Research Article

DCT-Based HDR Exposure Fusion Using Multiexposed Image Sensors

Geun-Young Lee,[1] Sung-Hak Lee,[1] and Hyuk-Ju Kwon[2]

[1]*School of Electronics Engineering, Kyungpook National University, 80 Deahakro, Buk-Gu, Daegu 702-701, Republic of Korea*
[2]*AJIN Industrial Co. Ltd., 40 Gongdan 8-ro 26-gil, Jillyang-eup, Gyeongsan-si, Gyeongsangbuk-do 38462, Republic of Korea*

Correspondence should be addressed to Sung-Hak Lee; shak2@ee.knu.ac.kr

It is difficult to apply existing exposure methods to a resource-constrained platform. Their pyramidal image processing and quality measures for interesting areas that need to be preserved require a lot of time and memory. The work presented in this paper is a DCT-based HDR exposure fusion using multiexposed image sensors. In particular, it uses the quantization process in JPEG encoding as a measurement of image quality such that the fusion process can be included in the DCT-based compression baseline. To enhance global image luminance, a Gauss error function based on camera characteristics is presented. In the simulation, the proposed method yields good quality images, which balance naturalness and object identification. This method also requires less time and memory. This qualifies our technique for use in resource-constrained platforms.

1. Introduction

In general, the range of luminance in a real scene is wider than the range of a digital camera. In addition, a commercial image format is capable of storing only 8 bits per channel; therefore, the range of storable luminance in a single image is limited. In order to capture all of the luminance information in a real scene, the information must be divided and allocated among several images with different exposures. This division not only uses more storage memory but also creates the inconvenience of scanning several images to recognize luminance information.

To solve such problems, methods of fusing differently exposed images into a single image have been proposed. There are two major fusion methods: high dynamic range (HDR) imaging [1] and exposure fusion [2]. In HDR imaging, an HDR image is first reconstructed from several low dynamic range (LDR) images using camera response function [3], and then an HDR-like LDR image including most of luminance information is produced from the HDR image using tone mapping operators (TMOs). Because an HDR image cannot be shown on general display devices that

do not support the HDR format, it is necessary to tone map an HDR image to an HDR-like LDR image. On the other hand, exposure fusion directly creates an HDR-like LDR image from several LDR images with different exposures. Exposure fusion is thus relatively simpler because it obviates the need to reconstruct an HDR image. The process of exposure fusion starts by defining and measuring image information, such as detail and contrast. The fusion methods then select informative parts from several LDR images and combine them into an HDR-like LDR image without redundancy.

Many methods for measuring and selecting informative parts of images have been researched. Mertens et al. [2] used a quality measure constructed by contrast, saturation, and well-exposedness and combined input images using a pyramid-based fusion technique. Song et al. [4] measured the visible contrast and the visual gradients in input images and synthesized input images based on a probabilistic model that can be transformed to a maximum a posteriori. Finally, block-based image fusion in [5] selected the most informative image for each block using entropy of the image.

These methods produce a reasonably good quality image; however, they are not qualified for resource-constrained platforms. It is time-consuming to obtain an HDR-like LDR image including most of the luminance information because of the computational complexity involved. Furthermore, pyramid-based fusions [2, 6] require more memory. Li and Kang [7] proposed relatively fast exposure fusion which can combine images with moving objects, but their approach is also quite time-consuming. Furthermore, the majority of digital cameras, including those used in resource-constrained platforms, which leads to additional steps to decode and encode compression data streams in a fusion process. Although Kakarala and Hebbalaguppe [8] avoid these extra steps in their proposed method of fusing two images in the JPEG domain, their results lack detailed information in dark areas because the boosted luminance channel of the short-exposure image is only used as a luminance channel of the result image. Finally, discrete cosine transform- (DCT-) based methods [9–11] compute local information using DCT coefficients, and this computation takes more time.

In this paper, we propose DCT-based HDR exposure fusion using dual exposed image sensors that have symmetric exposure values, +EV and −EV. The proposed fusion consists of two sections: image fusion in an encoding field for resource-constrained platforms and DC level reproduction in a decoding field for displaying a fusion image. In particular, to reduce computational complexity, which becomes a burden on resource-constrained platforms, our approach excludes additional measurements for image quality, such as contrast and entropy in DCT-based compressions. Instead, we assume that the quantization process in the JPEG baseline is sufficient for measuring image quality. As a result, we confirm that a proposed method is able to quickly yield a fusion image with equal to or higher than the quality of methods that can be used in resource-constrained platforms.

2. Image Compression Baseline

JPEG [12] is a widely used image compression standard. Owing to the simplicity of the processing and good compression performance for fair quality images, many kinds of digital cameras store images using the JPEG standard. In the JPEG baseline, an RGB color space of the image is first transformed to a YCbCr color space as follows:

$$
\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 0 \\ delta \\ delta \end{pmatrix},
$$

(1)

where delta has a different value according to the image data type. If the image data type is an unsigned 8-bit integer, the delta is set to 128.

After the forward transform from RGB to YCbCr, block-based JPEG compression is conducted. As shown in Figure 1, an image is divided into nonoverlapping 8×8 blocks, and then 64 pixels in the block are transformed into the frequency
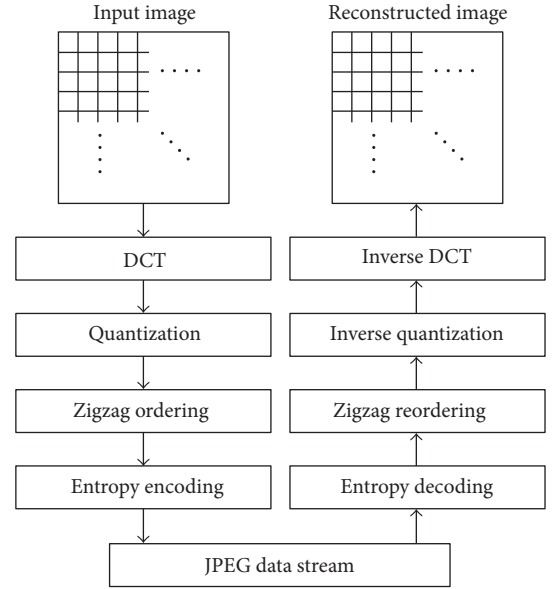

FIGURE 1: Block-based JPEG baseline: compression (left) and decompression (right).

domain using the DCT. The DCT of pixels in the 8×8 block is defined by

$$
F(u, v) = \frac{1}{4} \alpha(u)\alpha(v) \sum_{x=0}^{7} \sum_{y=0}^{7} f(x, y)
$$
$$
\cos\left[(2x + 1)\frac{\pi u}{16}\right] \cos\left[(2y + 1)\frac{\pi v}{16}\right],
$$

(2)

for $u = 0, 1, \dots, 7$ and $v = 0, 1, \dots, 7$, where

$$
\alpha(k) = \begin{cases} \dfrac{1}{\sqrt{2}} & \text{for } k = 0 \\ 1 & \text{otherwise,} \end{cases}
$$

(3)

and $f(x, y)$ is the pixel level in the spatial domain. The transformed 8×8 block consists of one DC coefficient and 63 AC coefficients. The DC coefficient, $F(0, 0)$, is the sum of the 64 pixels multiplied by the scale factor, 1/8. The quantization process is then dividing the DCT coefficients, $F(u, v)$, by the quantization matrix and rounding the values off to the nearest integer.

To encode the quantized 8×8 block, the DC coefficient in the current block is subtracted from the DC coefficient in the previous block, and the difference is encoded. In the case of AC coefficients, zigzag ordering is required to increase coding efficiency. Because high-frequency AC coefficients are quantized to zeros, the zigzag ordering might form a long sequence of zeros following low-frequency AC coefficients. Finally, the quantized data stream is encoded into the corresponding bit stream using run-length coding and Huffman coding.
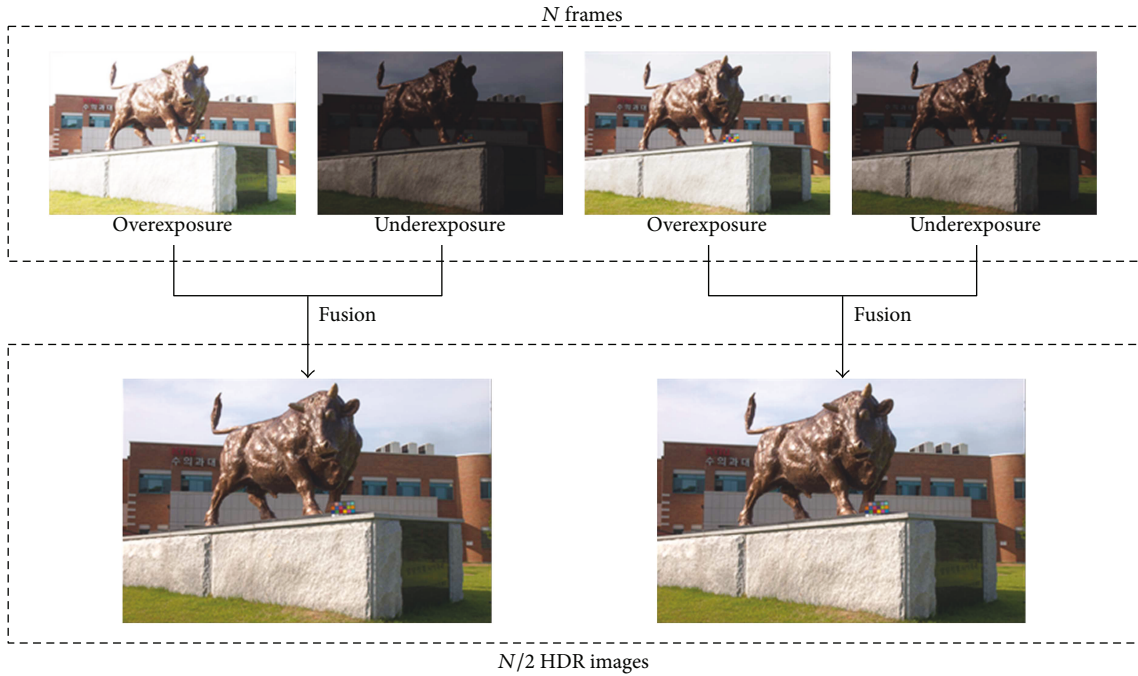
FIGURE 2: Illustration of HDR exposure fusion using dual exposed sensors.

3. DCT-Based HDR Exposure Fusion Using Dual Exposed Sensors

The bracketing mode of cameras typically produces images of symmetric exposure values, for instance, +EV, 0, and −EV, where EV is an exposure value. Because the luminance information of the scene is sufficiently present in +EV and −EV images, the proposed method utilizes only two symmetric exposed images with +EV and −EV. By alternatively capturing over- and underexposed images at N frames, it can generate $N/2$ HDR frames as shown in Figure 2. The process of exposure fusion is divided into image fusion in the JPEG compression and DC level reproduction in the JPEG decompression for a resource-constrained platform. Through the separation of the exposure fusion process, the two images can be quickly fused in the camera without computational complexity.

3.1. Image Fusion in the Compression Field

3.1.1. Quality Measurement Using the Length of DCT Coefficients. In general, the detail in bright regions appears best in the −EV image, whereas the detail in dark regions does so in the +EV image. In other words, in order to reproduce the detail of the scene in a fusion image, it is necessary to decide which of two images best represents the detail in each region. In the case of the JPEG data stream, the AC coefficients in the 8×8 DCT block correspond to the detail. In the quantization process, the insufficient level of high-frequency AC coefficients makes them converge to zeros, so that the low-frequency AC coefficients, which do not converge to zeros, represent the degree of the detail in the 8×8 block. Therefore, without additional steps, it is possible

to use the length of the AC coefficients as the quality measure in the fusion process [13].

Figure 3 presents one example of the encoding in JPEG. In this example, the bit stream encoding using Huffman code is skipped for clarity. First, Figure 3(a) shows the quantized 8×8 block in the DCT domain. This block has a DC coefficient and only a few low-frequency AC coefficients. Because of the quantization process, many AC coefficients have converged to zero. Coefficients in the block are arranged by zigzag ordering (as shown in Figure 3(b)). Finally, the arranged data is classified by run-length coding to reduce its length (as shown in Figure 3(c)). The run-length coded data stream consists of (*RUNLENGTH*, *CATEGORY*) and (*AMPLITUDE*) where *RUNLENGTH* is the number of consecutive zeros preceding the nonzero AC coefficient indicated by *AMPLITUDE*; *CATEGORY* is the number of bits to encode the nonzero AC coefficient. Therefore, the length without consecutive zeros, which correspond to converged high-frequency AC coefficients, can be directly estimated from the run-length coded data stream. In this example, the length without consecutive zeros is 5 (DC coefficient, 1 *RUNLENGTH*, −6 *AMPLITUDE*, 1 *RUNLENGTH*, and −4 *AMPLITUDE*).

3.1.2. Selective Fusion Rule in the Compression Field. Two image fusions follow the maximum selection rule; the block whose DCT coefficients have the maximum length belongs to the fusion image. Let $P = \{p_{x,y}; \ x = 0, \ldots, \ N-1 \ \text{and} \ y = 0, \ldots, \ M-1\}$ be an image which consists of $N \times M$ blocks of size 8×8. Suppose that $D_n = \{d_{n,u,v}; \ n = 0, \ldots, \ N \times M - 1 \ \text{and} \ 0 \leq u, \ v \leq 7\}$ be the corresponding DCT coefficients and $Q_n = \{q_{n,u,v}; \ n = 0, \ldots, \ N \times M - 1 \ \text{and} \ 0 \leq u, \ v \leq 7\}$ be the quantized DCT coefficients of each nth 8×8 block. In the

Quantized data in the 8 × 8 block

176	0	0	0	0	0	0	0
−6	−4	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Zigzag ordered data stream:

176 0 −6 0 −4 0 0 ⋯ 0

$\underbrace{\qquad}_{59 \text{ zeros}}$

Run-length coded data stream:

(0, 8)(176), (1, 3)(−6), (1, 3)(−4), (0, 0)

(a)   (b)   (c)

FIGURE 3: Example of an 8 × 8 block after quantization. (a) Quantized data in the block, (b) zigzag ordered data stream, and (c) run-length coded data stream.

672	33	60	0	−24	0	−51	0
−36	−24	0	19	26	0	0	0
0	0	−16	−24	0	0	0	0
14	34	22	0	0	0	0	0
18	22	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

176	0	0	0	0	0	0	0
−6	−4	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

(a)   (b)

FIGURE 4: Example of 8 × 8 blocks after the quantization. The lengths of the coefficients are (a) 37 (the last nonzero value in the zigzag ordering is −51) and (b) 4 (the last nonzero value is −4).

proposed image fusion, the nth block of the fusion image, $Q^F{}_n$, is obtained as follows:

$$Q^F{}_n = Q^K{}_n, \quad \text{where } K = \arg\max_k \left\{ L^k{}_n \right\}, \quad k = +\text{EV}, -\text{EV}.$$

(4)

$L^k{}_n$ is the length of the coefficients of the nth block of the kth image after the quantization. For example, the quantized DCT 8 × 8 blocks in the same position as the +EV and −EV images are shown in Figures 4(a) and 4(b), respectively. In this example, the block of the +EV image, $Q^{+\text{EV}}$, becomes that of the fusion image, Q^F, because $L^{+\text{EV}}$ (equal to 37) is longer than $L^{-\text{EV}}$ (equal to 4).

Fusion rule in (4) has an advantage that two images are fused in the JPEG data stream without complex computation or additional processing because the length of the coefficients can be derived simply from the JPEG data stream. Therefore, a result image can be easily fused in the camera and directly transmitted or stored because the result image is already a form of the JPEG data stream. Furthermore, as shown in Figure 5, the result is competitive with that of the variance-based fusion rule without consistency verification [10] in regard to detail selection.

3.2. DC Level Mapping in the Decoder. Although detail is reconstructed using the proposed image fusion, the transmitted or stored JPEG data stream of a fusion image requires the manipulation of local tone using DC coefficients. For faster processing in the JPEG compression, it is acceptable to take a simple average of two DC coefficients in the DCT 8 × 8 blocks; however, because of the significant level of difference between +EV and −EV images, a simple average of DC levels produces unpleasant local tones in the fusion image. In addition, as shown in Figure 6, detail in a fusion image does not appear clearly because the DC level is too dark or too high. Note that Sub1 is darker in the fusion image than in the +EV image and Sub2 is brighter in the fusion image than in the −EV image, so the detail is insufficient. To solve this problem, in the JPEG decompression, we estimate DC levels of the +EV image in dark regions and those of the −EV image in bright regions from an average DC level of the transmitted JPEG data stream.

3.2.1. Gauss Error Function for Estimating DC Levels. We conducted an experiment to determine the relationship between each input image and a transmitted average value of the two input images. A number of symmetrically exposed images using a linear gradient pattern were
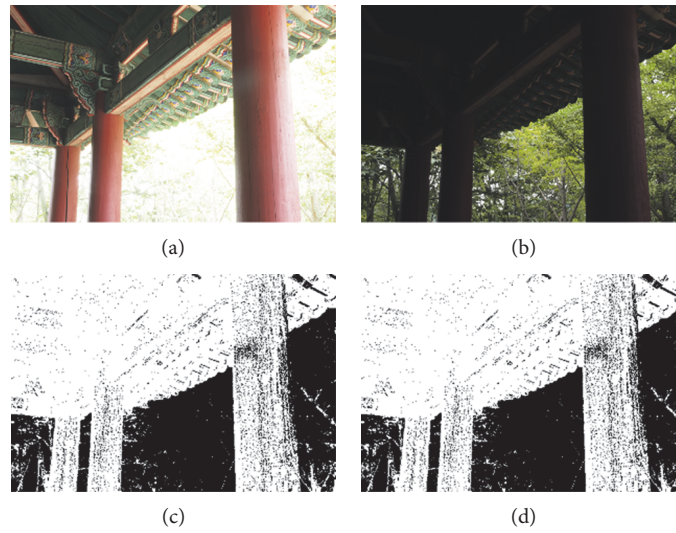
Figure 5: The results of fusion rules using (a) +EV and (b) −EV images (ΔEV = 4); (c) the variance-based fusion rule without consistency verification in [10] and (d) our proposed fusion rule. White pixels indicate that the +EV image is selected, and black pixels indicate that the −EV image is selected.
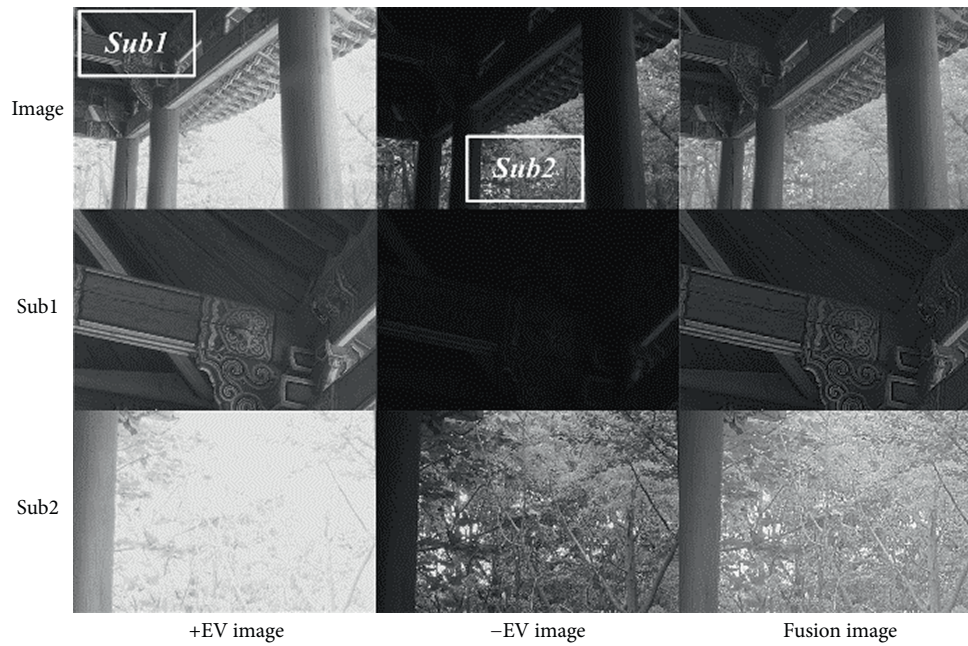

Figure 6: Gray images of +EV and −EV and fusion images using (4).

captured using a camera (model: Sony α6000) with ±0.3 EV, ±0.5 EV, ±0.7 EV, ±1.0 EV, ±1.3 EV, and ±2.0 EV. Then for each symmetric exposure value, the scatter graphs for the pixels of each test pattern image against the corresponding average pixel values of ±EV images were plotted as shown in Figure 7 (blue and green data). In our experiment, the maximum exposure is limited to ±2.0 EV because images with an EV value higher than 2.0 have too many saturated pixels. From the scatter graphs for each symmetric exposure value, we see that the scatter graphs exhibit point symmetry and can be estimated using the Gauss error function as follows:

$$I'_{+EV} = \mathrm{ERF}\left(\alpha I_{avg}\right), \tag{5}$$

$$I'_{-EV} = \mathrm{ERF}\left(\alpha\left(I_{avg} - 1\right)\right) + 1, \tag{6}$$

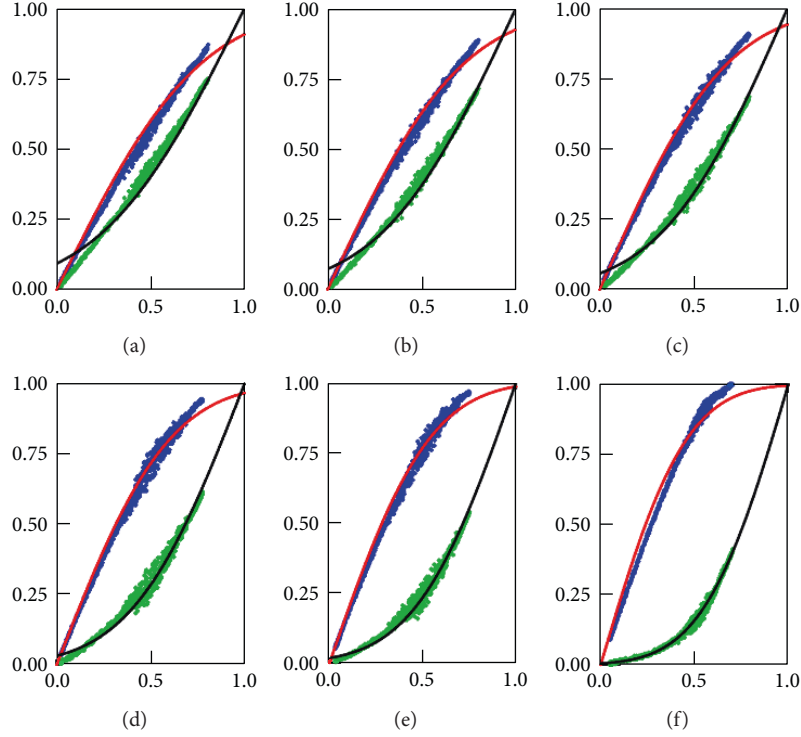$$\mathrm{ERF}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt\ (\mathrm{Gauss\ error\ function}), \tag{7}$$

FIGURE 7: Scatter graphs using a Sony α6000 and Gauss error function curves. Blue and green represent the scatter graphs for +EV and −EV images against the average image. Red and black lines represent the estimation curves using the Gauss error function. (a) EV = ±0.3 and $\alpha = 1.1895$ in (5) and (6); (b) EV = ±0.5 and $\alpha = 1.2665$; (c) EV = ±0.7 and $\alpha = 1.3435$; (d) EV = ±1.0 and $\alpha = 1.5150$; (e) EV = ±1.3 and $\alpha = 1.6845$; and (f) EV = ±2.0 and $\alpha = 1.9930$.

where I_{+EV}' and I_{-EV}' are the estimated image levels for the +EV and −EV images, respectively, and I_{avg} is the average level of the two images. Because the Gauss error function is an odd function, the scatter graphs can be estimated using the function and its translation. In addition, the parameter, α, in the Gauss error function correlates to the absolute exposure value of the images. We plot α against discrete EV data as a function of exposure value in Figure 8 and simply obtain the parameter α as follows:

$$\alpha = 0.5|EV| + 1. \qquad (8)$$

In Figure 7, we superimpose red and black lines obtained from (5) and (6) on the scatter graphs. Although there are deviations in the bright regions of the +EV image and the dark regions of the −EV image, they are acceptable because this luminance, which is generally saturated in the +EV and −EV images, is not included in the fusion image. In other words, each piece of luminance information in the dark and bright regions of the scene is estimated from the +EV and −EV images, respectively.

Similarly, Kakarala and Hebbalaguppe proposed the brightness transfer function (BTF) using a sigmoidal function to boost the intensity of a short exposure image up to that of a long exposure image [8]. However, the BTF for the image with large ΔEV has a high gradient, and for a fusion image, only the boosted pixel level in the short exposure image is used. On the contrary, our function,
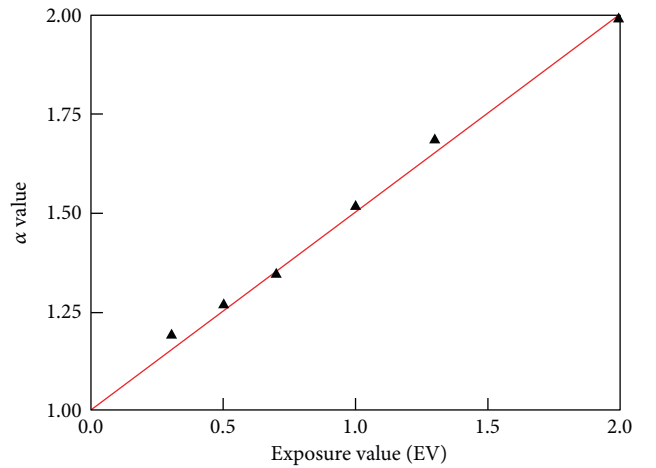

FIGURE 8: The parameter, α, as a function of exposure value. Black triangles represent the discrete EV data in Figure 5, and the red solid line represents (8).

which has a relatively low gradient, can estimate levels of both the +EV and the −EV images from an average level.

To confirm that the functions in (5), (6), (7), and (8) are available to different cameras, we captured the same pattern with ±1.0 EV and ±2.0 EV using an Olympus E-PM1 and the mobile phone cameras of the Nexus 5 and Galaxy S5. The phone cameras are considered resource-constrained platforms because they are relatively nonprofessional camera
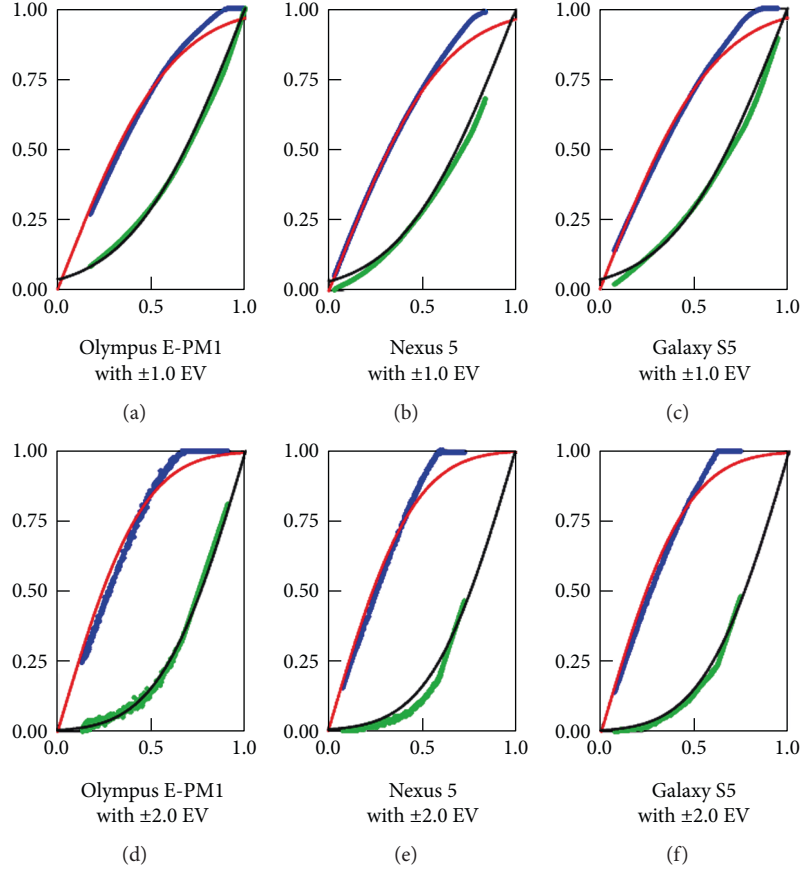
FIGURE 9: Scatter graphs using different cameras and Gauss error function curves. Blue and green represent the scatter graphs for +EV and −EV images against the average image. The red and black lines represent the estimation curves using the Gauss error function.

models. Similar to Figure 7, scatter graphs for these cameras and Gauss error function curves are shown in Figure 9. Although there are slight deviations, the estimation using the Gauss error function is successful.

In addition, based on the camera response function (CRF) constructed using the five images with EVs = 2, 1, 0, −1, and − 2, we verified our estimate for general images. The left graph in Figure 10 shows the CRF with irradiance, E, and exposure time, Δt. Assuming the exposure range of the camera with EV = +2 is [−2 4] in the log domain, the range of the camera with EV = −2 is [−4.77 1.23] because the exposure time of EV = +2 is sixteen times longer than that of EV = −2, and the gap is approximately 2.77 in the log domain. Similar to Figure 7, the pixel graphs against the corresponding average pixel values of ±EV images (blue and green) and the estimated graphs (red and black) are plotted in Figure 10(b). As in the test using a pattern image, the estimation for general images is successful.

3.2.2. Reproduction of Corresponding DC Levels. Our experiment demonstrates that the levels of ±EV images can be estimated using the Gauss error function of the average level with an exposure value. However, having the DC levels simply spatially switch between the estimated levels of ±EV images in the JPEG decompression produces level discontinuity in the result image. To smooth the

discontinuity, we apply the weighting map, w, to the sum of the estimated levels in (5) and (6) as follows:

$$
\begin{aligned}
DC_{fusion} = w\left[ERF\left(\alpha\left(DC_{avg} - 1\right)\right) + 1\right] \\
+ (1 - w)ERF\left(\alpha DC_{avg}\right),
\end{aligned}
\tag{9}
$$

where $ERF(\cdot)$ and α are derived from (7) and (8), respectively, and DC_{avg} is an average DC coefficient of the ±EV images. The weighting map, w, is obtained from blurring the subimage composed of DC_{avg} values so that the map varies spatially. For simplicity, the weighting map is constrained within the range [0 1]. Bright regions in the scene are indicated by $w = 1$. Thus, (9) estimates the level of the −EV image because bright regions appear best in the −EV image, whereas bright regions in the +EV image are saturated. On the other hand, dark regions in the scene are indicated by $w = 0$. This means that the level of the +EV image is obtained from (9) because the dark regions in the −EV image are too dark.

We show the function graphs of (9) in Figure 11(a). These graphs change smoothly between the estimated DC levels of ±EV images according to w values. Therefore, level discontinuity in the fusion image disappears. To illustrate this, in Figure 11(b), we show the five DC images: +EV, −EV, w, DC_{avg}, and DC_{fusion}. The dark regions in +EV and the bright regions in −EV are well expressed in DC_{fusion}, which is derived from (9) using DC_{avg} and w.
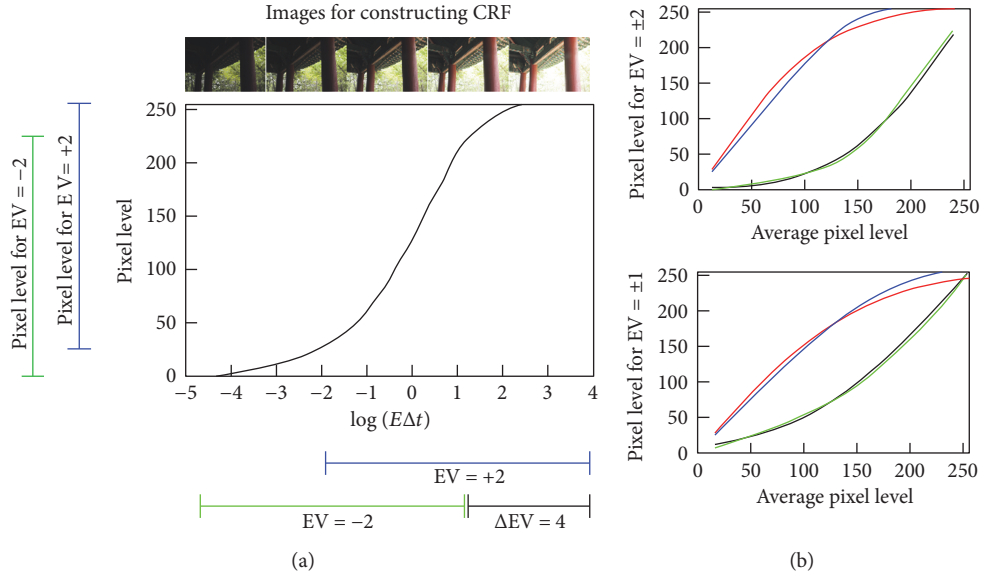
FIGURE 10: The CRF graph (a) and pixel curves (b) against the average pixel level. Blue and green represent the reference curves for +EV and −EV. The red and black lines represent the estimation curve using the Gauss error function. The CRF is constructed using Debevec's method [3].
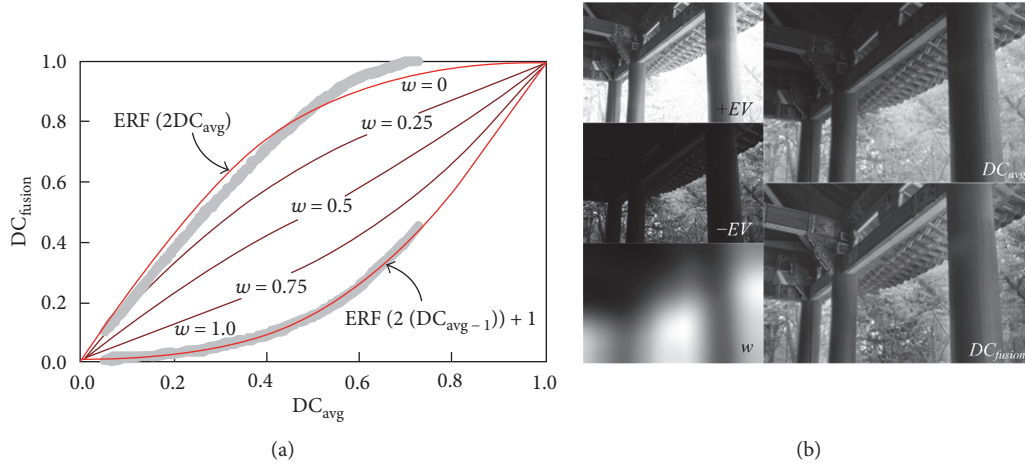

FIGURE 11: Proposed function graphs for a DC level reproduction (a) and example images (b). The red and dark red curves span five ws: 0, 0.25, 0.5, 0.75, and 1. The gray scatter graphs correspond to the blue and green scatters in Figure 7(f).

3.3. DCT-Based HDR Exposure Fusion. A block diagram of the proposed JPEG-based exposure fusion is shown in Figure 12; the blue and red lines indicate the manipulations of DC and AC coefficients, respectively. As shown in Figure 12(a), because the proposed fusion requires only two simple operations in the camera—the operation comparing length using AC coefficients and the average operation using DC coefficients—it is easily applied on a resource-constrained platform, such as a surveillance system. The reproduction of DC levels of the fusion image in the JPEG decompression is then enhanced when displaying the transmitted JPEG data stream as shown in Figure 12(b).

As an example, a specific JPEG data stream using image block data from the dark region of the scene is entered. In the JPEG compression within the camera of the resource-constrained platform, image fusion is conducted; AC coefficients are selected using the fusion rule, and DC coefficients are averaged. When displaying the fusion image in the JPEG decompression, the DC level is reproduced using an average value of DC coefficients and EV values. As a result, the fusion image block has a DC coefficient which is similar to that of the +EV image block and AC coefficients that are exactly the same as those of the +EV image block.

4. Simulations

4.1. Simulation Setup. Six image sets are used in simulation: "Building 1," "Building 2," "Gazebo," "Belgium house" [14], "Venice carnival" [15], and "Memorial church" [3]. For comparison, three existing fusion methods are used: exposure fusion (EF) [2], fast multiexposure image fusion (FMMR)

(a) Image fusion in JPEG compression
for resource constrained platform

(b) Reproduction of DC levels in JPEG decompression
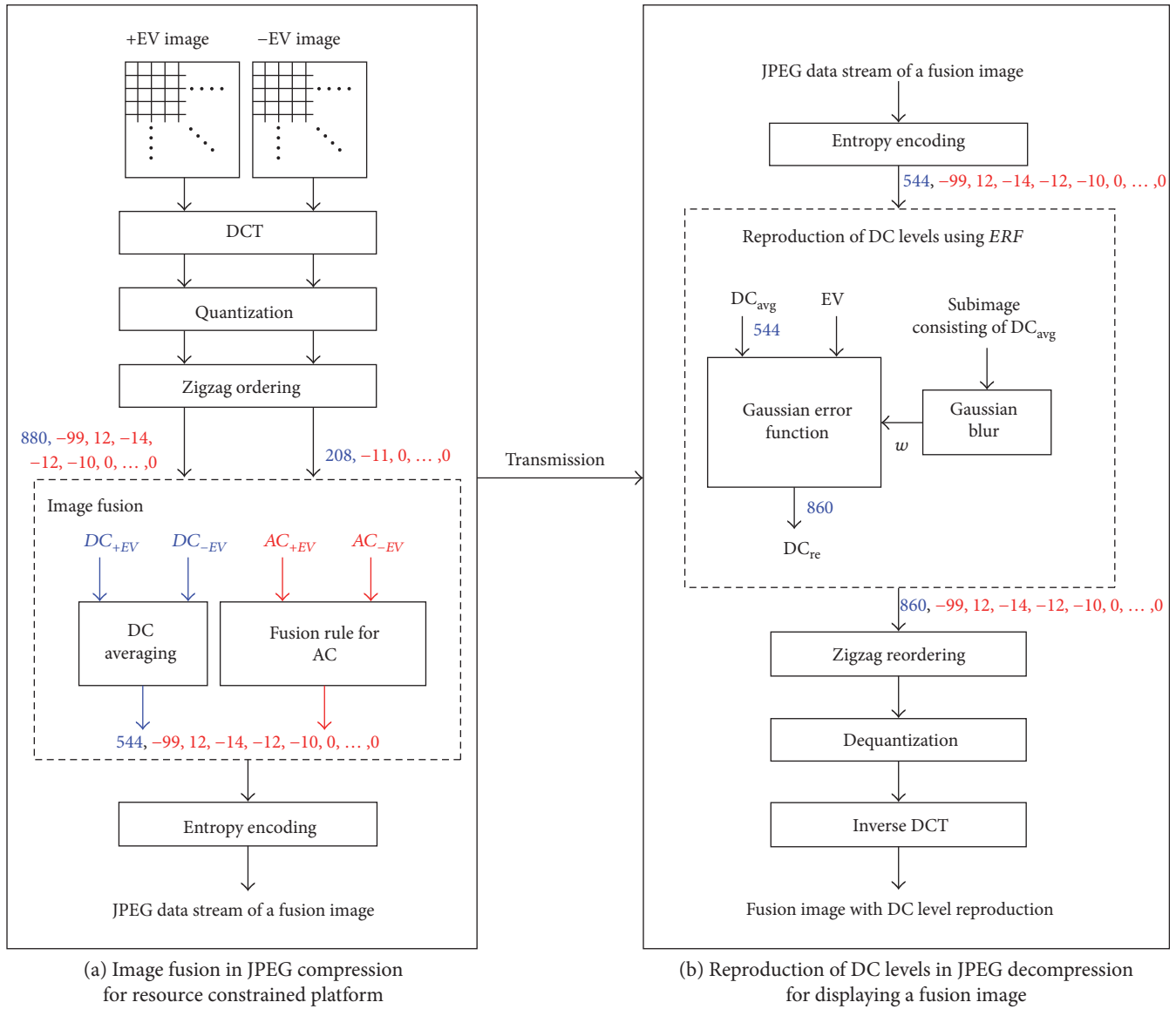for displaying a fusion image

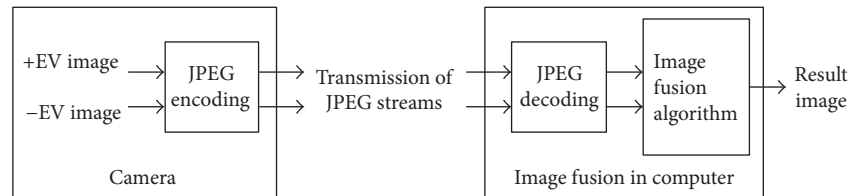FIGURE 12: Proposed JPEG-based exposure fusion diagram.

FIGURE 13: Test bed for a resource-constrained platform.

[7], and probabilistic model-based fusion using generalized random walks (GRW) [16]. Our approach considers only the Y channels of the two symmetric exposed images such that for color processing, we take the furthest value from the neutral point between the two images in the CbCr color domain.

The major advantage of our exposure fusion is its applicability to resource-constrained platforms. Furthermore, unlike our method, the existing methods are not able to take images in the JPEG stream without a JPEG decoder. Although raw image data may be available for fusion using the existing methods, the use of the raw data requires too much memory. Fusing images using the existing methods causes an increase of computational complexity in the camera that should be avoided. For this reason, we set the test bed that fuses images using the existing methods after JPEG encoding and decoding modules, as shown in Figure 13.
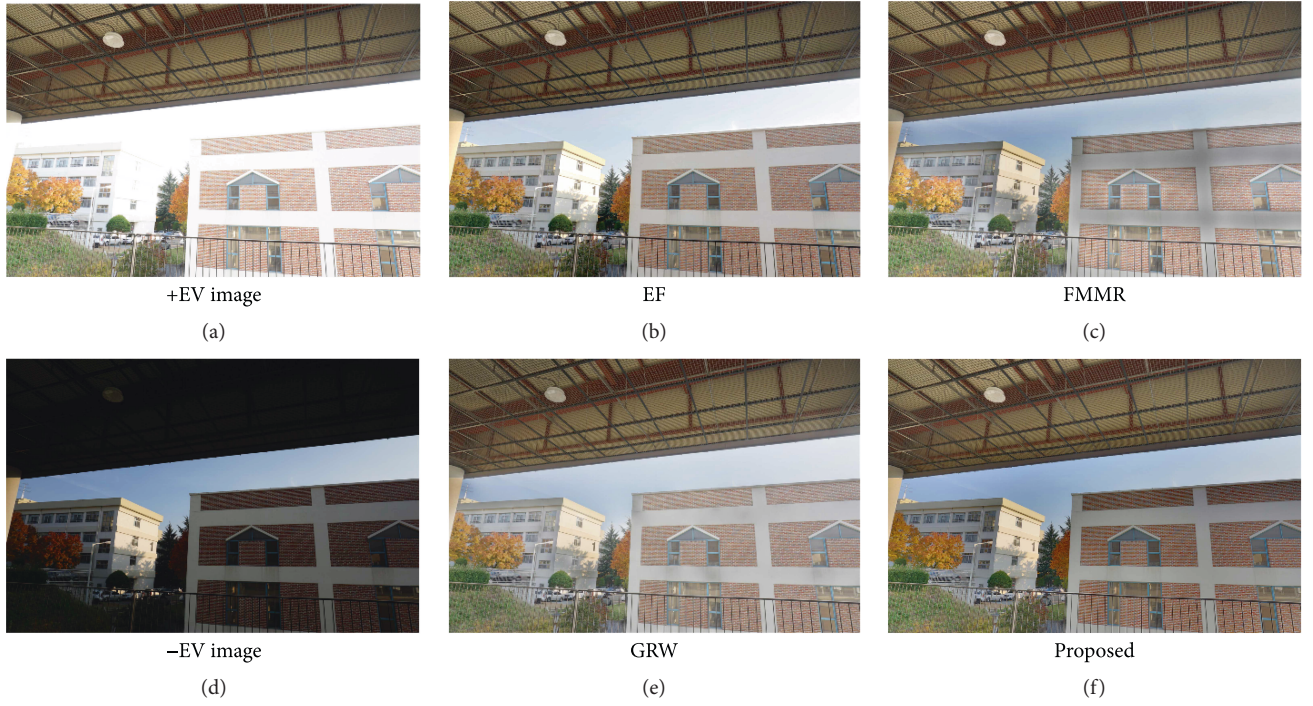
+EV image

(a)

EF

(b)

FMMR

(c)

−EV image

(d)

GRW

(e)

Proposed

(f)

Figure 14: Result images of "Building 1."

+EV image

(a)

EF
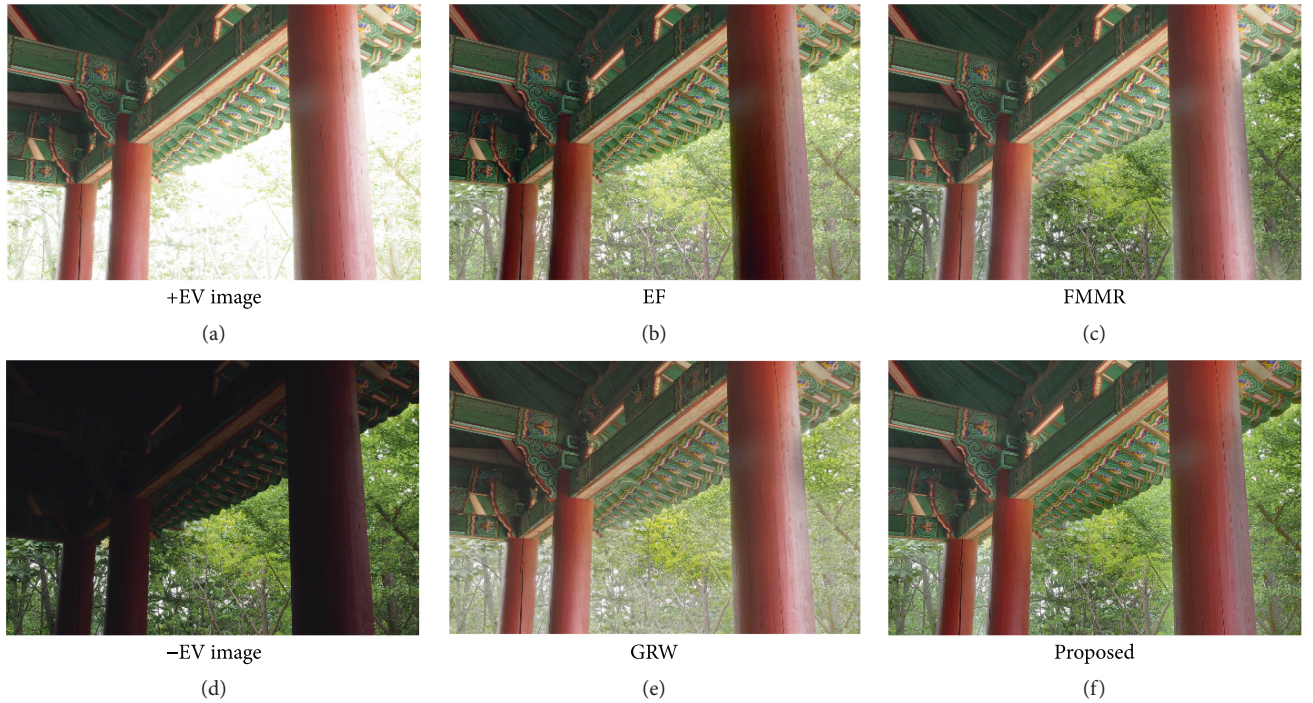
(b)

FMMR

(c)

−EV image

(d)

GRW

(e)

Proposed

(f)

Figure 15: Result images of "Gazebo."

4.2. Result Images. Figures 14–16 show the result images obtained with the existing and proposed methods using dual exposed capturing. In "Building 1" and "Gazebo," the results of FMMR and GRW are stained. Particularly in the FMMR result for "Buildings 1," the upper left of the red brick building is unnatural because the method has a subsampling process for reducing the computing and memory consumption.

Similarly, in the GRW result for "Gazebo," most of the green leaves have low chroma. While the results of EF seem more natural, it is generally hard to identify details in the darkest and brightest areas. In contrast, our proposed method produces natural images with good details in these areas. The enhanced results are surely confirmed in the cropped images shown in Figure 17.

(a) + EV image

(b) EF

(c) FMMR

(d) −EV image

(e) GRW

(f) Proposed

FIGURE 16: Result images of "Belgium house."

| EF | FMMR |
| GRW | Proposed |

FIGURE 17: Subimages of the result images.

To objectively verify the image quality, we use five metrics for quantitative assessments: SSIM [17], FSIM [18], FMI [19], MEF [20], and TMQI [21]. SSIM is a well-known quality metric based on structural similarity of images. FSIM is based on the salient low-level features of the perceived scene and shows higher consistency with the subjective evaluations. Because SSIM and FSIM are reference-based assessments, we crop dark and bright areas, respectively,

TABLE 1: Quantitative metric results for SSIM and FSIM.

	SSIM				FSIM			
	EF	FMMR	GRW	Proposed	EF	FMMR	GRW	Proposed
"Building 1"	0.9033	0.9219	0.9055	0.9312	0.9385	0.9516	0.9398	0.9585
"Building 2"	0.8878	0.8970	0.8392	0.9054	0.9404	0.9395	0.9230	0.9428
"Gazebo"	0.8578	0.9154	0.8602	0.9193	0.9218	0.9446	0.9192	0.9514
"Belgium house"	0.7926	0.8663	0.8496	0.8904	0.8964	0.9275	0.9197	0.9222
"Venice carnival"	0.9155	0.9220	0.8875	0.9317	0.9343	0.9405	0.9322	0.9450
"Memorial church"	0.8313	0.8695	0.8439	0.8856	0.9035	0.9233	0.9160	0.9304
Average	0.8647	0.8987	0.8643	0.9106	0.9225	0.9378	0.9250	0.9417
(rank)	(3)	(2)	(4)	(1)	(4)	(2)	(3)	(1)

TABLE 2: Quantitative metric results for FMI and MEF.

	FMI				MEF			
	EF	FMMR	GRW	Proposed	EF	FMMR	GRW	Proposed
"Building 1"	0.8539	0.8760	0.8695	0.8835	0.9772	0.9706	0.9694	0.9703
"Building 2"	0.9020	0.9110	0.8798	0.9099	0.9751	0.9668	0.9583	0.9729
"Gazebo"	0.7763	0.7803	0.8016	0.7747	0.9684	0.9598	0.9399	0.9577
"Belgium house"	0.8955	0.9074	0.9064	0.8895	0.9661	0.9661	0.9601	0.9629
"Venice carnival"	0.9303	0.9331	0.9329	0.9326	0.9757	0.9725	0.9696	0.9779
"Memorial church"	0.9022	0.8756	0.9174	0.9093	0.9712	0.9404	0.9669	0.9698
Average	0.8767	0.8806	0.8846	0.8832	0.9723	0.9627	0.9607	0.9686
(rank)	(4)	(3)	(1)	(2)	(1)	(3)	(4)	(2)

and the cropped images are used as reference images. FMI, which is a feature-based image fusion metric, calculates the amount of mutual information carried from the source image to the fused image. MEF is designed for multiexposure image fusion. MEF correlates particularly well with subjective judgement. Finally, TMQI is an image quality metric for tone-mapped images. In other words, using an HDR image as a reference, TMQI measures signal fidelity and naturalness of a tone-mapped image. We adopt this metric because a tone-mapped image is similar to an exposure-fused image. In our simulation, HDR reference images for TMQI are made using Adobe Photoshop CS6 with two source images.

Quantitative results are shown in Tables 1, 2, and 3. The proposed method ranks first in SSIM score (0.9106) and FSIM score (0.9417) and second in FMI (0.8832), MEF (0.9686), and TMQI (0.9394) scores. Although the proposed method is not the first in all of the individual metrics, it is the first in the overall ranking using all metrics (the sum of ranks for the four metrics; EF = 13, FMMR = 14, GRW = 15, proposed = 8). EF has the highest scores in MEF (0.9723) and TMQI (0.9497), but the lowest scores in SSIM (0.8647), FSIM (0.9225), and FMI (0.8767). This means that EF has good naturalness but bad signal fidelity, whereas the proposed method is more faithful to structural fidelity with a slight loss of naturalness. FMMR and GRW have relatively good signal fidelity but are lacking in naturalness. For example, their result images for "Building 1" and "Gazebo" have many halo

TABLE 3: Quantitative metric results for TMQI.

	TMQI			
	EF	FMMR	GRW	Proposed
"Building 1"	0.9372	0.9131	0.9216	0.9353
"Building 2"	0.9582	0.9325	0.9450	0.9559
"Gazebo"	0.9589	0.9539	0.9429	0.9330
"Belgium house"	0.9442	0.9552	0.9406	0.9311
"Venice carnival"	0.9427	0.9246	0.9190	0.9471
"Memorial church"	0.9568	0.9037	0.9512	0.9337
Average	0.9497	0.9305	0.9367	0.9394
(rank)	(1)	(4)	(3)	(2)

artifacts such that they are stained. In contrast, our proposed method yields well-balanced result images.

4.3. Computation Time. For resource-constrained platforms, computation time is one of the main points to be considered. Table 4 shows the computation time for each method in MATLAB on a 3.40 GHz (i7-2600K) CPU PC with 8.00 GB RAM. Because of the decision to use JPEG streams for a resource-constrained platform, the results include the times consumed by the JPEG modules as in the test bed in Figure 13. The proposed method has the fastest computational time. The brute force JPEG code causes the computational times of the JPEG decoding and encoding to

Table 4: Computation time of the methods in MATLAB on the test bed in Figure 13 (in seconds).

	EF	FMMR	GRW	Proposed method
"Building 1" (3000 × 2000)	50.94	185.20	47.94	38.35
"Building 2" (1200 × 800)	8.25	29.87	7.71	6.09
"Gazebo" (2016 × 1344)	22.95	83.58	21.77	17.22
"Belgium house" (1024 × 768)	6.78	24.40	6.35	5.00
"Memorial church" (512 × 768)	3.43	12.26	3.19	2.51

constitute a large portion of the result times. Nevertheless, considering that it does not take long (about 0.5 seconds) to write "Building 1" as a JPEG image file using the *imwrite* function in MATLAB, the proposed method can have very fast computational times.

If the JPEG modules in Figure 13 are removed, the memory requirements of the other fusion methods become excessive. For example, the memory for the two raw images of "Gazebo" is about 16.26 MB, while the memory for the two JPEG images is only about 1.95 MB. Therefore, in considering the memory requirement and the computation time together, the proposed method is superior to the existing methods.

5. Conclusions

In this paper, DCT-based HDR exposure fusion for resource-constrained platforms is proposed. To fuse two symmetric exposed images in the JPEG baseline, we demonstrate that the quantization process in the JPEG baseline qualifies for the quality measure in the fusion process and that the Gauss error function estimates the DC levels of the source images from average DC levels well. Furthermore, for resource-constrained platforms, two symmetric exposed images are fused in JPEG compression, and then the DC level of the fusion image is reproduced in the JPEG decompression. The simulation results indicate that the proposed method balances naturalness and detail in saturated regions for overall good image quality. In addition, the proposed method has a very fast computation time and requires less memory such that it satisfies the demands for exposure fusion in resource-constrained platforms.

Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High dynamic range imaging: acquisition, display, and image-based lighting, Morgan Kaufmann Publishers, 2005.CIE, Publication No. 15.2*, Central Bureau of the CIE, Vienna, Austria, 2nd edition, 1986.

[2] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: a simple and practical alternative to high dynamic range photography," *Computer Graphics Forum*, vol. 28, no. 1, pp. 161–171, 2009.

[3] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'97)*, pp. 369–378, 1997.

[4] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, "Probabilistic exposure fusion," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 341–357, 2012.

[5] A. A. Goshtasby, "Fusion of multi-exposure images," *Image and Vision Computing*, vol. 23, no. 6, pp. 611–618, 2005.

[6] T. Pu and G. Ni, "Contrast-based image fusion using the discrete wavelet transform," *Optical Engineering*, vol. 39, no. 8, pp. 2075–2082, 2000.

[7] S. Li and X. Kang, "Fast multi-exposure image fusion with median filter and recursive filter," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 626–632, 2012.

[8] R. Kakarala and R. Hebbalaguppe, "A method for fusing a pair of images in the JPEG domain," *Journal of Real-Time Image Processing*, vol. 9, no. 2, pp. 347–357, 2014.

[9] J. Tang, "A contrast based image fusion technique in the DCT domain," *Digital Signal Processing*, vol. 14, no. 3, pp. 218–226, 2004.

[10] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "Multi-focus image fusion for visual sensor networks in DCT domain," *Computers & Electrical Engineering*, vol. 37, no. 5, pp. 789–797, 2011.

[11] I. Zafar, E. A. Edirisinghe, and H. E. Bez, "Multi-exposure & multi-focus image fusion in transform domain," in *Proceeding of IET Conference on Visual Information Engineering*, pp. 606–611, Bangalore, India, September 2006.

[12] G. K. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp. 30–44, 1991.

[13] G. Y. Lee, S. H. Lee, H. J. Kwon, and K. I. Sohng, "Image fusion using two symmetric exposed images in the JPEG stream," in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition*, pp. 341-342, Athens, 2016.

[14] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 249–256, 2002.

[15] J. Joffre, "Venice Carnival 2007 at sunrise," https://www.hdrsoft.com/gallery/gallery.php?id=69&gid=0.

[16] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3634–3646, 2011.

[17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[18] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE*

Transactions on Image Processing, vol. 20, no. 8, pp. 2378–2386, 2011.

[19] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "A non-reference image fusion metric based on mutual information of image features," *Computers & Electrical Engineering*, vol. 37, no. 5, pp. 744–756, 2011.

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

[21] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 657–667, 2013.