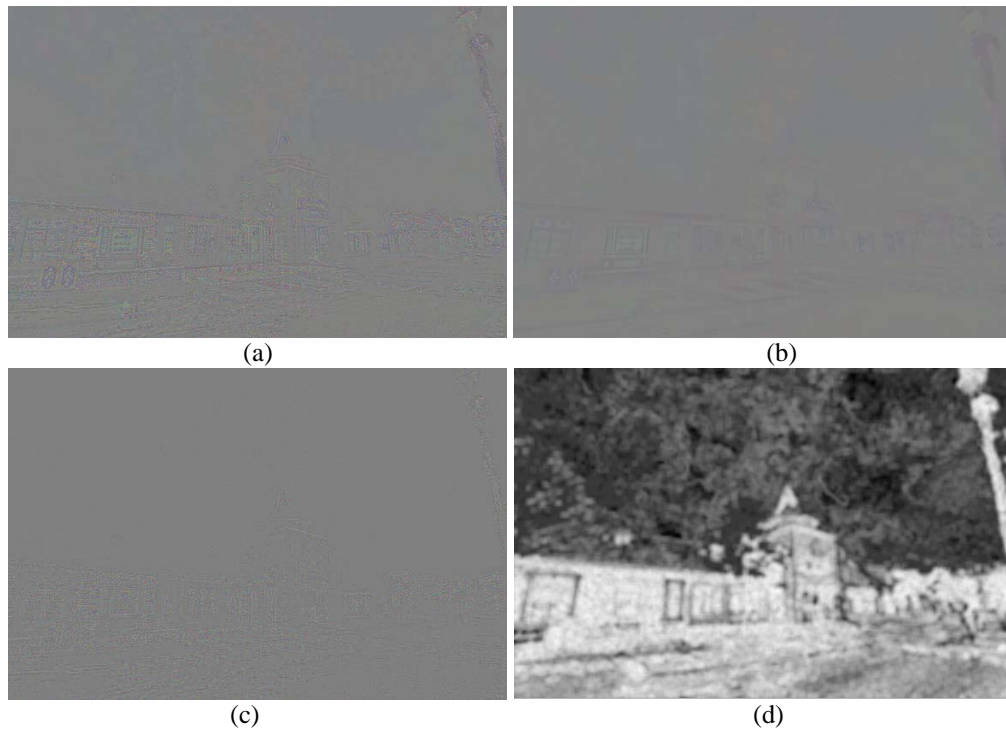
(e)

**Fig. 7**. Example of our entropy-based residual filtering. (a) $N^S$. (b) $S$. (c) $N$.
(d) Estimated entropy $\varepsilon$, which is scaled to gray level for display purpose. (e) $I^F$.
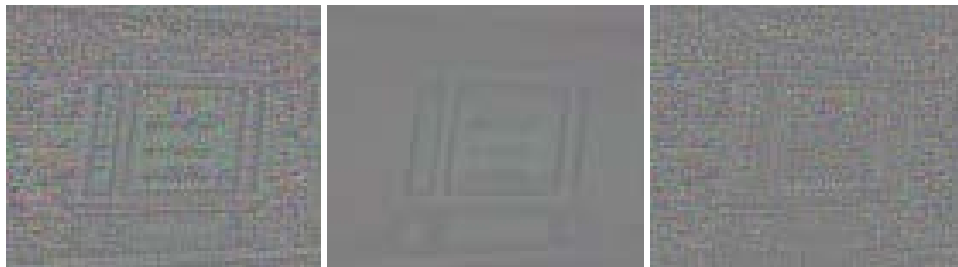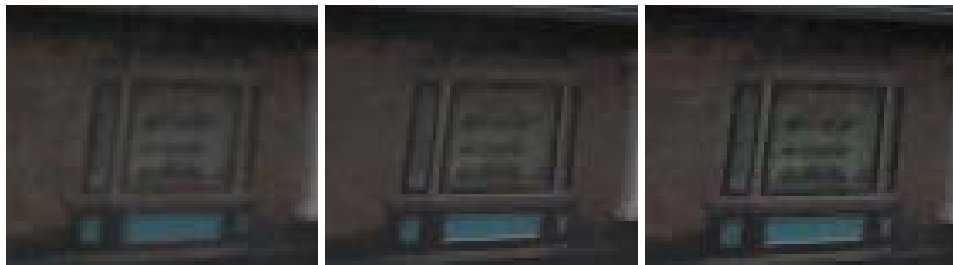


| (a) | (b) | (c) |

**Fig. 8**. Magnified views of the cropped region of the images in **Fig. 7.** (a) $N^S$. (b) $S$. (c) $N$.



| (a) | (b) | (c) |



| (d) | (e) | (f) |

**Fig. 9**. Magnified views of the cropped regions. (a) $I^L$. (b) $J^S$. (c) Result using [14].
(d) Result using Eq. (4). (e) $I^D$. (f) $I^F$.

# 4. Experimental Results and Analysis

In order to test the performance of our method, we performed various experiments for synthetic and real datasets.

## 4.1 Results for Synthetic datasets

For these experiments, we used optical flow datasets including "Dogdance," "Backyard," "Grove2," "Evergreen," and "Teddy" provided by the Middlebury dataset [26]. Each dataset consists of eight frames, where the first and the last frame are used as our input images in order to generate images that have the largest difference of motion between them. Thus, we blur the first frame in the input sequence using a $17 \times 17$ size nonlinear blur kernel; the last frame of each set is contaminated by zero-mean Gaussian noise with a standard deviation of 12.5 for each color channel. Note that these two input images have different properties, such as noise and blur, and there are significant object movements between them. Thus, we used these two images as input of our method.

**Table 1** and **Table 2** respectively show the quantitative PSNR and SSIM [27] results for the initial enhanced image $I^D$ and the final enhanced image $I^F$ using our method. First, we compared our DNL-means method with the NL-means method [24] that uses only a single image $J^S$. As shown in **Table 1** and **Table 2**, our method outperforms the method of [24] for obtaining initial enhanced images, because our method uses the dual exposures that contain more support pixels. The main difference of [24] and our method is in Eq. (5), where [24] uses only a single image, while our method is designed to use dual exposure images based on the estimated correspondence function $f(\cdot)$ using PMC in Eq. (4).

Meanwhile, we also compared our entropy-based residual filtering method with [25]. In this case, we used the same initial enhanced image $I^D$ as input. As shown in **Table 1** and **Table 2**, our method generates more enhanced results than those of [25] for most of the datasets. The method of [25] uses a fixed size of window for the residual filtering, which can cause over- or under-smoothing results for some regions. In contrast, our method adaptively controls the window size by means of the entropy in the window. That is the reason why our method outperforms the method of [25]. **Fig. 10** shows comparison results for the example images, and corresponding magnified views of the images are shown in **Fig. 11**. Note that the noise component of [25] still contains the structure component, whereas the noise component of our method is more random than that of [25], which indicates that our method estimates the noise component more accurately, enabling our method to generate better noise-free and blur-free results than those of [25].

**Table 1.** Performance comparison (PSNR) for synthetic datasets (dB)
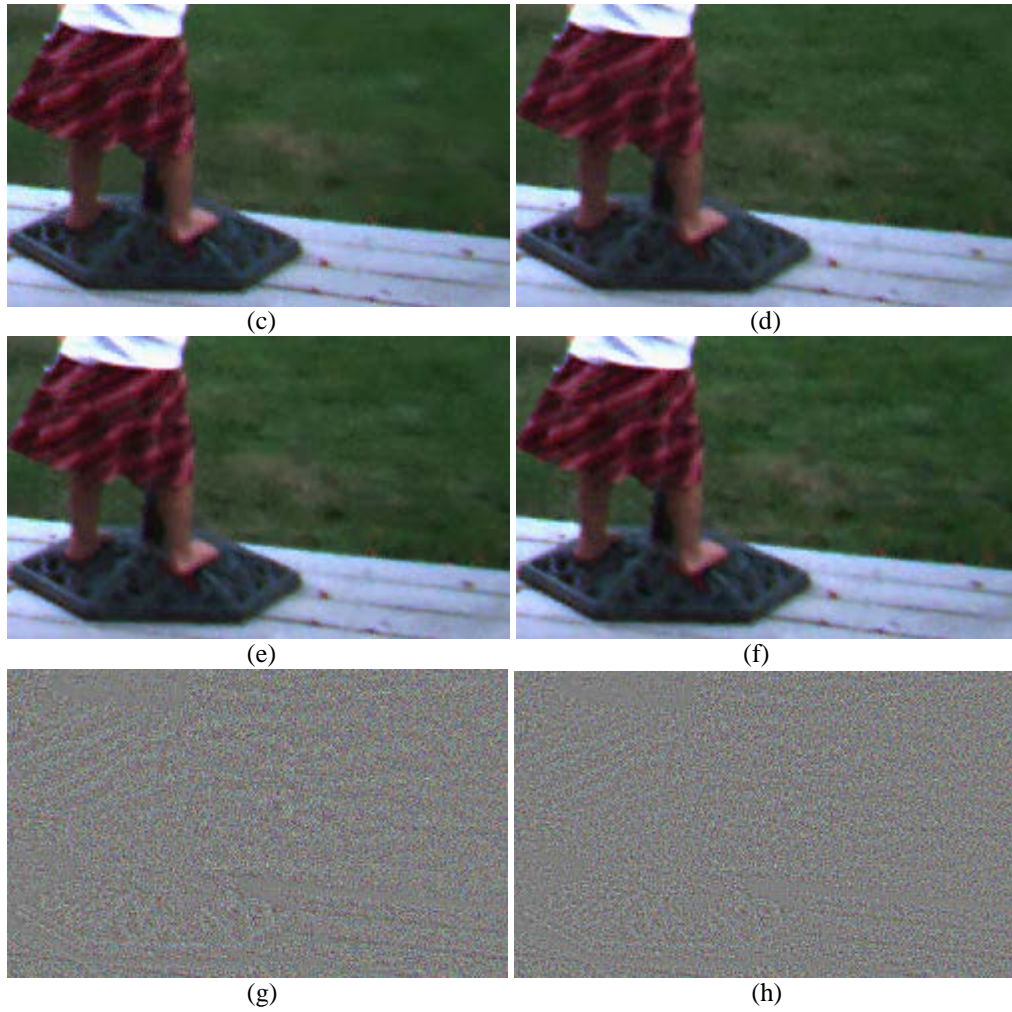
| Dataset | $J^S$ | $I^D$ using [24] | $I^D$ using ours | $I^F$ using [25] | $I^F$ using ours |
|---------|-------|-------------------|-------------------|-------------------|-------------------|
| Dogdance | 26.47 | 31.80 | 32.16 | 32.35 | 32.39 |
| Backyard | 26.43 | 31.14 | 31.34 | 31.52 | 31.57 |
| Grove2 | 26.25 | 29.61 | 29.80 | 29.93 | 30.01 |
| Evergreen | 26.27 | 31.04 | 31.86 | 32.04 | 32.08 |
| Teddy | 26.24 | 31.50 | 31.69 | 31.90 | 32.03 |

(a)          (b)

(c)          (d)

(e)          (f)

**Fig. 10.** Comparison results for dataset "Backyard". (a) $I^L$. (b) $J^S$. (c) $I^D$ using [24]. (d) $I^D$ using our method. (e) $I^F$ using [25]. (f) $I^F$ using our method.



(a)          (b)

**Fig. 11.** Magnified views of results for dataset "Backyard". (a) $J^S$. (b) Ground truth $I$. (c) $I^D$ using [24]. (d) $I^D$ using our method. (e) $I^F$ using [25]. (f) $I^F$ using our method. (g) $N$ using [25]. (h) $N$ using our method.

**Table 2.** Performance comparison (SSIM) for synthetic datasets

| Dataset | $J^S$ | $I^D$ using [24] | $I^D$ using ours | $I^F$ using [25] | $I^F$ using ours |
|---------|-------|------------------|------------------|------------------|------------------|
| Dogdance | 57.09 | 84.62 | 85.81 | 86.22 | 86.47 |
| Backyard | 60.11 | 84.29 | 84.88 | 85.38 | 85.69 |
| Grove2 | 65.68 | 84.42 | 84.30 | 85.26 | 85.33 |
| Evergreen | 60.27 | 85.62 | 87.67 | 88.17 | 88.39 |
| Teddy | 57.63 | 85.06 | 85.42 | 86.37 | 86.52 |

**Table 3.** Average time comparison (sec) of each process for synthetic datasets

| Process | $J^S$ | $f(\cdot)$ using PMC | $I^D$ using [24] | $I^D$ using ours | $I^F$ using [25] | $I^F$ using ours | Total Time (ours) |
|---------|-------|----------------------|------------------|------------------|------------------|------------------|-------------------|
| AvgTime | 0.4 | 61.3 | 29.6 | 54.9 | 54.7 | 15.1 | 131.7 |

On the other hand, we compared time duration of our method with those of other methods for the synthetic datasets in **Table 1** and **Table 2**. Each input image has size of 640 x 480, and we use a PC with Intel® i7-4770 CPU @3.40GHz for our experiments. **Table 3** shows the average time (sec) to compute each process for the synthetic datasets. The process of computing $J^S$ in Eq. (2) includes the histogram matching and the exposure fusion operation which are relatively faster than other processes. The process of computing $f(\cdot)$ using PMC in Eq. (4) takes relatively longer time, because it involves the Gaussian-weighted window matching in Eq. (3) which is somewhat computationally expensive. For the process of computing $I^D$ using the estimated function $f(\cdot)$, we compared time duration of [24] with that of our method. Although [24] takes less time than ours, it cannot fully utilize the dual exposure information. In contrast, our method combines the dual exposure to generate less noisy results. For the process of computing $I^F$, we compared time duration of [25] with that of our method. Note that our method adaptively determines the size of window for residual filtering based on the entropy, while [25] fixed the size of it. Thus, our method is faster and more efficient than [25].

### 4.2 Results for Real datasets

The main goal of our method is to generate a reliable result from a pair of different exposures. Thus, we also tested our method on real datasets where each dataset is composed of a short-exposure and long-exposure image taken from consecutive periods of time. These real datasets were taken using a Canon PowerShot G3 with auto exposure bracketing mode. These datasets contain various practical factors such as noise and blur as well as object and camera movements. **Fig. 12, Fig. 13**, and **Fig. 14** show the comparative results of our method and the method of [4], which is the state-of-the-art HDR method for dynamic scenes. The method of [4] is also based on the patch-match method [14] with optical flow initialization. For these datasets, [4] generates significantly noisy and blurry HDR images. In contrast, our method generates sharper and less noisy results. Thus, our method is more robust to a variety of image degrading factors including noise, blur, and object movements.
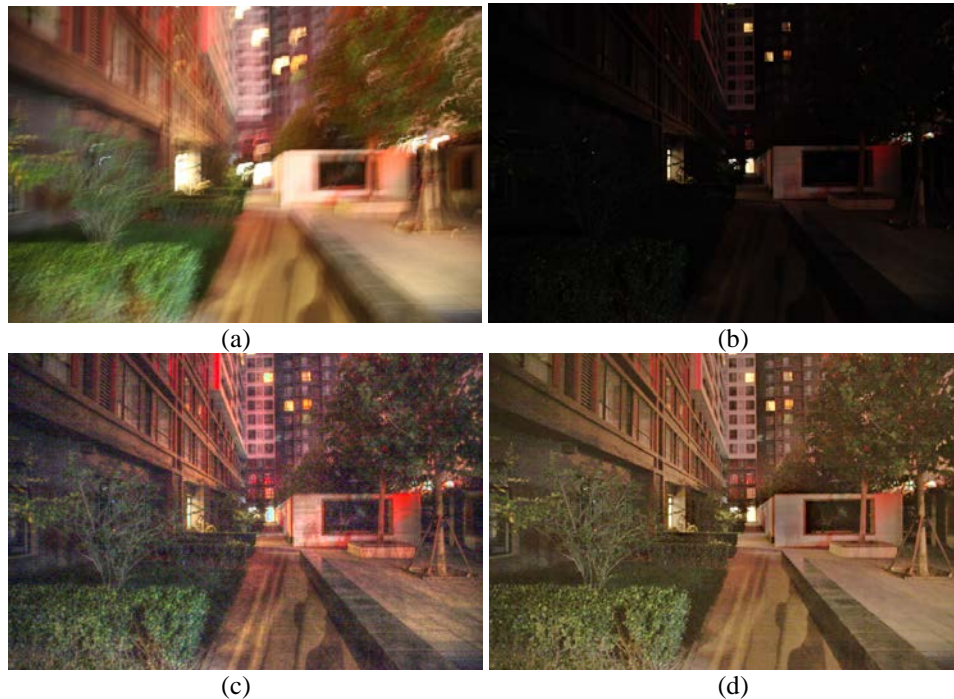


(a)                                                    (b)

(c)                                                          (d)
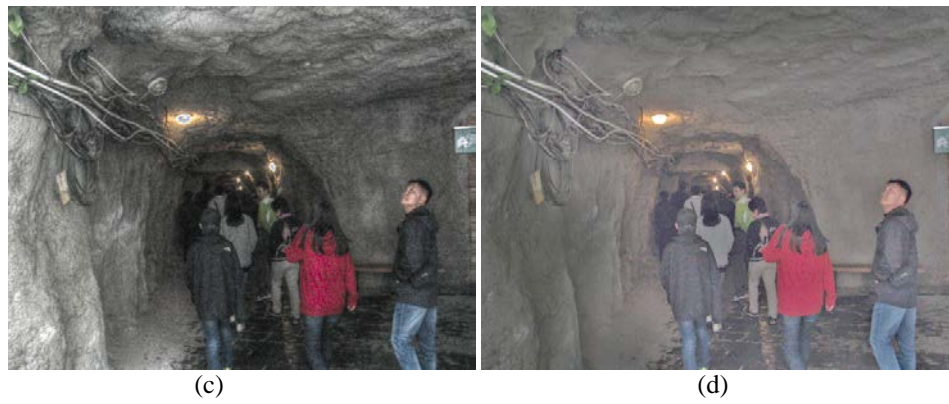
**Fig. 12.** Comparison results for dual exposure input images.
(a) Long exposure. (b) Short exposure. (c) Result of [4]. (d) Result of our method.



(a)                                                          (b)

(c)                                                          (d)
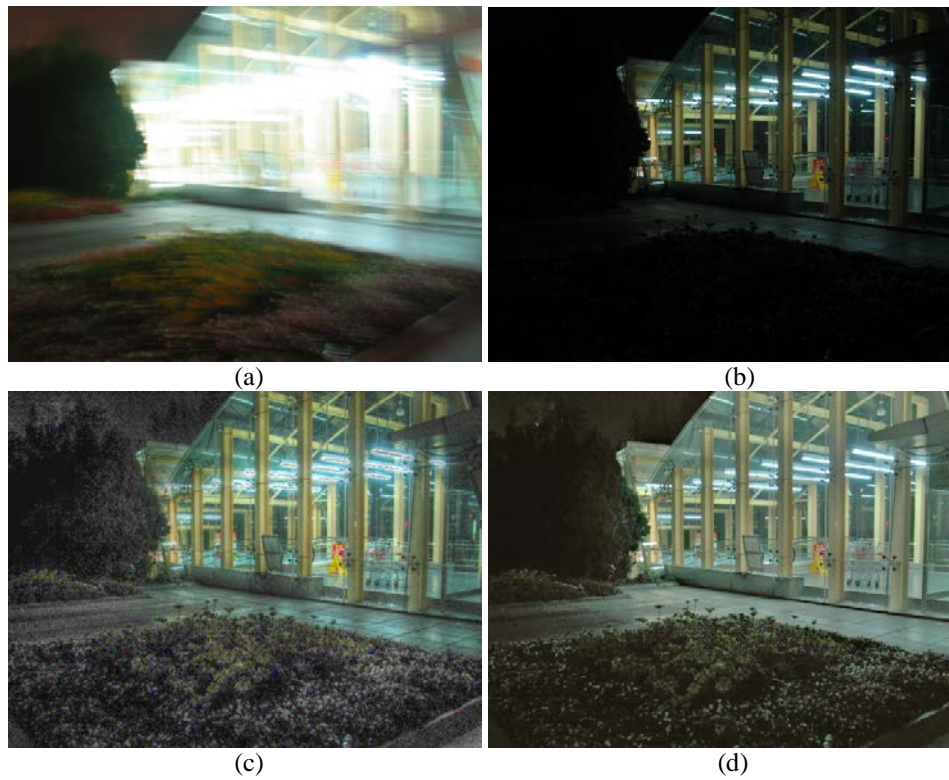
**Fig. 13.** Comparison results for dual exposure input images.
(a) Long exposure. (b) Short exposure. (c) Result of [4]. (d) Result of our method.

(a)                                         (b)

(c)                                         (d)

**Fig. 14.** Comparison results for dual exposure input images.
(a) Long exposure. (b) Short exposure. (c) Result of [4]. (d) Result of our method.

## 5. Conclusion

In this paper, we have proposed a dual exposure fusion method with entropy-based residual filtering. To this end, we adopt a patch-match method with an additional coherent position constraint, and a DNL means method is also proposed for obtaining initial enhancement results. To obtain a more enhanced image, we propose an entropy-based residual filtering method that effectively eliminates the structure component in the residual image. Experimental results show that our method robustly combines the short and long exposures and generates enhanced images from various dynamic scenes.

## Acknowledgment

## References

[1] U. S. Kim, J. M. Lee, Y. M. Kim, K. T. Park and Y. S. Moon, "Photographic Color Reproduction Based on Color Variation Characteristics of Digital Camera," *KSII Transactions on Internet and Information Systems*, vol. 5, no. 11, pp. 2160-2174, Nov. 2011. Article (CrossRef Link)

[2] B. Jin, Z. Jing, and H. Pan, "Multi-modality image fusion via generalized Riesz-wavelet transformation," *KSII Transactions on Internet and Information Systems*, vol. 8, no. 11, pp. 4118-4136, Nov. 2014. Article (CrossRef Link)

[3] F.-P. An, X.-W. Zhou, and D.-C. Lin, "Multiscale self-coordination of bidimensional empirical mode decomposition in image fusion," *KSII Transactions on Internet and Information Systems*, vol. 9, no. 4, pp. 1441-1456, Apr. 2015. Article (CrossRef Link)

[4] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B Goldman, and E. Shechtman, "Robust Patch-Based HDR Reconstruction of Dynamic Scenes," in *Proc. of SIGGRAPH Asia*, pp. 203:1-203:11, 2012. Article (CrossRef Link)

[5] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, "High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting," *Morgan Kauffman*, 2005.

[6] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. of SIGGRAPH*, pp. 369-378, 1997. Article (CrossRef Link)

[7] K. Jacobs and C. Loscos and G. Ward, "Automatic high-dynamic range image generation for dynamic scenes," *IEEE Computer Graphics and Applications*, vol. 28, no. 2, pp. 84-93, 2008. Article (CrossRef Link)

[8] T. Grosch, "Fast and Robust High Dynamic Range Image Generation with Camera and Object Movement," in *Proc. of Vision, modeling and visualization*, 2006. Article (CrossRef Link)

[9] O. Gallo, N. Gelfandz, W.C. Chenz, M. Tico, and K. Pulli, "Artifact-free High Dynamic Range Imaging," in *Proc. of IEEE Int'l Conf. on Computational Photography*, pp. 1-7, 2009. Article (CrossRef Link)

[10] S. Raman, V. Kumar, and S. Chaudhuri, "Blind De-ghosting for Automatic Multi-Exposure Compositing," in *Proc. of SIGGRAPH Asia Poster*, 2009. Article (CrossRef Link)

[11] E. Khan, A. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Proc. of IEEE Int'l Conf. on Image Processing*, pp.1-4, 2006. Article (CrossRef Link)

[12] Y. S. Heo, K. M. Lee, S. U. Lee, Y. Moon, and J. Cha, "Ghost-Free High Dynamic Range Imaging," in *Proc. of Asian Conf. Computer Vision*, pp. 486-500, 2010. Article (CrossRef Link)

[13] T.J. Park and I.K. Park, "High dynamic range image acquisition using multiple images with different apertures," *Optical Engineering,* vol. 51, no. 12, 127002, 2012. Article (CrossRef Link)

[14] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," in *Proc. of SIGGRAPH*, pp. 24:1-24:12, 2009. Article (CrossRef Link)

[15] T. Mertens, J. Kautz, and F. V. Reeth, "Exposure Fusion," in *Proc. of Pacific Conference on Computer Graphics and Applications*, pp. 1-9, 2007. Article (CrossRef Link)

[16] M. Tico, N. Gelfand, and K. Pulli, "Motion-blur-free Exposure Fusion," in *Proc. of IEEE Int'l Conf. on Image Processing*, pp. 3321-3324, 2010. Article (CrossRef Link)

[17] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive Digital Photomontage," in *Proc. of SIGGRAPH*, pp. 294-302, 2004. Article (CrossRef Link)

[18] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama, "Digital Photography with Flash and No-Flash Image Pairs," in *Proc. of SIGGRAPH*, pp. 664-672, 2004. Article (CrossRef Link)

[19] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Image Deblurring with Blurred/Noisy Image Pairs," in *Proc. of SIGGRAPH*, pp.1:1-1:10, 2007. Article (CrossRef Link)

[20] L. Zhang, A. Deshpande, and X. Chen, "Denoising vs. Deblurring: HDR Imaging Techniques Using Moving Cameras," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 522-529, 2010. Article (CrossRef Link)

[21] N. Joshi and M. F. Cohen, "Seeing Mt. Rainier: Lucky Imaging for Multi-Image Denoising, Sharpening, and Haze Removal,*"* in *Proc. of IEEE Int'l Conf. on Computational Photography*, pp. 1-8, 2010. Article (CrossRef Link)

[22] Y. HaCohen, E. Shechtman, D.B. Goldman, and D. Lischinski, "Non-Rigid Dense Correspondence with Applications for Image Enhancement," in *Proc. of SIGGRAPH*, pp. 70:1-70:10, 2011. Article (CrossRef Link)

[23] R. Gonzalez and R. Woods, "Digital Image Processing," *Second ed. Prentice Hall*, 2002.

[24] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1-6, 2005. Article (CrossRef Link)

[25] J. Chen, C.-K. Tang and J. Wang, "Noise Brush: Interactive High Quality Image-Noise Separation," in *Proc. of SIGGRAPH Asia*, pp. 146:1-146:10, 2009. Article (CrossRef Link)

[26] http://vision.middlebury.edu/stereo/data/, 2016. Article (CrossRef Link)

[27] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004. Article (CrossRef Link)

**Yong Seok Heo** received the B.S. degree in Electrical Engineering in 2005, and the M.S. and the Ph.D. degrees in Electrical Engineering and Computer Science in 2007 and 2012, respectively, from Seoul National University, Korea. During 2012 to 2014, he was with Samsung Electronics, in the Digital Media and Communications R&D Center. Currently, he is with the Department of Electrical and Computer Engineering at Ajou University as an assistant professor. His research interests include segmentation, stereo matching, 3D reconstruction and computational photography.



**Soochahn Lee** received his B.S. and Ph.D. degrees from Seoul National University in 2004 and 2011, respectively. During 2011 to 2014, he was with Samsung Electronics, in the Digital Media and Communications R&D Center. From 2014, he is with the Department of Electronic Engineering at Soonchunhyang University as an assistant professor. His research interests include computer vision and medical imaging.



**Ho Yub Jung** received his B.S. and Ph.D. degrees from Seoul National University in 2006 and 2012, respectively. During 2012 to 2014, he was with Samsung Electronics, in the Digital Media and Communications R&D Center. From 2014, he is with the Division of Computer and Electronics System Engineering at Hankuk University of Foreign Studies as an assistant professor. His research interests include computer vision, machine learning, and medical imaging.