

Article

Two-Exposure Image Fusion Based on Optimized Adaptive Gamma Correction

Yan-Tsung Peng , He-Hao Liao  and Ching-Fu Chen

Department of Computer Science, National Chengchi University, Taipei City 116, Taiwan; 110753115@nccu.edu.tw (H.-H.L.); ecpoem@gmail.com (C.-F.C.)

* Correspondence: ytpeng@cs.nccu.edu.tw

Abstract: In contrast to conventional digital images, high-dynamic-range (HDR) images have a broader range of intensity between the darkest and brightest regions to capture more details in a scene. Such images are produced by fusing images with different exposure values (EVs) for the same scene. Most existing multi-scale exposure fusion (MEF) algorithms assume that the input images are multi-exposed with small EV intervals. However, thanks to emerging spatially multiplexed exposure technology that can capture an image pair of short and long exposure simultaneously, it is essential to deal with two-exposure image fusion. To bring out more well-exposed contents, we generate a more helpful intermediate virtual image for fusion using the proposed Optimized Adaptive Gamma Correction (OAGC) to have better contrast, saturation, and well-exposedness. Fusing the input images with the enhanced virtual image works well even though both inputs are underexposed or overexposed, which other state-of-the-art fusion methods could not handle. The experimental results show that our method performs favorably against other state-of-the-art image fusion methods in generating high-quality fusion results.

Keywords: two-exposure image fusion; gamma correction; high dynamic imaging



Citation: Peng, Y.-T.; Liao, H.-H.; Chen, C.-F. Two-Exposure Image Fusion Based on Optimized Adaptive Gamma Correction. *Sensors* **2022**, *22*, 24. <https://doi.org/10.3390/s22010024>

Academic Editors: Gianni D'Angelo and Arcangelo Castiglione

Received: 14 November 2021

Accepted: 19 December 2021

Published: 22 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Image fusion has been a crucial low-level image processing task for various applications, such as multi-spectrum image fusion [1,2], multi-focus image fusion [3], multi-modal image fusion [4], and multi-exposure image fusion [5]. Among these applications, thanks to smartphones' prevalence with their built-in cameras, multi-exposure image fusion is one of the most common applications. Since most natural scenes have a larger ratio of light to dark than what a single camera shot can capture, a single-shot image usually cannot present details of high dynamic ranges, thus having under- or overexposed parts for the scene. When a camera captures an image, its sensors can only catch a limited luminance range during a specific exposure time, resulting in a so-called low-dynamic-range image. An image taken for short exposure tends to be dark, while it is bright for long exposure, as shown in Figure 1a. Fusing differently exposed low-dynamic-range (LDR) images to obtain a high-dynamic-range (HDR) image requires extracting well-exposed (highlighted) regions from each LDR image to generate an excellent fused image, which has been very challenging.

Several research works have been performed for Multi-scale Exposure Fusion (MEF) [6–8]. In general, it is common to fuse LDR images using a weighted sum, where the weight associated with each input LDR is determined in a pixel-wise fashion [6–8]. Mertens et al. [6] proposed the fusion of images in a multi-scale manner based on pixel contrast, saturation, and well-exposedness to ease content inconsistency issues in the fused results. However, this often yields halo artifacts in its fusion results. In [7,8], the authors addressed the artifacts by applying modified guided image filtering to weight maps to eliminate halos around edges.

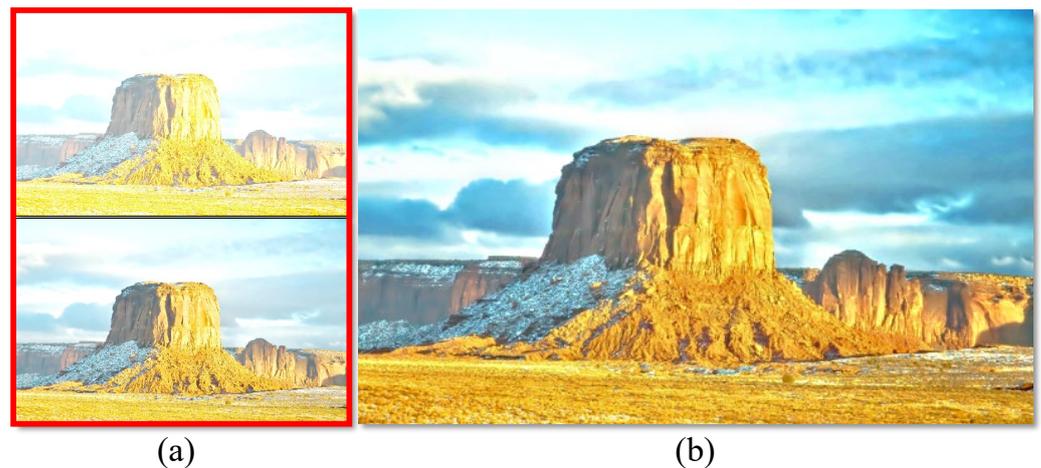


Figure 1. An example of the proposed two-exposure image fusion. (a) An input image pair with two exposures. (b) The fused image using the proposed method.

The abovementioned methods produce good results using a sequence of images exposed in a small interval of different exposure values (EV). Thanks to advanced sensor technology, a camera with Binned Multiplexed Exposure High-Dynamic-Range (BME-HDR) or Spatially Multiplexed Exposure High-Dynamic-Range (SME-HDR) technology can simultaneously capture an image pair with short- and long-exposure image sensors. The captured pair has only a negligible difference, possibly caused by local motion blur between them. The existing MEF methods may not work well with two exposure images, since none of the inputs may have well-exposed contents. In addition, weighted-sum fusion based on well-exposedness may not be able to deal with highlighted regions of a short-exposure image that are darker than dark parts in a long-exposure image, resulting in the method ignoring contents in the short-exposure image. Yang et al. [9] proposed the production of an intermediate virtual image with a medium exposure based on an image pair with two exposures to help generate better fusion results. Nevertheless, it does not work in situations where highlighted regions of both input LDR images are not well exposed.

In recent years, deep convolutional neural networks (CNNs) have gained tremendous success in low-level image processing works. In MEF, CNN-based methods [10,11] can better learn features from input multiple-exposure images and fuse them into a nice image. However, the fused images often lack image details [12], since spatial information may be lost when features pass through deep layers. Xu et al. [13] proposed a unified unsupervised image fusion network trained based on the importance and information carried by the two input images to generate fusion results. However, these learning-based methods can only produce a fused image based on the two input images' interpolation. They cannot deal with cases where both of the input images do not have highlighted regions/contents.

This paper presents a two-exposure fusion framework that generates a more helpful intermediate virtual image for fusion using the proposed Optimized Adaptive Gamma Correction (OAGC). The virtual image has better contrast, saturation, and well-exposedness, and it is not restricted to being an interpolated version of the two input images. Fusing the input images with their virtual image processed by OAGC works well even though both inputs have no well-exposed contents or regions. Figure 1b shows an example where the proposed framework can still generate a good fusion result for when both of the input images lack highlighted regions (Figure 1a). Our primary contributions are three-fold:

- Our image fusion framework adopting the proposed OAGC can produce better fusion results for two input images with various exposure ratios, even when both of the input images lack well-exposed regions.
- The proposed framework with OAGC can also adapt to single-image enhancement.

- We conduct an extensive experiment using a public multi-exposure dataset [14] to demonstrate that the proposed fusion framework performs favorably against the state-of-the-art image fusion methods.

2. Related Work

MEF-based methods produce fusion results using a weighted combination of the input images based on each pixel's "well-exposedness". In [15], fusion weight maps were calculated based on the correlation-based match and saliency measures of the input images. With the weight maps, one can fuse the input images into one by using the gradient pyramid.

Mertens et al. [6] constructed fusion weight maps based on contrast, saturation, and exposedness of the input images. Differently from [15], the fusion was performed with the Gaussian and Laplacian pyramids. The problem was that using the smoothed weight maps in fusion often causes halo artifacts, especially around the edges. The method proposed in [7] addressed this issue by applying an edge-preserving filter (weighted guided image filtering [16]) to fusion weight maps. Kou et al. [8] further proposed an edge-preserving gradient-domain guided image filter (GGIF) to avoid generating halo artifacts in the fused image. To extract image details, Li et al. [7] proposed a weighted structure tensor to manipulate details presented in a fused image. In general, MEF-based methods can generate decent fusion results.

General MEF algorithms [6,8] that require a sequence of images with different exposure ratios as the inputs may not work with only two input images. Yang et al. [9] proposed the use of the MEF algorithm for two-exposure-ratio image fusion, where an intermediate virtual image with a medium exposure is generated to help produce a better fusion result. However, the virtual image's intensity and exposedness are bounded by the two input images, which often fails to work for cases where two images are both underexposed and overexposed. Yang's method [9] can only generate both the intermediate and fusion results with approximate medium exposure between its two input images. The problem is that medium exposure between the inputs may still be under- or overexposure. Image fusion will not improve visual quality. We will discuss this issue more in the next section.

In the following paragraphs, we introduce the techniques adopted in the work of Yang et al., including the generation of the virtual image and fusion weights and the multi-scale image fusion. Before continuing, we define several notations that are used here. Let $\mathbf{I} \in \mathbb{R}^{M \times N \times 3}$ be a color image. We denote $\mathbf{I}^{(c)}$ as the color channel c , where $c \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}$ stand for the red, green, and blue channels. $I(m, n)$ represents the pixel located at (m, n) , where $0 \leq m < M$ and $0 \leq n < N$. M and N are the image width and height. Let \mathbf{Y} be the luminance component or the grayscale version of \mathbf{I} . Note that the values of images in this paper are normalized to $[0, 1]$.

2.1. Quality Measures and Fusion Weight Maps

In HDR imaging, an image taken at a certain exposure may contain underexposed or overexposed regions, which are less informative and should be assigned fewer weights in multi-exposure fusion. The input's contrast, saturation, and well-exposedness determine a pixel's weight at (m, n) [6]. The contrast of a pixel, denoted by $C(m, n)$, is obtained by applying a 3×3 Laplacian filter to a grayscale version of the image:

$$C(m, n) = |4Y(m, n) - Y(m-1, n) - Y(m, n-1) - Y(m+1, n) - Y(m, n+1)|. \quad (1)$$

Let $\mathbf{C} = (C(m, n))$ be the map of the contrast of \mathbf{I} ; therefore,

$$\mathbf{C} = |4\mathbf{Y} - \mathbf{Y}_l - \mathbf{Y}_r - \mathbf{Y}_u - \mathbf{Y}_d|,$$

where \mathbf{Y}_l , \mathbf{Y}_r , \mathbf{Y}_u , and \mathbf{Y}_d are obtained from \mathbf{I}_l , \mathbf{I}_r , \mathbf{I}_u , and \mathbf{I}_d ; i.e., shifting \mathbf{I} one pixel left, right, up, and down, respectively. The saturation of the pixel, denoted by $S(m, n)$, is obtained by computing the standard deviation across the red, green, and blue channels:

$$S(m, n) = \sqrt{\frac{1}{3} \sum_{c \in \{R, G, B\}} [I^{(c)}(m, n) - \bar{I}(m, n)]^2}, \quad (2)$$

where

$$\bar{I}(m, n) = \frac{1}{3} \sum_{c \in \{R, G, B\}} I^{(c)}(m, n).$$

The well-exposedness of the pixel, $E(m, n)$, is defined as:

$$E(m, n) = \exp \left[-\frac{1}{2\sigma^2} \sum_{c \in \{R, G, B\}} \left(I^{(c)}(m, n) - \zeta \right)^2 \right], \quad (3)$$

where $\sigma = 0.2$ and $\zeta = 0.5$. Essentially, E is a normal distribution centered at 0.5 with a standard deviation of 0.2. The maps of saturation and well-exposedness of \mathbf{I} can, respectively, be represented as $\mathbf{S} = (S(m, n))$ and $\mathbf{E} = (E(m, n))$. Next, the weight of the pixel for fusion is computed using:

$$\tilde{W}(m, n) = C^{\omega_c}(m, n) S^{\omega_s}(m, n) E^{\omega_e}(m, n), \quad (4)$$

where ω_c , ω_s , and ω_e can be adjusted to emphasize or ignore one or more measures. Considering a set of P images $\mathbf{I}_1, \dots, \mathbf{I}_P$ for image fusion, the weight of this pixel in the p_{th} image is normalized by the sum of the weights across all the images at the same pixel:

$$W_p(m, n) = \frac{\tilde{W}_p(m, n)}{\sum_{p'=1}^P \tilde{W}_{p'}(m, n)}. \quad (5)$$

The weight map of the image \mathbf{I}_p is represented as $\mathbf{W}_p = (W_p(m, n))$.

2.2. Multi-Scale Fusion

In the MEF algorithm [6], a fusion image, $\hat{\mathbf{I}}$, is obtained through multi-scale image fusion based on the standard Gaussian and Laplacian pyramids. For each input image \mathbf{I}_p in the set of $\{\mathbf{I}_p\}_{p=1}^P$, the Laplacian pyramid, $\mathcal{L}^{(l)}\{\mathbf{I}_p\}$, and the Gaussian pyramid of its weight map, $\mathcal{G}^{(l)}\{\mathbf{W}_p\}$, in the l_{th} level are constructed by applying the Gaussian pyramid generation [17]. In this level, the overall Laplacian pyramid is collapsed by performing weighted averaging on the Laplacian pyramids from all of the input images in the set:

$$\mathcal{L}^{(l)}\{\hat{\mathbf{I}}\} = \sum_{p=1}^P \left(\mathcal{G}^{(l)}\{\mathbf{W}_p\} \odot \mathcal{L}^{(l)}\{\mathbf{I}_p\} \right), \quad (6)$$

where \odot denotes element-wise multiplication. Finally, the fusion image, $\hat{\mathbf{I}}$, is reconstructed by collapsing the Laplacian pyramids $\mathcal{L}^{(l)}\{\hat{\mathbf{I}}\}$.

Applying edge-preserving filtering to preserve edges in the weight maps before averaging the Laplacian pyramids in Equation (6) can reduce halo artifacts in fused images. In [9], the GGIF [18] was adopted to smooth the weight maps \mathbf{W}_p and to preserve the significant change as well. Let $\Omega_\rho(m_0, n_0)$ be the square local patch with a radius of ρ centered at (m_0, n_0) , and let (m, n) be a pixel in the patch. In $\Omega_\rho(m_0, n_0)$, the weight map in the l_{th} level of the p_{th} image, $W_p^{(l)}(m, n)$, is the linear transform of the luminance component, $Y_p^{(l)}(m, n)$:

$$Y_p^{(l)}(m, n) = a_{p,(m_0,n_0)}^{(l)} W_p^{(l)}(m, n) + b_{p,(m_0,n_0)}^{(l)}, \quad (7)$$

where $a_{p,(m_0,n_0)}^{(l)}$ and $b_{p,(m_0,n_0)}^{(l)}$ are the coefficients and are assumed to be constant in $\Omega_\rho(m_0, n_0)$. $a_{p,(m_0,n_0)}^{(l)}$ and $b_{p,(m_0,n_0)}^{(l)}$ can be obtained by minimizing the objective function:

$$\Lambda_{GIF} = \sum_{\substack{m,n \in \\ \Omega_\rho(m_0,n_0)}} \left[\left(a_{p,(m_0,n_0)}^{(l)} W_p^{(l)}(m,n) + b_{p,(m_0,n_0)}^{(l)} - Y_p^{(l)}(m,n) \right)^2 - \epsilon \left(a_{p,(m_0,n_0)}^{(l)} \right)^2 \right], \quad (8)$$

where ϵ is a constant for regularization. The variance of the intensities within this local patch, $\sigma_{Y_p^{(l)},\Omega}^2$, is computed when solving for the coefficients in Equation (8).

In GGIF, a 3×3 local window, Ψ , is applied to the pixels within $\Omega_\rho(m_0, n_0)$ for capturing the structure within $\Omega_\rho(m_0, n_0)$ by computing the variance within Ψ , $\sigma_{Y_p^{(l)},\Psi}^2$ [18].

This local window makes GGIF a content-adaptive filter; thus, GGIF produces fewer halos and better preserves the edge than the GIF. In GGIF, the regularization term is designed to yield:

$$\Lambda_{GG} = \sum_{\substack{m,n \in \\ \Omega_\rho(m_0,n_0)}} \left[\left(a_{p,(m_0,n_0)}^{(l)} W_p^{(l)}(m,n) + b_{p,(m_0,n_0)}^{(l)} - Y_p^{(l)}(m,n) \right)^2 - \lambda \frac{\left(a_{p,(m_0,n_0)}^{(l)} - \zeta_{(m_0,n_0)} \right)^2}{\Gamma_{Y_k}(m_0, n_0)} \right], \quad (9)$$

where $\Gamma_{Y_k}(m_0, n_0)$ and $\zeta_{(m_0,n_0)}$ are computed according to the product of $\sigma_{Y_p^{(l)},\Omega}$ and $\sigma_{Y_p^{(l)},\Psi}$ (the standard deviations of the pixels within $\Omega_\rho(m_0, n_0)$ and Ψ), and λ is a constant for regularization. The filter coefficients $a_{p,(m_0,n_0)}^{(l)}$ and $b_{p,(m_0,n_0)}^{(l)}$ can be solved by minimizing Λ_{GG} in Equation (9).

The fused image $\hat{\mathbf{I}}$ can be obtained by fusing the Laplacian pyramids of the input images taken at different exposures using the weight maps retrieved from the Gaussian pyramids, $\mathcal{G}^{(l)}\{\mathbf{W}_p\}$. Note that the weight maps are filtered using GGIF, as described in Equation (9), to preserve edges.

2.3. Virtual Image Generation

In [9], Yang et al. proposed the modification of two differently exposed images to have the same medium exposure using the intensity mapping function based on the cross-histogram between two images, called the comparagram (Ref. [19]), and fused them to produce an intermediate virtual image. Let I_1 and I_2 be the two input images and let \mathcal{F}_{12} and \mathcal{F}_{21} be the intensity mapping functions (IMFs) that map I_1 to I_2 and I_2 to I_1 . Based on [19], the IMFs that map the two images to the same exposure, denoted as \mathcal{F}_{13} and \mathcal{F}_{23} , are computed as

$$\mathcal{F}_{13}(z)(I_i) = (z\mathcal{F}_{12}(z))^{0.5}, \quad \mathcal{F}_{23}(z) = (z\mathcal{F}_{21}(z))^{0.5}, \quad (10)$$

where z is a pixel intensity. The two modified images with the same exposure are $\tilde{I}_i = \mathcal{F}_{i3}(I_i)$, $i = 1, 2$. The desired virtual image I_v is computed by fusing \tilde{I}_1 and \tilde{I}_2 using the weighting functions adopted in [9]. The two-exposure-fusion image in [9] is obtained by fusing I_1 , I_2 , and I_v based on the MEF algorithm [8].

As described previously, Yang's method often fails to produce a satisfying fusion result when the medium exposure between inputs is still under- or overexposure. The proposed method addresses this issue by improving the contrast, saturation, and well-exposedness for the intermediate virtual image to generate better fusion results under different input conditions.

3. Proposed Method

The algorithm in [9] can work for two images with a large difference between their exposure ratios. In this case, the intermediate virtual image with medium exposure helps bridge the dynamic range gap between the two inputs. Thus, it can improve the quality of the fusion result. However, if the two inputs' exposure is under- or overexposure, the generated virtual image would not help fusion. Thus, the quality of the fused image is not improved much.

For example, to fuse Figure 2a,b, both of which look overexposed, the virtual image I_v (Figure 2c) generated by [9] with medium exposure between the inputs is still overexposed and, thus, not helpful for the fusion result (Figure 2e). We propose Optimized Adaptive Gamma Correction (OAGC) to enhance the intermediate virtual image to have better contrast, saturation, and well-exposedness (Figure 2d) so that it can improve the fusion quality and produce a better result (Figure 2f).



Figure 2. Comparison of intermediate virtual images. (a,b): Input images I_1 and I_2 ; (c,d) are the intermediate virtual images using [9] and our method; (e,f) are the fusion results using [9] and ours.

In OAGC, we derive an optimal γ based on the input's contrast, saturation, and well-exposedness by formulating an objective function based on these image quality metrics and apply it to the input image using gamma correction. Let $Y(m, n)$ be the luminance of a pixel. One can gamma-correct the image \mathbf{Y} to alter its luminance through the power function as follows:

$$Y_\gamma(m, n) = \eta Y(m, n)^\gamma, \quad (11)$$

where \mathbf{Y}_γ is the corrected image, η and γ are positive scalars, and η is usually set to 1 [20]. Here, the notation \mathbf{Y}_γ in bold represents the entire image, while $Y_\gamma(m, n)$ stands for the pixel located at (m, n) . If $\gamma < 1$, it stretches the contrast of shadow regions (pixel intensities less than the mid-tone of 0.5), and features in these regions become discernible, whereas if $\gamma > 1$, it stretches the contrast of bright regions (intensities larger than 0.5), and features in the regions become perceptible. For $\gamma = 1$, it is linear mapping.

To derive the optimal gamma, we design an objective function as follows:

$$f(\gamma) = q_1 \|\hat{\mathbf{c}}(\gamma)\|^2 + q_2 \|\hat{\mathbf{s}}(\gamma)\|^2 + q_3 \|\hat{\mathbf{e}}(\gamma)\|^2 + \delta \|\hat{\mathbf{r}}(\gamma)\|^2, \quad (12)$$

where $\hat{\mathbf{c}}(\gamma) = k_c \mathbf{1} - \text{vec}(\mathbf{C}_\gamma)$, $\hat{\mathbf{s}}(\gamma) = k_s \mathbf{1} - \text{vec}(\mathbf{S}_\gamma)$, $\hat{\mathbf{e}}(\gamma) = k_e \mathbf{1} - \text{vec}(\mathbf{E}_\gamma)$, and $\hat{\mathbf{r}}(\gamma) = \text{vec}(\mathbf{I}_v - \mathbf{I}_\gamma)$, where \mathbf{C}_γ , \mathbf{S}_γ , and \mathbf{E}_γ are the maps of quality measures computed based on the gamma-corrected version of the input image, denoted as \mathbf{I}_γ . Here, the virtual image I_v is used as the input, which is $\mathbf{I}_\gamma := \mathbf{I}_v^\gamma$. We set k_c , k_s , and k_e to 4, 0.5, and 1 according to the upper bounds of the corresponding quality measures (contrast, saturation, and well-exposedness; refer to the Appendix A for the derivation). The term with $\hat{\mathbf{r}}(\gamma)$ in the objective function prevents the corrected image from deviating the input too much. Hence, minimizing the objective function $f(\gamma)$ is to maximize all three quality measures: the contrast, saturation, and well-exposedness. q_1 , q_2 , and q_3 are the weighting factors for the contributions from different quality measures (independent from ω_c , ω_s , and ω_e in Equation (4) and are all set to $\frac{1}{3}$. δ is a small, fixed scalar and is set to 0.1 in the present study. $\mathbf{1}$ is the vector of 1s, $\text{vec}(\cdot)$ is the vectorization of a matrix, and $\|\cdot\|$ represents the 2-norm of a vector. The regularization term is added to avoid possible color distortion caused by gamma correction.

The optimal gamma, γ^* , which aims to increase contrast, saturation, and well-exposedness simultaneously, can be obtained by minimizing the optimization function $f(\gamma)$:

$$\gamma^* = \arg \min_{\gamma} f(\gamma). \quad (13)$$

Since there is no closed-form solution for Equation (13), we apply the gradient descent to iteratively approximate it:

$$\begin{aligned} \gamma^{(k+1)} &= \gamma^{(k)} + \alpha^{(k)} \frac{d}{d\gamma} f(\gamma) \\ &= \gamma^{(k)} + \alpha^{(k)} \left[q_1 \left(\frac{d}{d\gamma} \hat{\mathbf{c}}(\gamma) \right)^{\top} \hat{\mathbf{c}}(\gamma) + q_2 \left(\frac{d}{d\gamma} \hat{\mathbf{s}}(\gamma) \right)^{\top} \hat{\mathbf{s}}(\gamma) \right. \\ &\quad \left. + q_3 \left(\frac{d}{d\gamma} \hat{\mathbf{e}}(\gamma) \right)^{\top} \hat{\mathbf{e}}(\gamma) + \delta \left(\frac{d}{d\gamma} \hat{\mathbf{r}}(\gamma) \right)^{\top} \hat{\mathbf{r}}(\gamma) \right], \end{aligned} \quad (14)$$

where

$$\begin{aligned} \frac{d}{d\gamma} \hat{\mathbf{c}}(\gamma) &= -\text{sgn} \left[\sum_{c \in \{\text{R,G,B}\}} t^{(c)} \left(4\mathbf{i}_v^{(c)\gamma} - \mathbf{i}_{v,l}^{(c)\gamma} - \mathbf{i}_{v,r}^{(c)\gamma} - \mathbf{i}_{v,u}^{(c)\gamma} - \mathbf{i}_{v,d}^{(c)\gamma} \right) \right] \\ &\times \left[\sum_{c \in \{\text{R,G,B}\}} t^{(c)} \left(4\mathbf{i}_v^{(c)\gamma} \odot \log \mathbf{i}_v - \mathbf{i}_{v,l}^{(c)\gamma} \odot \log \mathbf{i}_{v,l} - \mathbf{i}_{v,r}^{(c)\gamma} \odot \log \mathbf{i}_{v,r} \right. \right. \\ &\quad \left. \left. - \mathbf{i}_{v,u}^{(c)\gamma} \odot \log \mathbf{i}_{v,u} - \mathbf{i}_{v,d}^{(c)\gamma} \odot \log \mathbf{i}_{v,d} \right) \right], \end{aligned} \quad (15)$$

with $\mathbf{i}_{v,l}$, $\mathbf{i}_{v,r}$, $\mathbf{i}_{v,u}$, and $\mathbf{i}_{v,d}$ being the vectorization of $\mathbf{I}_{v,l}$, $\mathbf{I}_{v,r}$, $\mathbf{I}_{v,u}$, and $\mathbf{I}_{v,d}$, as well as $t^{(\text{R})}$, $t^{(\text{G})}$, and $t^{(\text{B})}$ being 0.299, 0.587, and 0.114 respectively.

$$\begin{aligned} \frac{d}{d\gamma} \hat{\mathbf{s}}(\gamma) &= -\frac{1}{9} \left[2 \sum_{c \in \{\text{R,G,B}\}} \left(\mathbf{i}_v^{(c)(2\gamma)} \odot \log \mathbf{i}_v^{(c)} \right) \right. \\ &\quad - \mathbf{i}_v^{(\text{R})\gamma} \odot \mathbf{i}_v^{(\text{G})\gamma} \odot \log \left(\mathbf{i}_v^{(\text{R})\gamma} + \mathbf{i}_v^{(\text{G})\gamma} \right) \\ &\quad - \mathbf{i}_v^{(\text{G})\gamma} \odot \mathbf{i}_v^{(\text{B})\gamma} \odot \log \left(\mathbf{i}_v^{(\text{G})\gamma} + \mathbf{i}_v^{(\text{B})\gamma} \right) \\ &\quad \left. - \mathbf{i}_v^{(\text{B})\gamma} \odot \mathbf{i}_v^{(\text{R})\gamma} \odot \log \left(\mathbf{i}_v^{(\text{B})\gamma} + \mathbf{i}_v^{(\text{R})\gamma} \right) \right] \oslash \mathbf{s}_{\gamma}, \end{aligned} \quad (16)$$

with \oslash being the element-wise division,

$$\frac{d}{d\gamma} \hat{\mathbf{e}}(\gamma) = \frac{\mathbf{e}_{\gamma}}{2\sigma^2} \odot \sum_{c \in \{\text{R,G,B}\}} \left[\left(2\mathbf{i}_v^{(c)(2\gamma)} - \mathbf{i}_v^{(c)\gamma} \right) \odot \log \mathbf{i}_v^{(c)} \right], \quad (17)$$

$$\frac{d}{d\gamma} \hat{\mathbf{r}}(\gamma) = \sum_{c \in \{\text{R,G,B}\}} \left[\left(\mathbf{i}_v^{(c)(2\gamma)} - \mathbf{i}_v^{(c)(\gamma+1)} \right) \odot \log \mathbf{i}_v^{(c)} \right], \quad (18)$$

and $\alpha^{(k)}$ is the adjustable learning rate.

Figure 3 shows the flowchart of the presented two-exposure image fusion framework, where the two inputs are taken in the same scene at different exposure ratios. The virtual image is first generated using the intensity mapping function [9]. Next, we solve Equation (12) to find the optimal gamma value γ^* for the virtual image, which enhances the contrast, saturation, and well-exposedness of \mathbf{I}_v . The final fused image, $\hat{\mathbf{I}}$, is obtained by applying the MEF algorithm [8,9] to the fusion of two input images and \mathbf{I}_{γ} .

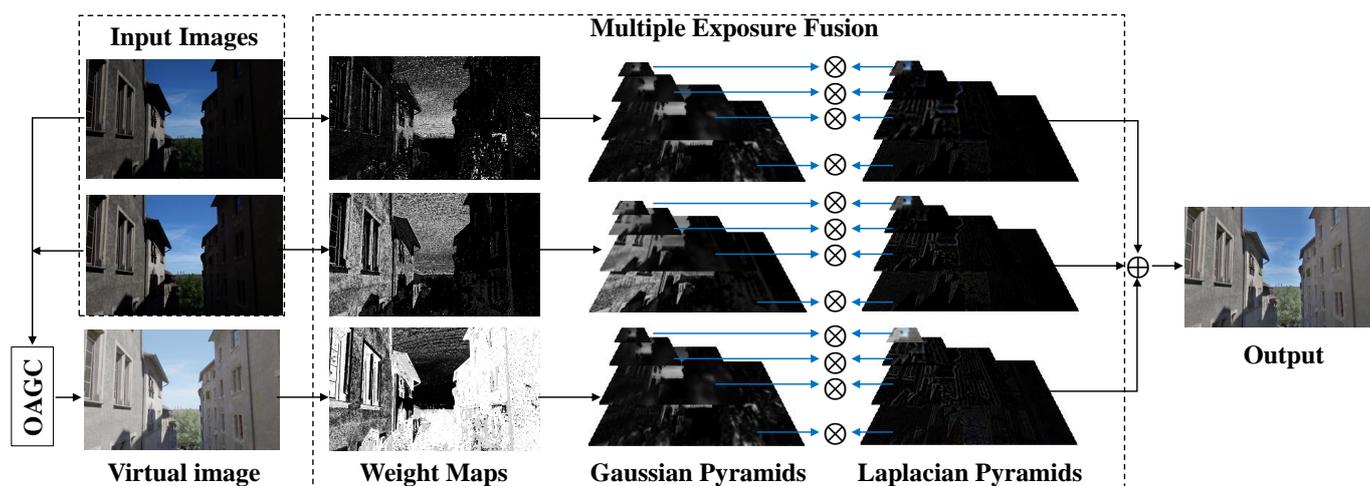


Figure 3. Flowchart of the proposed method. Note that the images of Gaussian and Laplacian pyramids are contrast-enhanced for display.

4. Experimental Results

In the experiment, we compared the proposed method against state-of-the-art image fusion methods, which included Kou’s method [8], DeepFuse [10], Yang’s method [9], and U2Fusion [13]. We adopted the SICE dataset [14] and collected 116 image pairs that consisted of various scenes to evaluate the performance. The presented algorithm was implemented using MATLAB R2019b on a MacBook Pro with an Intel i5 dual-core processor at 2.7 GHz and 8 GB 1867 MHz DDR3 RAM. We present a performance evaluation with a qualitative visual comparison and a quantitative objective assessment in the following.

4.1. Qualitative Assessment

We compared different fusion results under various input conditions. First, Figure 4 shows the fusion results of using the compared image fusion algorithms [8–10,13] and our presented framework. As can be seen, one input image is underexposed, and the other is overexposed in the two cases, where the fusion results should have middle exposure between the two inputs. All of the compared methods worked fine in such cases, although U2fusion’s [13] fusion results were a little darker than the others’ results.



Figure 4. Comparison of the results obtained using different fusion methods with an underexposed and an overexposed input. (a,b) show the input images squared in red. The fusion results were obtained using (c) Kou’s method [8], (d) DeepFuse [10], (e) Yang’s method [9], (f) U2Fusion [13], and (g) the proposed method.

Figure 5 shows image fusion cases where the difference between the two input images’ EVs was not large. Thus, fusion methods that can only produce results with medium exposure between the inputs do not work. As shown, all of the compared methods except for the proposed framework output fusion results similar to the input images, and were thus unable to reveal more details than the inputs. In contrast, the proposed framework produced an intermediate virtual image enhanced by OAGC with additional well-exposed

highlighted contents and generated better fusion results. Therefore, we can further improve the overall image visibility by revealing details in regions that are too dark or bright.

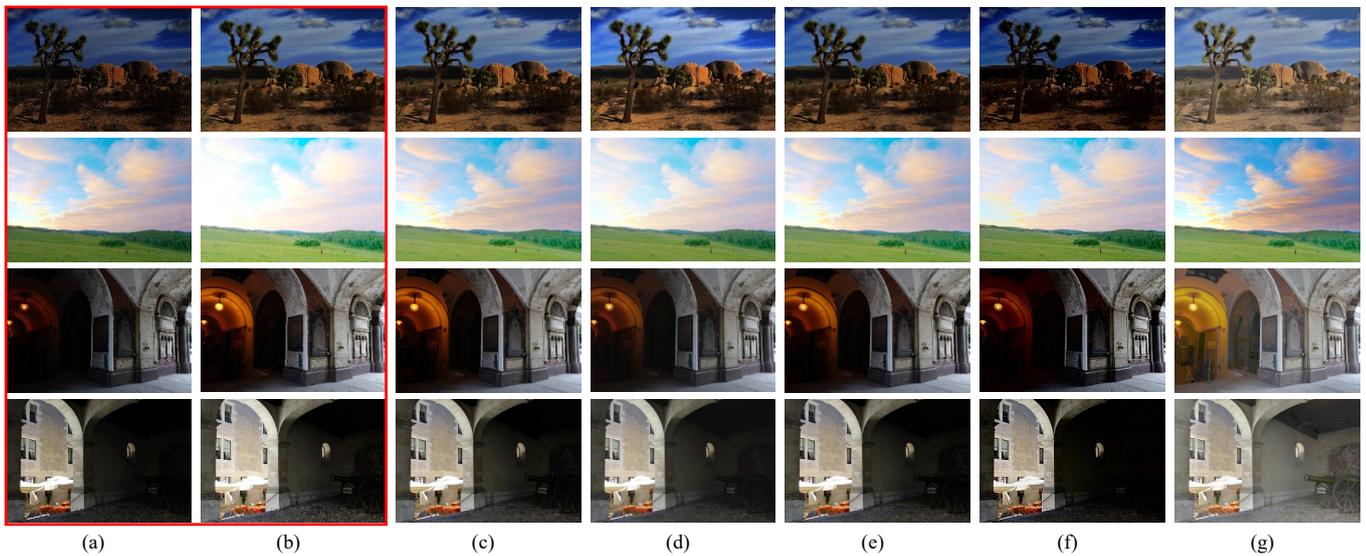


Figure 5. Comparisons of the fusion results using different algorithms, where the two input images had smaller exposure differences. (a,b) show the input images squared in red. The fusion results were obtained using (c) Kou’s method [8], (d) DeepFuse [10], (e) Yang’s method [9], (f) U2Fusion [13], and (g) the proposed method.

4.2. Quantitative Assessment

Objectively, we compare the performance of our presented framework against other image fusion methods using five benchmark metrics: the Naturalness Image Quality Evaluator (NIQE) [21], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [22], No-Reference Image Quality Assessment (NR-IQA) [23], and discrete entropy (DE) [24]. The NIQE [21] is a no-reference image quality metric that is trained on pristine images without subjective scores from humans. Therefore, it can measure image quality degradation if any distortions exist, but is correlated little with human perceptuality. A smaller value means a better quality. BRISQUE [22] is a natural scene statistics-based distortion-generic no-reference image quality assessment model that is trained on images with known distortions and subjective quality scores. It can evaluate losses of naturalness of an image caused by possible distortions. A BRISQUE value ranges between 0 and 100. A smaller value means worse visual quality. NR-IQA [23] is another no-reference image quality metric for HDR images that is constructed using deep CNNs while considering image saliency, and it constructs deep CNNs to extract quality features across the HDR and LDR domains. DE [24] can represent the information contained in an information source, i.e., if an image has higher entropy, it contains more information. Consequently, it is often used to measure the richness of image details. It is defined as:

$$DE(I) = - \sum_{l=0}^{L-1} p_I(l) \log(p_I(l)), \quad (19)$$

where I is a grayscale image, L represents the largest pixel intensity value, and $p_I(l)$ is the probability density function of a given grayscale intensity l .

Table 1 shows the quantitative performance of different fusion methods, where the scores are averaged over all of the test images. As can be seen, the results demonstrate that the presented framework achieved the best scores in all four categories, meaning that our fusion results looked natural with the fewest distortions (having the lowest NIQE [21] and lowest BRISQUE values [22]). In assessing the HDR image quality (NR-IQA [23]),

our method performed favorably against other fusion methods. Our method could also preserve the most image details in the fusion results (largest DE value [24]).

Table 1. Quantitative comparisons of different fusion methods. The scores are averaged over all of the test images. The values in bold represent the best scores.

	NIQE ↓	BRISQE ↓	NR-IQA ↑	DE ↑
Kou [8]	2.7866	22.7140	43.6202	6.9171
DeepFuse [10]	2.9555	31.5365	32.9696	6.9564
Yang [9]	2.7895	23.0660	43.5312	6.8617
U2Fusion [13]	3.421	35.5939	34.5449	6.4195
Proposed	2.7339	21.7120	43.8968	6.9814

4.3. Extension to Single Image Enhancement

As stated previously, the proposed framework works well for two-exposure image fusion in cases where the difference between the two input images' EVs varies. In recent years, fusion-based single image enhancement methods have attracted much attention [14]. We can also extend our framework to single image enhancement by applying OAGC to the input image, I , to yield a quality-improved image I_γ . Then, both I and I_γ are fused to obtain an enhancement result.

Figure 6 compares the results obtained using various single image enhancement methods, including global histogram equalization (HE) [25], CVC [26], AGCWD [27], EPMP [28], SICE [14], and the proposed method. As shown, the conventional HE tended to over-enhance/introduce noise to the processed images (Figure 6b), since the input images had over- and underexposed regions. SICE [14] only performed well for the second row of Figure 6, where the input image was underexposed. For the other cases, it tended to overexpose the input images. The other methods [26–28] could only enhance the contrast of the input images, while the proposed framework not only did that, but also revealed unseen details from the input and increased the color vividness (Figure 6g).

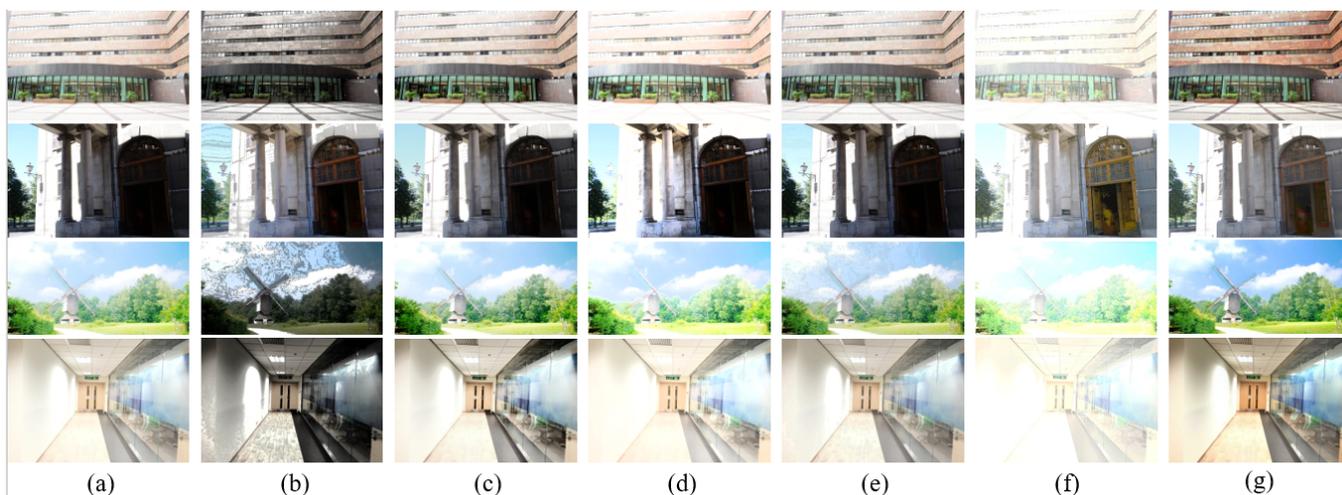


Figure 6. Comparison of single image enhancement results using different algorithms. (a) Input image; the enhanced results obtained using (b) HE [25], (c) CVC [26], (d) AGCWD [27], (e) EPMP [28], (f) SICE [14], and (g) the proposed method (fusing I and I_γ).

4.4. Analysis of OAGC

Convergence of Gradient Descent: To further analyze the process of attaining the target gamma value γ^* in OAGC, we take the case in the top row of Figure 5 as an example to show the iterative steps of finding γ^* for the intermediate virtual image I_v . Figure 7 shows that it takes about 66 steps for the objective function $f(\gamma)$ to converge with gradient descent,

and it attains $\gamma^* = 0.434$. The value of the objective function changes from the initial 2.690 to 2.688. As $\gamma = \gamma^*$, $\hat{\mathbf{e}}(\gamma)$ reaches its minimum, and $\hat{\mathbf{c}}(\gamma)$ is close to its minimum while $\hat{\mathbf{s}}(\gamma)$ is at its maximum, indicating that contrast, saturation, and well-exposedness are all maximized. This also shows that solving the objective function strikes a balance among these three measures.

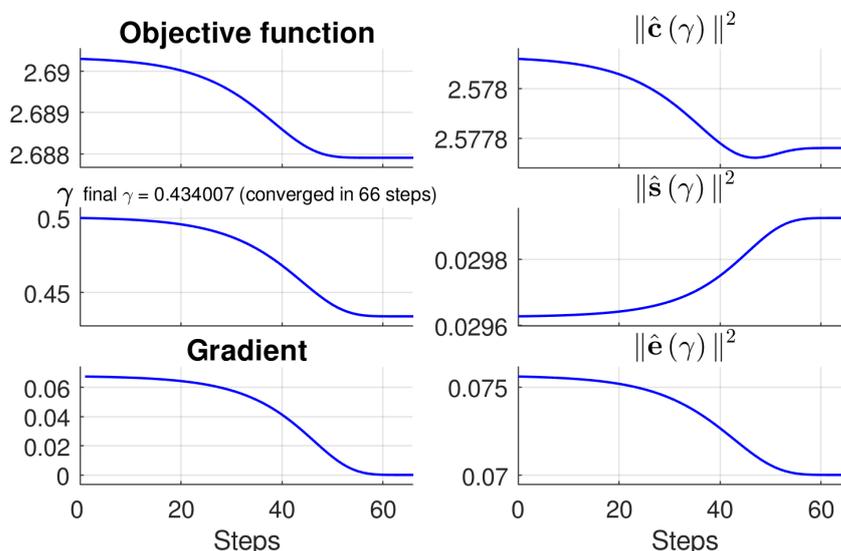


Figure 7. The progressive process of attaining γ^* for an intermediate virtual image using Equation (12) (taking I_v in the top row of Figure 5 as an example).

To further attest to the effectiveness of OAGC, Figure 8 shows the trend of values of the objective function and its quality measure terms using the grid-search method on γ . As seen, the minimum of the objective function is 2.688 when $\gamma = 0.434$, consistently with the γ^* obtained using gradient descent.

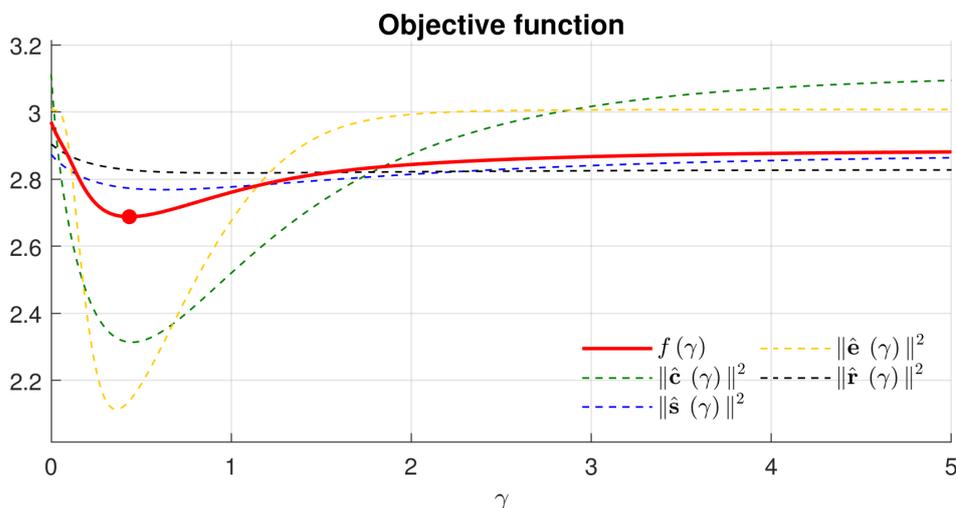


Figure 8. The grid search of γ^* (taking I_v in the top row of Figure 5 as an example).

Limitation of OAGC: Using OAGC, we can attain a gamma coefficient γ^* from the input image by optimizing the objective function in Equation (12), and we can then apply gamma correction to the input to generate a corrected image whose contrast, saturation, and well-exposedness are improved. However, if both of the input images have no content at all for the same regions due to extremely low or high exposure, even OAGC cannot help generate or restore those regions from nothing. Figure 9 shows a failure case of OAGC, where both of the input images are very underexposed and bear little content. Figure 9c shows the intermediate virtual images obtained using the intensity mapping algorithm

described in [9], which is similar to the interpolation of the inputs, and the result is still very dark and lacks content. After applying OAGC to it ($\gamma^* = 0.2015$), the processed virtual image (Figure 9d) presents more details than before. Still, it inevitably has noise in some regions (such as the door and the banner on the façade).



Figure 9. Limitation of OAGC. (a,b) Input images. (c) Virtual image I_v using [9]; (d) virtual image processed using OAGC I_γ .

5. Conclusions

This paper presented a two-exposure image fusion framework that utilizes the proposed OAGC to bring out additional well-exposed contents from an intermediate virtual image derived from the two inputs. It can work better for the input images with various combinations of exposure ratios and can produce more well-exposed fusion results. In addition, the proposed framework with OAGC can easily adapt to single image enhancement. The experimental results have demonstrated that the proposed method performs favorably against the state-of-the-art image fusion methods.

Author Contributions: Conceptualization, Y.-T.P.; methodology, Y.-T.P.; software, Y.-T.P. and C.-F.C.; validation, H.-H.L. and C.-F.C.; formal analysis, Y.-T.P.; investigation, Y.-T.P. and C.-F.C.; resources, Y.-T.P.; data curation, C.-F.C.; writing—original draft preparation, Y.-T.P. and C.-F.C.; writing—review and editing, Y.-T.P.; visualization, Y.-T.P. and C.-F.C.; supervision, Y.-T.P.; project administration, Y.-T.P.; funding acquisition, Y.-T.P. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was supported in part by the Ministry of Science and Technology, Taiwan under Grants MOST 110-2221-E-004-010, 110-2622-E-004-001, 109-2622-E-004-002, 110-2634-F-019-001, 110-2221-E-019-062, 110-2622-E-019-006, and also in part by Qualcomm Technologies, Inc. through a Taiwan University Research Collaboration Project under Grant NAT-414673.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Upper Bounds of the Quality Measures

Since the luminance of the pixel is $Y(m, n) \in [0, 1]$, the upper bound of its contrast value $C(m, n)$ is:

$$C(m, n) = |4Y(m, n) - Y(m-1, n) - Y(m, n-1) - Y(m+1, n) - Y(m, n+1)| \leq 4. \quad (\text{A1})$$

For exposedness, $E(m, n)$ is shown as:

$$E(m, n) = \exp \left[-\frac{1}{2\sigma^2} \sum_{c \in \{R, G, B\}} \left(I^{(c)}(m, n) - \xi \right)^2 \right] \leq 1, \quad (\text{A2})$$

where $E(m, n) = 1$ when $I^{(c)}(m, n) = \xi$.

The saturation $S(m, n)$ is defined as in Equation (2). Let $\bar{I}(m, n)$ be the mean of all the channels of this pixel; i.e.,

$$\bar{I}(m, n) = \frac{1}{3} \sum_{c \in \{R, G, B\}} I^{(c)}(m, n).$$

The upper bound of $S(m, n)$ can be obtained from:

$$\begin{aligned} S^2(m, n) &= \frac{1}{3} \sum_{c \in \{R, G, B\}} \left[I^{(c)}(m, n) - \bar{I}(m, n) \right]^2 \\ &= \frac{1}{3} \left[\sum_{c \in \{R, G, B\}} I^{(c)}(m, n)^2 - 2\bar{I}(m, n) \sum_{c \in \{R, G, B\}} I^{(c)}(m, n) + 3\bar{I}(m, n)^2 \right] \\ &= \frac{1}{3} \left[\sum_{c \in \{R, G, B\}} I^{(c)}(m, n)^2 - 3\bar{I}(m, n)^2 \right]. \end{aligned}$$

Because $I^{(c)}(m, n) \in [0, 1]$,

$$\sum_{c \in \{R, G, B\}} I^{(c)}(m, n)^2 \leq \sum_{c \in \{R, G, B\}} I^{(c)}(m, n) = 3\bar{I}(m, n).$$

This indicates that

$$\begin{aligned} S(m, n)^2 &\leq \bar{I}(m, n) - \bar{I}(m, n)^2 \\ &= -(\bar{I}(m, n) - 0.5)^2 + 0.25 \leq 0.25. \end{aligned}$$

Therefore, the upper bound of $S(m, n) = 0.5$.

References

- Ma, Y.; Chen, J.; Chen, C.; Fan, F.; Ma, J. Infrared and visible image fusion using total variation model. *Neurocomputing* **2016**, *202*, 12–19. [[CrossRef](#)]
- Meng, F.; Guo, B.; Song, M.; Zhang, X. Image fusion with saliency map and interest points. *Neurocomputing* **2016**, *177*, 1–8. [[CrossRef](#)]
- Yin, H.; Li, Y.; Chai, Y.; Liu, Z.; Zhu, Z. A novel sparse-representation-based multi-focus image fusion approach. *Neurocomputing* **2016**, *216*, 216–229. [[CrossRef](#)]
- Du, J.; Li, W.; Lu, K.; Xiao, B. An overview of multi-modal medical image fusion. *Neurocomputing* **2016**, *215*, 3–20. [[CrossRef](#)]
- Kou, F.; Wei, Z.; Chen, W.; Wu, X.; Wen, C.; Li, Z. Intelligent Detail Enhancement for Exposure Fusion. *IEEE Trans. Multimedia* **2018**, *20*, 484–495. [[CrossRef](#)]
- Mertens, T.; Kautz, J.; Reeth, F.V. Exposure Fusion. In Proceedings of the 15th Pacific Conference on Computer Graphics and Applications (PG'07), Maui, HI, USA, 29 October–2 November 2007; pp. 382–390.
- Li, Z.; Wei, Z.; Wen, C.; Zheng, J. Detail-Enhanced Multi-Scale Exposure Fusion. *IEEE Trans. Image Process.* **2017**, *26*, 1243–1252. [[CrossRef](#)] [[PubMed](#)]
- Kou, F.; Li, Z.; Wen, C.; Chen, W. Multi-scale exposure fusion via gradient domain guided image filtering. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 10–14 July 2017; pp. 1105–1110.
- Yang, Y.; Cao, W.; Wu, S.; Li, Z. Multi-Scale Fusion of Two Large-Exposure-Ratio Images. *Signal Process. Lett.* **2018**, *25*, 1885–1889. [[CrossRef](#)]
- Prabhakar, K.R.; Srikanth, V.S.; Babu, R.V. DeepFuse: A Deep Unsupervised Approach for Exposure Fusion with Extreme Exposure Image Pairs. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4714–4722.
- Prabhakar, K.R.; Arora, R.; Swaminathan, A.; Singh, K.P.; Babu, R.V. A fast, scalable, and reliable deghosting method for extreme exposure fusion. In Proceedings of the IEEE International Conference on Computational Photography (ICCP), Seoul, Korea, 27 October–2 November 2019; pp. 1–8.
- Yin, J.L.; Chen, B.H.; Peng, Y.T.; Tsai, C.C. Deep Prior Guided Network For High-Quality Image Fusion. In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME), London, UK, 6–10 July 2020; pp. 1–6.
- Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2Fusion: A unified unsupervised image fusion network. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 502–518. [[CrossRef](#)] [[PubMed](#)]
- Cai, J.; Gu, S.; Zhang, L. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. *IEEE Trans. Image Process.* **2018**, *27*, 2049–2062. [[CrossRef](#)] [[PubMed](#)]
- Burt, P.J.; Kolczynski, R.J. Enhanced image capture through fusion. In Proceedings of the 1993 (4th) International Conference on Computer Vision, Berlin, Germany, 11–14 May 1993; pp. 173–182.
- Li, Z.; Zheng, J.; Zhu, Z.; Yao, W.; Wu, S. Weighted guided image filtering. *IEEE Trans. Image Process.* **2014**, *24*, 120–129. [[PubMed](#)]
- Burt, P.; Adelson, E. The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.* **1983**, *31*, 532–540. [[CrossRef](#)]

18. Kou, F.; Chen, W.; Wen, C.; Li, Z. Gradient domain guided image filtering. *IEEE Trans. Image Process.* **2015**, *24*, 4528–4539. [[CrossRef](#)] [[PubMed](#)]
19. Grossberg, M.D.; Nayar, S.K. Determining the camera response from images: What is knowable? *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1455–1467. [[CrossRef](#)]
20. Gonzalez, R.C.; Woods, R.E. Image processing. In *Digital Image Processing*; Prentice Hall: Hoboken, NJ, USA, 2007.
21. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “completely blind” image quality analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212. [[CrossRef](#)]
22. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *20*, 209–212. [[CrossRef](#)] [[PubMed](#)]
23. Jia, S.; Zhang, Y.; Agrafiotis, D.; Bull, D. Blind high dynamic range image quality assessment using deep learning. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 765–769.
24. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
25. Hummel, R. Image enhancement by histogram transformation. *Comput. Graph. Image Process.* **1975**, *6*, 184–195. [[CrossRef](#)]
26. Celik, T.; Tjahjadi, T. Contextual and variational contrast enhancement. *IEEE Trans. Image Process.* **2011**, *20*, 3431–3441. [[CrossRef](#)] [[PubMed](#)]
27. Huang, S.C.; Cheng, F.C.; Chiu, Y.S. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Trans. Image Process.* **2012**, *22*, 1032–1041. [[CrossRef](#)] [[PubMed](#)]
28. Chen, B.H.; Wu, Y.L.; Shi, L.F. A fast image contrast enhancement algorithm using entropy-preserving mapping prior. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *29*, 38–49. [[CrossRef](#)]