Perceptual Quality Assessment for Multi-Exposure Image Fusion

Kede Ma, *Student Member, IEEE*, Kai Zeng, and Zhou Wang, *Fellow, IEEE*

Abstract—Multi-exposure image fusion (MEF) is considered an effective quality enhancement technique widely adopted in consumer electronics, but little work has been dedicated to the perceptual quality assessment of multi-exposure fused images. In this paper, we first build an MEF database and carry out a subjective user study to evaluate the quality of images generated by different MEF algorithms. There are several useful findings. First, considerable agreement has been observed among human subjects on the quality of MEF images. Second, no single state-of-the-art MEF algorithm produces the best quality for all test images. Third, the existing objective quality models for general image fusion are very limited in predicting perceived quality of MEF images. Motivated by the lack of appropriate objective models, we propose a novel objective image quality assessment (IQA) algorithm for MEF images based on the principle of the structural similarity approach and a novel measure of patch structural consistency. Our experimental results on the subjective database show that the proposed model well correlates with subjective judgments and significantly outperforms the existing IQA models for general image fusion. Finally, we demonstrate the potential application of the proposed model by automatically tuning the parameters of MEF algorithms.[1]

Index Terms—Multi-exposure image fusion (MEF), image quality assessment, structural similarity, luminance consistency, subjective evaluations, perceptual image processing.

I. INTRODUCTION

MULTI-EXPOSURE image fusion (MEF) is considered an effective quality enhancement technique that is widely adopted in consumer electronics [1]. MEF takes a sequence of images with different exposure levels as inputs and synthesizes an output image that is more informative and perceptually appealing than any of the input images [2], [3]. MEF fills the gap between high dynamic range (HDR) natural scenes and low dynamic range (LDR) pictures captured by normal digital cameras. Comparing with typical HDR imaging techniques which first construct an HDR image from the source sequence and then tone-map it to an LDR image, MEF bypasses the intermediate HDR image construction step

and directly yields an LDR image that can be displayed on standard viewing devices.

Since first introduced in 1980's [2], MEF has been an active research topic and attracted an increasing amount of attention in recent years [3]–[13]. With many MEF algorithms at hand, it becomes pivotal to compare their performance, so as to find the best algorithm as well as directions for further advancement. Because the human visual system (HVS) is the ultimate receiver in most applications, subjective evaluation is a straightforward and reliable approach to evaluate the quality of fused images [7], [13]–[16]. Although expensive and time consuming [17], a comprehensive subjective user study has several benefits. First, it provides useful data to study human behaviors in evaluating perceived quality of fused images. Second, it supplies a test set to evaluate and compare the relative performance of classical and state-of-the-art MEF algorithms. Third, it is useful to validate and compare the performance of existing objective image quality assessment (IQA) models in predicting the perceptual quality of fused images. This will in turn provide insights on potential ways to improve them.

Over the past decade, substantial effort has been made to develop objective IQA models for image fusion applications [18]–[26]. Most of them are designed for general-purpose image fusion applications, not specifically for MEF, and some of them can only work with the case of two input images. Furthermore, little has been done to compare them with (or calibrate against) subjective data that contains a wide variety of source sequences and MEF algorithms.

In this work, we aim to tackle the problem of perceptual quality assessment of MEF images. We build one of the first databases dedicated to subjective evaluation of MEF images. The database contains 17 source sequences with multiple exposure levels ($\geq$ 3) and the fused images generated by 8 classical and state-of-the-art MEF algorithms. Based on the database, we carry out a subjective user study to evaluate and compare the quality of the fused images. We observe considerable agreement between human subjects, and not a single MEF algorithm produces the best quality for all test images. More importantly, we find that existing objective quality models for general image fusion are very limited in predicting perceived quality of MEF images. This motivates us to develop a novel objective IQA model for MEF images. Our model is inspired by the structural similarity (SSIM) index [27], whose philosophy is that the HVS is highly adapted for extracting structural information from natural scenes. To compare the structures of multiple

[1]The subjective database and the MATLAB code of the proposed model will be made available online. Preliminary results of Section III were presented at the *6th International Workshop on Quality of Multimedia Experience*, Singapore, 2014.

patches from different exposures, we introduce a novel measure of patch structural consistency. Furthermore, to balance between finer-scale detail preservation and coarser-scale luminance consistency [16], we adopt a multi-scale approach [28], where with the scale shifting from fine to coarse, SSIM-based structural comparison captures image distortions from fine details to large-scale luminance variations. Experimental results show that the proposed model well correlates with subjective judgments and significantly outperforms existing objective IQA models for general image fusion. The value of objective models are beyond measuring and comparing MEF images and algorithms. A reliable objective model can play a key role in the design and optimization of novel MEF algorithms. To demonstrate this potential, we apply the proposed model to automatic parameter tuning of an advanced MEF algorithm. Numerical experiments show that the proposed model provides a useful tool to exploit the parameter space and to pick the optimal parameter that produces MEF images of the best perceptual quality.

II. RELATED WORK

The problem of MEF can be generally formulated as

$$\mathbf{Y}(i) = \sum_{k=1}^{K} \mathbf{W}_k(i)\mathbf{X}_k(i), \qquad (1)$$

where K is the number of multi-exposure input images in the source sequence, $\mathbf{X}_k(i)$ and $\mathbf{W}_k(i)$ represent the luminance value (or the coefficient amplitude in the transform domain) and the weight at the i-th pixel in the k-th exposure image, respectively. $\mathbf{Y}$ denotes the fused image. The weight factor $\mathbf{W}_k(i)$ is often spatially adaptive and bears information regarding the relative structural detail and perceptual importance at different exposure levels. Depending on the specific models for structural information and perceptual importance, MEF algorithms differ in the computation of $\mathbf{W}_k$.

A significant number of MEF algorithms have been proposed, ranging from simple weighted averaging to sophisticated methods based on advanced statistical image models. Local and global energy weighting approaches are the simplest ones, which employ the local or global energy in the image to determine $\mathbf{W}_k$. Dated back to 1984, Burt [2] first employed Laplacian pyramid decomposition for binocular image fusion. Later in 1994, Burt and Kolczynski applied this decomposition to MEF, where they selected the local energy of pyramid coefficients and the correlation between pyramids within the neighborhood to compute $\mathbf{W}_k$. Laplacian pyramid turns out to be an effective scheme in image fusion to avoid unnatural appearance and unwanted artifacts introduced by fusion in the spatial domain [5]. Goshtasby [4] partitioned each source image into several non-overlapping blocks and selected the block with the highest entropy to construct the fused image. Due to the non-overlapping partition, the method inevitably suffers from blocking artifacts. Mertens *et al.* [5] adopted proper contrast, high saturation and well exposure as quality measures to guide the fusion process in a multiresolution fashion. Using the same weighting map to guide the fusion in the spatial domain, the method tends to introduce

artificial edges and color distortions in the fused image. Bilateral filter is used in [6] to calculate edge information, which is subsequently employed to compute the weights. This method puts no constraints on global luminance consistency and often produces dark appearance of the fused image. Song *et al.* [7] firstly estimated the initial image by maximizing the visual contrast and scene gradient, and synthesized the fused image by suppressing reversals in image gradients. In practice, this method tends to produce color saturated images. Zhang and Cham [8] constructed visibility and consistency measures from gradient information and used them as the weighting factors. The adoption of gradient direction enables the method to fuse a source sequence captured in a dynamic scene that has moving objects. A similar gradient-based MEF method is proposed in [9]. Based on [5], Li *et al.* [10] enhanced the details of a given fused image by solving a quadratic optimization problem. A median filter and recursive filter based MEF method is developed in [11] by taking local contrast, brightness and color dissimilarity into consideration. The use of median filter also enables the method to handle dynamic scenes. More recently, Li *et al.* [12] proposed a guided filter to control the roles of pixel saliency and spatial consistency when constructing $\mathbf{W}_k$. Shen *et al.* [13] embedded perceived local contrast and color saturation into a conditional random field and derived $\mathbf{W}_k$ based on maximum a posteriori estimation.

Despite the increasing interests in developing objective IQA models for various image fusion applications, systematic and comprehensive evaluation and comparison of these models have been largely lacking. To validate the performance of objective IQA models, subjective user study is necessary. Toet and Franken [14] examined the perceptual quality of multi-scale image fusion schemes, where only night-time outdoor scenes and very simple fusion methods were included in the study. Petrović [15] reported subjective assessment results for multi-sensor image fusion algorithms. However, the number of input images was limited to 2 and most test images were monochrome aerial pictures. Moreover, state-of-the-art image fusion algorithms are missing from the experiment. To demonstrate the effectiveness of their fusion algorithm, Song *et al.* [7] conducted two groups of paired comparison tests through both on-site and Web platforms, where the subjective experimental results only include few examples. Shen *et al.* [13] reported subjective evaluation results to verify the performance of their algorithm. However, the number of test images involved is also limited. To the best of our knowledge, a comprehensive subjective user study that contains sufficient test sequences and compares a wide variety of MEF algorithms has not been reported in the literature.

An excellent survey on objective IQA models for image fusion applications can be found in [29]. Here we only provide a brief overview: Qu *et al.* [18] combined the mutual information between the fused and multiple input images to evaluate image quality. Xydeas and Petrovic [19] extracted edge information using the Sobel operator and employed edge strength as the main feature in assessing the quality of fused images. A similar idea was employed in [20], where Wang and Liu retrieved edge strength using a two-scale Haar wavelet.

Fig. 1. Input source image sequences contained in the database. Each image sequence is represented by one image, which is a fused image of the sequence that has the best quality in the subjective test.

TABLE I
INFORMATION ABOUT SOURCE INPUT IMAGE SEQUENCES

Source sequence	Size	Image courtesy
House	$340 \times 512 \times 4$	Tom Mertens
Light house	$340 \times 512 \times 3$	HDRsoft
Chinese garden	$340 \times 512 \times 3$	Bartlomiej Okonek
Madison capitol	$384 \times 512 \times 30$	Chaman Singh Verma
Tower	$512 \times 341 \times 3$	Jacques Joffre
Balloons	$339 \times 512 \times 9$	Erik Reinhard
Kluki	$341 \times 512 \times 3$	Bartlomiej Okonek
Cave	$384 \times 512 \times 4$	Bartlomiej Okonek
Belgium house	$384 \times 512 \times 9$	Dani Lischinski
Landscape	$341 \times 512 \times 3$	HDRsoft
Office	$340 \times 512 \times 6$	MATLAB
Venice	$341 \times 512 \times 3$	HDRsoft
Lamp1	$384 \times 512 \times 15$	Martin Ĉadík
Memorial	$512 \times 381 \times 16$	Paul Debevec
Lamp2	$342 \times 512 \times 6$	HDR projects
Farmhouse	$341 \times 512 \times 3$	HDR projects
Candle	$364 \times 512 \times 10$	HDR projects

Zheng *et al.* [21] computed spatial frequency using multi-directional gradient filters and estimated the quality of fused images based on activity levels. Inspired by the SSIM index [27] for general-purpose IQA, Piella and Heijmans [22] developed three models to predict fused image quality based on the universal quality index [30]. Cvejic *et al.* [23] and Yang *et al.* [24] also built their quality measures upon structural information theory. Chen and Varshney [25] estimated local saliency based on edge intensities and combined saliency with global contrast sensitive function. Chen and Blum [26] applied contrast sensitivity filter in the frequency domain and then pool local information preservation scores to produce a global quality measure. All aforementioned models are designed for general-purpose image fusion applications, not specifically for MEF. In most cases, the performance of existing models were demonstrated using limited examples only, without being thoroughly validated based on subjective user studies. Most of the algorithms were elaborated with the source sequence containing two input images only. However, most of these algorithms can be extended to the cases of multiple input images except [23] and [24].

III. SUBJECTIVE QUALITY ASSESSMENT OF MEF IMAGES

A. *Image Database and Subjective User Study*

Seventeen high-quality natural source image sequences are selected to cover diverse image content including indoor and outdoor views, natural sceneries, and man-made architectures. All source image sequences are shown in Fig. 1 and listed in Table I. All of them contain at least 3 input images that represent underexposed, overexposed, and in-between cases. For visualization purpose, in Fig. 1, we select the best quality fused image in terms of subjective evaluations to represent each source sequence.

Eight MEF algorithms are selected, which include simple operators such as 1) local energy weighted linear combination and 2) global energy weighted linear combination, as well as advanced MEF algorithms such as 3) Raman09 [6], 4) Gu12 [9], 5) ShutaoLi12 [11], 6) ShutaoLi13 [12], 7) Li12 [10], and 8) Mertens07 [5]. These algorithms are chosen to cover a diverse types of MEF methods in terms of methodology and behavior. In all cases, default parameter

settings are adopted without tuning for better quality. Eventually, a total of 136 fused images are generated, which are divided into 17 image sets of 8 images each, where the images in the same set are created from the same source sequence. An example is shown in Fig. 2, which includes a source sequence at three exposure levels (Fig. 2(a1-a3)) and eight fused images (Fig. 2(b-i)). Note that different MEF algorithms produce substantially different fused images in terms of perceptual appearance and quality. Thus, quality assessment of fused images are desirable to pick the one with the best quality.

The subjective testing environment was setup as a normal indoor office workspace of ordinary illumination level, with no reflecting ceiling walls and floor. All images are displayed on an LCD monitor at a resolution of 2560×1600 pixel with Truecolor (32bit) at 60Hz. The monitor was calibrated in accordance with the recommendations of ITU-T BT.500 [31]. A customized MATLAB figure window was used to render the images on the screen. During the test, all 8 fused images from the same set are shown to the subject at the same time on one computer screen at actual pixel resolution but in random spatial order. The study adopted a multi-stimulus quality scoring strategy without showing the reference sequence. A total of 25 naïve observers, including 15 male and 10 female subjects aged between 22 and 30, participated in the subjective experiment. The subjects are allowed to move their positions to get closer or further away from the screen for better observation. All subject ratings were recorded with pen and paper during the study. To minimize the influence of visual fatigue, the length of a session was limited to a maximum of 30 minutes. For each image set, the subject was asked to give an integer score that best reflects the perceptual quality of each fused image. The score ranges from 1 to 10, where 1 denotes the worst quality and 10 the best.

B. *Subjective Data Analysis*

After the subjective experiment, 2 outlier subjects were removed based on the outlier removal scheme in [31], resulting in 23 valid subjects. The final quality score for each

Fig. 2.   An example of multi-exposure source image sequence (a1, a2, a3) and fused images (b)-(i) created by different MEF algorithms.

individual image is computed as the average of subjective scores, namely the mean opinion score (MOS), from all valid subjects. Considering the MOS as the "ground truth", the performance of individual subjects can be evaluated by calculating the correlation coefficient between individual subject ratings and MOS values for each image set, and then averaging the correlation coefficients of all image sets. Pearson linear correlation coefficient (PLCC) and Spearman's rand-order correlation coefficient (SRCC) are employed as the evaluation criteria [32]. Both criteria range from 0 to 1, where higher values indicate better performance. The mean and standard deviation (std) of the results are depicted in Fig. 3. It can be seen that each individual subject performs quite

consistently with relatively low variations for different image content. The average performance across all individual subjects is also given in the rightmost columns of Fig. 3.

C. Performance of Existing MEF Algorithms

We use the MOS values of the 8 MEF algorithms described in Section III-A to evaluate and compare their performance. The mean and std of MOS values over all 17 image sets are summarized in Fig. 4. It is worth mentioning that this only provides a rough comparison of the relative performance of the MEF algorithms, where default parameters are used without fine tuning. Besides, computational complexity is not a factor under consideration.
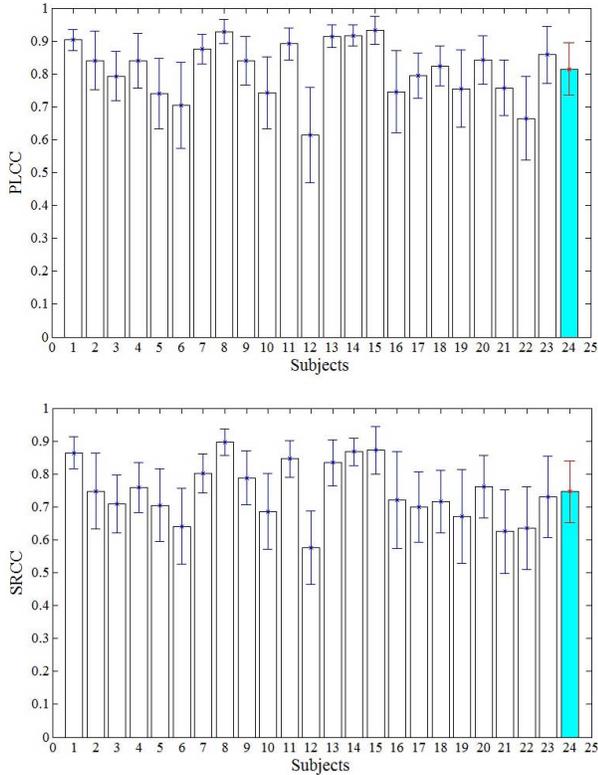
Fig. 3. PLCC and SRCC between individual subject rating and MOS. Rightmost column: performance of an average subject.
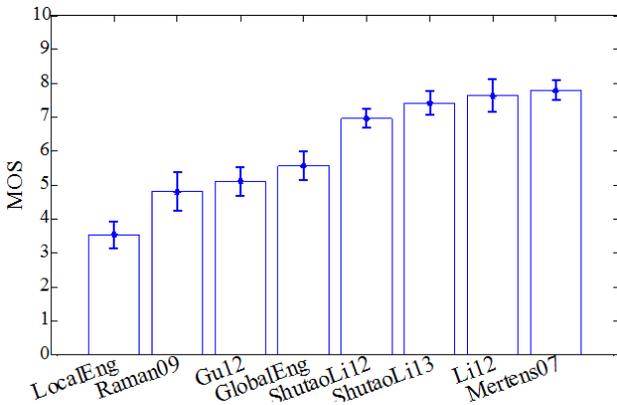
Fig. 4. Mean and std of subjective rankings of individual image fusion algorithms across all image sets.

From the subjective test results, we have several observations. First, from the sizes of the error bars, we observe that subjects agree with each other to a significant extent on the performance of any individual MEF algorithm, but the performance difference between different MEF algorithms is sometimes small (when compared with the error bars). Second, Mertens's method [5] achieves the best performance on average, while Li's method [10], which is the second best on average, is actually a detail-enhanced algorithm built upon Mertens's method [5]. It has very similar average performance and a larger error-bar than Mertens's method [5]. This suggests that detail enhancement might be useful to create perceptually appealing results on some images, but may also

create unwanted artifacts in some other images, and the overall performance gain is not always guaranteed in the current approaches. Third, comparing local energy weighting with global energy weighting approaches, the former focuses more on enhancing local structures while the latter emphasizes more on global luminance consistency. The large performance gap between them indicates that maintaining large-scale luminance consistency may be an important factor in determining the quality of the fused images. Fourth, not a single algorithm produces fused images with the best perceptual quality for all image sets. This suggests that there is still room for future improvement, and proper combination of the ideas used in different MEF algorithms has the potential to further improve the performance.

D. Performance of Existing Objective IQA Models

Using the above database, we test the performance of 9 existing objective IQA models for image fusion. Since the source sequences in the database consist of at least 3 input images, models that can only work with the case of 2 input images [23], [24] are excluded from the comparison. For the purpose of fairness, all models are tested using their default parameter settings. Note that to obtain a reasonable result, we take the absolute value of the objective score in [21]. Three implementations of the algorithm in [22] were proposed and the one with the best performance is reported here.

Tables II and III summarize the evaluation results, which is somewhat disappointing because state-of-the-art IQA models do not seem to provide adequate predictions of perceived quality of fused images. Even the models with the best performance, such as Xydeas's [19] and Wang's [20] methods, are only moderately correlated with subjective scores.

The above test results also provide some useful insights regarding the general approaches used in IQA models. First, models based on entropy computations of pixel intensity values and transform coefficients [33], [34] have poor correlation with perceptual quality. The reason may be that the quality of fused images is highly content dependent and only entropy of image intensity/coefficient histogram is insufficient in capturing the perceptual distortions introduced by MEF processes. Second, local structure-preservation based models, such as SSIM and gradient based approaches applied in spatial or transform domain [19]–[22], provide the most promising results so far. However, they are often unsuccessful in capturing the degradations of luminance consistency across the image space. This suggests that more accurate objective IQA models may be developed by achieving a good balance between assessing local structure preservation and evaluating large-scale luminance consistency.

IV. OBJECTIVE QUALITY ASSESSMENT OF MEF IMAGES

A. Proposed Objective Quality Assessment Model

Following the general construction of SSIM [27], we first examine how the information in the multi-exposure image sequence is preserved in the fused image at each spatial location. Direct use of the SSIM algorithm [27], however, is impossible, which requires a single perfect quality reference image. How to work with multiple input images in MEF is

TABLE II

PLCC PERFORMANCE EVALUATION OF THE PROPOSED MODEL AGAINST 9 EXISTING MODELS

Source sequence	[33]	[34]	[35]	[19]	[20]	[21]	[22]	[25]	[26]	Mean subject	Proposed model
Balloons	-0.5417	0.7605	-0.5283	0.7050	0.4389	0.6648	0.5038	-0.3417	0.2765	0.9019	0.9297
Belgium house	-0.3853	0.1744	-0.4001	0.8017	0.6261	0.5608	0.5020	-0.9610	-0.5719	0.8207	0.9312
Lamp1	-0.1205	-0.4785	-0.1655	0.7294	0.7281	0.4017	0.4320	-0.9180	0.0095	0.8595	0.8907
Candle	0.2654	-0.7290	0.1253	0.9388	0.8917	0.1063	0.1789	-0.9563	-0.8006	0.7461	0.9513
Cave	-0.2143	0.0526	-0.1574	0.6946	0.8136	0.6206	0.6301	-0.8479	0.1562	0.6774	0.7719
Chinese garden	-0.2238	-0.2942	-0.2687	0.7684	0.8358	0.4812	0.4087	-0.5263	0.3800	0.6858	0.9563
Farmhouse	-0.6407	0.5039	-0.7019	0.6408	0.5997	0.6929	0.2163	-0.4077	-0.2729	0.7574	0.8632
House	-0.2885	-0.5237	-0.3076	0.6207	0.5956	0.4762	0.4805	-0.8928	0.4154	0.8693	0.8414
Kluki	-0.0912	0.0209	-0.0480	0.3914	0.3585	-0.1124	-0.0489	-0.7126	0.5952	0.7721	0.8242
Lamp2	-0.3870	0.6205	-0.4039	0.8445	0.7518	0.6493	0.5955	-0.6828	0.6131	0.7960	0.8291
Landscape	-0.2107	0.5393	-0.1148	0.3204	0.4479	0.0806	0.0306	0.4677	0.9015	0.7165	0.7457
Lighthouse	-0.2955	-0.2613	-0.2411	0.8383	0.6553	0.2462	-0.0226	0.1694	0.6680	0.7330	0.9420
Madison capitol	-0.4061	0.0313	-0.4434	0.6278	0.4225	0.5414	0.6184	-0.7133	0.4900	0.7620	0.9141
Memorial	-0.4182	0.4454	-0.4446	0.8276	0.6780	0.5877	0.7332	-0.9493	-0.2164	0.8184	0.8981
Office	-0.2030	0.3022	-0.1817	0.4980	0.4725	0.3161	0.3238	-0.5301	0.5309	0.8590	0.9628
Tower	-0.4777	-0.1155	-0.4822	0.7719	0.8347	0.5718	0.5941	-0.1729	0.6806	0.7112	0.9561
Venice	-0.3584	-0.0222	-0.3226	0.7949	0.6543	0.4786	0.2796	-0.4498	0.6790	0.6705	0.9699
Average	**-0.2939**	**0.0604**	**-0.2992**	**0.6949**	**0.6356**	**0.4332**	**0.3798**	**-0.5544**	**0.2667**	**0.7739**	**0.8928**

TABLE III

SRCC PERFORMANCE EVALUATION OF THE PROPOSED MODEL AGAINST 9 EXISTING MODELS

Source sequence	[33]	[34]	[35]	[19]	[20]	[21]	[22]	[25]	[26]	Mean subject	Proposed model
Balloons	-0.4286	0.7143	-0.3810	0.6667	0.5000	0.5952	0.4524	-0.3810	0.3333	0.9005	0.8333
Belgium house	-0.2994	0.0000	-0.2515	0.7785	0.7545	0.5389	0.4671	-0.9102	-0.7066	0.7872	0.9701
Lamp1	-0.0714	-0.3810	-0.0238	0.7857	0.6190	0.4762	0.4048	-0.9286	0.1905	0.8445	0.9762
Candle	0.3571	-0.6667	0.4048	0.9762	0.7857	0.1667	0.5476	-0.9762	-0.4524	0.6761	0.9286
Cave	-0.1190	0.0238	-0.0238	0.7143	0.8095	0.6429	0.5714	-0.8571	0.3333	0.6675	0.8333
Chinese garden	-0.2143	-0.2857	-0.1905	0.6905	0.7857	0.5476	0.5238	-0.7143	0.4048	0.6920	0.9286
Farmhouse	-0.4524	0.5000	-0.4524	0.7381	0.8095	0.5000	0.2857	-0.6905	-0.1905	0.6945	0.9286
House	-0.0476	-0.6905	-0.0238	0.5952	0.4524	0.5238	0.4048	-0.9048	0.4524	0.8641	0.8571
Kluki	-0.2381	0.1667	-0.2381	0.2619	0.2857	0.0476	0.1190	-0.7381	0.7381	0.7429	0.7857
Lamp2	-0.4286	0.8333	-0.4762	0.7619	0.6190	0.6905	0.5476	-0.6905	0.8810	0.7859	0.7143
Landscape	-0.7381	0.5476	-0.7381	0.0238	0.4048	0.1429	0.1429	0.7143	0.8333	0.6562	0.5238
Lighthouse	-0.8333	-0.4286	-0.8333	0.5000	0.4286	0.3810	0.0714	0.2381	0.6905	0.6971	0.8810
Madison capitol	-0.2143	-0.3095	-0.1429	0.5238	0.3571	0.5238	0.4762	-0.7381	0.5476	0.7655	0.8810
Memorial	0.0000	0.8095	0.0238	0.7619	0.5476	0.5238	0.6667	-0.8571	-0.2381	0.8059	0.8571
Office	-0.1928	0.0843	-0.1566	0.2771	0.3976	0.3856	0.4579	-0.6266	0.4940	0.8011	0.7832
Tower	-0.4762	-0.2143	-0.4524	0.5714	0.5238	0.5952	0.5714	-0.0238	0.5952	0.6873	0.9524
Venice	-0.3353	0.2994	-0.3353	0.9102	0.7306	0.5629	0.3114	-0.4551	0.6587	0.6127	0.9341
Average	**-0.2784**	**0.059**	**-0.2524**	**0.6198**	**0.5771**	**0.4614**	**0.4131**	**-0.5612**	**0.3274**	**0.7459**	**0.8570**

a major challenge here. Let $\{\mathbf{x}_k\} = \{\mathbf{x}_k | 1 \leq k \leq K\}$ denote the set of image patches extracted from the same spatial location in the source image sequence of K multi-exposure images, and let $\mathbf{y}$ be the corresponding patch in the fused image, respectively. Here $\mathbf{x}_k$ for all k and $\mathbf{y}$ are N dimensional column vectors, where N is the number of pixels in the patch and each entry is given by the intensity value of a pixel in the patch. Given the patch set $\{\mathbf{x}_k\}$ as a reference, the goal here is to define a quality measure of the fused image patch $\mathbf{y}$.

A useful approach we learn from the SSIM approach is to look at an image patch from three separate aspects: luminance, contrast and structure. An easy approach to implement this is to decompose any given image patch into three components:

$$
\begin{aligned}
\mathbf{x}_k &= \|\mathbf{x}_k - \mu_{\mathbf{x}_k}\| \cdot \frac{\mathbf{x}_k - \mu_{\mathbf{x}_k}}{\|\mathbf{x}_k - \mu_{\mathbf{x}_k}\|} + \mu_{\mathbf{x}_k} \\
&= \|\tilde{\mathbf{x}}_k\| \cdot \frac{\tilde{\mathbf{x}}_k}{\|\tilde{\mathbf{x}}_k\|} + \mu_{\mathbf{x}_k} \\
&= c_k \cdot \mathbf{s}_k + l_k,
\end{aligned} \tag{2}
$$

where $\|\cdot\|$ denotes the l^2 norm of a vector, $\mu_{\mathbf{x}_k}$ is the mean value of the patch, and $\tilde{\mathbf{x}}_k = \mathbf{x}_k - \mu_{\mathbf{x}_k}$ is a mean-removed patch, or a zero-mean patch that contains the contrast and structure information only. The scalar $l_k = \mu_{\mathbf{x}_k}$, the scalar $c_k = \|\tilde{\mathbf{x}}_k\|$, and the unit-length vector $\mathbf{s}_k = \tilde{\mathbf{x}}_k / \|\tilde{\mathbf{x}}_k\|$ roughly represent the luminance, contrast and structure components of $\mathbf{x}_k$, respectively.

In the case of MEF, direct preservation of the luminance of the local source image patches (that are all somewhat under- or over-exposed) is of low relevance with regard to the overall image quality, and thus we exclude it from the local patch comparison described below. This differentiates our model from other SSIM-based quality models such as [22]–[24], and the more critical differences are in the way we work with the contrast and structure components of multiple source images.

The visibility of the local patch structure largely depends on local contrast. On one hand, the higher the contrast, the better the visibility. On the other hand, too large contrast may lead to unrealistic representation of the local structure. Considering all input source image patches as realistic

capturing of the scene, the patch that has the highest contrast among them would correspond to the best visibility under the realisticity constraint. Therefore, the desired contrast of the fused image patch is determined by the highest contrast of all source image patches:

$$\hat{c} = \max_{\{1 \leq k \leq K\}} c_k = \max_{\{1 \leq k \leq K\}} \|\tilde{\mathbf{x}}_k\|. \tag{3}$$

Different from contrast, the structures of local image patches are denoted by unit-length vectors $\mathbf{s}_k$ for $1 \leq k \leq K$, each of which points to a different direction in the vector space. The desired structure of the fused image patch corresponds to another direction in the same vector space that best represents the structures of all source image patches. A simple model to account for this relationship is given by

$$\bar{\mathbf{s}} = \frac{\sum_{k=1}^{K} w(\tilde{\mathbf{x}}_k) \mathbf{s}_k}{\sum_{k=1}^{K} w(\tilde{\mathbf{x}}_k)} \quad \text{and} \quad \hat{\mathbf{s}} = \frac{\bar{\mathbf{s}}}{\|\bar{\mathbf{s}}\|}, \tag{4}$$

where $w(\cdot)$ is a weighting function that determines the contribution of each source image patch in the structure of the fused image patch. Intuitively, the contribution should increase with the strength of the image patch. A straightforward approach that conforms with such intuition is to employ a power weighting function given by

$$w(\tilde{\mathbf{x}}_k) = \|\tilde{\mathbf{x}}_k\|^p, \tag{5}$$

where $p \geq 0$ is an exponent parameter. With various choices of the value of p, this general formulation leads to a family of weighting functions with different physical meanings. The larger the p value, the more emphasis is put on the patches that have relatively larger strength. Specifically, $p = 0$ corresponds to straightforward direction average (where low and high contrast source patches are accounted for equally); $p = 1$ corresponds to length-weighted direction average; $p = 2$ corresponds to energy-weighted direction average; and $p = \infty$ corresponds to picking the direction corresponding to the patch that has the largest vector length among all patches.

It remains to determine the value of p. Instead of fixing p to be a constant, here we propose an automatic approach that chooses p at each spatial location adaptively. The motivation is to adjust the relative weighting factors in (4) based on the consistency between the structures of the source image patches. To implement the idea, we first need to quantify the consistency between the set of structural vectors $\{\tilde{\mathbf{x}}_k\}$. Each of the vectors points to a specific direction in the vector space. In the extreme case, when one vector is a contrast enhanced variation of another (i.e., there is no structural changes between the vectors), both vectors will point to the same direction. Therefore, we define a structure consistency measure between a set of vectors based on the degree of direction agreement between them. In particular, we compute

$$R(\{\tilde{\mathbf{x}}_k\}) = \frac{\|\sum_{k=1}^{K} \tilde{\mathbf{x}}_k\|}{\sum_{k=1}^{K} \|\tilde{\mathbf{x}}_k\|}. \tag{6}$$

By triangular inequality, $0 \leq R \leq 1$, and a lager R indicates stronger consistency between the set of vectors. To better understand this, two examples are illustrated in Fig. 5. The sum of a set of source patch vectors corresponds to connecting
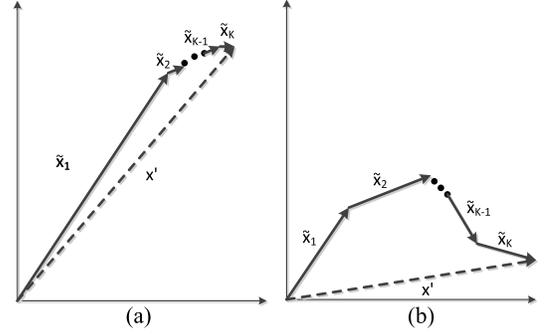
Fig. 5. Illustration of structure consistency measure in the vector space. The vectors in (a) have much stronger structure consistency than those in (b).

the starts and ends of these vectors one by one, which creates a new vector $\mathbf{x}' = \sum_{k=1}^{K} \tilde{\mathbf{x}}_k$ pointing from the origin to the end of the last patch vector. The length of the new vector $\|\mathbf{x}'\|$ (the numerator of Eq. (6)) is usually smaller than adding the length of all patch vectors together (the denominator of Eq. (6)), but when all the source patch vectors point to exactly the same direction, these two quantities are equal, leading to the maximal possible value of $R = 1$. Fig. 5(a) and 5(b) are two examples showing the cases of stronger (larger R) and weaker (smaller R) patch structure consistencies, respectively. When the value of R is small, the multiple source patches convey different structures with similar strength, and thus it is more appropriate to assign them similar weights, resulting in a smaller desired p value. On the other hand, when R is large and the structures of the patches are similar, a stronger patch would have higher contrast and be more resistent to distortions such as noise, and thus should be given a higher weight, leading to a larger desired p value. We propose an empirical approach to account for this, which is given by

$$p = \tan \frac{\pi R}{2}. \tag{7}$$

In the extreme case when all patches agree with each other ($R = 1$), a value of $p = \infty$ is chosen, while at the other extreme when there is no consistency in patch structures ($R = 0$), $p = 0$ is selected.

Once the value of p is determined at each spatial location, Eq. (4) is employed to compute $\hat{\mathbf{s}}$, which is subsequently combined with $\hat{c}$ in Eq. (3) to yield a new vector

$$\hat{\mathbf{x}} = \hat{c} \cdot \hat{\mathbf{s}}. \tag{8}$$

Following the construction of the SSIM approach [27], we use a simplified definition of the SSIM index to evaluate local image quality:

$$S(\{\mathbf{x}_k\}, \mathbf{y}) = \frac{2\sigma_{\hat{\mathbf{x}}\mathbf{y}} + C}{\sigma_{\hat{\mathbf{x}}}^2 + \sigma_{\mathbf{y}}^2 + C}, \tag{9}$$

where $\sigma_{\hat{\mathbf{x}}}^2$, $\sigma_{\mathbf{y}}^2$ and $\sigma_{\hat{\mathbf{x}}\mathbf{y}}$ denote the local variances of $\hat{\mathbf{x}}$ and $\mathbf{y}$, and the local covariance between $\hat{\mathbf{x}}$ and $\mathbf{y}$, respectively. C is a small positive stabilizing constant that accounts for the saturation effects of the visual system at low contrast [27]. Note that since luminance is not considered a highly relevant component in the current scenario, the luminance term is not included in the above structural comparison assessment.
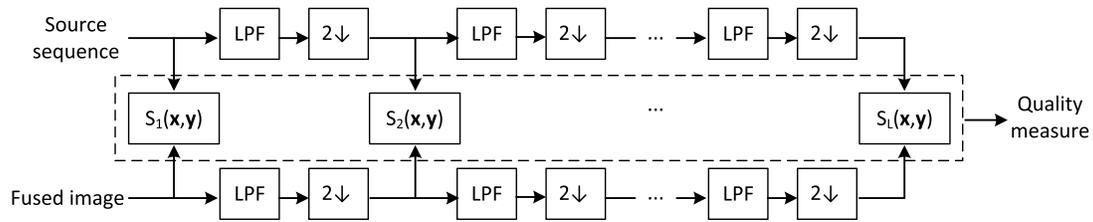
Fig. 6. Diagram of multi-scale structure comparison. LPF: low-pass filtering; 2 ↓: downsampling by a factor of 2.
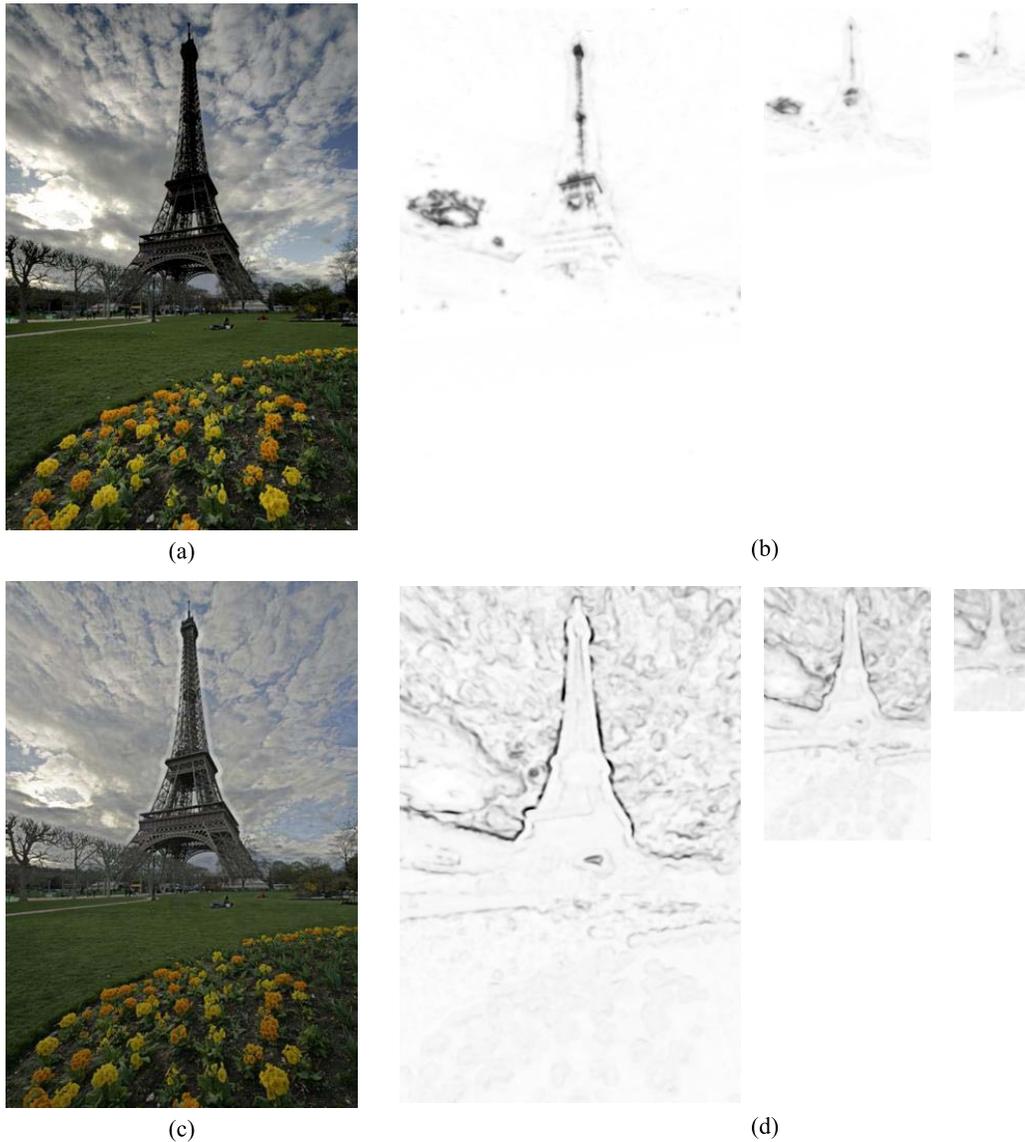
Fig. 7. Examples of multi-scale structural quality maps. (a) Fused image created by Mertens' algorithm [5]. (b) The structural quality maps of (a). (c) Fused image created by local energy weighting. (d) The structural quality maps of (c). In (b) and (d), higher brightness indicates better quality.

The local structure comparison of Eq. (9) is applied using a sliding window approach across the entire image, resulting in a quality map indicating how the structural details are preserved at each spatial location. Fig. 7 provides a visual demonstration of the quality maps of two fused images created from the same source sequence, where brighter regions in the maps indicate better structure preservation. It appears that the quality maps accurately predict local structure preservations in the fused images. For instance, Fig. 7(a) generated by Mertens' algorithm [5] fails to preserve the fine details in the tower and the brightest cloud region in the left part of the image. Such detail loss is well reflected in the corresponding quality maps. By contrast, Fig. 7(c) computed by the local energy weighting algorithm creates unnatural artifacts near the strong edges of the Tower. Such structural distortion is again clearly indicated in the corresponding structural quality maps.

TABLE IV

STATISTICAL SIGNIFICANCE MATRIX BASED ON QUALITY PREDICTION RESIDUALS. A SYMBOL "1" MEANS THAT THE PERFORMANCE OF THE ROW MODEL IS STATISTICALLY BETTER THAN THAT OF THE COLUMN MODEL, A SYMBOL "0" MEANS THAT THE ROW MODEL IS STATISTICALLY WORSE, AND A SYMBOL "-" MEANS THAT THE ROW AND COLUMN MODELS ARE STATISTICALLY INDISTINGUISHABLE

	[33]	[34]	[35]	[19]	[20]	[21]	[22]	[25]	[26]	Proposed
[33]	-	-	-	-	-	-	-	-	-	0
[34]	-	-	-	0	-	-	-	-	-	0
[35]	-	-	-	-	-	-	-	-	-	0
[19]	-	1	-	-	-	1	-	-	-	0
[20]	-	-	-	-	-	-	-	-	-	0
[21]	-	-	-	0	-	-	-	-	-	0
[22]	-	-	-	-	-	-	-	-	-	0
[25]	-	-	-	-	-	-	-	-	-	0
[26]	-	-	-	-	-	-	-	-	-	0
Proposed	1	1	1	1	1	1	1	1	1	-

The quality map is averaged to obtain an overall structural quality measure of the fused image

$$Q(\mathbf{Y}) = \frac{1}{M} \sum_{j=1}^{M} S(\{\mathbf{x}_k\}(j), \mathbf{y}(j)), \tag{10}$$

where j is the spatial patch index and M is the total number of patches.

Based on the subjective test results, analysis and discussions presented in Section III, an ideal quality MEF image should not only preserve as much as possible the local structural details in the source sequence, but also achieve good luminance consistency across space. Although it is difficult to directly determine the exact luminance value at each spatial location in the fused image, the relative luminance/brighness values of different regions across space in the source input images provide useful references. For example, Fig. 2(a1)-(a3) show the three input images of the source sequence "Venice", where the absolute local luminance of the sky, the buildings and the water areas in the right part of the image is difficult to decide, but their relative luminance values are apparent — the sky is brighter than the buildings, and the buildings are brighter than the water area. Such larger-scale relative luminance structure cannot be properly captured if only a local windowing approach like the one described above is adopted. Instead, the image content needs to be observed in larger windows or at coarser scales. The principle behind the multi-scale SSIM (MS-SSIM) approach [27], [28] provides a natural fit to what we need, where fine-scale structure comparisons capture local structural detail loss and distortions, while at the coarse scales, the same comparisons reveal the consistency of large-range luminance patterns. As illustrated in Fig. 6, we iteratively apply a low-pass filter followed by downsampling the filtered images by a factor of 2. We index the original scale of the image or the finest scale as Scale 1, and the coarser scale after $l-1$ iterations as Scale l. At the l-th scale, the structure comparison (9) is conducted and denoted as $S_l(\hat{\mathbf{x}}, \mathbf{y})$. When this is computed for all scales, a set of multi-scale quality maps are obtained, as exemplified in Fig. 7. By pooling the quality map at each scale using Eq. (10), we obtain a set of scale-level quality scores $\{Q_l(\mathbf{Y})\}$. The overall quality is then computed by combining scale-level quality scores using a similar method as in [28]:

$$Q(\mathbf{Y}) = \prod_{l=1}^{L} [Q_l(\mathbf{Y})]^{\beta_l}, \tag{11}$$

where L is the total number of scales and β_l is the weight assigned to the l-th scale.

The proposed model does not involve any training process or introduce any new parameter. All parameters are inherited from previous publications. These include $C = (0.03D)^2$ from [27], where D is the dynamic range of intensity values (For 8 bits/pixel gray-scale images, $D = 255$), $L = 3$, and the normalized fine-to-coarse scale weights are given by $\{\beta_1, \beta_2, \beta_3\} = \{0.0710, 0.4530, 0.4760\}$ [28].

B. Validation

We validate the proposed objective model using the database presented in Section III and compare its performance against the nine objective IQA models for image fusion described in Section II. Tables II and III summarize the PLCC and SRCC evaluation results, where we have also included the performance of an average subject (as discussed in Section III). It can be seen that the proposed method delivers the best performance in predicting subjective quality of fused images on almost every set of test images. On average, it not only performs much better than existing IQA models (which are designed for quality assessment of general image fusion), but also outperforms an average subject.

To ascertain that the improvement of the proposed model is statistically significant, we carried out a statistical significance analysis by following the approach introduced in [36]. First, a nonlinear regression function is applied to map the objective quality scores to predict the subjective scores. We observe that the prediction residuals all have zero-mean, and thus the model with lower variance is generally considered better than the one with higher variance. We conduct a hypothesis testing using F-statistics. Since the number of samples exceeds 40, the Gaussian assumption of the residuals approximately hold based on the central limit theorem [37]. The test statistic is the ratio of variances. The null hypothesis is that the prediction residuals from one quality model come from the same distribution and are statistically indistinguishable (with 95% confidence) from the residuals from another model.
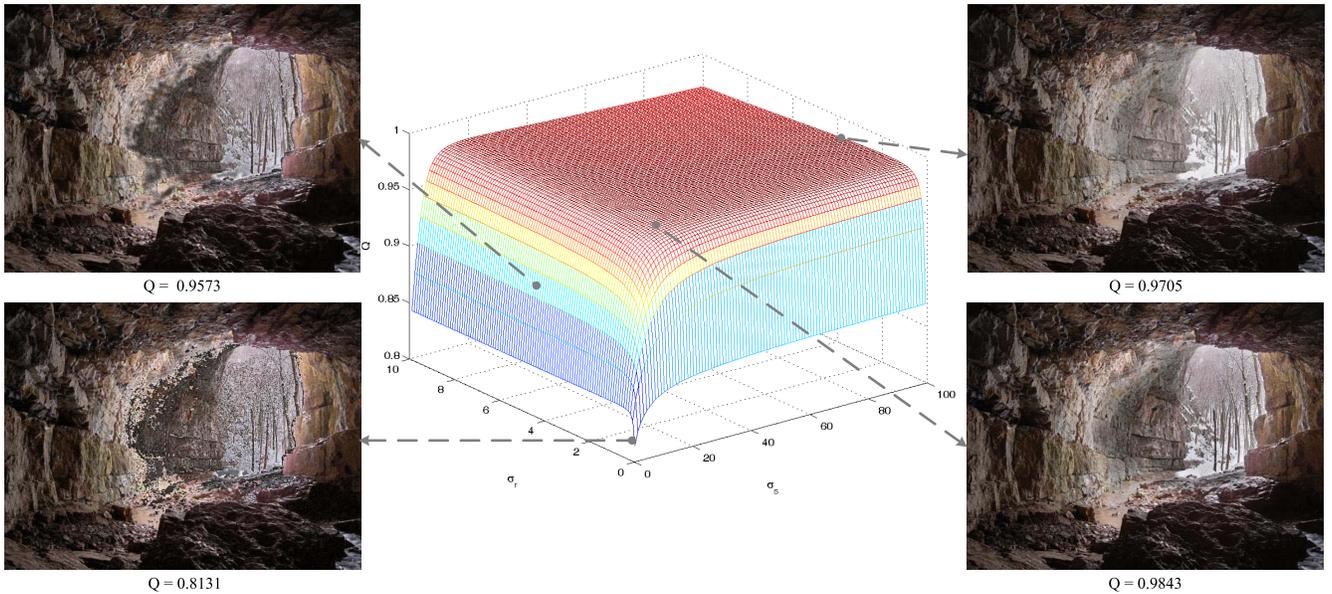
Fig. 8. "Cave" image sequence fused by Li's method [11] over a wide range of σ_s and σ_r values. Higher surface in the middle plot indicates a higher value of the proposed objective quality model, suggesting better quality of the fused image.

After comparing every possible pairs of objective models, the results are summarized in Table IV, where a symbol "1" means the row model performs significantly better than the column model, a symbol "0" means the opposite, and a symbol "-" indicates that the row and column models are statistically indistinguishable. It can be observed that most existing IQA models for image fusion are statistically indistinguishable from each other, while the proposed model is statistically better than all other methods.

C. Potential Application: Automatic Parameter Tuning of MEF Algorithms

The application scope of objective IQA models is much broader than performance assessment and comparison of MEF algorithms. A more interesting application area is to use them to guide the design and optimization of novel MEF algorithms. Here we use automatic parameter tuning of MEF models as an example to demonstrate the potential values of the proposed IQA model.

Many MEF algorithms include one or more free parameters whose best values largely depend on the image content. In practice, these values often need to be selected manually, which is difficult, inconvenient, and time-consuming. Objective quality models that can automatically tune the parameters in MEF algorithms are highly desirable, especially when the volume of images being processed is large. To demonstrate the usefulness of the proposed method, here we apply the proposed model to automatically tune the parameters of the MEF algorithm proposed in [11].

The algorithm in [11] extracts local contrast and brightness for static scenes and color dissimilarity for dynamic scenes to compute the initial weighting map. A recursive filter is then employed to smooth the map while preserving the edge information. The space and range supports of the recursive filter are controlled by two parameters σ_s and σ_r. The default

values of σ_s and σ_r are chosen under the guidance of the model in [19], which results in a fixed set of values given by $\sigma_s = 100$ and $\sigma_r = 4$. In Fig. 8, we plot the quality score of the proposed model as a function of σ_s and σ_r for the "Cave" sequence fused by the chosen algorithm [11], where higher surface in the plot indicates a higher quality of the fused image in terms of the proposed model. From the figure, we have several interesting observations. First, the perceptual quality of the fused image varies significantly with σ_s and σ_r when they are relatively small. Second, careful inspections and comparisons of the fused images together with their corresponding quality scores suggest that the proposed model is a good perceptual quality indicator and provides a useful tool to automatically choose the optimal values of σ_s and σ_r. Third, the upper-right image corresponds to the default parameters chosen by the model in [19], while the lower-right image corresponds to the optimal parameter values chosen by the proposed model. Comparing the two images, we can see that the best quality image indicated by the proposed model better preserves the fine detail and color information of the tree trunks and the rocks in the middle part of the image.

V. CONCLUSION AND DISCUSSION

In this work, we tackle the problem of perceptual quality assessment of MEF images. Our major contributions are threefold. First, we construct a database dedicated to MEF IQA, which is perhaps the first of its kind that contains a wide variety of image content and fused images created by different kinds of MEF algorithms. We carry out subjective experiment and data analysis on the database, which lead to three useful findings: 1) although the behaviors of individual subjects varies, there is generally a considerable agreement between them on the quality of fused images; 2) there is still room for designing better MEF algorithms

that are more robust to the variation of image content; 3) state-of-the-art IQA models for general image fusion applications do not seem to provide adequate predictions of perceived quality of fused images. Second, we propose an objective IQA model for MEF images based on the multiscale SSIM principle [27], [28] and a novel measure of structural patch consistency. The proposed model measures the local structure preservation at the fine scales and also captures luminance consistency at coarser scales. Experiments on the benchmark database show that the proposed model significantly outperforms existing IQA models designed for image fusion. Third, we demonstrate the usefulness of the proposed model by applying it to automatic parameter tuning of a state-of-the-art MEF algorithm and obtained promising initial results.

For future research, MEF IQA still has many challenging and interesting problems to be solved. First, the proposed metric is mainly based on the principle of structural similarity. Other principles that have been successfully used in full-reference IQA may be exploited in the context of MEF. These include information theoretic and natural scene statistical approaches [38]–[40], the adaptive linear system decomposition framework [41]–[43], the feature similarity method [44], and visual attention and saliency-based approaches [39], [45]–[47]. Second, most existing objective quality models, including the proposed one, work with the luminance component only. Proper accounting for color distortions have great potentials to improve the performance of the objective quality model [42], [48]. Third, all source sequences that constitute the subjective database are nearly static, but the natural scenes we encounter in practice are often dynamic and contain moving objects [49]. It is useful to generalize the proposed model to account for dynamic scenes. Fourth, how to integrate the quality model into consumer electronics to capture high quality fused image in real-time is another challenging problem yet to be explored.

Besides quality assessment of MEF images, objective quality models for other image fusion applications are largely lacking. These include images fused from multi-focus images in photography applications [50]; images fused from hyperspectral images in remote sensing applications; and images generated by merging different imaging modalities in medical imaging applications. They are related to MEF IQA but also have fundamentally interesting differences. It is interesting to explore the common characteristics shared by these applications as well as the unique features of each individual application.

ACKNOWLEDGMENT

REFERENCES

[1] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. San Mateo, CA, USA: Morgan Kaufmann, 2010.

[2] P. J. Burt, "The pyramid as a structure for efficient computation," in *Multiresolution Image Processing and Analysis*. Berlin, Germany: Springer-Verlag, 1984.

[3] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *Proc. 4th IEEE Int. Conf. Comput. Vis.*, May 1993, pp. 173–182.

[4] A. A. Goshtasby, "Fusion of multi-exposure images," *Image Vis. Comput.*, vol. 23, no. 6, pp. 611–618, 2005.

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

[6] S. Raman and S. Chaudhuri, "Bilateral filter based compositing for variable exposure photography," in *Proc. Eurographics*, 2009, pp. 1–4.

[7] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, "Probabilistic exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 341–357, Jan. 2012.

[8] W. Zhang and W.-K. Cham, "Gradient-directed multiexposure composition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2318–2323, Apr. 2012.

[9] B. Gu, W. Li, J. Wong, M. Zhu, and M. Wang, "Gradient field multi-exposure images fusion for high dynamic range image visualization," *J. Vis. Commun. Image Represent.*, vol. 23, no. 4, pp. 604–610, 2012.

[10] Z. G. Li, J. H. Zheng, and S. Rahardja, "Detail-enhanced exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4672–4676, Nov. 2012.

[11] S. Li and X. Kang, "Fast multi-exposure image fusion with median filter and recursive filter," *IEEE Trans. Consum. Electron.*, vol. 58, no. 2, pp. 626–632, May 2012.

[12] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[13] R. Shen, I. Cheng, and A. Basu, "QoE-based multi-exposure fusion in hierarchical multivariate Gaussian CRF," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2469–2478, Jun. 2013.

[14] A. Toet and E. M. Franken, "Perceptual evaluation of different image fusion schemes," *Displays*, vol. 24, no. 1, pp. 25–37, 2003.

[15] V. Petrović, "Subjective tests for image fusion evaluation and objective metric validation," *Inf. Fusion*, vol. 8, no. 2, pp. 208–216, Apr. 2007.

[16] K. Zeng, K. Ma, R. Hassen, and Z. Wang, "Perceptual evaluation of multi-exposure image fusion algorithms," in *Proc. 6th Int. Workshop Quality Multimedia Exper.*, Sep. 2014, pp. 7–12.

[17] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[18] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, pp. 313–315, 2002.

[19] C. S. Xydeas and V. S. Petrovic, "Objective pixel-level image fusion performance measure," *Proc. SPIE, Sensor Fusion, Archit., Algorithms, Appl. IV*, vol. 4051, pp. 89–98, Apr. 2000.

[20] P.-W. Wang and B. Liu, "A novel image fusion metric based on multi-scale analysis," in *Proc. IEEE 9th Int. Conf. Signal Process.*, Oct. 2008, pp. 965–968.

[21] Y. Zheng, E. A. Essock, B. C. Hansen, and A. M. Haun, "A new metric based on extended spatial frequency and its application to DWT based fusion algorithms," *Inf. Fusion*, vol. 8, no. 2, pp. 177–192, 2007.

[22] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Sep. 2003, pp. III-173–III-176.

[23] N. Cvejic, A. Loza, D. Bull, and N. Canagarajah, "A similarity metric for assessment of image fusion algorithms," *Int. J. Signal Process.*, vol. 2, no. 3, pp. 178–182, 2005.

[24] C. Yang, J.-Q. Zhang, X.-R. Wang, and X. Liu, "A novel similarity based quality metric for image fusion," *Inf. Fusion*, vol. 9, no. 2, pp. 156–160, 2008.

[25] H. Chen and P. K. Varshney, "A human perception inspired quality metric for image fusion based on regional information," *Inf. Fusion*, vol. 8, no. 2, pp. 193–207, 2007.

[26] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," *Image Vis. Comput.*, vol. 27, no. 10, pp. 1421–1432, 2009.

[27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[28] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th IEEE Asilomar Conf. Signals, Syst., Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.

[29] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 94–109, Jan. 2012.

[30] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[31] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document Rec. ITU-R BT.500-12, Nov. 1993.

[32] VQEG. (Apr. 2000). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment.* [Online]. Available: http://www.vqeg.org

[33] M. Hossny, S. Nahavandi, and D. Creighton, "Comments on 'information measure for performance of image fusion,'" *Electron. Lett.*, vol. 44, no. 18, pp. 1066–1067, Aug. 2008.

[34] N. Cvejic, C. N. Canagarajah, and D. R. Bull, "Image fusion metric based on mutual information and Tsallis entropy," *Electron. Lett.*, vol. 42, no. 11, pp. 626–627, 2006.

[35] Q. Wang, Y. Shen, and J. Jin, "Performance evaluation of image fusion techniques," in *Image Fusion: Algorithms and Applications*. New York, NY, USA: Academic, 2008, pp. 469–492.

[36] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[37] D. C. Montgomery, *Applied Statistics and Probability for Engineers*, 6th ed. New York, NY, USA: Wiley, 2013.

[38] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[39] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[40] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.

[41] Z. Wang and E. P. Simoncelli, "An adaptive linear system framework for image distortion analysis," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Sep. 2005, pp. III-1160–III-1163.

[42] U. Rajashekar, Z. Wang, and E. P. Simoncelli, "Quantifying color image distortions based on adaptive spatio-chromatic signal decompositions," in *Proc. 16th IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 2213–2216.

[43] R. M. Figueras i Ventura, U. Rajashekar, Z. Wang, and E. P. Simoncelli, "Contextually adaptive signal representation using conditional principal component analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar./Apr. 2008, pp. 877–880.

[44] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[45] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 2945–2948.

[46] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 193–201, Apr. 2009.

[47] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014.

[48] U. Rajashekar, Z. Wang, and E. P. Simoncelli, "Perceptual quality assessment of color images using adaptive signal representation," *Proc. SPIE, Human Vis. Electron. Imag. XV*, vol. 7527, p. 75271L, Feb. 2010.

[49] Z. Li, J. Zheng, Z. Zhu, and S. Wu, "Selectively detail-enhanced fusion of differently exposed images with moving objects," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4372–4382, Oct. 2014.

[50] R. Hassen, Z. Wang, and M. M. A. Salama, "Objective quality assessment for multiexposure multifocus image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2712–2724, Sep. 2015.
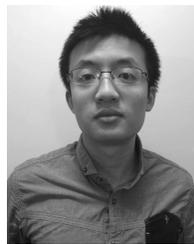
Kede Ma (S'13) received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2012, and the M.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. His research interest lies in perceptual image processing.

Kai Zeng received the B.E. and M.A.Sc. degrees in electrical engineering from Xidian University, Xian, China, in 2009, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2013. He is currently a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo. His research interests include computational video and image pattern analysis, multimedia communications, and image and video processing (coding, denoising, analysis, and representation), with an emphasis on image and video quality assessment and corresponding applications. He was a recipient of the IEEE Signal Processing Society Student Travel Grant at the 2010 and 2012 IEEE International Conference on Image Processing, and the prestigious 2013 Chinese Government Award for Outstanding Students Abroad.

Zhow Wang (S'99–M'02–SM'12–F'14) received the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin, in 2001. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image processing, coding, and quality assessment; computational vision and pattern analysis; multimedia communications; and biomedical signal processing. He has over 100 publications in these fields with over 24 000 citations (Google Scholar). He was a member of the IEEE Multimedia Signal Processing Technical Committee (2013–2015). He served as an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2009–2014), *Pattern Recognition* (2006–present), and the IEEE SIGNAL PROCESSING LETTERS (2006–2010), and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (2013–2014 and 2007–2009), the *EURASIP Journal of Image and Video Processing* (2009–2010), and *Signal, Image and Video Processing* (2011–2013). He was a recipient of the 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, the 2013 IEEE Signal Processing Best Magazine Paper Award, the 2009 IEEE Signal Processing Society Best Paper Award, the 2009 Ontario Early Researcher Award, and the ICIP 2008 IBM Best Student Paper Award (as senior author).