Human Visual Perception-Based Multi-Exposure Fusion Image Quality Assessment

Yueli Cui, Aihua Chen *, Benquan Yang, Shiqing Zhang and Yang Wang

School of Electronic and Information Engineering, Taizhou University, Taizhou 318017, China; cuiyueli@tzc.edu.cn (Y.C.); yangbq266@sina.com (B.Y.); tzczsq@163.com (S.Z.); wangyang@tzc.edu.cn (Y.W.)

* Correspondence: chen_aihua1980@163.com

Received: 19 November 2019; Accepted: 6 December 2019; Published: 9 December 2019

Abstract: Compared with ordinary single exposure images, multi-exposure fusion (MEF) images are prone to color imbalance, detail information loss and abnormal exposure in the process of combining multiple images with different exposure levels. In this paper, we proposed a human visual perception-based multi-exposure fusion image quality assessment method by considering the related perceptual features (i.e., color, dense scale invariant feature transform (DSIFT) and exposure) to measure the quality degradation accurately, which is closely related to the symmetry principle in human eyes. Firstly, the L1 norm of chrominance components between fused images and the designed pseudo images with the most severe color attenuation is calculated to measure the global color degradation, and the color saturation similarity is added to eliminate the influence of color over-saturation. Secondly, a set of distorted images under different exposure levels with strong edge information of fused image is constructed through the structural transfer, thus DSIFT similarity and DSIFT saturation are computed to measure the local detail loss and enhancement, respectively. Thirdly, Gauss exposure function is used to detect the over-exposure or under-exposure areas, and the above perceptual features are aggregated with random forest to predict the final quality of fused image. Experimental results on a public MEF subjective assessment database show the superiority of the proposed method with the state-of-the-art image quality assessment models.

Keywords: multi-exposure image quality assessment; color saturation; dense scale invariant feature transform (DSIFT); guided filtering; perceptual symmetry principle

1. Introduction

Natural scenes usually have a wide brightness range from $10^{-5}$ cd/m$^2$ to $10^8$ cd/m$^2$, but it is difficult for the existing imaging devices to acquire all parts of scene information at the single exposure situation due to the limitation of its own dynamic range [1]. Multi-exposure fusion (MEF), as an effective quality enhancement technology, is able to integrate multiple low dynamic range (LDR) images under different exposure levels captured by the normal cameras into a perceptually attractive image and has been successfully applied in various multimedia fields, such as remote sensing, medical imaging, panoramic imaging and HDTV [2,3].

Generally, the performance differences between several MEF algorithms are mainly reflected in the solving process of fusion weights. The simplest local and global energy weighting algorithms obtain weights by measuring the local and global energy among source images. Mertens et al. [4] constructed the weights by considering contrast, saturation and good exposure, and fused multiple images by the multi-scale pyramid model. On this basis, Li et al. [5] made the fused image more realistic subjectively by solving the quadratic optimization problem to enhance the detail information. Gu et al. [6] extracted gradient information from the structural tensor of source images to design the initial weights and smoothed them with edge-preserving filter to prevent the artifacts. Remarkably, different kinds of
filters (i.e., bilateral filter [7], recursive filter [8] and guided filter [9]) all solve the problem of spatial consistency from several aspects but produce inconsistent visual experience, respectively. In brief, the above-mentioned MEF algorithms cannot guarantee the achievement of perfect quality of fused images, which are mainly reflected in color, detail and exposure. Hence, in order to further improve the existing MEF algorithms, it is urgent to develop the corresponding MEF image quality assessment (MEF-IQA) metrics in accordance with the human visual system (HVS).

Up to now, a number of IQA methods had been developed for image coding application, which can be divided into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR) [10]. The FR methods are guided by a distortion-free reference image. The RR methods require a part of reference image information, while the NR methods do not. Evidently, due to the particularity in the process of imaging for MEF images, MEF-IQA can be regarded as the special FR-IQA method with multiple reference images. Therefore, the main challenges in developing the effective MEF-IQA models are how to accurately obtain the real reference information from source image sequences with different exposure levels. However, most existing IQA models in the field of image fusion are suitable for general image fusion, not dedicated to MEF. Hossny et al. [11] predicted the image quality by measuring the mutual information between source images and fused image on different scales. Cvejic et al. [12] utilized the Tsallis entropy to form the quality metric. Inspired by the edge information, Xydeas et al. [13] evaluated the fused image by exploring the edge preservation from source images. Wang et al. [14] and Zheng et al. [15] extracted the multi-directional gradient feature in wavelet domain and spatial domain, respectively. Considering that HVS is highly sensitive to structural degradation with salient objects, Piella et al. [16] combined the visual saliency with structural similarity (SSIM) index to predict the quality of fused image. Unfortunately, most above-mentioned general-purpose fusion IQA (GF-IQA) metrics will be invalid on the condition of more than two source images.

In recent years, with the further exploration in the field of MEF, several perceptual IQA metrics specialized for MEF images have been proposed one after another. Ma et al. [17] constructed the first MEF subjective assessment database, including 17 multi-exposure image sequences and the corresponding fused images generated by 8 classical MEF algorithms. In addition, an objective MEF-IQA method based on SSIM was presented by measuring small-scale structural consistency and large-scale luminance consistency. Xing et al. [18] proposed a multi-scale contrast-based model by integrating the structure similarity and saturation similarity into contrast feature. Taking into account the importance of color for MEF images, Deng et al. [19] combined numerous perceptual features such as color, texture and structure, and carried out the quality regression by extreme learning machine. Although these existing MEF-IQA metrics can more accurately evaluate the performance of different MEF algorithms compared with GF-IQA metrics, the robustness of models in several extreme conditions (e.g., over-saturated color, detail enhancement and under-exposure) cannot be guaranteed. It largely depends on the specific theory that the excessive hue or detail information may appear the unnecessary artifacts in some fused images, and under-exposed areas of images will not only result in information loss but generate unnatural shadows. These terrible quality degradation phenomena will make the perceptual symmetry in human eyes unbalanced.

To overcome these shortcomings, an effective human visual perception-based multi-exposure fusion image quality assessment method is proposed in this paper. Unlike the existing GF-IQA and MEF-IQA metrics, the presented perceptual model thoroughly simulates human visual physiology, including the impact of color, detail and exposure on MEF images. Its main contributions are described as follows:

1. The difference of chrominance components between fused images and the defined pseudo images with the most severe color attenuation is calculated to measure the global color degradation, and the color saturation similarity is added to eliminate the influence of over-saturated color.

2. A set of distorted source images with strong edge information of fused image is constructed by the structural transfer characteristic of guided filter; thus, structure similarity and structure saturation are computed to measure the local detail loss and enhancement, respectively.
(3) The Gauss function is designed to accurately detect the over-exposed or under-exposed areas of images; then, the local luminance of each source images and the global luminance of fused image are used to measure the luminance consistency between them.

The remaining of this paper is organized as follows: The proposed MEF-IQA method is investigated in Section 2. The performance comparison between the proposed method and the state-of-the-art ones is described in Section 3. Finally, the conclusion and future work are drawn in Section 4.

2. Proposed Human Visual Perception-Based MEF-IQA Method

It is generally acknowledged that MEF images with ideal quality are rich in color, detail information and exposure, so single perceptual feature is inadequate for accurately evaluating MEF images. In this paper, a human visual perception-based MEF-IQA method is proposed, which mainly consists of color, detail and good exposure metrics, and Figure 1 shows its flowchart. In the first stage, local saturation similarity and global color distortion metric are designed for detecting the unbalanced chrominance of fused images. In the second stage, dense scale invariant feature transform (DSIFT) descriptor is adopted to obtain the local structure information along different directions for each pixel in the images, and DSIFT similarity and DSIFT saturation between the source images and the pseudo reference images with fused image’s strong edge information are calculated at different scales to measure the distortion such as detail loss and detail enhancement. In the third stage, local exposure similarity and global exposure metric are presented by combining the luminance of source images and fused image with Gauss function, respectively. The specific implementation details of the proposed MEF-IQA method are stated in the following four subsections.

![Figure 1. The proposed human visual perception-based multi-exposure fusion–image quality assessment (MEF-IQA) method.](image)

2.1. Local and Global Color Metrics

Since HVS is more sensitive to luminance than chrominance, most existing GF-IQA and MEF-IQA models focus on grayscale information, while ignoring the importance of color. Unfortunately, there are distinct color differences between the MEF images formed by different algorithms, and Figure 2 depicts an example of MEF images of sequence “Lamp1” from MEF database. From Figure 2, it can be seen that the image in Figure 2a has a bright and vivid color, while the MEF images in Figure 2b,c are very dim, causing a terrible visual experience. Therefore, the quality assessment for the colorful MEF
images will be more reliable, and we adopt the global and local ways to evaluate the color distortion of MEF images in this section.

![Figure 2. An example of multi-exposure fusion images of sequence “Lamp1” generated by three MEF algorithms. (a) Fused image created by Li’s algorithm [5]; (b) fused image created by Gu’s algorithm [6]; (c) fused image created by Raman’s algorithm [7].](image)

2.1. Local and Global Color Metrics

Since HVS is more sensitive to luminance than chrominance, most existing GF-IQA and MEF-IQA models focus on grayscale information, while ignoring the importance of color. Unfortunately, there are distinct color differences between the MEF images formed by different algorithms, and Figure 2b,c are very dim, causing a terrible visual experience. Therefore, the quality assessment for the colorful MEF images will be more reliable, and we adopt the global and local ways to evaluate the color distortion of MEF images in this section.

2.1.1. Global Color Distortion Metric

To extract the chrominance components of MEF images, the RGB color space is first transformed into YCbCr color space that is more in line with human visual characteristics, which is expressed as

\[
\begin{bmatrix}
Y_f \\
C_{b_f} \\
C_{r_f}
\end{bmatrix}
= \begin{bmatrix}
0.299 & 0.587 & 0.114 \\
-0.169 & -0.331 & 0.500 \\
0.500 & -0.419 & -0.081
\end{bmatrix}
\begin{bmatrix}
R_f \\
G_f \\
B_f
\end{bmatrix}
+ \begin{bmatrix}
0 \\
128 \\
128
\end{bmatrix}
\] (1)

where \(Y_f\) is the luminance information, and \(C_{b_f}\) and \(C_{r_f}\) are the two chrominance components.

Evidently, the inverse process from YCbCr color space to RGB space according to Equation (1) can be deduced as

\[
\begin{align*}
R_f &= Y_f + 1.402(C_{r_f} - 128) \\
G_f &= Y_f - 0.344(C_{b_f} - 128) - 0.714(C_{r_f} - 128) \\
B_f &= Y_f + 1.772(C_{b_f} - 128)
\end{align*}
\] (2)

where \(R_f\), \(G_f\) and \(B_f\) are the R, G and B channels of fused image, respectively. From Equation (2), it can be found that the farther away the value of \(C_{b_f}\) and \(C_{r_f}\) are from 128, the more colorful subjectively the fused image will be.

Hence, we utilize the L1 norm between the chrominance components and 128 to approximately measure the global color distortion of fused image, which is calculated by

\[
|C_{Cb}^G, C_{Cr}^G| = \frac{1}{N} \left( \|C_{b_f} - 128\|_1, \|C_{r_f} - 128\|_1 \right)
\] (3)

where \(N\) is the number of pixels for each chrominance component, \(\| \cdot \|_1\) is the L1 norm operator, and \(C_{Cb}^G\) and \(C_{Cr}^G\) are the two global color distortion metrics for \(C_{b_f}\) and \(C_{r_f}\), respectively. However, the presented \(C_{Cb}^G\) and \(C_{Cr}^G\) will be worthless in the case of over-saturated color, so it is necessary to eliminate this negative effect.

2.1.2. Local Saturation Similarity

According to the related visual psychology research, color saturation can accurately indicate the natural response of HVS to color information. Generally, pixels with high saturation have more vivid
color, while pixels with low saturation are dim. Moreover, color saturation $S$ can be simply measured by calculating the standard deviation in the RGB space, which is defined as

$$S = \sqrt{\frac{(R - \mu)^2 + (G - \mu)^2 + (B - \mu)^2}{3}}$$ (4)

where $\mu$ is the mean value of $R$, $G$ and $B$ channels. Thus, the color saturation maps for each source image $I_k$ and fused image $I_f$ can be computed by Equation (4), which are denoted as $\{S_k\}(k = 1, 2, \ldots, K)$ and $S_f$, respectively.

Then, a maximum color saturation map $S_{\text{max}}$ can be calculated from $\{S_k\}$ as the pseudo reference image with the optimal chrominance and is expressed as

$$S_{\text{max}}(p) = \max(S_1(p), S_2(p), \ldots, S_k(p)), \quad k = 1, 2, \ldots, K$$ (5)

where $S_k(p)$ is the maximum saturation at the position of $p$ for the $k$-th source image, $\max(\cdot)$ is the “select max” operation, and $K$ is the number of source images.

Finally, the local color distortion is evaluated by calculating the similarity between $S_{\text{max}}$ and $S_f$, thereby eliminating the impact caused by over-saturated color, and the local saturation similarity $C_{\text{SIM}}^L$ is defined as

$$C_{\text{SIM}}^L = \text{mean}\left(\frac{2S_{\text{max}}(p) \cdot S_f(p) + c_1}{S_{\text{max}}(p)^2 \cdot S_f(p)^2 + c_1}\right)$$ (6)

where $\text{mean}(\cdot)$ is the mean operator, and $c_1$ is a constant to control the denominator not to be zero.

2.2. Structure Similarity and Saturation Metric

The structural information of images usually carries the essential visual contents of scenes, and HVS is highly adaptable to extract structures for visual perception. Moreover, DSIFT descriptor [20], as an effective means for obtaining local gradient information of each pixel in eight directions, has been successfully applied in the field of computer vision, such as image registration and image fusion. Compared with gradient magnitude, DSIFT is more accurate and robust to extract the structural information of images. Figure 3a–c shows three fused images of sequence “Tower” created by Mertens’ algorithm [4], local energy weighting and Li’s algorithm [5], respectively. From Figure 3, we can have the following observations: Figure 3a cannot preserve the fine details in the center of tower and the brightest cloud region; Figure 3b produces unnatural artifacts near the edges of sky or tower and also known as pseudo contour; Figure 3c can be regard as the ones after performing the edge enhancement operation on Figure 3a. Generally, the detail enhancement algorithm may create more appealing results perceptually, but it also introduces some unnatural shadow into fused image. Therefore, in this section, DSIFT similarity and DSIFT saturation are designed for precisely detecting the edge distortion in the fused image, which is mainly reflected in three aspects, i.e., detail loss, pseudo contour and detail enhancement.
I with the dimension of \( r \) where \( I \)

2.2.1. DSIFT Similarity

As an effective edge-preserving filter, guided filter \( f_G(\cdot) \) is determined by the input image \( I \) and guided image \( I_G \), and the specific filtering process is defined as

\[
I_o = f_G(I, I_G, r, \varepsilon)
\]

where \( I_o \) is the filtering output, and \( r \) and \( \varepsilon \) are the filtering radius and regularization parameter, respectively. In addition, when \( I_G \) is different from \( I_o \), guided filter is equivalent to the structure transfer filter, that is, \( I_o \) will retain the strong edge information of \( I_G \), and the size of \( r \) and \( \varepsilon \) limits the strength of the retained edge information.

According to the characteristic, we first choose the multi-exposure source images \( \{I_k\} \) as the filtering input, and the fused image \( I_f \) is used as the guided image. Thus, a set of pseudo multi-exposure images with the strong edge information of fused image are constructed and denoted as \( \{I_k^{s}\} \) \((k = 1, 2, \ldots, K) \). At the same time, in order to eliminate the influence caused by filtering, \( I_k \) is selected as the input image and guided image once again to generate a set of filtered multi-exposure source images \( \{I_k^{s}\} \) \((k = 1, 2, \ldots, K) \). The above filtering process can be expressed as

\[
\begin{align*}
I_k^{s} &= f_G(I_k, I_k, r_s, \varepsilon_s) \\
I_k^{d,s} &= f_G(I_k, I_k, r_s, \varepsilon_s)
\end{align*}
\]

where \( r_s \) and \( \varepsilon_s \) are set as the small values to guarantee that the constructed pseudo multi-exposure images retain all edge information of fused image.

Then, DSIFT descriptor is applied to each pixel point in \( I_k^{s} \) and \( I_k^{d,s} \), and the related DSIFT feature with the dimension of \( M \) is extracted, which can be defined as

\[
\begin{align*}
D_k^{s} &= f_D(I_k^{s}) \\
D_k^{d,s} &= f_D(I_k^{d,s})
\end{align*}
\]
where $D^{r,s}_{k,1:M}$ and $D^{d,s}_{k,1:M}$ are the obtained DSIFT feature of the $k$-th source image and pseudo image after filtering, respectively. $f_D(\cdot)$ is the operator for calculating DSIFT feature, and $M$ is the dimension of feature.

Finally, DSIFT feature similarity between $I^{r,s}_{k}$ and $I^{d,s}_{k}$ is calculated to measure the detail loss or pseudo contour in the fused image, which is expressed as

$$D^{k,1:M}_{SIM} = \frac{2D^{r,s}_{k,1:M} \cdot D^{d,s}_{k,1:M} + c_2}{D^{r,s}_{k,1:M}^2 + D^{d,s}_{k,1:M}^2 + c_2}$$

where $D^{k,1:M}_{SIM}$ is the obtained $k$-th DSIFT similarity map, and $c_2$ is a constant to control the denominator not to be zero.

Since $D^{k,1:M}_{SIM}$ symbolizes the edge distortion areas of fused image in different gradient orientations, the final $k$-th DSIFT similarity map $D^{k}_{SIM}$ can be computed by the simple average operation.

$$D^{k}_{SIM} = \frac{1}{M} \sum_{m=1}^{M} D^{k,m}_{SIM}$$

where $m$ is the dimension index of the $k$-th DSIFT similarity map.

Furthermore, we define the Gauss exposure weighting function $f^{E}_{s}(\cdot)$ to integrate all DSIFT similarity maps under different exposure levels, and it is expressed as

$$f^{E}_{s} = \exp\left(-\frac{(L^s_k - 0.5)^2}{2\sigma^2}\right)$$

$$D_{SIM} = \sum_{k=1}^{K} D^{k}_{SIM} \cdot f^{E}_{s}$$

where $L^s_k$ is a set of images after smoothing source images $I_k$ by the mean filter with the window size of $7 \times 7$, and it is mainly based on the spatial consistency principle. Specially, DSIFT similarity map is calculated from $I^{r,s}_{k}$ and $I^{d,s}_{k}$ with smaller smoothness level, thus the corresponding weight should be also smooth in space. $D_{SIM}$ is the aggregated DSIFT similarity map, and the final DSIFT similarity feature is simply calculated by the mean operation for $D_{SIM}$.

Figure 4a–c depicts the corresponding DSIFT similarity maps of Figure 3a–c, respectively. From Figure 4, it can be found that the edge distortion areas in Figure 3a,b (e.g., the center of tower, the brightest cloud and the sky) are well reflected in the calculated quality maps. However, Figure 4c shows the false results at the position of enhanced pixels in Figure 3c, because detail enhancement usually makes the edge information of images change strongly. Actually, an appropriate detail enhancement algorithm will produce more attractive results subjectively, which is just opposite of the impact caused by DSIFT similarity. Therefore, it is necessary to add the DSIFT saturation on the basic of DSIFT similarity to eliminate this effect.
2.2.2. DSIFT Saturation

Similar to the calculation of DSIFT similarity, a set of filtered pseudo images $I_k^{d,b}$ and source images $I_k^{r,b}$ can be generated by guided filter, and the filtering process is expressed as

$$
I_k^{r,b} = f_C(I_k, I_{k'}, r_b, \epsilon_b) \\
I_k^{d,b} = f_C(I_k, I_{k'}, r_b, \epsilon_b)
$$

(14)

where $r_b$ and $\epsilon_b$ are set as the large values to guarantee that the constructed pseudo multi-exposure images only retain some large intensity edge information, i.e., detail enhancement area.

Then, DSIFT descriptor is also applied to each pixel point in $I_k^{r,b}$ and $I_k^{d,b}$, and the related DSIFT feature with the dimension of $M$ is extracted as follows

$$
D_{k,1:M}^{r,b} = f_D(I_k^{r,b}) \\
D_{k,1:M}^{d,b} = f_D(I_k^{d,b})
$$

(15)

where $D_{k,1:M}^{r,b}$ and $D_{k,1:M}^{d,b}$ are the obtained DSIFT feature of the $k$-th source image and pseudo image after filtering, respectively.

Finally, DSIFT saturation between $I_k^{r,b}$ and $I_k^{d,b}$ is calculated to measure the detail enhancement, which is expressed as

$$
D_{SA}^{k,1:M} = \frac{4}{\pi} \arctan\left(\frac{D_{k,1:M}^{d,b} + c_3}{D_{k,1:M}^{r,b} + c_3}\right)
$$

(16)

where $D_{SA}^{k,1:M}$ is the obtained $k$-th DSIFT saturation map, $\arctan$ is the inverse tangent function, and $c_3$ is a constant to control the denominator not to be zero. Evidently, when $D_{k,1:M}^{d,b}$ is greater than $D_{k,1:M}^{r,b}$, the saturation is greater than 1, indicating the high visual perception of HVS on detail enhancement areas. Similarly, the $k$-th DSIFT saturation map $D_{SA}^{k}$ can be obtained by averaging the feature on each dimension, which is expressed as

$$
D_{SA}^{k} = \frac{1}{M} \sum_{m=1}^{M} D_{SA}^{k,m}
$$

(17)
Moreover, we also use the Gauss exposure weighting function $f_E^b(\cdot)$ to integrate all DSIFT saturation maps under different exposure levels, and it is denoted as

$$f_E^b = \exp\left(-\frac{(L^b_k - 0.5)^2}{2\sigma^2}\right)$$

(18)

$$D_{SA} = \sum_{k=1}^{K} D_{SA}^k \cdot f_E^b$$

(19)

where $L^b_k$ is a set of images after smoothing source images $I_k$ by the mean filter with the window size of $15 \times 15$ according to the spatial consistency principle. $D_{SA}$ is the aggregated DSIFT saturation map, and the final DSIFT saturation feature is simply calculated by the mean operation for $D_{SA}$.

Figure 5a–c shows the corresponding DSIFT saturation maps of Figure 3a–c, respectively. From Figure 5, it can be seen that Figure 5c has higher saturation at the position of pixels with enhanced detail information in Figure 3c, which eliminates the undesired effects caused by DSIFT similarity. Fortunately, the detail loss areas in Figure 3c are also reflected on Figure 5c, demonstrating the validity of the proposed DSIFT saturation metric.

2.3. Local and Global Exposure Metrics

There are usually under-exposed or over-exposed areas in MEF images, which is mainly caused by the luminance inconsistency among adjacent pixels, thus resulting in a poor visual experience. In this section, we constructed the local and global exposure metrics to evaluate these luminance distortion phenomena.

2.3.1. Local Exposure Similarity

Similar to the definition of Gauss exposure weighting function in Equations (12) and (18), the local exposure map of an image can be defined by measuring the distance between normalized pixel intensity
and 0.5, i.e., when the pixel intensity is close to 0 or 1, the pixel point is considered as under-exposed or over-exposed. Therefore, local exposure maps $E_b^k$ for each source image are calculated by

$$E_b^k = \exp\left(-\frac{(I_k - 0.5)^2}{2\sigma^2}\right)$$

(20)

Then, the best exposure areas in each source image are selected to form a good exposure reference image $I_{br}^f$, which is defined as

$$I_{br}^f = I_k \text{ if } E_b^k = \max(E_b^1, E_b^2, \ldots, E_b^k)$$

(21)

Finally, the local exposure distortion areas can be detected by calculating the similarity between $I_{br}^f$ and $I_f$.

$$E_{SIM}^b = \frac{2I_{br}^f \cdot I_f + c_4}{(I_{br}^f)^2 + (I_f)^2 + c_4}$$

(22)

where $c_4$ is a constant to control the denominator not to be zero, and the final local exposure similarity feature is simply calculated by the mean operation for $E_{SIM}^b$.

Figure 6a–c depicts three fused images of sequence “Candle” created by different algorithms, and Figure 6d–f are the local exposure similarity maps of Figure 6a–c, respectively. Obviously, there are several under-exposed areas (e.g., the teacup in Figure 6a and the shadows in Figure 6c) that are inconsistent with surrounding areas in terms of luminance. Such exposure distortion is clearly indicated in the corresponding quality maps. Moreover, Figure 6b has the uniform luminance distribution, but it is darker on the whole space than Figure 6a,c, which still results in the bad visual perception. Therefore, it is essential to consider the impact by the overall luminance.

Figure 6. Examples of local exposure similarity maps. (a) Fused image created by Li’s algorithm [5]. (b) fused image created by Raman’s algorithm [7]. (c) fused image created by Li’s algorithm [9]. (d–f) Corresponding local exposure similarity maps of images in the first row.
2.3.2. Global Exposure Metric

Similar to the local exposure metric, we combine the average luminance of fused image with Gauss function to design the global exposure metric $E_g^m$, and it can be expressed as

$$E_g^m = \exp\left( -\frac{\left( I_f - 0.5 \right)^2}{2\sigma^2} \right)$$  \hspace{1cm} (23)

where $I_f$ is average luminance of fused image. When the luminance value is close to 0 or 1, the fused image looks entirely dark or bright.

2.4. Quality Prediction

In addition, multi-scale characteristics in spatial domain can acquire the image content from the fine level to the coarse level, which is consistent with the processing mechanism of low-level retina and cortex in primate visual system. As illustrated in Figure 7, the original scale of multi-exposure images and fused images are marked as scale 1. By iteratively applying a low-pass filter and a down-sampling operation with a factor of 2 on the original images, the filtered images at scale 1 can be obtained after $l$-1 iterations. Then, the above-mentioned feature extraction method is conducted in the multi-scale space, thus generating the final feature vector $F_l = [F_1, F_2, F_3]$, where $F_1$, $F_2$, and $F_3$ are color, structure and exposure features, respectively. After feature extraction, the quality regression from feature space to image quality is conducted, which can be denoted as

$$Q = f_Q(F_l)$$  \hspace{1cm} (24)

where $f_Q(\cdot)$ is a quality regression function achieved by random forest (RF) algorithm, and $Q$ is the quality of fused image.

![Figure 7. The multi-scale scheme used in the proposed method, where the LPF denotes a low-pass filter and 2↓ means the down-sampling operation with a factor of 2.](image)

3. Experimental Results

3.1. Experimental Settings

3.1.1. Database

To compare the performance of the proposed MEF-IQA method with other state-of-the-art IQA models, the experiments were performed on the public MEF subjective assessment database [21] provided by Waterloo IVC. Specially, it consists of 17 multi-exposure source image sequences, and each image sequence contains the corresponding fused images generated by 8 MEF algorithms. Therefore, a total of 136 fused images with associated mean opinion score (MOS) are included in this database, and more details can be referred to Figure 8 and Table 1.
3.1.2. Evaluation Criteria

According to the related standard in the field of image quality assessment formulated by video quality expert group (VQEG) [22], three evaluation criteria, i.e., Pearson linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SROCC) and root mean square error (RMSE), are selected to evaluate the performance of IQA models, and the most excellent model will be generated when the values of PLCC and SROCC are 1, RMSE is 0. Moreover, a 5-parametric logistic regression process is employed to make the predicted quality closer to subjective scores before calculating PLCC and RMSE.

3.1.3. Experimental Parameters

In terms of the proposed MEF-IQA method, there are several experimental parameters to be fixed in the process of feature extraction. Specifically, the filtering radius and regularization parameter of guided filter (i.e., $r_s$, $\varepsilon_s$, $r_b$ and $\varepsilon_b$) are mainly applied for the structure transfer, so we set $r_s = 11$, $\varepsilon_s = 10^{-6}$, $r_b = 21$ and $\varepsilon_b = 0.3$, respectively, which are in accordance with the advice in [9]. Moreover, the feature dimension $M$ of DSIFT descriptor is the same as the ones in [20], that is, $M = 32$. Generally, the above parameters cannot make a significant impact on the final performance of MEF-IQA models, so we strictly follow the recommendation in the previous study rather than setting the parameters.

Figure 8. Source image sequences contained in the MEF database [21]. Each image sequence is represented by one image, which is a fused image with the best quality in the subjective test.

Table 1. Information about source image sequence in the MEF database [21].

| No. | Source Sequences | Size           | Image Source           |
|-----|------------------|----------------|------------------------|
| 1   | Balloons         | $339 \times 512 \times 9$ | Erik Reinhard          |
| 2   | Belgium house    | $512 \times 384 \times 9$ | Dani Lischinski        |
| 3   | Lamp1            | $512 \times 384 \times 15$ | Martin Cadik           |
| 4   | Candle           | $512 \times 364 \times 10$ | HDR Projects           |
| 5   | Cave             | $512 \times 384 \times 4$  | Bartlomiej Okonek      |
| 6   | Chinese garden   | $512 \times 340 \times 3$  | Bartlomiej Okonek      |
| 7   | Farmhouse        | $512 \times 341 \times 3$  | HDR Projects           |
| 8   | House            | $512 \times 340 \times 4$  | Tom Mertens            |
| 9   | Kluki            | $512 \times 341 \times 3$  | Bartlomiej Okonek      |
| 10  | Lamp2            | $512 \times 342 \times 6$  | HDR Projects           |
| 11  | Landscape        | $512 \times 341 \times 3$  | HDRsoft                |
| 12  | Lighthouse       | $512 \times 340 \times 3$  | HDRsoft                |
| 13  | Madison capitol  | $512 \times 384 \times 30$ | Chaman Singh Verma     |
| 14  | Memorial         | $341 \times 512 \times 16$ | Paul Debevec           |
| 15  | Office           | $512 \times 340 \times 6$  | Matlab                 |
| 16  | Tower            | $341 \times 512 \times 3$  | Jacques Joffre         |
| 17  | Venice           | $512 \times 341 \times 3$  | HDRsoft                |
randomly. Finally, considering that feature extraction is conducted in the multi-scale space, and the scale number, \( l \), will affect the performance evidently. Therefore, we select the optimal scale value (i.e., \( l = 3 \)) to achieve a trade-off between complexity and accuracy, and the specific details about analyzing the impact of the scale number on performance will be discussed in the following sections.

### 3.2. Performance Comparison

To verify the performance of the proposed MEF-IQA method, we compare it on MEF database [21] with nine existing state-of-the-art IQA metrics, including six GF-IQA metrics [11–16] and three MEF-IQA metrics [17–19]. Remarkably, the proposed MEF-IQA method adopts the supervised learning approach to obtain the image quality, thus the MEF database is first divided into training and testing subsets. Then, 17-fold cross validation is used to evaluate the performance of model, that is, for each train-test stage, 16 distorted image sequences and the rest ones are used for training and testing, respectively. The results of performance comparison for each source image sequence are tabulated in Tables 2 and 3, and only the values of PLCC and SROCC are presented for brevity, where the best two performances are highlighted in boldface. Furthermore, we simultaneously record the corresponding hit count of performance highlighted in boldface to discriminate the performance difference among ten IQA metrics more intuitively. From Tables 2 and 3, we can have the following findings. First, compared with six GF-IQA metrics, three MEF-IQA metrics specially designed for multi-exposure images achieve a more outstanding performance, which indicates that MEF images have the special perceptual characteristics due to the distortion introduced in the imaging process, such as color imbalance, structure degradation and abnormal exposure. Second, the related color distortion metric is extra considered in [19], which makes the performance of MEF-IQA metric in [19] slightly improved against the other two existing MEF-IQA metrics [17,18]. It is mainly due to the fact that although HVS is more sensitive to luminance than chrominance in real situation, an image with serious color distortion also causes the bad visual experience. Therefore, color information cannot be ignored in terms of MEF-IQA. Finally, the proposed MEF-IQA method achieves 0.952 and 0.897 on PLCC and SROCC on average, respectively. Obviously, it outperforms all the competing MEF-IQA and GF-IQA metrics, because it considers three perceptual factors, namely color, structure and exposure.

### Table 2. Pearson linear correlation coefficient (PLCC) performance evaluation of ten IQA models.

| No. | GF-IQA | MEF-IQA |
|-----|--------|---------|
|     | [11]  | [12]    |
| 1   | −0.542 | 0.761   |
| 2   | −0.385 | 0.174   |
| 3   | −0.121 | −0.479  |
| 4   | 0.265  | −0.729  |
| 5   | −0.214 | 0.053   |
| 6   | −0.224 | −0.294  |
| 7   | −0.641 | 0.504   |
| 8   | −0.289 | −0.524  |
| 9   | −0.091 | 0.021   |
| 10  | −0.387 | 0.621   |
| 11  | −0.211 | 0.359   |
| 12  | −0.296 | −0.261  |
| 13  | −0.406 | 0.031   |
| 14  | −0.418 | 0.445   |
| 15  | −0.203 | 0.302   |
| 16  | −0.478 | −0.116  |
| 17  | −0.358 | −0.022  |
| Average | −0.294 | 0.060   |
| Hit count | 0     | 0       |


Table 3. Spearman rank-order correlation coefficient (SROCC) performance evaluation of ten IQA models.

| No. | GF-IQA | MEF-IQA |
|-----|--------|---------|
|     | [11]   | [12]    | [13]   | [14]   | [15]   | [16]   | [17]   | [18]   | [19]   | Proposed |
| 1   | −0.429 | 0.714   | 0.667  | 0.500  | 0.595  | 0.452  | 0.833  | 0.952  | 0.935  | 0.929    |
| 2   | −0.299 | 0.000   | 0.779  | 0.755  | 0.539  | 0.467  | 0.970  | 0.958  | 0.934  | 0.934    |
| 3   | −0.071 | −0.381  | 0.786  | 0.619  | 0.476  | 0.405  | 0.976  | 1.000  | 0.954  | 0.976    |
| 4   | 0.357  | −0.667  | 0.976  | 0.786  | 0.167  | 0.548  | 0.927  | 0.952  | 0.927  | 0.905    |
| 5   | −0.119 | 0.024   | 0.714  | 0.810  | 0.643  | 0.571  | 0.833  | 0.619  | 0.851  | 0.762    |
| 6   | −0.214 | −0.286  | 0.691  | 0.786  | 0.548  | 0.524  | 0.929  | 0.762  | 0.946  | 0.881    |
| 7   | −0.452 | 0.500   | 0.738  | 0.810  | 0.500  | 0.286  | 0.929  | 0.810  | 0.883  | 0.952    |
| 8   | −0.048 | −0.691  | 0.595  | 0.452  | 0.524  | 0.405  | 0.857  | 0.905  | 0.909  | 0.976    |
| 9   | −0.238 | 0.167   | 0.262  | 0.286  | 0.048  | 0.119  | 0.786  | 0.905  | 0.867  | 0.929    |
| 10  | −0.429 | 0.833   | 0.762  | 0.619  | 0.691  | 0.548  | 0.714  | 0.905  | 0.844  | 0.714    |
| 11  | −0.738 | 0.548   | 0.024  | 0.405  | 0.143  | 0.143  | 0.524  | 0.881  | 0.760  | 0.810    |
| 12  | −0.833 | −0.429  | 0.500  | 0.429  | 0.381  | 0.071  | 0.881  | 0.691  | 0.815  | 0.881    |
| 13  | −0.214 | 0.310   | 0.524  | 0.357  | 0.524  | 0.476  | 0.881  | 0.881  | 0.955  | 0.881    |
| 14  | 0.000  | 0.810   | 0.762  | 0.548  | 0.524  | 0.667  | 0.857  | 0.857  | 0.907  | 0.857    |
| 15  | −0.193 | 0.084   | 0.277  | 0.398  | 0.386  | 0.458  | 0.785  | 0.988  | 0.907  | 0.988    |
| 16  | −0.476 | −0.214  | 0.571  | 0.524  | 0.595  | 0.571  | 0.952  | 0.929  | 0.941  | 0.857    |
| 17  | −0.335 | 0.299   | 0.910  | 0.731  | 0.563  | 0.311  | 0.934  | 0.934  | 0.893  | 0.934    |
|     | −0.278 | 0.059   | 0.620  | 0.577  | 0.461  | 0.413  | 0.857  | 0.878  | 0.896  | 0.897    |

Hit count 0 0 1 0 0 0 7 8 9 11

3.3. Impacts of Multi-Scale Scheme and Different Feature

To analyze how much of the contributions coming from each kind of feature in the proposed MEF-IQA model, the evaluation performance resulted from three perceptual attributes (i.e., color, structure and exposure) are investigated on the MEF database [21]. The corresponding results averaged on all source image sequences in the database are reported in Table 4, where $F_1$, $F_2$ and $F_3$ denote the extracted feature vector about color, structure and exposure, respectively. From Table 4, it can be found that the structure feature has a more significant influence on the final performance compared with the color and exposure features, which demonstrates that HVS is highly sensitive to the structure degradation in an image. Moreover, we also explore the impact of the scale number on the final performance in the multi-scale scheme. Specially, we assign $l = \{1, 2, 3, 4\}$ and calculate the PLCC, SROCC and RMSE for each situation, respectively, and the results are shown in Table 5. Evidently, when the value of $l$ is 3, the proposed MEF-IQA model achieves an optimal performance, so we set the scale number as 3 in the paper to guarantee the accuracy of method.

Table 4. Performance resulted from each perceptual attribute of the proposed. MEF-IQA method averaged on the MEF database [21].

| Feature Category | PLCC | SROCC | RMSE |
|------------------|------|-------|------|
| $F_1$            | 0.711| 0.591 | 1.029|
| $F_2$            | 0.890| 0.793 | 0.608|
| $F_3$            | 0.663| 0.566 | 1.096|

Table 5. Performance comparison results for different numbers of scale.

| Scale ($l$) | PLCC | SROCC | RMSE |
|-------------|------|-------|------|
| 1           | 0.933| 0.868 | 0.495|
| 2           | 0.945| 0.880 | 0.465|
| 3           | **0.982**| **0.897**| **0.442**|
| 4           | 0.947| 0.889 | 0.453|
4. Conclusions

In this paper, a human visual perception-based multi-exposure fusion image quality assessment (MEF-IQA) method is proposed by considering three perceptual features about color, structure and exposure, and the superiority of our approach is mainly reflected in the following three aspects. First, the chrominance information, which is usually ignored in the existing IQA models for image fusion, is utilized to form the local color saturation similarity and global color distortion metric. Second, dense scale invariant feature transform (DSIFT) descriptor is used for obtaining the structure information of image from multiple orientations because it is more robust and accurate than gradient amplitude. Third, Gauss exposure function is designed to evaluate the luminance inconsistency among the adjacent pixels, and multi-scale scheme is adopted in the process of feature extraction to explore the perceptual difference from the fine level to the coarse level. Extensive experiments have indicated that the proposed method is more effective in predicting the quality of MEF images compared with other state-of-the-art metrics. However, the natural scene is almost moving in practice, and these moving objects in the imaging process will make the final fused images appear ghosting. In future work, we will focus on the MEF-IQA models in dynamic scene, which are more practical than those in static scene.

Author Contributions: Y.C. and A.C. provided the initial idea for this work; Y.C. and A.C. designed the algorithm. Y.C., A.C., B.Y., S.Z. and Y.W. contributed to the analyses of results. Y.C. contributed to the collection and analysis of field test data. Y.C. and A.C. wrote the paper.

Funding: This work was supported by Zhejiang Provincial National Science Foundation of China and National Science Foundation of China (NSFC) under Grant No. LZ20F020002, No. LY18F010005 and 61976149. Taizhou Science and Technology Project under Grant No. 1803gy08 and 1802gy06, and Outstanding Youth Project of Taizhou University under Grant No. 2018JQ003 and 2017PY026.

Acknowledgments: Thanks to Aihua Chen, Yang Wang and the members of image research team for discussions about the algorithm. Thanks also to anonymous reviewers for their comments.

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

References

1. Yang, K.F.; Li, H.; Kuang, H.L.; Li, C.Y.; Li, Y.J. An adaptive method for image dynamic range adjustment. *IEEE Trans. Circuits Syst. Video Technol.* 2019, 29, 640–652. [CrossRef]
2. Yue, G.H.; Hou, C.P.; Zhou, T.W. Blind quality assessment of tone-mapped images considering colorfulness, naturalness, and structure. *IEEE Trans. Ind. Electron.* 2019, 66, 3784–3793. [CrossRef]
3. Wang, J.G.; Zhou, L.B. Traffic light recognition with high dynamic range imaging and deep learning. *IEEE Trans. Intell. Transp. Syst.* 2019, 20, 1341–1352. [CrossRef]
4. Mertens, T.; Kautz, J.; Reeth, F.V. Exposure fusion: A simple and practical alternative to high dynamic range photography. *Comput. Graph. Forum* 2009, 28, 161–171. [CrossRef]
5. Li, Z.G.; Zheng, J.H.; Rahardja, S. Detail-enhanced exposure fusion. *IEEE Trans. Image Process.* 2012, 21, 4672–4676. [PubMed]
6. Gu, B.; Li, W.J.; Wong, J.T.; Zhu, M.Y.; Wang, M.H. Gradient field multi-exposure images fusion for high dynamic range image visualization. *J. Vis. Commun. Image Represent.* 2012, 23, 604–610. [CrossRef]
7. Raman, S.; Chaudhuri, S. Bilateral filter based compositing for variable exposure photography. In Proceedings of the Eurographics (Short Papers), Munich, Germany, 1–3 April 2009.
8. Li, S.T.; Kang, X.D. Fast multi-exposure image fusion with median filter and recursive filter. *IEEE Trans. Consum. Electron.* 2012, 58, 626–632. [CrossRef]
9. Li, S.T.; Kang, X.D.; Hu, J.W. Image fusion with guided filtering. *IEEE Trans. Image Process.* 2013, 22, 2864–2875. [PubMed]
10. Zhang, C.; Cheng, W.; Hirakawa, K. Corrupted reference image quality assessment of denoised images. *IEEE Trans. Image Process.* 2019, 28, 1731–1747. [CrossRef] [PubMed]
11. Hossny, M.; Nahavandi, S.; Creighton, D. Comments on information measure for performance of image fusion. *Electron. Lett.* 2008, 44, 1066–1067. [CrossRef]
12. Cvejic, N.; Canagarajah, C.N.; Bull, D.R. Image fusion metric based on mutual information and Tsallis entropy. *Electron. Lett.* **2006**, *42*, 626–627. [CrossRef]

13. Xydeas, C.S.; Petrovic, V.S. Objective pixel-level image fusion performance measure. In *Sensor Fusion: Architectures, Algorithms, and Applications IV*; International Society for Optics and Photonics: Orlando, FL, USA, 2000.

14. Wang, P.W.; Liu, B. A novel image fusion metric based on multi-scale analysis. In Proceedings of the 2008 9th International Conference on Signal Processing, Beijing, China, 26–29 October 2008.

15. Zheng, Y.; Essock, E.A.; Hansen, B.C.; Haun, A.M. A new metric based on extended spatial frequency and its application to DWT based fusion algorithms. *Inf. Fusion* **2007**, *8*, 177–192. [CrossRef]

16. Piella, G.; Heijmans, H. A new quality metric for image fusion. In Proceedings of the 2003 International Conference on Image Processing, Barcelona, Spain, 14–17 September 2003.

17. Ma, K.D.; Zeng, K.; Wang, Z. Perceptual quality assessment for multi-exposure image fusion. *IEEE Trans. Image Process.* **2015**, *24*, 3345–3356. [CrossRef]

18. Xing, L.; Cai, L.; Zeng, H.G.; Chen, J.; Zhu, J.Q.; Hou, J.H. A multi-scale contrast-based image quality assessment model for multi-exposure image fusion. *Signal Process.* **2018**, *145*, 233–240. [CrossRef]

19. Deng, C.W.; Li, Z.; Wang, S.G.; Liu, X.; Dai, J.H. Saturation-based quality assessment for colorful multi-exposure image fusion. *Int. J. Adv. Robot. Syst.* **2017**, *14*, 1–15. [CrossRef]

20. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–100. [CrossRef]

21. Multi-exposure Fusion Image Database. Available online: http://ivc.uwaterloo.ca/database/MEF/MEF-Database.php (accessed on 11 July 2015).

22. Antkowiak, J.; Baina, T.J. *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*; ITU-T Standards Contributions COM: Geneva, Switzerland, March 2000.

© 2019 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 (http://creativecommons.org/licenses/by/4.0/).