Displays 74 (2022) 102188

Contents lists available at ScienceDirect

Displays

journal homepage: www.elsevier.com/locate/displa

Quality assessment of multi-exposure image fusion by synthesizing local and global intermediate references☆

Jiawu Xu [a,1], Wei Zhou [b,1], Hong Li [c], Fucui Li [d], Qiuping Jiang [a,*]

[a] School of Information Science and Engineering, Ningbo University, Ningbo 315211, China
[b] Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[c] College of Science and Technology, Ningbo University, Ningbo 315211, China
[d] School of Modern Information Technology, Zhejiang Institute of Mechanical and Electrical Engineering, Hangzhou, Zhejiang, China

ARTICLE INFO

Keywords:
Multi-exposure fusion
Image quality assessment
Local reference information
Global reference information

ABSTRACT

Multi-exposure fusion (MEF) takes a sequence of images with different exposure levels as input and generates a fused image that is more informative and perceptually appealing than any of the input images as output. During the past decades, many MEF algorithms have been proposed. Therefore, how to effectively compare the performance of different MEF algorithms is of great significance. Despite of this, research efforts on objective image quality assessment (IQA) of MEF images remain limited. In this paper, we propose a novel full-reference (FR) IQA method for MEF images by generating Local and Global Intermediate References (LGIR) from the input multiple images. Specifically, the intermediate reference features are synthesized in gradient domain, structural tensor domain, and global perception domain, respectively. The gradient and structure tensor domains reflect the local structural perception of the human visual system (HVS), while the global perception domain integrately considers the overall perception. In each domain, a single quality measure is estimated to reflect the visual quality of the fused image from a specific perspective. In addition, on considering the multi-scale property of the HVS, we estimate those quality measures at multiple scales, and fuse them together to predict the final quality score. Experimental results demonstrate the superiority of LGIR, achieving higher consistency with subjective quality scores than existing relevant FR-IQA methods.

1. Introduction

Multi-exposure fusion (MEF) aims to fuse a sequence of low dynamic range (LDR) images with different exposure levels to obtain a fused image with high dynamic range (HDR) imaging effects [1,2,3]. With MEF, the fused image contains richer information such as finer details and more appealing colors than any of the input image. In general, the MEF technology has a wide range of applications. For example, it can enhance the image display effects and the fused image can be displayed on common devices without expensive HDR displays. Compared with tone mapping [4,5,6], which firstly constructs HDR images from multi-exposure image sequences and then converts HDR images back into LDR images for display, the MEF bypasses constructing HDR images and has more advantages in computational efficiency.

Due to the success of the pyramid structure proposed in [7], MEF has gradually attracted lots of research attentions and many MEF algorithms have been proposed during the past decades [8,9,10,11,12,13]. With many MEF algorithms at hand, it becomes vital to compare their performance, so as to select the best-performing algorithm as well as directions for further advancement. In addition, it is found that none of the existing MEF algorithms can achieve ideal results for all multi-exposure image sequences [14]. Therefore, it is important to fairly compare the performance of different MEF algorithms and flexibly select the best fused result under different scenarios. Since the human visual system (HVS) is the ultimate receiver in most applications, subjective evaluation is considered as the most straightforward and reliable way for MEF image quality evaluation [14]. However, subjective evaluation is always time-consuming and laborious, which are also hard to be embedded into online systems for optimizing MEF algorithms. Thus, it is important to design efficient objective image quality assessment (IQA) methods to

☆ This paper was recommended for publication by Prof. Guangtao Zhai.
* Corresponding author.
E-mail address: jiangqiuping@nbu.edu.cn (Q. Jiang).
[1] Authors contributed equally.

https://doi.org/10.1016/j.displa.2022.102188
Received 25 December 2021; Received in revised form 25 February 2022; Accepted 10 March 2022
Available online 14 March 2022

replace the role of subjective evaluation.

According to the participation of the reference information, objective IQA methods can be divided into three types: full-reference (FR) [15,16,17,18,19,20,21,22,23,24,25], reduced-reference (RR) [26,27], and no-reference (NR) [28,29,30,31,32,33,34,35,36,37,38,39]. Although many FR-IQA algorithms have been proposed during the last two decades [40], they cannot be directly applied to MEF images because there are multiple reference images and only one single fused image [41]. It means that the dimensions of the reference image and the distorted/fused image (to-be-evaluated) are not consistent. Therefore, the fused image cannot be directly compared with the reference images and the key to the success of FR-IQA metrics for MEF images is to extract and aggregate effective intermediate reference information from multiple reference images, so as to enable the FR comparison between the fused image and reference images [41].

In this paper, inspired by the global-to-local perception mechanism of the HVS, we propose a FR-IQA method for MEF images by generating Local and Global Intermediate References (LGIR) from the input multiple reference images. Specifically, for the input image sequence with different exposure levels, proper intermediate reference information is extracted in gradient domain, structure tensor domain, and global perception domain, respectively. The gradient and structure tensor domains reflect the local structural perception of the HVS, while the global perception domain integrately considers the overall perception. In each domain, a single quality measure is derived to reflect the visual quality of the fused image from a specific perspective. On considering the multi-scale perception of the HVS, we estimate those quality measures at multiple scales, and fuse all the quality measures together to predict the final quality score via support vector regression (SVR). Experimental results show that compared with the existing FR-IQA algorithms, the objective quality scores predicted by LGIR are more consistent with subjective quality scores. Besides, we also apply LGIR to automatic parameter tuning of an existing MEF algorithm. Numerical experiments indicate that LGIR provides a useful tool to exploit the parameter space and to pick the optimal parameter set that produces MEF image with the best visual quality.

2. Related work

In this section, we review the existing related work, mainly introduce some existing image fusion algorithms and image quality evaluation algorithms.

2.1. Multi-exposure image fusion algorithm

Some MEF algorithms have been proposed. For example, a simple method using local and global energy weighting, Burt et al., [42] applied Laplacian pyramid decomposition to multi-exposure image fusion, the weights are obtained by calculating the local energy of the image pyramid. The Laplacian pyramid is widely applied to the field of image fusion. It can effectively avoid unnatural artifacts in the spatial domain of the fused image. Similarly, Wang et al., [43] calculates the weight map from three aspects of exposure, contrast and saturation, decompose the multi-exposure image by the Laplacian pyramid and decompose the weighting map by Laplacian pyramid, Finally, multiply the multi-exposure image by the weighting map to get the fusion image. Goshtasby et al., [44] divided images with different exposures into many different small patches, and then selected the patch with the largest image entropy as the patch corresponding to the fused image. However, due to the limitations of the segmentation method, the fused image will have a certain block effect. thereby affecting the image quality. Zhang et al., [45] guided the generation of the weight map by calculating the gradients of the multi-exposure images and using the gradient information to calculate the correlation between the images with different exposures. This algorithm can enhance the details in the fused images. According to the characteristics of multi-exposure images, Ma et al., [46] proposed a structure patch decomposition method, which decomposes the image block into three components: signal intensity, signal structure and average intensity. Then fuses these three components to reconstruct image patch. This method has better robustness to image ghosting.

2.2. Fusion image quality evaluation algorithm

Although many image fusion algorithms have been proposed, some databases have also been established gradually [47]. There is still a lack of research on the complete and comprehensive quality evaluation of these algorithms. Usually, in order to verify the effectiveness of the fusion method, it is necessary to manually judge the quality of the fusion image. In order to avoid this time-consuming and laborious manual method, it is important to develop a multi-exposure fusion image quality evaluation algorithm. For example, Qu et al., [48] evaluates the quality of the fused image by calculating the mutual information between the original image and the fusion image. Considering that the human visual system is very sensitive to the edge of the image, some evaluation methods based on edge information have been proposed one after another. Xydea et al., [49] uses the Sobel edge operator to extract the edge information of the original image and the fused image respectively, then calculates the edge loss between different original images and the fused image as features, finally fuses the features to calculate the quality score. Piella et al., [50] considered the influence of structural information such as image clarity and image entropy on image quality, and calculated the saliency map to weight the feature map and proposed a new quality evaluation model. However, the above assessment algorithms for fusion images are not consider particularity of MEF images. For accurate assessment the visual quality of MEF images. Ma et al., [41] established a MEF image quality evaluation database and proposed an objective quality evaluation algorithm for MEF images. First, the author believes that images can be decomposed into three parts: contrast, structure and brightness. Then, the reference image is reconstructed by enhanced contrast and structural components of the multi-exposure image. Finally, calculate the SSIM between the reference image and the fused image to get the quality score. This algorithm shows good performance on the newly established MEF image quality evaluation database.

3. Proposed method

3.1. Problem formulation

In the context of IQA, given a distorted (to-be-evaluated) image I_d and its corresponding reference image I_r, the general pipeline of FR-IQA involves two key steps: 1) feature extraction (F) from I_d and I_r; and 2) similarity/distance measure (S) between $F(I_d)$ and $F(I_r)$, to derive the final quality score Q_d as follows:

$$Q_d = S(F(I_d), F(I_r)) \tag{1}$$

For the FR-IQA of MEF images, the distorted image is a single image fused by a certain MEF algorithm. Here, we denote this to-be-evaluated MEF image as I_{mef}. However, the reference information is not a single image. Instead, it is a sequence of multi-exposure images with different exposure levels, which is represented by $\{I_m | m = 1, 2, \cdots, M\}$ where M is the number of reference images or exposure levels. In order to conduct FR-IQA, we need to aggregate the features extracted from multiple reference images to build proper intermediate reference information which can be easily compared with the same kind of features extracted from the (single) fused image. Afterwards, the similarity/distance between the aggregated intermediate reference information and the corresponding information extracted from the fused image is calculated. The above processes can be formulated as follows:

$$Q_{mef} = S\big(F\big(I_{mef}\big), A(F(I_1), F(I_2), \cdots, F(I_M))\big) \tag{2}$$

here A represents a specific feature aggregation scheme. For better understanding, the comparison between traditional FR-IQA and MEF FR-IQA pipelines are illustrated in Fig. 1.

Considering that different image features have different influences on perceptual quality, it is critical to design effective aggregation strategies (A) for different features. In our proposed method, we consider two kinds of features (*local and global*) and design three different feature aggregation schemes (*maximum gradient, structure tensor, and pyramid fusion*) to build intermediate reference information from the input multi-exposure image sequence. In what follows, we first give a brief overview of the proposed MEF FR-IQA method and then illustrate each step with details.

3.2. Overview

In this work, we propose a novel FR-IQA metric called LGIR for MEF images from the perspective of characterizing the global-to-local image perception mechanism of the HVS. In the proposed method, we extract image features from both local and global aspects and apply different feature aggregation schemes to generate corresponding intermediate references. Specifically, the image gradient magnitude map and Jacobian matrix are built to represent the local features, while the multi-exposure image sequence itself is directly regarded as the global feature. For the gradient magnitude maps, we compute the maximum gradient to aggregate the gradient magnitude maps of the multiple reference images to obtain the maximum gradient feature G_{max} as one local reference. For the Jacobian matrix, we apply the structure tensor operator to convert the Jacobian matrix of the corresponding pixels in multiple reference images into a unified structure tensor representation Z_r as another local reference. For the multi-exposure image sequence itself, we aggregate them via pyramid fusion to obtain a global reference image I_{gr}. In a similar way, these features are also extracted from the MEF image, as denoted by G_{mef}, Z_{mef} , and I_{mef}, respectively. Then, the perceptual

similarity/distance scores between the generated intermediate reference features and the corresponding features extracted from the MEF image are calculated, yielding three quality measure scores (i.e.,Q_{lgs}, Q_{lst}, and Q_{gps}). In addition, on considering the multi-scale perception of the HVS, we estimate those quality measure scores at multiple scales, and fuse all the quality measure scores to predict the final quality via SVR. The diagram of our proposed LGIR is shown in Fig. 2. For conciseness, only the single-scale version is depicted and the details regarding the multiple-scale version will be illustrated in Section II-E.

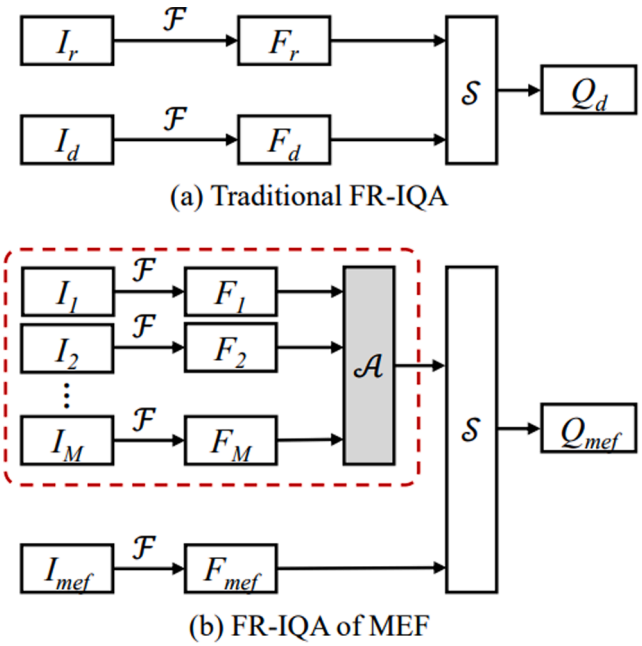

Fig. 1. Comparison between traditional FR-IQA and MEF FR-IQA pipelines.

3.3. Local reference information

3.3.1. Maximum gradient

Image gradient reflects the change degree/intensity of local neighborhood pixel values. The direction of image gradient indicates the fastest direction of pixel change. Moreover, the edge positions of images usually have larger gradient values. At the same time, in regions where image details are less and smoother, image pixel values change less, and the corresponding gradient values of these areas could also decrease. In image processing field, the gradient of an image usually refers to gradient modulus. To obtain the edge or texture information of reference image sequences and reduce the influence of brightness, commonly-used operators to calculate image gradient include Prewitt, Roberts, Scharr, and Sobel operators. In general, we use these operators to calculate the difference between neighboring pixels to approximate the derivative of the image. In our work, the horizontal gradient map $\mathbf{G}_x$ is obtained by the convolution operation: $\mathbf{G}_x = \mathbf{S}_x \otimes \mathbf{I}$ where $\mathbf{I}$ represents the target image and $\otimes$ is the convolution opera-tion. $\mathbf{S}_x$ is the Sobel operator:

$$\mathbf{S}_x = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{3}$$

Similarly, we use the transposition of $\mathbf{S}_x$ to calculate the vertical gradient map $\mathbf{G}_y$ by $\mathbf{G}_y = \mathbf{S}_y \otimes \mathbf{I}$where $\mathbf{S}_y = \mathbf{S}_y^T$ After obtaining the horizontal and vertical gradient magnitudes, the gradient magnitude of the pixel (i,j) is defined by:

$$\mathbf{G}(i,j) = \sqrt{\mathbf{G}_x(i,j)^2 + \mathbf{G}_y(i,j)^2}, \tag{4}$$

where $\mathbf{G}_x(i,j)$ and $\mathbf{G}_y(i,j)$ are the pixel values of horizontal and vertical gradient maps at the pixel (i,j). Using Eq. (4), we can compute the gradient magnitude maps of all reference image and the to-be-evaluated MEF image, as denoted by $\{\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_M\}$ and $\mathbf{G}_{mef}$, respectively.

In general, the visibility of image details is closely related to the gradient magnitude, i.e., larger magnitude of pixel gradient indicates higher visibility and vice versa. That is, the image regions with larger magnitude of pixel gradient is usually more visible and clearer. Therefore, the maximum gradient of pixels corresponding to multi-exposure image sequences is taken as the best gradient magnitude under the real scene conditions, which is computed as follows:

$$\mathbf{G}_{\max}(i,j) = \max\{\mathbf{G}_1(i,j), \mathbf{G}_2(i,j), \cdots, \mathbf{G}_M(i,j)\} \tag{5}$$

The derived maximum gradient map $\mathbf{G}_{\max}$ can be regarded as the reference/ideal gradient information which is expected to be well retained in the MEF image. Thus, by measuring the similarity between the maximum gradient map derived from the input multi-exposure image sequences and the gradient map derived from the to-be-evaluated MEF image, the quality score of the MEF image in terms of the local gradient information is thus obtained.

Since the structural similarity is well-known to estimate the perceptual similarity between two images, here we calculate the similarity between the maximum gradient map $\mathbf{G}_{\max}$ and the gradient map of the to-be-evaluated MEF image $\mathbf{G}_{mef}$ as:

$$S(\mathbf{G}_{\max}, \mathbf{G}_{mef}) = \frac{2\mu_{\mathbf{G}_{\max}}\mu_{\mathbf{G}_{mef}} + c_1}{\mu_{\mathbf{G}_{\max}}^2 + \mu_{\mathbf{G}_{mef}}^2 + c_1} \cdot \frac{2\sigma_{\mathbf{G}_{\max}\mathbf{G}_{\mathrm{mef}}} + c_2}{\sigma_{\mathbf{G}_{\max}}^2 + \sigma_{\mathbf{G}_{\mathrm{mef}}}^2 + c_2}, \tag{6}$$

where $\mu_{\mathbf{G}_{\max}}$, $\mu_{\mathbf{G}_{mef}}$, $\sigma_{\mathbf{G}_{\max}}^2$, $\sigma_{\mathbf{G}_{mef}}^2$, and $\sigma_{\mathbf{G}_{\max}\mathbf{G}_{mef}}$ represent the local mean of $\mathbf{G}_{\max}$ and $\mathbf{G}_{mef}$, the local variance of $\mathbf{G}_{\max}$ and $\mathbf{G}_{mef}$, and the local

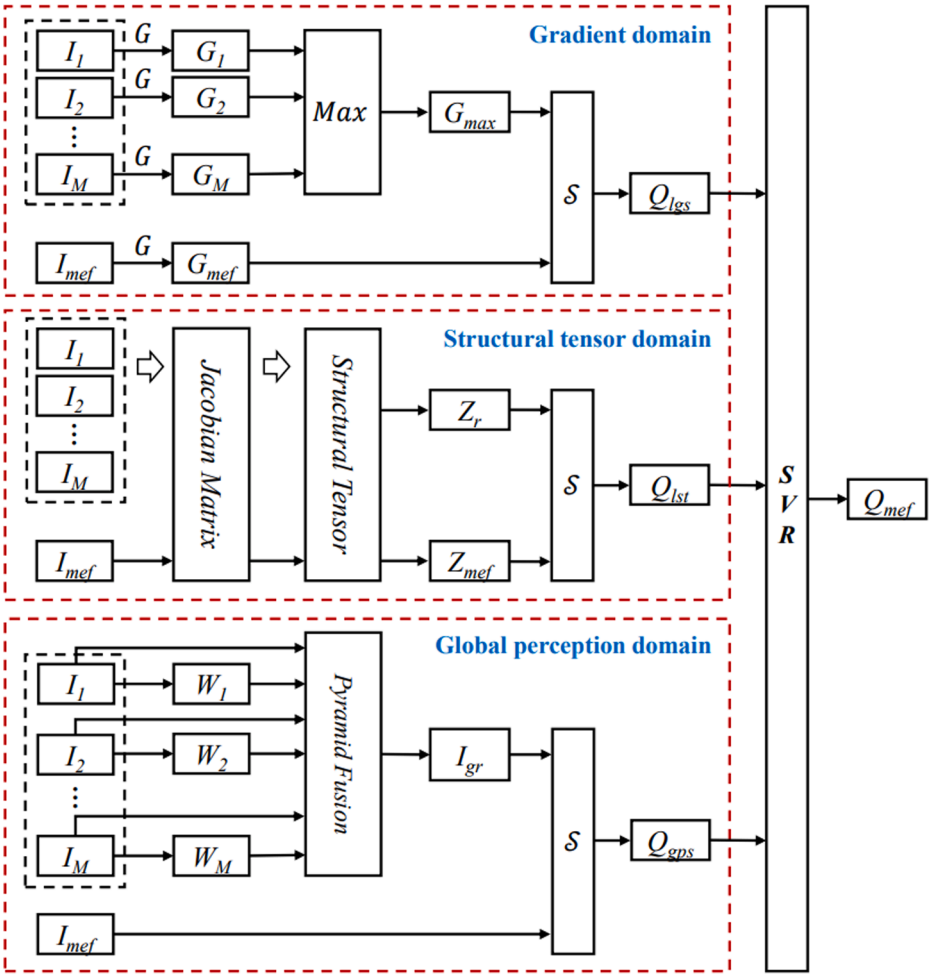

Fig. 2. Diagram of the single-scale LGIR.

covariance between $\mathbf{G}_{\max}$ and $\mathbf{G}_{mef}$, respectively. In addition, c_1 and c_2 are small positive stability constants that account for the saturation effects of the HVS at extremely low luminance and contrast.

With this similarity, we can assess the local gradient quality of the fused image. That is, by averaging the quality map, we have the gradient quality as follows:

$$Q_{lgs} = \frac{1}{H \times W} \sum_{i,j} S\left(\mathbf{G}_{\max}(i,j), \mathbf{G}_{mef}(i,j)\right), \tag{7}$$

where H and W denote the image height and width, respectively. The gradient quality derived above is thus considered as one local quality measure score of an MEF image.

As shown in Fig. 3, the maximum gradient map can effectively reflect 175 the optimal gradient based on pixel visibility criterion under different exposure levels. Therefore, by computing the similarity between the gradient map of the fused image and the maximum gradient map of the corresponding multi-exposure image sequence, we can effectively capture the pixel degradations caused by various distortions existing in the edge positions of the fused image and the unnatural artifacts, which have significant effects on the perceived quality.

3.3.2. Structure tensor

Image structure is one of the important features in evaluating image quality. In general, image structures should be recovered as much as possible for a high-quality MEF image. Thus, how to effectively extract the structural features of images is important. However, when evaluating the perceptual quality of MEF images, to deal with the high-dimensional image sequences composed of multiple images with different exposure levels, it is necessary to reconsider how to combine the structural information of different reference images. Inspired by the structural tensor theory [51]

we first use Jacobian matrix to represent the local structural features of multiple reference images with different exposure levels. Specially, the structural information of pixel (i, j) corresponding to image sequences with different exposure levels can be expressed by:

$$\mathbf{J}(i,j) = \begin{bmatrix} \nabla_x I_1(i,j) & \nabla_y I_1(i,j) \\ \vdots & \vdots \\ \nabla_x I_M(i,j) & \nabla_y I_M(i,j) \end{bmatrix}, \tag{8}$$

where $\mathbf{J}(i,j)$ represents the combination of structural information at pixel (i, j) for the multi-exposure image sequence. $\nabla_x I_1(i,j)$ and $\nabla_y I_1(i,j)$ are the first-order derivatives of pixel (i, j) along the x and y directions, respectively. Additionally, M is the number of the multi-exposure images.

After constructing the Jacobian matrix, we use the structure tensor operator to aggregate the Jacobian matrix of pixels with different exposure levels as follows:

$$\mathbf{Z}(i,j) = \mathbf{J}(i,j)^T \mathbf{J}(i,j)$$

$$= \begin{bmatrix} \sum_{k=1}^{M} (\nabla_x I_k(i,j))^2 & \sum_{k=1}^{M} (\nabla_x I_k(i,j))(\nabla_y I_k(i,j)) \\ \sum_{k=1}^{M} (\nabla_y I_k(i,j))(\nabla_x I_k(i,j)) & \sum_{k=1}^{M} (\nabla_y I_k(i,j))^2 \end{bmatrix}. \tag{9}$$

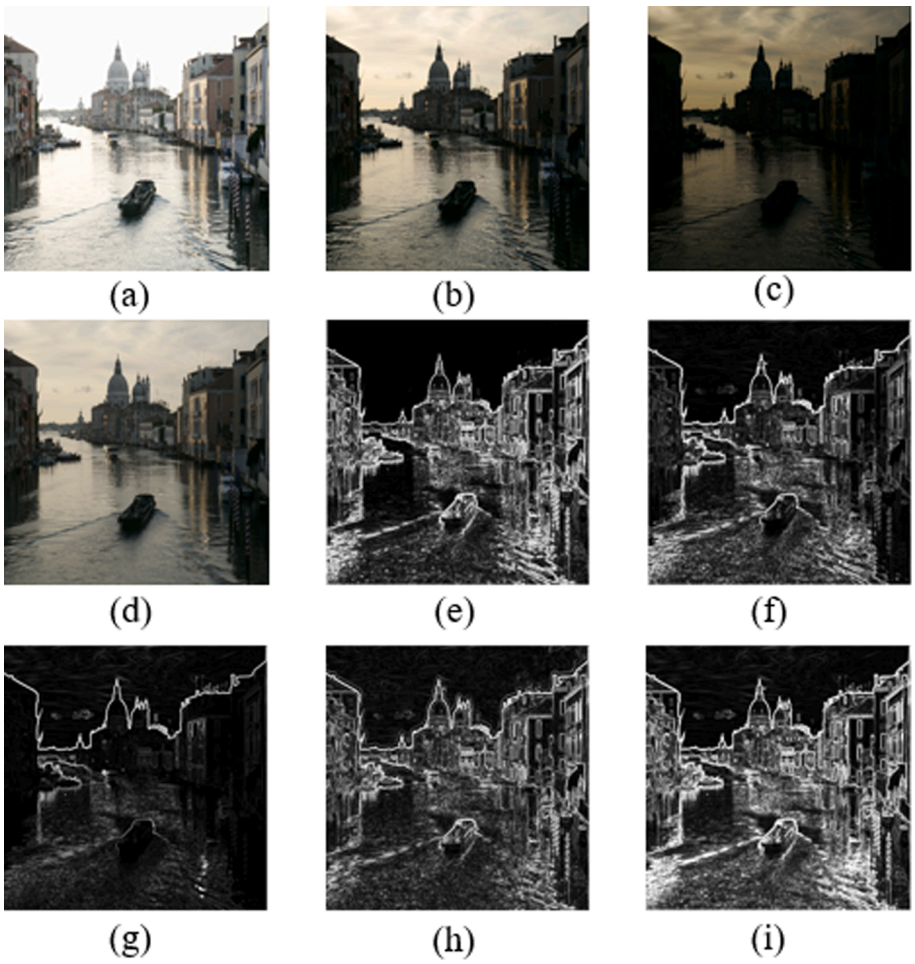

Fig. 3. (a)-(c) are three images with different exposure levels, (d) is the fused image, (e)- (h) are the gradient maps of (a)-(d), respectively, (i) is the generated maximum gradient map from (e)-(h) via Eq. (8).

According to Eqs. (8) and (9), each corresponding pixel in the multi-exposure image sequence and the MEF image results in a matrix of structure tensor. In order to measure the difference between the structure tensors of the multi-exposure image sequence and the MEF image, we first convert the structure tensor matrix $\mathbf{Z}(i,j)$ into a vector form $\overrightarrow{Z}(i,j)$. Then, we calculate the cosine distance between the structure tensors of the corresponding pixels in multi-exposure image sequence $\overrightarrow{Z_r}(i,j)$ and the MEF image $\overrightarrow{Z_{mef}}(i,j)$ as follows:

$$D_{cos}(i,j) = \frac{\overrightarrow{Z_r}(i,j)\overrightarrow{Z_{mef}}(i,j)}{\|\overrightarrow{Z_r}(i,j)\|\|\overrightarrow{Z_{mef}}(i,j)\|}. \tag{10}$$

Finally, we calculate the average value of the structure tensor distance of all pixels as another local quality measure score:

$$Q_{lst} = \frac{1}{H \times W}\sum_{i,j} D_{cos}(i,j). \tag{11}$$

3.4. Global reference information

Although the image gradient magnitude and structure tensor can effectively capture image local structures, they cannot characterize the global perception of the HVS. Therefore, in order to make our proposed quality assessment model more consistent with human visual perception, we additionally propose a global perception measurement. Specifically, the globally perceived quality of the MEF image is measured by first synthesizing a global reference image from the multi-exposure image sequences with pyramid fusion and then calculate the similarity between the global reference image and the to-be-evaluated MEF image. In what follows, we illustrate how to construct the global reference image by taking human global perception characteristics into consideration.

First, considering that the HVS is very sensitive to the color and brightness changes of images, we transform the image from RGB space to YUV space, where Y channel represents the brightness and UV channels convey the color information. In the process of multi-exposure fusion, different exposure levels have great influence on the fused result. Moreover, image brightness and saturation also affect the global perception of humans. Therefore, we construct weighting maps from three aspects, i.e., exposure, contrast, and saturation, to fuse the multi-exposure image sequences to obtain the global reference image in the YUV space. The exposure, contrast, and saturation are represented by E, C and S, respectively.

3.4.1. Exposure

In multi-exposure image sequences, underexposed and overexposed pixels usually fail to capture the details genuinely, while well-exposed pixels can well retain fine details. Therefore, compared to the underexposed and overexposed pixels, the normally exposed pixels should be encouraged to have larger weights in the process of multi-exposure image fusion. We use a Gaussian kernel to derive the weights of different pixels in terms of their corresponding exposure levels:

$$E = \exp\left(-\frac{(Y-\mu)^2}{2\sigma^2}\right), \tag{12}$$

where Y is the normalized value of the $\mathbf{Y}$ channel. Two parameters of the Gaussian kernel, i.e., standard deviation σ and expected value μ. We believe that the brightness of well-exposed pixels tend to be close to 0.5, in order to assign a higher weight to the pixels with an exposure of 0.5, μ is set to 0.5. At the same time, in order to reduce the weight obtained by pixels with too high exposure and too low exposure as much as possible, a Gaussian function with a smaller width should be selected. After some attempts, we find that setting σ to 0.2 is the most appropriate.

3.4.2. Contrast

Generally, the visibility of edges will be reduced in over-/under-exposed areas [52]. Thus, to assign larger weights to clear edges in the fusion process, we define the contrast measure C by calculating the absolute value of the convolution of Laplacian operator $\mathbf{L}$ and luminance channel $\mathbf{Y}$

$$C = |\mathbf{L} \otimes \mathbf{Y}|, \tag{13}$$

where $\otimes$ denotes the 2-D convolution operation.

3.4.3. Saturation

A well-exposed pixel usually captures color saturation well. In the RGB color space, the standard deviation within the R, G and B channel of each pixel is taken as the saturation. While in the YUV space, the color saturation [53] is defined as follows:

$$S = |\mathbf{U}| + |\mathbf{V}| + 1. \tag{14}$$

3.4.4. Weight combination and normalization

The final weighting maps are obtained by calculating the product of the above three measures as follows:

$$W(i,j)m = E(i,j)_m^{\omega_E} \times C(i,j)_m^{\omega_C} \times S(i,j)_m^{\omega_S}, \tag{15}$$

where $(i,j)_m$ denotes the pixel (i,j) in the m-th image with different exposure levels. ω_C, ω_S and ω_E represent the weights of contrast, saturation and expo- sure, respectively. The default values of ω_C, ω_S and ω_E are all set to 1. In order to ensure that the sum of weights for each pixel in reference images with different exposure levels equals to 1, the weighting maps are further normalized as follows:

$$\widehat{W}(i,j)_m = \left[\sum_{M=1}^{M} W(i,j)_m\right]^{-1} W(i,j)_m, \tag{16}$$

where M is the total number of reference images and $\widehat{W}(i,j)_m$ represents thenormalized weighting map of the m-th reference image.

3.4.5. Pyramid fusion

The product of multi-exposure image sequences and weighting maps are used for pyramid fusion. That is, we first decompose the reference images with different exposure levels and weighting maps into Laplacian and Gaussian pyramids, respectively. Then, the decomposed images are fused to produce the global reference image, as illustrated in Fig. 4. We denote the fused global reference image as $\mathbf{I}_{gr}$ hereinafter.

Similar to the local gradient quality, the structural similarity is also applied to the global reference image $\mathbf{I}_{gr}$ and the to-be-evaluated MEF image $\mathbf{I}_{mef}$ as follows:

$$S(\mathbf{I}_{gr}, \mathbf{I}_{mef}) \frac{2\mu_{\mathbf{I}_{gr}}\mu_{\mathbf{I}_{mef}} + c_1}{\mu_{\mathbf{I}_{gr}}^2 + \mu_{\mathbf{I}_{mef}}^2 + c_1} \bullet \frac{2\sigma_{\mathbf{I}_{gr}\mathbf{I}_{\mathrm{mef}}} + \mathrm{c}_2}{\sigma_{\mathbf{I}_{\mathrm{gr}}}^2 + \sigma_{\mathbf{I}_{\mathrm{mef}}}^2 + \mathrm{c}_2}, \tag{17}$$

where $\mu_{\mathbf{I}_{gr}}$, $\mu_{\mathbf{I}_{mef}}$, $\sigma_{\mathbf{I}_{gr}}^2$, $\sigma_{\mathbf{I}_{mef}}^2$, and $\sigma_{\mathbf{I}_{gr}\mathbf{I}_{mef}}$ denote the local mean of $\mathbf{I}_{gr}$ and $\mathbf{I}_{mef}$, the local variance of $\mathbf{I}_{gr}$ and $\mathbf{I}_{mef}$, and the local covariance between $\mathbf{I}_{gr}$ and $\mathbf{I}_{mef}$, respectively. c_1 and c_2 are small positive stability constants that account for the saturation effects of the HVS at extremely low luminance and contrast. As can be seen in Fig. 5, there exist some obvious unnatural visual artifacts around the edges of the tower and clouds in the to-be-evaluated MEF image shown in (d) and these artifacts are well identified by the quality map shown in (f). Finally, the global perception quality score of the MEF image is derived by averaging the quality map as:

$$Q_{gps} = \frac{1}{H \times W} \sum_{i,j} S\left(\mathbf{I}_{\mathrm{gr}}(i,j), \mathbf{I}_{\mathrm{mef}}(i,j)\right). \tag{18}$$

3.5. Multi-scale feature extraction and quality regression

Since the recent research has revealed that HVS perceives image in a coarse-to-fine strategy [54], a multi-scale strategy is also taken into account. In practice, the coarser scale is obtained by low-pass filtering and then down sampled by a factor of two. As shown in [55], an appropriate scaling factor z is close to the square root of the ratio of the focused visual scope and the image size:

$$z = \sqrt{\frac{1}{4 \bullet \tan\left(\frac{\theta_h}{2}\right) \bullet} \tan\left(\frac{\theta_w}{2}\right) \bullet \left(\frac{\mathrm{h}}{\mathrm{d}}\right)^2 \bullet \frac{\mathrm{w}}{\mathrm{h}}}, \tag{19}$$

where h, w, and d denote image height, image width, and viewing distance, respectively; θ_h and θ_w represent horizontal and vertical visual angles. As reported in [55], it is found that z = 0.4955($\approx$ 0.5) works well. Thus, the used downsampling factor "2″ can be viewed as approximately optimal in given viewing conditions. To consider varying image resolutions and viewing distances, our quality measure scores (i.e., Q_{lgs}, Q_{lst}, and Q_{gps}) are estimated in three coarse to fine scales. To be specific, the coarser scale is first processed by a low-pass filter, followed by a down sampling operation with a factor of 2. Since there are three quality measure scores in each scale, thus a total number of nine quality

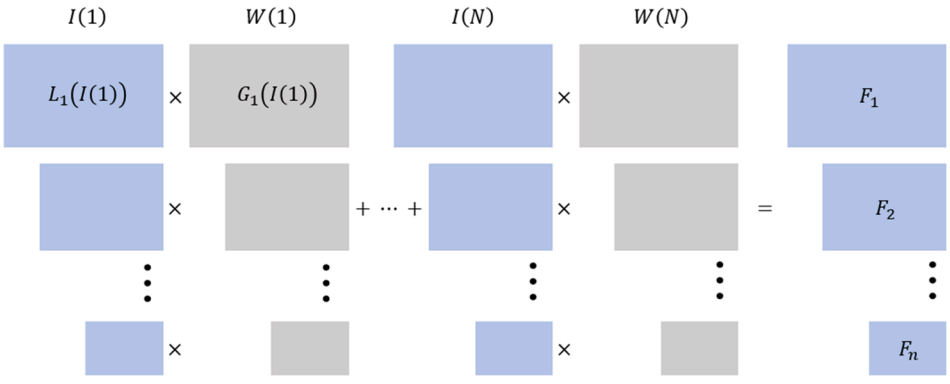

Fig. 4. Diagram of the construction of the global reference image via pyramid fusion. $I(1)$-$I(N)$ are N images with different exposure levels, W (1)-W (N) are the corresponding weighting maps, F_1-F_n are the corresponding fused results.

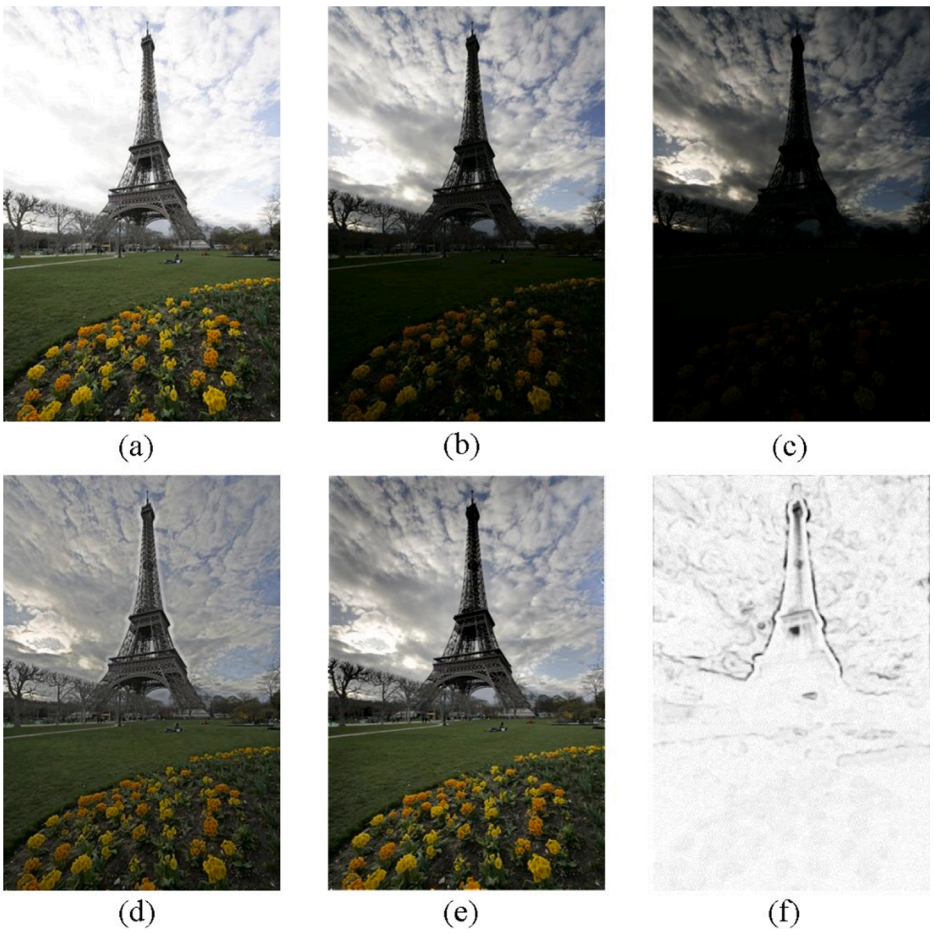

Fig. 5. (a), (b), and (c) are the input image sequence with different exposure levels, (d) is the to-be-evaluated MEF image, (e) is the constructed global reference image, (f) is the quality map estimated by Eq. (17).

measure scores are derived for quality regression.

Specifically, we learn a regression model via SVR for mapping the nine quality measure scores into a single quality score. The LIBSVM package [56] is used to implement SVR with the radial basis function kernel. Once the SVR model is built, it can be used for quality prediction of a test MEF image with its corresponding quality measure scores at three scales as input.

4. Experimental results

4.1. Database and evaluation metrics

4.1.1. Database

In our experiments, we evaluate the performance of the proposed LGIR 250 model on the database established by Ma et al. [41]. This database consists of 17 multi-exposure image sequences with different scenes suitable for image quality prediction. e.g., In BelgiumHouse and Cave, the brightness of the indoor scene contrasts sharply with the outdoor scene. Image brightness of Farmhouse in night scene is no satisfactory and Madison has many intricate texture details. Each sequence contains at least three images corresponding to under-exposure, normal-exposure, and over-exposure. Eight representative MEF algorithms are applied to generate a total number of 136 MEF images. These MEF algorithms include local energy linear weighting, global energy linear weighting, Mertens09 [8], Raman09 [9], Gu12 [10], Li12 [11], ShutaoLi12 [12], and ShutaoLi13 [13]. In order to obtain the subjective quality scores of these MEF images, 25 subjects (including 10 males and 15 females) were asked to participate the subjective quality assessment experiments where each subject was asked to give his/her subjective quality score (ranging from 1 to 10) for each observed MEF image. The final subjective score of an MEF image is derived in the form of mean opinion score (MOS) which is obtained by averaging the corresponding subjective scores from all subjects.

4.1.2. Evaluation metrics

To measure the performance of our proposed LGIR model, we use the Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank-order Correlation Coefficient (SRCC) as the evaluation criteria. Specifically, we compute the PLCC and SRCC values between the predicted quality scores by LGIR and the MOSs. In theory, larger PLCC and SRCC values indicate higher consistency between the predicted objective scores and subjective MOSs. In the process of SVR training and testing, we adopt the leave-one-out validation (LOOV) strategy, i.e., 128 MEF images corresponding to 16 sequences are used as the training set and 8 MEF images corresponding to remaining one sequence are used for testing set at a time. After 17 train-test trails, all the 136 MEF images have been tested.

4.2. Overall performance comparison

To demonstrate the superiority of our proposed LGIR method, we compare it with eight existing objective IQA methods that are widely-used for image fusion quality evaluation. These compared IQA methods include [41,50,57,58,59,60,61,62]. Among the compared methods, [60] is a learning method based on SVR feature fusion, [62] uses a learning method of back-propagation to adjust parameters, and the rest of the compared methods are not learning based methods. The PLCC and SRCC results of these methods are reported in Table 1 and

Table 1
PLCC results of different MEF IQA algorithms.

Source sequence sequence sequence	[57]	[58]	[50]	[59]	[41]	[60]	[61]	[62]	LGIR
Balloons	0.7050	0.4389	0.5038	0.2765	0.9297	**0.9493**	0.7605	0.6648	0.8358
Belgium house	0.8017	0.6261	0.5020	−0.5719	0.9312	0.9897	0.1744	0.5608	**0.9931**
Lampl	0.7294	0.7281	0.4320	0.0095	0.8907	0.8091	−0.4785	0.4017	**0.9218**
Candle	0.9388	0.8917	0.1789	−0.8006	**0.9513**	0.8547	−0.7290	0.1063	0.8643
Cave	0.6946	0.8136	0.6301	0.1562	0.7719	0.5511	0.0526	0.6206	**0.9421**
Chinese garden	0.7684	0.8358	0.4087	0.3800	0.9563	0.4697	−0.2942	0.4812	**0.9683**
Farmhouse	0.6408	0.5997	0.2163	−0.2729	0.8632	0.7652	0.5039	0.6929	**0.8759**
House	0.6207	0.5956	0.4805	0.4154	0.8414	0.9656	−0.5237	0.4762	**0.9601**
Kluki	0.3914	0.3585	−0.0489	0.5952	**0.8242**	−0.1624	0.0209	−0.1124	0.7305
Lamp2	0.8445	0.7518	0.5955	0.6131	0.8291	0.7938	0.6205	0.6493	**0.9748**
Landscape	0.3204	0.4479	0.0306	**0.9015**	0.7457	0.6156	0.5393	0.0806	0.8838
Lighthouse	0.8383	0.6553	−0.0226	0.6680	0.9420	0.8173	−0.2613	0.2462	**0.9879**
Madison capitol	0.6278	0.4225	0.6184	0.4900	0.9141	0.7006	0.0313	0.5414	**0.9433**
Memorial	0.8276	0.6780	0.7332	−0.2164	0.8981	0.8300	0.4454	0.5877	**0.9680**
Office	0.4980	0.4725	0.3238	0.5309	**0.9628**	0.5742	0.3022	0.3161	0.9553
Tower	0.7719	0.8347	0.5941	0.6806	0.9561	0.7839	−0.1155	0.5718	**0.9580**
Venice	0.7949	0.6543	0.2796	0.6790	**0.9699**	0.6075	−0.0222	0.4786	0.9668
Average	0.6950	0.6356	0.3798	0.2667	0.8928	0.7009	0.0604	0.4332	**0.9252**

Table 2
SRCC results of different MEF IQA algorithms.

Source sequence	[57]	[58]	[50]	[59]	[41]	[60]	[61]	[62]	LGIR
Balloons	0.6667	0.5000	0.4524	0.3333	0.8333	0.9286	0.7143	0.5952	0.8095
Belgium house	0.7785	0.7545	0.4671	−0.7066	**0.9701**	0.9222	0.0000	0.5389	**0.9701**
Lampl	0.7857	0.6190	0.4048	0.1905	**0.9762**	0.8095	−0.3810	0.4762	0.9048
Candle	0.9762	0.7857	0.5476	−0.4524	0.9286	0.7615	−0.6667	0.1667	**0.9762**
Cave	0.7143	0.8095	0.5714	0.3333	0.8333	0.6190	0.0238	0.6429	**0.8623**
Chinese garden	0.6905	0.7857	0.5238	0.4048	**0.9286**	0.5714	−0.2857	0.5476	0.7857
Farmhouse	0.7381	0.8095	0.2857	−0.1905	0.9286	0.5714	0.5000	0.5000	**0.9341**
House	0.5952	0.4524	0.4048	0.4524	**0.8571**	0.9762	−0.6905	0.5238	0.8333
Kluki	0.2619	0.2857	0.1190	0.7381	0.7857	−0.1667	0.1667	0.0476	**0.8810**
Lamp2	0.7619	0.6190	0.5476	0.8810	0.7143	0.7381	0.8333	0.6905	**0.9524**
Landscape	0.0238	0.4048	0.1429	**0.8333**	0.5238	0.5000	0.5476	0.1429	0.7619
Lighthouse	0.5000	0.4286	0.0714	0.6905	0.8810	0.7857	−0.4286	0.3810	**0.9222**
Madison capitol	0.5238	0.3571	0.4762	0.5476	**0.8810**	0.6429	−0.3095	0.5238	0.8095
Memorial	0.7619	0.5476	0.6667	−0.2381	0.8571	0.8810	0.8095	0.5238	**0.9102**
Office	0.2771	0.3976	0.4579	0.4940	0.7832	0.1687	0.0843	0.3856	**0.8555**
Tower	0.5714	0.5238	0.5714	0.5952	**0.9524**	0.7381	−0.2143	0.5952	0.8571
Venice	0.9102	0.7306	0.3114	0.6587	**0.9341**	0.5868	0.2994	0.5629	0.8623
Average	0.6198	0.5771	0.4131	0.3274	0.8570	0.6491	0.0590	0.4614	**0.8597**

Table 2, respectively.

It can be seen that most competing IQA methods cannot predict the perceived quality of MEF images with high accuracy. The reason is that the used features by the existing IQA methods are too simple and not sufficiently comprehensive to reflect the characteristics of MEF image quality. For example, in [61], using entropy and coefficient histogram of pixel values as features has poor correlation with subjective perception, which may be due to the fact that only using entropy and histogram of pixel intensity cannot accurately predict the perceptual changes of the fused image content. Moreover, based on structural similarity, gradient, structure and other information, [57], [58] and [41] show relatively good results for MEF image quality prediction, but these evaluation models which only consider the information such as image texture/structure have poor perception of image brightness and color change. Overall, the performance results still have much room for further improvement. By contrast, we observe that the PLCC and SRCC valus of our proposed LGIR reach to 0.9252 and 0.8758, respectively, indicating a high consistency with subjective quality ratings and the best performance

among all the competitors when considering the average performance over all sequences. When considering each image sequence set, the proposed LGIR method delivers he best performance on 11 out of 17 sequences in terms of PLCC while Ma et al. [41] ranks the second place, which achieves the best performance on only 4 out of 17 sequences in terms of PLCC. The is mainly attributed to that we jointly take both the local and global image perception characteristics into account for feature representation towards a more comprehensive and accurate quality evaluation. We further draw the scatter plots between the predicted quality scores and the subjective MOSs in Fig. 6. It also suggests that the proposed LGIR method can predict the quality of MEF images in a highly consistent manner with subjective perception.

4.3. Validity of each individua

Our proposed LGIR method evaluates the quality of MEF images from both local and global aspects. Specifically, three different types of quality measures are estimated at three scales for quality regression. In order to understand the contribution of each individual type of quality measurement in our proposed LGIR, we further conduct ablation studies. The ablation study results are shown in Table 3 and Table 4. We can find that the PLCC and SRCC values of the single index for the maximum gradient feature Q_{lgs} can reach 0.8730 and 0.7503, respectively, which are quite competitive. This is because the edge information, as an important image feature, will make the observer very sensitive to the

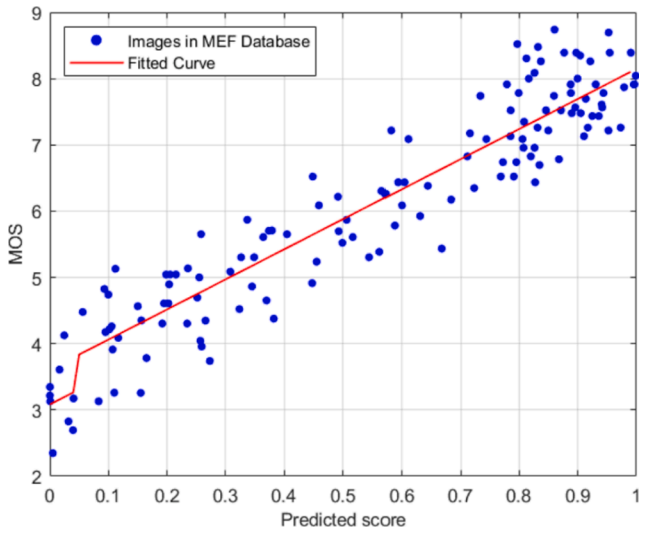

Fig. 6. Scatter plot between predicted quality scores and subjective MOSs.

subtle changes of the local edge feature of an image. Therefore, the proposed maximum gradient in the gradient domain can obtain the most obvious edge information in all different exposure images, which effectively reflects the edge quality of fused images. In addition, our global perceptual quality measure Q_{gps} and its combination variations have also achieved competitive PLCC and SRCC values since the generated global reference image is highly consistent with the human overall perception of high-quality images. The global perceptual quality measure Q_{gps} can not only reflect the details such as structure and texture locally, but also represent the overall characteristics such as contrast and brightness of images globally. Third, although the structure tensor distance measure Q_{lst} delivers moderate performance in an average level, it achieves the highest PLCC and SRCC values on some image sequences such as "Farmhouse" and "Kluki". It suggests that the structure tensor distance measure Q_{lst} is complementary to Q_{lgs} and Q_{gps}. Finally, by jointly considering these three complementary quality measurements at multiple scales, the best performance can be achieved, which well demonstrate the effectiveness of each quality measure component in the proposed LGIR method. In the future, the research on MEF FR-IQA model could focus on evaluating the preservation degree of the fused image by considering more comprehensive information so as to develop a more accurate FR-IQA model for MEF images.

4.4. Application to parameter tuning of MEF algorithms

Besides automatic quality evaluation, an effective objective MEF image quality evaluation metric should be able to guide the optimization of MEF algorithms. In this section, we demonstrate this idea by applying the proposed LGIR metric to automatic parameter tuning of MEF algorithms. There are always one or several parameters in MEF algorithms whose optimal values vary with different image contents. It is challenging and time-consuming to handpick a set of parameters that work well for all scenes. Our proposed MEF image quality evaluation metric is able to replace the role of humans in this task, especially when the volume of images to be processed is large. Here, we use the MEF algorithm proposed in [12] as an example, which involves two tunable parameters σ_s and σ_r. The visual quality of the final fused image is highly sensitive to these two parameters. The default values settled in [12] are: $\sigma_s = 100$ and $\sigma_r = 4$.

Fig. 7 shows the fused images generated with different σ_s and σr values with a step of 1 and 0.1, respectively. By varying σ_s and σ_r, we can obtain MEF images with significantly different quality. In the figure, warmer color indicates higher quality score. The corresponding predicted quality scores by our proposed LGIR method are also shown under each image. It can be seen that the MEF result obtained by default parameter values (i.e., $\sigma_s = 100$ and$\sigma_r = 4$) delivers worse visual quality than that obtained by the optimal parameters chosen by our proposed MEF quality evaluation model, i.e., LGIR. Specifically, the upper-right image (chosen by our proposed MEF image quality evaluation model, i.e., $\sigma_r = 95$ and $\sigma_r = 3.2$) preserves much more finer details and exhibits more natural color appearance than the upper-left image (obtained by the default parameter values with $\sigma_r = 100$ and $\sigma_r = 4$). In addition, our proposed LGIR method can also successfully identify the fused results with low visual quality, as indicated by the lower-left and lower-right two examples. All these examples can well demonstrate the effectiveness of LGIR for flexible parameter tuning of MEF algorithms.

5. Conclusion

In this paper, inspired by the global-to-local perception mechanism of the HVS, we propose a novel FR-IQA method called LGIR for MEF images by synthesizing both local and global intermediate references from the input multiple reference images. More specifically, the intermediate references are synthesized in gradient domain, structural tensor

Table 3
PLCC results of each individual quality measure and its combination variations.

Q_{lgs}	✓			✓	✓		✓
Q_{lst}		✓		✓		✓	✓
Q_{gps}			✓		✓	✓	✓
Balloons	0.8260	0.8116	**0.8947**	0.8150	0.8599	0.8147	0.8358
Belgium house	0.9517	0.8381	0.9759	0.9608	0.9820	0.9776	**0.9931**
Lampl	0.8318	0.8590	0.8578	0.8674	0.8767	0.8639	**0.9218**
Candle	0.9390	0.8355	**0.9770**	0.9519	0.9731	0.9590	0.8643
Cave	0.9077	0.3578	0.9293	0.9144	0.9391	0.9293	**0.9421**
Chinese garden	0.8809	0.6531	0.9596	0.8891	0.9556	0.9596	**0.9683**
Farmhouse	0.8159	**0.9787**	0.8616	0.8218	0.8729	0.8617	0.8759
House	0.8947	0.7171	0.9553	0.8993	0.9505	0.9556	**0.9601**
Kluki	0.5157	**0.8402**	0.7446	0.7947	0.7233	0.7418	0.7305
Lamp2	0.9279	0.7933	0.8432	0.9127	0.9432	0.8436	**0.9748**
Landscape	**0.8997**	0.8110	0.7210	0.9335	0.9525	0.7208	0.8838
Lighthouse	0.9753	0.9417	0.9500	0.9729	0.9812	0.9470	**0.9879**
Madison capitol	0.8378	0.7791	0.9368	0.9241	**0.9509**	0.9364	0.9433
Memorial	0.9567	0.9579	0.9689	0.9659	0.9642	**0.9700**	0.9680
Office	0.8689	0.8507	0.8904	0.9429	0.8913	0.8902	**0.9553**
Tower	0.9148	0.8701	0.9370	0.9190	0.9428	0.9480	**0.9580**
Venice	0.8976	0.8925	0.8764	0.8842	0.8793	0.8765	**0.9668**
Average	0.8730	0.8110	0.8987	0.9048	0.9199	0.8938	**0.9252**

Table 4
SRCC results of each individual quality measure and its combination variations.

Q_{lgs}	✓			✓	✓		✓
Q_{lst}		✓		✓		✓	✓
Q_{gps}			✓		✓	✓	✓
Balloons	0.6667	0.6249	0.6667	0.6905	0.7857	0.6667	**0.8095**
Belgium house	0.8503	0.6467	0.9222	0.9222	0.9461	0.9461	**0.9701**
Lampl	0.8571	0.7619	0.8571	0.8810	0.8810	0.8571	**0.9048**
Candle	0.9524	0.7143	0.9286	0.9524	0.9524	0.8810	**0.9762**
Cave	0.7143	0.3429	0.7619	0.7143	0.7619	0.7619	**0.8333**
Chinese garden	**0.8095**	0.2143	0.7143	**0.8095**	0.7857	0.7143	0.7857
Farmhouse	0.8571	0.5476	0.8333	0.8571	**0.9286**	0.8333	**0.9286**
House	0.6667	0.6190	**0.8571**	0.7381	0.8333	**0.8571**	0.8333
Kluki	0.5000	0.4048	0.7143	0.6190	0.6429	0.7143	**0.7381**
Lamp2	0.8095	0.4286	0.6905	0.7619	0.7619	0.6905	**0.9524**
Landscape	0.5714	0.7143	0.6667	0.5952	**0.8095**	0.6667	0.7619
Lighthouse	0.8095	**0.9286**	0.8571	0.8095	0.8095	0.8571	0.8810
Madison capitol	0.7619	0.4048	0.8333	0.8095	**0.8810**	0.8333	0.8095
Memorial	0.8095	0.6667	0.8333	0.7857	0.7619	0.8333	**0.8571**
Office	0.6988	0.6386	0.6747	0.7711	0.7350	0.6747	**0.8555**
Tower	0.7143	0.5238	**0.8571**	0.8095	0.9048	**0.8571**	**0.8571**
Venice	0.7066	0.4671	0.8024	0.7545	0.7066	0.8024	**0.8623**
Average	0.7503	0.5675	0.7923	0.7812	0.8169	0.7909	**0.8597**

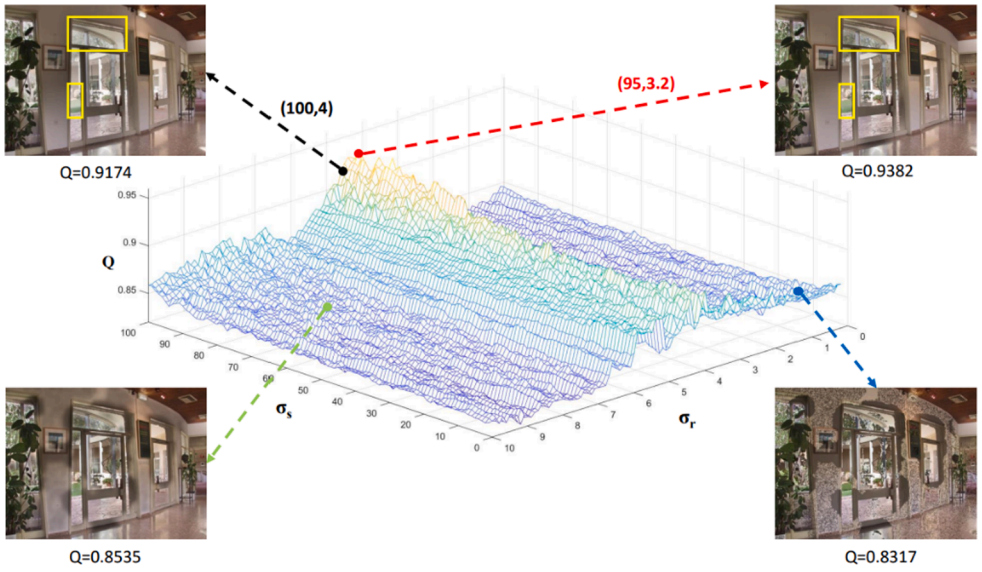

Fig. 7. Predicted quality scores of the fused images generated with different σ_s and σ_r values in [12].

domain, and global perception domain, respectively. In each domain, a single quality measure is derived to reflect the visual quality of the fused image from a specific perspective. In addition, we estimate those quality measures at multiple scales, and fuse all the quality measures together to predict the final quality score via SVR. Experimental results confirm the superiority of the proposed LGIR over some existing relevant algorithms. Besides, we also demonstrate that LGIR is a useful tool to exploit the parameter space and to pick the optimal parameter set that produces MEF image with the best visual quality.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the Zhejiang Natural Science Foundation under Grant LR22F020002; in part by the Natural Science Foundation of China under Grant 61901236; in part by the Natural Science Foundation of Ningbo under Grant 202003N4155, and in part by the Fundamental Research Funds for the Provincial Universities of Zhejiang under Grant SJLZ2020003.

References

[1] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, High Dynamic Range Imaging: Acquisition, in: High Dynamic Range Imaging, Elsevier, 2006, pp. 367–462, https://doi.org/10.1016/B978-012585263-0/50010-5.
[2] R. Yonesaka, Y. Lee, P. Xia, T. Tahara, Y. Awatsuji, K. Nishio, O. Matoba, High dynamic range digital holography and its demonstration by off-axis configuration, IEEE Trans. Ind. Inf. 12 (5) (2016) 1658–1663.
[3] G. Yue, W. Yan, T. Zhou, Referenceless quality evaluation of tone-mapped hdr and multiexposure fused images, IEEE Trans. Ind. Inf. 16 (3) (2020) 1764–1775, https://doi.org/10.1109/TII.2019.2927527.
[4] Z. Li, J. Zheng, Visual-salience-based tone mapping for high dynamic range images, IEEE Trans. Ind. Electron. 61 (12) (2014) 7076–7082, https://doi.org/10.1109/TIE.2014.2314066.

[5] I.R. Khan, S. Rahardja, M.M. Khan, M.M. Movania, F. Abed, A tone-mapping technique based on histogram using a sensitivity model of the human visual system, IEEE Trans. Ind. Electron. 65 (4) (2018) 3469–3479.

[6] K. Gu, S. Wang, G. Zhai, S. Ma, X. Yang, W. Lin, W. Zhang, W. Gao, Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure, IEEE Trans. Multimedia 18 (3) (2016) 432–443, https://doi.org/10.1109/TMM.2016.2518868.

[7] P.J. Burt, The pyramid as a structure for efficient computation, in: Multi-resolution image processing and analysis, Springer, 1984, pp. 6–35.

[8] T. Mertens, J. Kautz, F. Van Reeth, Exposure fusion: A simple and practical alternative to high dynamic range photography, in: Computer Graphics Forum, vol. 28, 2009, pp. 161–171.

[9] S. Raman, S. Chaudhuri, Bilateral filter based compositing for variable exposure photography, Proc. Eurographics (2009).

[10] B.o. Gu, W. Li, J. Wong, M. Zhu, M. Wang, Gradient field multi-exposure images fusion for high dynamic range image visualization, J. Vis. Commun. Image Represent. 23 (4) (2012) 604–610.

[11] Z.G. Li, J.H. Zheng, S. Rahardja, Detail-enhanced exposure fusion, IEEE Trans. Image Process. 21 (11) (2012) 4672–4676, https://doi.org/10.1109/TIP.2012.2207396.

[12] S. Li, X. Kang, Fast multi-exposure image fusion with median filter and recursive filter, IEEE Trans. Consum. Electron. 58 (2) (2012) 626–632, https://doi.org/10.1109/TCE.2012.6227469.

[13] S. Li, X. Kang, J. Hu, Image fusion with guided filtering, IEEE Trans. Image Process. 22 (7) (2013) 2864–2875.

[14] X. Zhang, Benchmarking and comparing multi-exposure image fusion algorithms, Information Fusion.

[15] Z. Wang, Applications of objective image quality assessment methods, IEEE Signal Process Mag. 28 (6) (2011) 137–142.

[16] G. Zhai, X. Wu, X. Yang, W. Lin, W. Zhang, A psychovisual quality metric in free-energy principle, IEEE Trans. Image Process. 21 (1) (2012) 41–52, https://doi.org/10.1109/TIP.2011.2161092.

[17] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang, W. Zhang, Saliency guided quality assessment of screen content images, IEEE Trans. Multimedia 18 (6) (2016) 1098–1110, https://doi.org/10.1109/TMM.2016.2547343.

[18] X. Min, G. Zhai, K.e. Gu, X. Yang, X. Guan, Objective quality evaluation of dehazed images, IEEE Trans. Intell. Transp. Syst. 20 (8) (2019) 2879–2892, https://doi.org/10.1109/TITS.2018.2868771.

[19] T. Li, X. Min, H. Zhao, G. Zhai, Y. Xu, W. Zhang, Subjective and objective quality assessment of compressed screen content videos, IEEE Trans. Broadcast. 67 (2) (2021) 438–449, https://doi.org/10.1109/TBC.2020.3028335.

[20] X. Min, K. Ma, K.e. Gu, G. Zhai, Z. Wang, W. Lin, Unified blind quality assessment of compressed natural, graphic, and screen content images, IEEE Trans. Image Process. 26 (11) (2017) 5462–5474.

[21] H. Duan, G. Zhai, X. Min, Y. Zhu, Y. Fang, X. Yang, Perceptual quality assessment of omnidirectional images, in: 2018 IEEE international symposium on circuits and systems (ISCAS), IEEE, 2018, pp. 1–5.

[22] G. Zhai, W. Sun, X. Min, J. Zhou, Perceptual quality assessment of low-light image enhancement, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 17 (4) (2021) 1–24.

[23] Z. Peng, Q. Jiang, F. Shao, W. Gao, W. Lin, Lggd+: Image retargeting quality assessment by measuring local and global geometric distortions, IEEE Trans. Circuits Syst. Video Technol., doi: 10.1109/TCSVT.2021.3112933.

[24] Q. Jiang, Z. Peng, F. Shao, K. Gu, Y. Zhang, W. Zhang, W. Lin, Stereoars: Quality evaluation for stereoscopic image retargeting with binocular inconsistency detection, IEEE Trans. Broadcast., doi:10.1109/TBC. 460 2021.3113280.

[25] Q. Jiang, W. Zhou, X. Chai, G. Yue, F. Shao, Z. Chen, A full-reference stereoscopic image quality measurement via hierarchical deep feature de-gradation fusion, IEEE Trans. Instrum. Meas. 69 (12) (2020) 9784–9796, https://doi.org/10.1109/TIM.2020.3005111.

[26] M. Liu, K. Gu, G. Zhai, P. Le Callet, W. Zhang, Perceptual reduced reference visual quality assessment for contrast alteration, IEEE Trans. Broadcasting 63 (1) (2017) 71–81, https://doi.org/10.1109/TBC. 2016. 2597545.

[27] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, W. Gao, Reduced-reference image quality assessment in free-energy principle and sparse representation, IEEE Trans. Multimedia 20 (2) (2018) 379–391, https://doi.org/10.1109/TMM.2017.2729020.

[28] G. Zhai, Y. Zhu, X. Min, Comparative perceptual assessment of visual signals using free energy features, IEEE Trans. Multimedia 23 (2021) 3700–3713.

[29] X. Sun, G. Min, K. Zhai, H. Gu, S. Duan, Ma, Mc360iqa: A multi-channel cnn for blind 360-degree image quality assessment, IEEE J. Sel. Top. Signal Process. 14 (1) (2019) 64–77.

[30] Z. Zhang, W. Sun, X. Min, W. Zhu, T. Wang, W. Lu, G. Zhai, A no-reference evaluation metric for low-light image enhancement, in: 2021 IEEE Interna- tional Conference on Multimedia and Expo (ICME), IEEE, 2021, pp. 1–6.

[31] M. ur Rehman, I.F. Nizami, M. Majid, Deeprpn-biqa: Deep architectures with region proposal network for natural-scene and screen-content blind image quality assessment, Displays 71 (2022) 102101.

[32] Y. Huang, H. Xu, Z. Ye, Ye, Image quality evaluation for oled-based smart-phone displays at various lighting conditions, Displays 70 (2021) 102115, https://doi.org/10.1016/j.displa.2021.102115.

[33] R. Hu, Y. Liu, Z. Wang, X. Li, Blind quality assessment of night-time image, Displays 69 (2021) 102045, https://doi.org/10.1016/j.displa.2021.102045.

[34] L. Xu, J. Li, W. Lin, Y. Zhang, Y. Zhang, Y. Yan, Pairwise comparison and rank learning for image quality assessment, Displays 44 (2016) 21–26, https://doi.org/10.1016/j.displa.2016.06.002.

[35] T. Li, X. Min, W. Zhu, Y. Xu, W. Zhang, No-reference screen content video quality assessment, Displays 69 (2021) 102030, https://doi.org/10.1016/j.displa.2021.102030.

[36] K. Gu, X. Xu, J. Qiao, Q. Jiang, W. Lin, D. Thalmann, Learning a unified blind image quality metric via on-line and off-line big training instances, IEEE Trans. Big Data 6 (4) (2020) 780–791.

[37] Q. Jiang, Z. Peng, G. Yue, H. Li, F. Shao, No-reference image contrast evaluation by generating bidirectional pseudoreferences, IEEE Trans. Ind. Inf. 17 (9) (2021) 6062–6072.

[38] Q. Jiang, F. Shao, W. Lin, G. Jiang, Blique-tmi: Blind quality evaluator for tone-mapped images based on local and global feature analyses, IEEE Trans. Circuits Syst. Video Technol. 29 (2) (2019) 323–335, https://doi.org/10.1109/TCSVT.2017.2783938.

[39] X. Wang, Q. Jiang, F. Shao, K.e. Gu, G. Zhai, X. Yang, Exploiting local degradation characteristics and global statistical properties for blind quality assessment of tone-mapped hdr images, IEEE Trans. Multimedia 23 (2021) 692–705, https://doi.org/10.1109/TMM.2020.2986583.

[40] S. Athar, Z. Wang, A comprehensive performance evaluation of image quality assessment algorithms, IEEE Access 7 (2019) 140030–140070.

[41] K. Ma, K. Zeng, Z. Wang, Perceptual quality assessment for multi-exposure image fusion, IEEE Trans. Image Process. 24 (11) (2015) 3345–3356.

[42] P. J. Burt, R. J. Kolczynski, Enhanced image capture through fusion, in: 1993 (4th) international Conference on Computer Vision, IEEE, 1993, pp. 173–182.

[43] Q. Wang, W. Chen, X. Wu, Z. Li, Detail-enhanced multi-scale exposure fusion in yuv color space, IEEE Trans. Circuits Syst. Video Technol. 30 (8) (2020) 2418–2429, https://doi.org/10.1109/TCSVT.2019. 2919310.

[44] A.A. Goshtasby, Fusion of multi-exposure images, Image Vis. Comput. 23 (6) (2005) 611–618.

[45] W. Zhang, W.K. Cham, Gradient-directed multiexposure composition, IEEE Trans. Image Process. 21 (4) (2012) 2318–2323.

[46] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, L. Zhang, Robust multi-exposure image fusion: a structural patch decomposition approach, IEEE Trans. Image Process. 26 (5) (2017) 2519–2532.

[47] H. Duan, G. Zhai, X. Yang, D. Li, W. Zhu, Ivqad 2017: An immersive video quality assessment database, in: 2017 International Conference on Systems, Signals and Image Processing (IWSSIP), IEEE, 2017, pp. 1–5.

[48] G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion, Electron. Lett. 38 (7) (2002) 313–315.

[49] C. S. Xydeas, V. S. Petrovic, Objective pixel-level image fusion performance measure, in: Sensor Fusion: Architectures, Algorithms, and Applications IV, vol. 4051, SPIE, 2000, pp. 89–98.

[50] G. Piella, H. Heijmans, A new quality metric for image fusion, in: International Conference on Image Processing, 2003.

[51] S. Di Zenzo, A note on the gradient of a multi-image, Computer vision, graphics, and image processing 33 (1) (1986) 116–125.

[52] D. Marr, E. Hildreth, Theory of edge detection, Proc. R. Soc. Lond. B Biol. Sci. 207 (1167) (1980) 187–217.

[53] J.K.T. Mertens, F.V. Reeth, Exposure fusion, in: 15th Pacific Conference on Computer Graphics and Applications, IEEE, 2007, pp. 382–390.

[54] H. Hughes, G. Nozawa, F. Kitterle, Global precedence, spatial frequency channels, and the statistics of natural images, J. Cognitive Neuro-science 8 (3) (1996) 197–230.

[55] K.e. Gu, M. Liu, G. Zhai, X. Yang, W. Zhang, Quality assessment considering viewing distance and image resolution, IEEE Trans. Broadcast. 61 (3) (2015) 520–531.

[56] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 1–27.

[57] C. Xydeas, V. Petrovic, Objective image fusion performance measure, Elec-tronics Letters 36 (4) (2000) 308–309.

[58] P.-W. Wang, B. Liu, A novel image fusion metric based on multi-scale analysis, in: International Conference on Signal Processing, 2008, pp. 965–968.

[59] Y. Chen, R.S. Blum, A new automated quality assessment algorithm for image fusion, Image Vis. Comput. 27 (10) (2009) 1421–1432.

[60] D. Kundu, D. Ghadiyaram, A.C. Bovik, B.L. Evans, No-reference quality assessment of tone-mapped HDR pictures, IEEE Trans. Image Process. 26 (6) (2017) 2957–2971.

[61] N. Cvejic, C. Canagarajah, D. Bull, Image fusion metric based on mutual information and tsallis entropy, Electron. Lett. 42 (11) (2006) 626–627.

[62] Y. Zheng, E.A. Essock, B.C. Hansen, A.M. Haun, A new metric based on extended spatial frequency and its application to DWT based fusion algorithms, Information Fusion 8 (2) (2007) 177–192.