

# Perceptual Evaluation for Multi-Exposure Image Fusion of Dynamic Scenes

Yuming Fang<sup>1b</sup>, Senior Member, IEEE, Hanwei Zhu, Student Member, IEEE, Kede Ma<sup>1b</sup>, Member, IEEE, Zhou Wang, Fellow, IEEE, and Shutao Li<sup>1b</sup>, Fellow, IEEE

**Abstract**—A common approach to high dynamic range (HDR) imaging is to capture multiple images of different exposures followed by multi-exposure image fusion (MEF) in either radiance or intensity domain. A predominant problem of this approach is the introduction of the ghosting artifacts in dynamic scenes with camera and object motion. While many MEF methods (often referred to as deghosting algorithms) have been proposed for reduced ghosting artifacts and improved visual quality, little work has been dedicated to perceptual evaluation of their deghosting results. Here we first construct a database that contains 20 multi-exposure sequences of dynamic scenes and their corresponding fused images by nine MEF algorithms. We then carry out a subjective experiment to evaluate fused image quality, and find that none of existing objective quality models for MEF provides accurate quality predictions. Motivated by this, we develop an objective quality model for MEF of dynamic scenes. Specifically, we divide the test image into static and dynamic regions, measure structural similarity between the image and the corresponding sequence in the two regions separately, and combine quality measurements of the two regions into an overall quality score. Experimental results show that the proposed method significantly outperforms the state-of-the-art. In addition, we demonstrate the promise of the proposed model in parameter tuning of MEF methods. The subjective database and the MATLAB code of the proposed model are made publicly available at <https://github.com/h4nwei/MEF-SSIMd>.

**Index Terms**—High dynamic range imaging, multi-exposure image fusion, ghosting, image quality assessment, structural similarity.

## I. INTRODUCTION

ONE of the major bottlenecks of current sensors and displays lies in their limited dynamic range [1], which are unable to reproduce the full luminance levels of many

Manuscript received February 27, 2018; revised March 28, 2019; accepted August 23, 2019. Date of publication September 16, 2019; date of current version November 4, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61822109 and Grant 61571212 and in part by the Henry Fok Education Foundation under Grant 161061. This article was presented in part at the IEEE International Conference on Image Processing, Beijing, China, September 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Patrick Le Callet. (Corresponding author: Kede Ma.)

Y. Fang and H. Zhu are with the School of Information Management, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: fa0001ng@e.ntu.edu.sg; hanwei.zhu@outlook.com).

K. Ma is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: kede.ma@cityu.edu.hk).

Z. Wang is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: zhou.wang@uwaterloo.ca).

S. Li is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: shutao\_li@hnu.edu.cn).

Digital Object Identifier 10.1109/TIP.2019.2940678



Fig. 1. Ghosting artifacts due to object motion. (a) Source image sequence by courtesy of Fabrizio Pece. (b) Pece10 [10]. (c) SPD-MEF [11].

realistic natural scenes. During the past decade, various high dynamic range (HDR) imaging techniques have been developed to steadily improve the dynamic range of sensors. A common theme in computational HDR imaging is to capture multiple pictures of the same scene at different exposure levels, followed by multi-exposure image fusion (MEF) in either radiance or intensity domain. If the fusion is performed in radiance domain (*i.e.*, HDR reconstruction), tone mapping operators are necessary to display the reconstructed HDR images on devices with a low dynamic range (LDR). If the fusion is performed in intensity domain after applying a camera response function (CRF), a visually appealing LDR image with abundant details may be directly produced, ready for display.

The main technical impediments to most MEF algorithms are dynamic scenes, which contain moving objects. A small misalignment between two exposures would easily result in the ghosting artifacts during fusion (see Fig. 1). In recent years, extensive effort has been put on developing MEF methods for dynamic scenes (*i.e.*, the deghosting algorithms) [2], [3]. In contrast, only limited work has been done to evaluate the visual quality of deghosting results. Early subjective experiments [4], [5] only involved a small set of dynamic scenes and a limited number of MEF algorithms, whose results become less relevant with many new algorithms being proposed recently. For objective quality assessment, Ma *et al.* developed the MEF-SSIM index [6] for MEF of static scenes, which has been successfully applied to perceptual optimization of MEF algorithms [7], [8]. Tursun *et al.* [9] proposed one of the first quality models for MEF of dynamic scenes, which,

however, requires the camera parameters as input, limiting its practical applications.

Motivated by the lack of perceptual quality assessment for MEF of dynamic scenes, we first establish a database that contains 20 multi-exposure sequences of dynamic scenes, together with 180 fused images generated by nine state-of-the-art MEF algorithms. A subjective experiment is carried out using two-alternative forced choice (2AFC), where a considerable consensus among subjects is observed. Inspired by the philosophy behind MEF-SSIM [6], we describe an objective quality model to account for dynamic scenes with emphasis on quantifying the ghosting artifacts, leading to the MEF-SSIM<sub>d</sub> index. Specifically, we divide the test image into static and dynamic regions, measure structural similarity (SSIM) between the image and the corresponding sequence in the two regions separately, and average quality measurements of the two regions to obtain an overall quality score. Experimental results show that the proposed MEF-SSIM<sub>d</sub> significantly outperforms state-of-the-art quality models for MEF. In addition, we demonstrate that MEF-SSIM<sub>d</sub> is useful in guiding the parameter tuning of MEF methods, resulting in fused images with reduced ghosting artifacts and improved visual quality.

## II. RELATED WORK

In this section, we provide a review of recent MEF methods for dynamic scenes in both radiance and intensity domains. Subjective and objective quality assessment for MEF is also shortly discussed.

### A. MEF Methods for Dynamic Scenes

HDR reconstruction algorithms perform MEF in radiance domain, where the value is linear with respect to the exposure time. Eden *et al.* [3] stitched multiple images at varying orientations and exposures by selecting proper luminance values to construct the HDR image. Gallo *et al.* [12] made use of reference patches to detect inconsistent motion, and fused consistent patches in gradient domain to reduce boundary artifacts. Sen *et al.* [2] and Simakov *et al.* [13] adopted a bidirectional similarity measure for HDR reconstruction in an energy minimization framework. Lee *et al.* [14] cast HDR reconstruction as a rank minimization problem, and estimated a binary ghost indication matrix with sparsity and connectivity constraints. A similar rank minimization method was proposed in [15], where humans were involved to help detect moving objects. Kalantari and Ramamoorthi [16] applied optical flow techniques to align the input sequence, and then fed it into a convolutional neural network for HDR reconstruction. Wu *et al.* [17] described an end-to-end solution to HDR reconstruction and tone mapping based on deep neural networks.

Several other MEF methods for dynamic scenes work in intensity domain after applying CRFs. Median threshold bitmap [18] for image alignment was adopted in [10] to detect motion and to select the best exposure for fusion. Li and Kang [19] used a median filter to remove moving objects. Hu *et al.* [20] performed motion estimation for forward warping, and filled holes in the warped sequence due to

erroneous and incomplete matches. Qin *et al.* [21] performed a more flexible motion estimation, where patches can be rotated or scaled for better searching. The intensity mapping function (IMF) [22] has also been used to reduce ghosting artifacts. Li *et al.* [23] used a bidirectional normalization-based method to detect motion-inconsistent pixels, and applied IMF to correct them. Ma *et al.* [11] proposed a structural patch decomposition for MEF, and combined it with IMF to reject inconsistent motion.

### B. Quality Assessment of MEF Methods

Although many MEF algorithms for dynamic scenes have been proposed, research in perceptual quality assessment of multi-exposure fused images is quite limited. Srikantha and Sidibé [4] summarized 17 deghosting methods proposed before 2011, and compared them qualitatively. Hadziabdic *et al.* [24] conducted a subjective test to evaluate four deghosting algorithms on nine natural scenes. A larger database of 36 dynamic scenes was built recently [25] with subjective data collected from image processing experts. Tursun *et al.* [5] built a separate dataset consisting of ten dynamic scenes with seven deghosting algorithms, and carried out a subjective user study based on the 2AFC method.

Liu *et al.* [26] conducted an excellent review of objective quality models for general-purpose image fusion. Here we only review closely related studies that are used for comparison in Section V. Cvejic *et al.* [27] adopted Tsallis entropy to quantify the perceived quality of fused images, while Hossny *et al.* [28] made use of mutual information. Xydeas and Petrovic [29] explored edge strength and orientation preservation in fused images. Wang and Liu [30] measured edge preservation in wavelet domain at two scales. Zheng *et al.* [31] computed spatial frequency errors based on gradient information along four orientations. Inspired by the SSIM index [32], Piella and Heijmans [33] found a way to measure structural similarity between the input sequence and fused image. Chen and Varshney [34] applied the human perception theory in image fusion quality assessment. It has been experimentally verified that the above general-purpose quality models may not work well for the specific MEF application [35]. Ma *et al.* [6] built the first MEF database of static scenes, and developed the objective MEF-SSIM model by decomposing an image patch into three components: signal strength, signal structure and mean intensity. Tursun *et al.* [9] proposed an objective quality model for MEF of dynamic scenes, which produces three quality maps to localize ghosting, gradient inconsistency and visual difference artifacts respectively. The requirement of knowing the camera settings (*i.e.*, exposure time, ISO, f-number, and CRF) limits the applicability of the model to real-world scenarios, where only the fused image and the corresponding LDR sequence are available.

## III. SUBJECTIVE QUALITY ASSESSMENT FOR MEF OF DYNAMIC SCENES

In this section, we detail the construction of our MEF database of dynamic scenes. We then describe our subjective





Fig. 2. Multi-exposure sequences of dynamic scenes used in the subjective experiment. Each sequence is represented by the best-quality fused image. All images are resized for neat display.

experiment and data analysis, resulting in a number of useful observations [36].

#### A. MEF Database of Dynamic Scenes

We collect 20 multi-exposure sequences of dynamic natural scenes, including indoor and outdoor environments, rigid and deformable structures, clean and noisy regions, and small and large object motion patterns (see Fig. 2). Each sequence contains at least three images representing under-exposure, over-exposure, and in-between captures. We do not consider camera motion in our experiment, as it is usually small and uniform in practice, and can be well handled by setting a tripod or using image registration algorithms [18], [37]. In other words, all sequences in the database are perfectly aligned.

We select nine state-of-the-art MEF algorithms for dynamic scenes to generate fused images, including Pece10 [10], Sen12 [2], Li12 [19], Hu13 [20], Lee14 [14], Li14 [23], Photomatix [38], Qin15 [21], and SPD-MEF [11]. These methods are chosen to represent a variety of design philosophies, including pixel-based and patch-based methods, fusion in radiance domain followed by tone mapping and fusion in intensity domain directly, and SSIM based and low-rank based algorithms. The dehazing results are either generated by the original authors or using the publicly available implementations with default settings. Specifically, for algorithms that perform MEF in radiance domain (*i.e.*, HDR reconstruction), the Debevec and Malik’s approach [39] is used to recover the CRF (including gamma correction) and generate radiance sequences. Lee14 [14] adopts the MATLAB function `tonemap()` to obtain LDR images. Sen12, Pece10, and Hu13 fuse aligned LDR sequences using Mertens09 [40]. Photomatix is a commercial HDR software. Eventually, 180 fused images are produced with samples shown in Fig. 3.

#### B. Subjective Experiment

The subjective experiment is setup in a normal indoor environment. The display is a true-color LCD monitor at

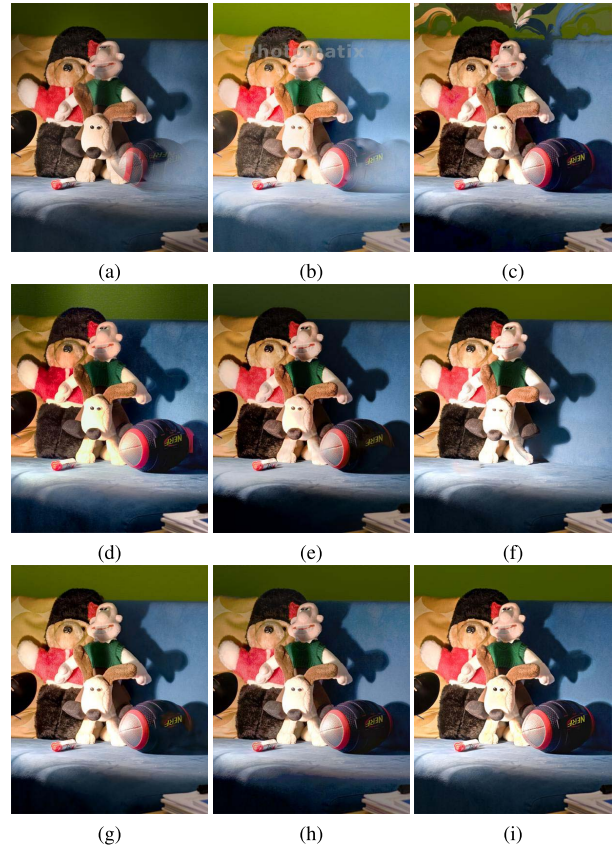


Fig. 3. Samples fused images from nine MEF algorithms in the subjective experiment. (a) Li12 [19]. (b) Photomatix [38]. (c) Qin15 [21]. (d) Lee14 [14]. (e) Sen12 [2]. (f) Pece10 [10]. (g) Hu13 [20]. (h) Li14 [23]. (i) SPD-MEF [11].

a resolution of  $1920 \times 1080$  pixels, and is calibrated in accordance with the ITU-T BT.500 recommendations [41]. The luminance of the background behind the monitor is about  $20 \text{ cd/m}^2$ . The ambient illumination does not directly reflect off the display.

We collect human opinions using the 2AFC method, where participants are shown two fused images along with three input exposures, and are asked to choose the one of higher visual quality (see Fig. 4). We invite 60 subjects (40 males and 20 females) with normal or corrected-to-normal visual acuity, aged from 18 to 40, to participate in the experiment. They have no previous experience in image quality assessment. The experiment starts with a training session using ten image pairs independent of the test session. The subjects are instructed to focus on visual artifacts and structural information preservation, and are allowed to move their positions to get closer or further away from the screen for better view experience. In the test session, a complete experiment includes  $20 \times \binom{9}{2} = 720$  paired comparisons. To reduce the fatigue effect, we divide the experiment into three sub-sessions, in which participants are able to compare 240 image pairs within 30 minutes. Each subject takes part in one sub-session, and each image pair is ranked exactly 20 times.

#### C. Subjective Data Analysis

The resulting subjective data is a preference tensor  $\mathbf{C}$  with  $\mathbf{C}(i, j, k)$  representing the number of times the fused image

TABLE I  
PREFERENCE MATRIX ACCUMULATED ACROSS 20 NATURAL SCENES

Algorithm	Li12	Lee14	Photomatix	Qin15	Pece10	Sen12	Hu13	Li14	SPD-MEF	SUM
Li12 [19]	0	134	86	68	52	60	52	54	15	521
Lee14 [14]	266	0	137	119	141	71	99	73	53	959
Photomatix [38]	314	263	0	189	184	115	108	102	59	1334
Qin15 [21]	332	281	211	0	211	143	123	109	52	1431
Pece10 [10]	348	259	216	220	0	127	160	117	94	1541
Sen12 [2]	340	329	285	257	273	0	157	177	91	1909
Hu13 [20]	348	301	292	277	240	243	0	159	111	1971
Li14 [23]	346	327	298	291	283	223	241	0	111	2120
SPD-MEF [11]	385	347	341	348	306	309	289	289	0	2614

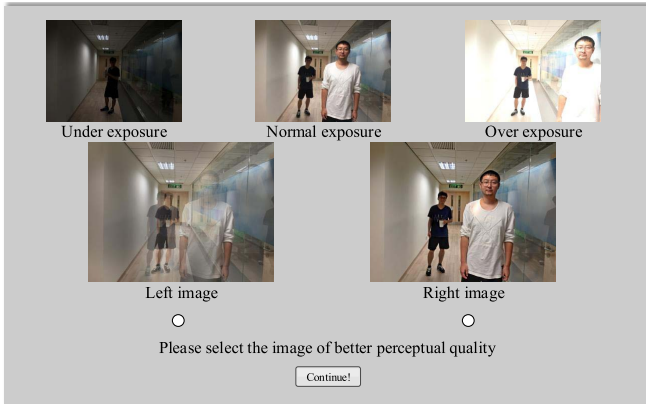


Fig. 4. User interface in the subjective experiment.

generated by  $i$ -th algorithm is preferred over that of the  $j$ -th algorithm on the  $k$ -th natural scene. The diagonals of  $\mathbf{C}$  in the first two dimensions are zeros, excluding self comparison. We apply the least squares method to aggregate pairwise rankings under the Bradley-Terry model [42]. The Gumbel random variable is assumed for the visual quality of each fused image, whose difference is a logistic random variable. Therefore,  $P(i \geq j|k)$  can be calculated from the logistic cumulative distribution function and has a closed-form solution [43]

$$P(i \geq j|k) = \frac{\exp(\boldsymbol{\mu}(i, k)/s)}{\exp(\boldsymbol{\mu}(i, k)/s) + \exp(\boldsymbol{\mu}(j, k)/s)}, \quad (1)$$

which can be estimated by the empirical count proportion  $\frac{\mathbf{C}(i, j, k)}{\mathbf{C}(i, j, k) + \mathbf{C}(j, i, k)}$ .  $\boldsymbol{\mu}(i, k)$  represents the location parameter to be estimated using least squares, namely the mean opinion score (MOS) of the fused image generated by the  $i$ -th algorithm on the  $k$ -th scene.  $s = \frac{\sqrt{3}}{\pi}$  is the scale parameter set to match the Thurstone's Case V model. We also try to aggregate pairwise ranking information in  $\mathbf{C}$  under the Thurstone model (assuming Gaussian for the quality of each fused image) and obtain very similar results, as evidenced by a mean Spearman's rank-order correlation coefficient (SRCC) of 0.998 across different scenes.

We randomly choose three subjects from three different subsections to form a subject group, and evaluate its performance by computing SRCC between the global rankings from each group and the MOSs. The mean and standard deviation (std) results are depicted in Fig. 5, where we find that individual subject groups achieve considerable agreement with each

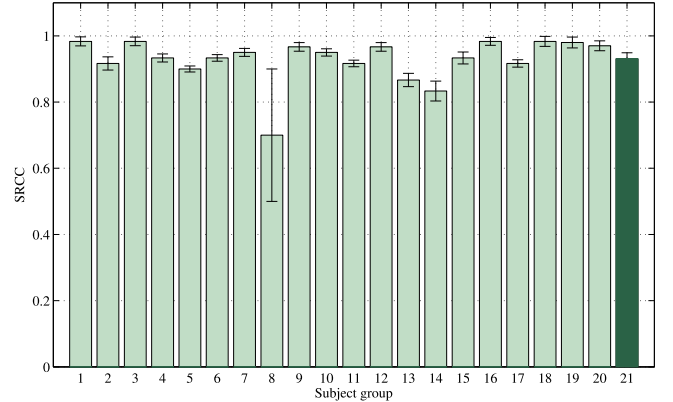


Fig. 5. SRCC between the global rankings of individual subject groups and the MOSs. Rightmost column: performance of the mean subject group.

other. The performance of the “mean subject group” (averaged across all subject groups) is also given in the rightmost column.

#### D. Performance of MEF Algorithms for Dynamic Scenes

We first conduct pairwise comparison of MEF algorithms by summing the preference tensor  $\mathbf{C}$  over the third dimension, resulting in an accumulated preference matrix  $\mathbf{P}$  (see Table I). A larger  $\mathbf{P}(i, j) - \mathbf{P}(j, i)$  indicates that the  $i$ -th algorithm produces better-quality fused images than the  $j$ -th algorithm. We also aggregate the pairwise comparisons into global ranking results as shown in Fig. 6, from which we have several useful observations. First, SPD-MEF [11] performs the best on average, whose success may be attributed to the exposure-invariant features and IMF for robust motion estimation. Second, patch-based methods such as SPD-MEF [11], Hu13 [20] and Sen12 [2] generally perform better than pixel-based algorithms such as Pece10 [10], Li12 [19] and Lee14 [14]. This is not surprising because patch-based algorithms take neighbouring information into account, leading to more robust motion alignment. As the second best algorithm, the pixel-based Li14 [23] also considers neighbouring pixels in extreme cases to refine the motion rejection process. Third, the low-rank based method, Lee14 [14] is subpar in our test largely due to the failure of preventing ghosting artifacts, especially on sequences with small object motion. Low-rank based schemes assume that the static background dominates the scene and therefore large foreground motion can be

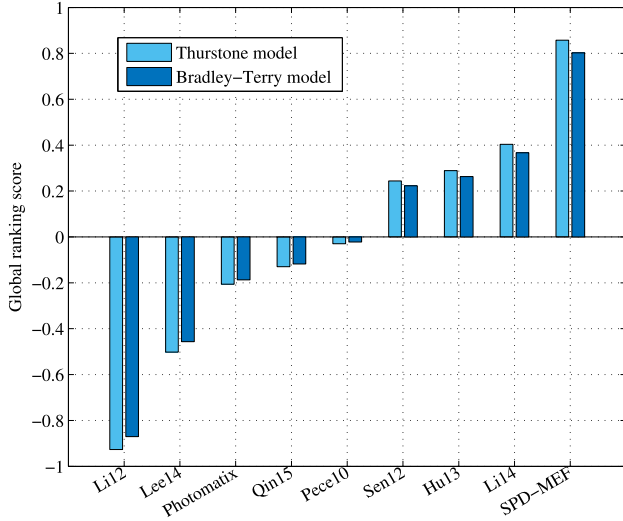


Fig. 6. Global ranking results of MEF algorithms for dynamic scenes under two statistical models.

sparingly coded. However, this is not the case for small and random motion, which occurs frequently in realistic natural scenes. Other methods (*e.g.*, Pece10 [10] and Li12 [19]) that rely on the static background assumption do not perform well either. Fourth, we find several types of distortions that may strongly affect human judgments of perceptual quality. For example, in certain extreme cases, the ghosting artifacts cannot be avoided when the moving objects are under-/over-exposed, whose structures cannot be properly retrieved from other exposures. This turns out to be the main challenge of the current MEF algorithms for dynamic scenes. Besides, halo artifacts near strong object boundaries would likely appear if the dynamic range difference between the foreground and background is large. Color speckle noise may arise in fused images due to inaccurate CRF estimation. Blurring artifacts may emerge if there are small errors in motion estimation.

#### IV. OBJECTIVE QUALITY ASSESSMENT FOR MEF OF DYNAMIC SCENES

In this section, we present in detail our objective quality model - MEF-SSIM<sub>d</sub> for MEF of both static and dynamic scenes. When the input natural scene is static, MEF-SSIM<sub>d</sub> reduces gracefully to MEF-SSIM [6], which is verified to be consistent with human perception of fused image quality. Specifically, we first divide the test fused image into static and dynamic regions, quantify the perceived quality of static and dynamic regions separately, and combine the measurements to obtain an overall quality score.

##### A. Region Segmentation

Before region segmentation, we need to make sure that the input exposures are aligned for camera motion. Among existing image registration algorithms, we employ a feature-based method, which relies on SIFT [37] matching to estimate an affine transformation with an  $l_2$ -norm loss. It works well for test sequences in our database. The  $l_2$ -norm loss is robust

to mismatched points, and can be efficiently implemented using an iteratively reweighted least squares method.

Our region segmentation approach starts by decomposing an image patch into three conceptually independent components [6]

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

where  $\|\tilde{\mathbf{x}}_k\|$  denotes the  $l_2$ -norm of a mean-removed patch from the  $k$ -th exposure.  $c_k$ ,  $\mathbf{s}_k$ , and  $l_k$  stand for contrast, structure, and intensity of  $\mathbf{x}_k$ , respectively.

The mean-removed and contrast-normalized  $\mathbf{s}_k$  contains ideal information for object motion detection in the test sequence. Specifically, we compute the cross-correlation (*i.e.*, the inner product) of two signal structures co-located in two exposures

$$\rho_{k,k'} = \mathbf{s}_k^T \mathbf{s}_{k'} \approx \frac{\tilde{\mathbf{x}}_k^T \tilde{\mathbf{x}}_{k'} + \epsilon}{\|\tilde{\mathbf{x}}_k\| \|\tilde{\mathbf{x}}_{k'}\| + \epsilon}, \quad (3)$$

where  $k, k' \in \{1, \dots, K\}$ ,  $k \neq k'$ , and  $K$  is the number of exposures in the sequence.  $\epsilon$  is a small positive constant to improve numerical stability of motion estimation against sensor noise [11].  $\rho_{k,k'}$  lies in  $[-1, 1]$  with a smaller value indicating less structural consistency between  $\mathbf{s}_k$  and  $\mathbf{s}_{k'}$ . By applying Eq. (3) across different spatial locations and different distinctive pairs of exposures, we obtain  $\binom{K}{2}$  structural consistency maps  $\{\mathbf{M}_{k,k'}\}$  (see Fig. 7, when  $K = 3$ ). We binarize  $\mathbf{M}_{k,k'}$  with a predefined threshold  $T$

$$\mathbf{B}_{k,k'}(i) = \begin{cases} 1 & \text{if } \mathbf{M}_{k,k'}(i) \geq T \\ 0 & \text{if } \mathbf{M}_{k,k'}(i) < T, \end{cases} \quad (4)$$

where  $i$  is the spatial location index. We combine the binary maps into one by

$$\mathbf{B} = \bigcap \mathbf{B}_{k,k'}, \quad (5)$$

where  $\bigcap$  is the AND operation, meaning that a region is detected to be static if it is static in all  $\{\mathbf{B}_{k,k'}\}$  maps. An example of  $\mathbf{B}$  is shown in Fig. 7, where white represents static regions and black represents dynamic regions.

##### B. Quality Assessment of Static Regions

For quality assessment of static regions, we make use of the MEF-SSIM [6] index. Specifically, we first perform structural patch decomposition on the input sequence using Eq. (2). The contrast of a desired patch is determined by the highest contrast among all co-located patches across exposures

$$\hat{c} = \max_{1 \leq k \leq K} c_k = \max_{1 \leq k \leq K} \|\tilde{\mathbf{x}}_k\|, \quad (6)$$

assuming that higher contrast implies better visibility under the realistic capture constraint. The structure of the desired patch is computed by a weighted average of the input structure vectors

$$\hat{\mathbf{s}} = \frac{\bar{\mathbf{s}}}{\|\bar{\mathbf{s}}\|}, \quad \text{where } \bar{\mathbf{s}} = \frac{\sum_{k=1}^K w(\tilde{\mathbf{x}}_k) \mathbf{s}_k}{\sum_{k=1}^K w(\tilde{\mathbf{x}}_k)}, \quad (7)$$



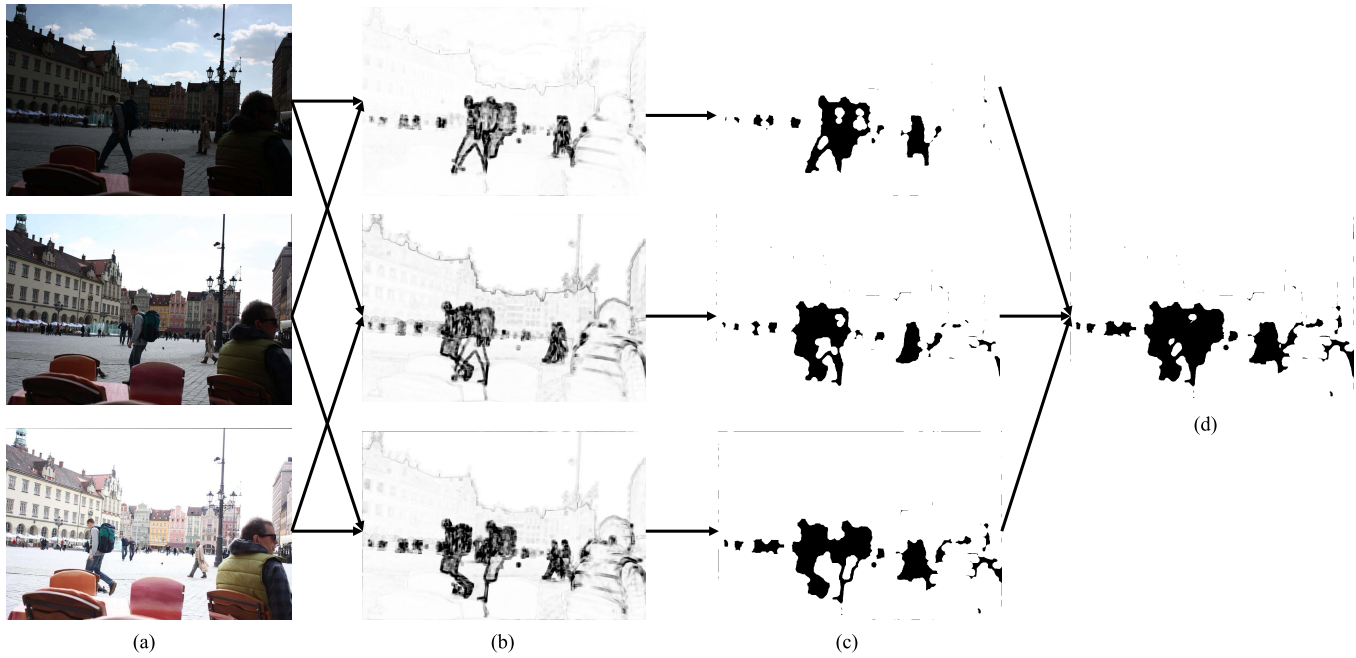


Fig. 7. Demonstration of region segmentation. (a) Input multi-exposure sequence by courtesy of Fabrizio Pece. (b) Structural consistency maps using Eq. (3). (c) Corresponding binary maps using Eq. (4). (d) Final binary map using Eq. (5).

and  $w(\cdot)$  is an  $\ell_p$ -norm weighting function with  $p$  adaptively determined based on a patch consistency measure [6]. In MEF-SSIM [6], the desired intensity is not computed due to its low relevance to perceived quality.

Once  $\hat{c}$  and  $\hat{s}$  are determined, we combine them to obtain the desired mean-removed patch

$$\hat{\mathbf{x}} = \hat{c} \cdot \hat{\mathbf{s}}. \quad (8)$$

A simplified version of the SSIM index [32] is used to assess the local image quality between the desired patch  $\hat{\mathbf{x}}$  and the corresponding patch  $\mathbf{y}$  for static regions

$$q^s(\hat{\mathbf{x}}, \mathbf{y}) = \frac{2\sigma_{\hat{\mathbf{x}}\mathbf{y}} + \epsilon}{\sigma_{\hat{\mathbf{x}}}^2 + \sigma_{\mathbf{y}}^2 + \epsilon}, \quad (9)$$

where  $\sigma_{\hat{\mathbf{x}}}^2$ ,  $\sigma_{\mathbf{y}}^2$ , and  $\sigma_{\hat{\mathbf{x}}\mathbf{y}}$  are the local variance of  $\hat{\mathbf{x}}$  and  $\mathbf{y}$ , and the local covariance between  $\hat{\mathbf{x}}$  and  $\mathbf{y}$ , respectively. The local quality scores are averaged to obtain an overall quality measure of the static regions

$$q^s = \frac{1}{N} \sum_{i=1}^N q^s(\mathbf{R}_i \hat{\mathbf{X}}, \mathbf{R}_i \mathbf{Y}), \quad (10)$$

where  $\hat{\mathbf{X}}$  and  $\mathbf{Y}$  stand for the desired image for reference and the test fused image, respectively.  $N$  is the total number of static patches.  $\mathbf{R}_i$  is a binary matrix with the number of columns equal the image dimension and the number of rows equal to the patch size. It serves as an operator that extracts the  $i$ -th local patch from the image.

### C. Quality Assessment of Dynamic Regions

To achieve high accuracy in object motion detection, many dehosing algorithms pick one exposure as the reference

motion to appear in the fused images [6], [12], [19]. However, for quality assessment, it is difficult to know beforehand whether the motion of the test fused image has a correspondence in the sequence. Moreover, different dehosing algorithms may have different strategies to select the ideal exposure as the reference, which result in different motion appearances in the fused images. To possibly match the motion in the test image and to successfully detect the ghosting artifacts, we exhaust all  $K$  exposures in the sequence, each of which is used as the reference to generate a quality map, and the one with the highest overall quality survives. Specifically, we select the  $k$ -th exposure as the reference, and create  $(K - 1)$  latent images with the help of IMF [22] by mapping the intensity values of the reference image to the rest  $(K - 1)$  exposures. This results in a pseudo static sequence that contains the same motion in the reference exposure, and Eq. (10) can be directly adopted to evaluate the perceptual quality

$$q_k^s = \frac{1}{N} \sum_{i=1}^N q^s(\mathbf{R}_i \hat{\mathbf{X}}_k, \mathbf{R}_i \mathbf{Y}), \quad (11)$$

where  $\hat{\mathbf{X}}_k$  is the desired image generated by using the  $k$ -th exposure as the reference. After sweeping all exposures, we obtain  $K$  quality scores, over which we compute the maximum as the final quality score of dynamic regions

$$q^d = \max_{1 \leq k \leq K} q_k^s. \quad (12)$$

Some examples of quality maps generated by MEF-SSIM<sub>d</sub> are shown in Fig. 8, where higher brightness indicates better quality. For static regions, we observe that the visual appearance of the sky and buildings in (d) is slightly unnatural, resulting in relatively low  $q^s$  value. The static

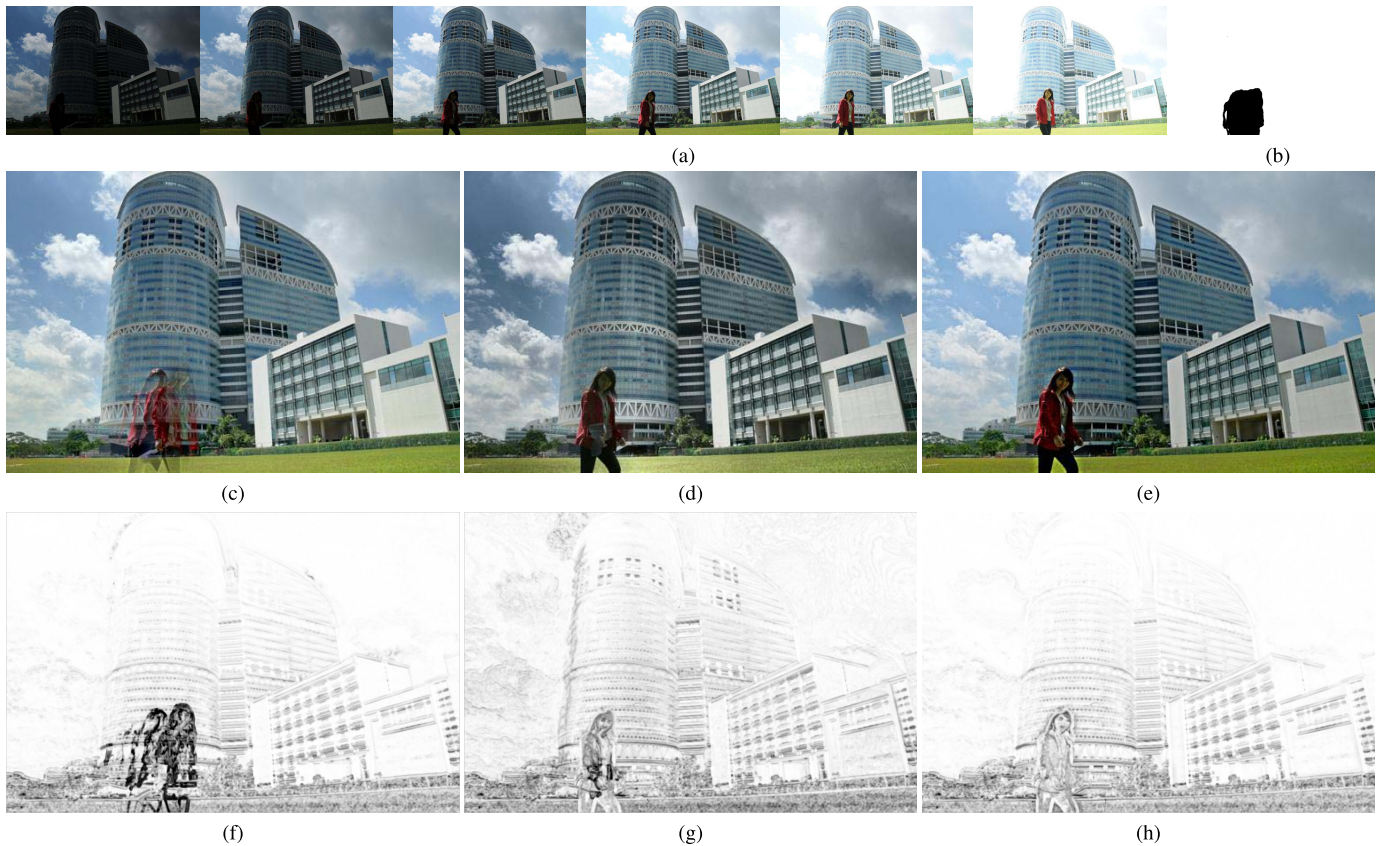


Fig. 8. Demonstration of quality maps generated by the proposed MEF-SSIM<sub>d</sub>. (a) Input multi-exposure sequence by courtesy of Zhengguo Li. (b) Binary map for region segmentation. (c) Fused image by Pece10 [10]. (d) Fused image by Lee14 [14]. (e) Fused image by SPD-MEF [11]. (f) Quality map of (c) with  $q^s = 0.937$ ,  $q^d = 0.558$ , and  $q = 0.748$ . (g) Quality map of (d) with  $q^s = 0.908$ ,  $q^d = 0.768$ , and  $q = 0.838$ . (h) Quality map of (e) with  $q^s = 0.939$ ,  $q^d = 0.829$ , and  $q = 0.884$ . Higher brightness in the quality map indicates better quality.

regions in (c) and (e) are visually similar, resulting in close  $q^s$  values. For dynamic regions, we observe that MEF-SSIM<sub>d</sub> successfully captures the ghosting artifacts in (c), which are strongly penalized. Although ghosting is significantly reduced in (d), some structural details of the girl in red are not well preserved compared to (e). All of these are clearly reflected in the corresponding quality maps.

The overall quality score is obtained by simple averaging the two quality measurements from static and dynamic regions

$$q = \frac{q^s + q^d}{2}. \quad (13)$$

We have tried other advanced pooling strategies, *e.g.*, weighting by distortion and information content [45], but obtain similar performance.

## V. EXPERIMENTS

In this section, we compare the performance of MEF-SSIM<sub>d</sub> to a set of objective quality models for image fusion on two databases, and verify that the performance improvements are statistically significant using a recent statistical analysis tool [46], [47].

### A. Experimental Setup

We validate the proposed MEF-SSIM<sub>d</sub> on two databases - the one described in Section III-A and the other introduced

in [25]. The latter consists of 36 multi-exposure sequences of dynamic scenes, which is divided into four sets according to different motion patterns. Six dehazing algorithms are used to produce 216 HDR images. In order to generate the corresponding LDR images for quality evaluation, we adopt the same tone mapping operator in [48]. We compare MEF-SSIM<sub>d</sub> against nine objective models, including Hossny08 [28], Cvejic06 [27], Wang04 [44], Xydeas00 [29], Wang08 [30], Zheng07 [31], Piella03 [33], Chen07 [34], and MEF-SSIM [6]. The implementation of MEF-SSIM is obtained from the original authors, while others are based on the implementations in [26] with respective parameters set by default.

Most parameters of MEF-SSIM<sub>d</sub> are inherited from MEF-SSIM [6]. These include a window size of  $11 \times 11$ , and  $\epsilon = (0.03L)^2$  in Eq. (3) and Eq. (9), where  $L = 255$  represents the dynamic range of intensity value for 8-bit images. The threshold  $T$  for static and dynamic region segmentation in Eq. (4) is set to 0.5.

### B. Validation

We first use SRCC between MOSs and model predictions as the quantitative measure, which lies in  $[0, 1]$  with a higher value indicating better performance. The results for each sequence in our database and for each set (consisting of nine sequences) in [25] are listed in Tables II and III, respectively.

TABLE II  
SRCC RESULTS OF MEF-SSIM<sub>d</sub> AGAINST NINE EXISTING MODELS ON OUR DATABASE

Sequence	Zheng07 [31]	Cvejic06 [27]	Chen07 [34]	Piella03 [33]	MEF-SSIM [6]	Xydeas00 [29]	Wang08 [30]	Hossny08 [28]	Wang04 [44]	MEF-SSIM <sub>d</sub>
Men	-0.250	-0.066	-0.133	0.166	0.500	0.483	0.600	0.666	0.583	0.866
Arch	-0.666	-0.283	-0.350	0.533	0.766	0.816	0.766	0.400	0.400	0.533
Llandudno	-0.350	-0.166	-0.366	0.533	0.566	0.183	0.400	0.333	0.400	0.800
Square	-0.466	-0.016	0.316	0.033	-0.050	-0.066	-0.033	0.466	0.466	0.933
Tate3	-0.433	0.133	0.133	-0.016	-0.033	0.016	0.216	0.850	0.766	0.667
Forest	-0.283	-0.633	-0.416	0.566	0.233	0.733	0.600	0.666	0.666	0.783
Horse	-0.366	0.166	0.166	-0.133	-0.300	-0.100	0.000	0.650	0.650	0.667
Corridor	-0.250	0.300	-0.066	0.533	0.450	0.333	0.400	0.750	0.750	0.700
Office	-0.550	-0.133	-0.433	0.283	0.433	-0.066	-0.183	0.583	0.600	0.350
Russ1	-0.400	-0.583	-0.266	0.216	0.366	0.466	0.616	0.866	0.866	0.833
Puppets	-0.833	-0.150	0.433	0.250	0.066	0.066	0.250	0.616	0.616	0.783
Cliff	-0.400	0.333	-0.716	0.283	0.616	0.266	0.233	0.583	0.600	0.466
Sculpture	-0.300	-0.583	-0.233	0.300	0.150	0.200	0.016	0.716	0.550	0.683
Wroclav	-0.150	-0.816	0.000	-0.150	-0.250	-0.116	-0.133	0.533	0.650	0.383
ProfJeon	-0.150	-0.416	-0.216	0.333	0.016	0.350	0.433	0.750	0.800	0.867
NoiseCam	-0.283	-0.033	-0.400	0.450	0.416	0.766	0.683	0.516	0.750	0.767
Campus	-0.133	0.216	-0.333	-0.266	-0.150	0.060	0.300	0.216	0.216	0.933
Brunswick	-0.616	-0.300	-0.033	-0.100	0.100	0.133	0.366	0.500	0.533	0.883
YWFusion	-0.066	0.216	0.083	-0.233	-0.116	-0.100	0.083	0.617	0.616	0.917
Lady	0.200	-0.516	-0.316	0.133	0.250	0.066	0.033	0.883	0.883	0.817
Average	<b>-0.338</b>	<b>-0.167</b>	<b>-0.158</b>	<b>0.186</b>	<b>0.202</b>	<b>0.225</b>	<b>0.283</b>	<b>0.608</b>	<b>0.618</b>	<b>0.730</b>

TABLE III  
SRCC RESULTS OF MEF-SSIM<sub>d</sub> AGAINST NINE EXISTING MODELS ON THE DATABASE INTRODUCED IN [25]

Set	Zheng07 [31]	Cvejic06 [27]	Chen07 [34]	Piella03 [33]	MEF-SSIM [6]	Xydeas00 [29]	Wang08 [30]	Hossny08 [28]	Wang04 [44]	MEF-SSIM <sub>d</sub>
#01	0.591	0.296	0.182	0.589	0.601	0.689	0.430	0.479	0.435	0.690
#02	0.442	0.165	0.458	0.804	0.757	0.585	0.413	0.471	0.553	0.782
#03	0.632	0.219	0.429	0.765	0.486	0.702	0.441	0.473	0.619	0.784
#04	0.575	0.271	0.357	0.454	0.556	0.771	0.644	0.581	0.581	0.765
Average	<b>0.560</b>	<b>0.357</b>	<b>0.238</b>	<b>0.653</b>	<b>0.600</b>	<b>0.687</b>	<b>0.482</b>	<b>0.501</b>	<b>0.547</b>	<b>0.755</b>

We find that the proposed MEF-SSIM<sub>d</sub> achieves the best performance for most visual scenes in both databases, and is a significant improvement over its baseline model MEF-SSIM [6]. Other objective models, mainly based on mutual information [28], gradient magnitude [29], or SSIM [33] are not effective at measuring structural preservation in static regions and the ghosting artifacts in dynamic regions. Nevertheless, nearly all models fail on the “Arch” and “Office” sequences, whose object motion is large and discontinuous with extremely under-/over-exposed regions in the background. This casts grand challenges to region segmentation and subsequent quality assessment because little consistent information across exposures can be gathered for quality evaluation of the fused image. In addition, when comparing the performance of MEF-SSIM<sub>d</sub> with that of the mean subject group (Fig. 5), we find that there is still quite some room for further development of better objective quality models.

We also adopt two recent quantitative measures [46], [47], which are specifically designed for human data from 2AFC experiments, to compare the performance of objective quality models. The test image pairs are first classified into pairs with and without significant differences according to subjective preference judgments collected from different participants. For all image pairs, the receiver operating characteristic (ROC) curve can be drawn to illustrate the model capability to

discriminate between different and similar image pairs. For the subset of image pairs that are of significant differences, the better/worse ROC measures the model ability to determine which image in a pair has better visual quality. The area under curve (AUC) is applied to summarize the performance of competing models, with a higher value indicating better performance. The associated statistical significance can be demonstrated using a hypothesis testing approach based on t-statistics [49]. We show the different/similar and the better/worse AUC results in Fig. 9 and Fig. 10, respectively, which are computed with a 95% confidence interval. The statistical significance results are also shown, where a black square indicates that the row model is significantly worse than the column model, and a white square means the opposite. A gray square indicates that the row and column models are statistically indistinguishable. From the results, we find that the AUC performance improvement of MEF-SSIM<sub>d</sub> is statistically significant for both different/similar ROC and better/worse ROC analysis.

In addition, we analyze the sensitivity of MEF-SSIM<sub>d</sub> with respect to  $T$  in Eq. (4), a critical parameter that segments the image into static and dynamic regions. A large  $T$  would misclassify some static regions as dynamic, whose quality assessment depends on IMF and is prone to error. A small  $T$  may fail in capturing the ghosting artifacts due to inac-



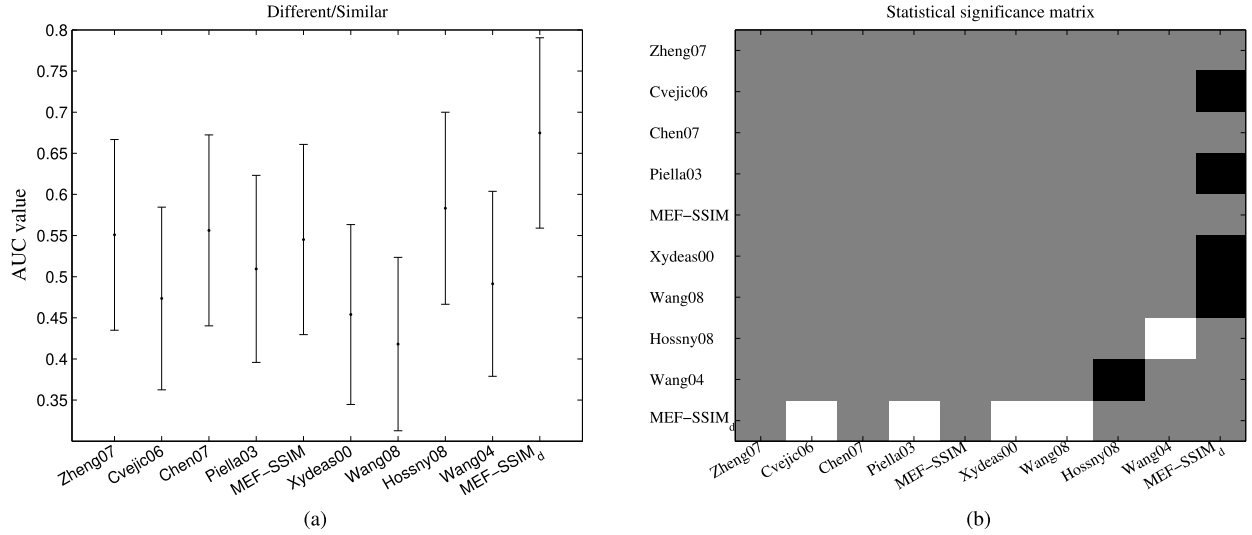


Fig. 9. Different/similar ROC analysis. (a) AUC values of different quality models, which range from 0 to 1 with higher values indicating better performance. (b) Statistical significance matrix based on t-statistic. A black square indicates that the row model is significantly worse than the column model, and a white square means the opposite. A gray square indicates that the row and column models are statistically indistinguishable. We follow the style in [47].

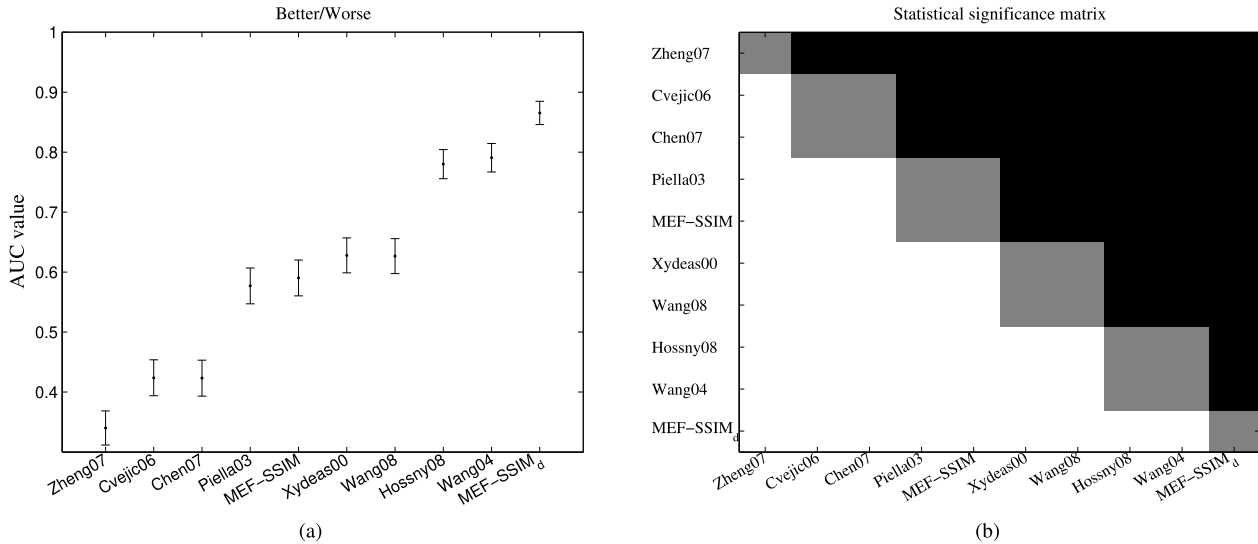


Fig. 10. Better/worse ROC analysis. (a) AUC values of different quality models, which range from 0 to 1 with higher values indicating better performance. (b) Statistical significance matrix based on t-statistic.

curate motion estimation. We find that  $\text{MEF-SSIM}_d$  is quite robust to variations of  $T$ , especially within small values. The default setting generates reasonable region segmentation maps (see Fig. 7) and achieves satisfactory performance on our database (see Table IV).

## VI. APPLICATION: AUTOMATIC PARAMETER TUNING OF MEF ALGORITHMS OF DYNAMIC SCENES

Besides objective quality evaluation of fused images, an effective objective quality model should be able to guide the design and optimization of MEF algorithms for dynamic scenes. In this section, we demonstrate this idea by applying  $\text{MEF-SSIM}_d$  to automatic parameter tuning of MEF algorithms.

There are often one or multiple parameters in MEF algorithms for dynamic scenes, whose optimal values are

TABLE IV  
SENSITIVITY ANALYSIS OF THE THRESHOLD  $T$  IN TERMS OF SRCC ON OUR DATABASE. THE DEFAULT SETTING IS HIGHLIGHTED IN BOLD

$T$	0.1	0.3	<b>0.5</b>	0.7	0.9
Average	0.725	0.731	<b>0.730</b>	0.677	0.595

image-dependent. Therefore, it is a challenging and time-consuming task to handpick a set of parameter values that work well for all images.  $\text{MEF-SSIM}_d$  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 SPD-MEF [11] as the example MEF algorithm, which contains two user-specified parameters  $T_s$  and  $T_m$  to detect inconsistent object motion.  $T_s$  rejects inconsistent patches across exposures with respect to the reference patch.  $T_m$  identifies large pixel-wise

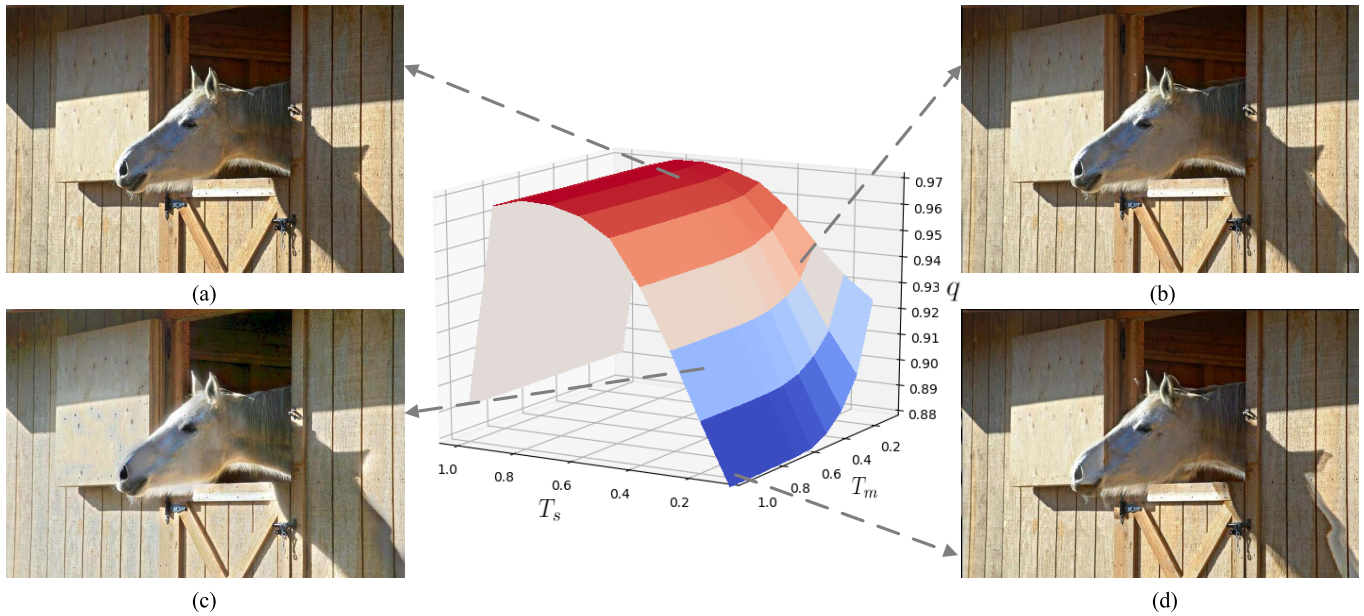


Fig. 11. Parameter tuning using MEF-SSIM<sub>d</sub>. Warmer color in the surface plot indicates better predicted quality. (a)  $q = 0.971$ . (b)  $q = 0.934$ . (c)  $q = 0.901$ . (d)  $q = 0.885$ .

differences between each exposure and its latent correspondence generated by IMF. Both  $T_s$  and  $T_m$  lie in  $[0, 1]$  with default values of  $T_s = 0.8$  and  $T_m = 0.1$  [11], but the perceptual quality of fused images could be highly sensitive to these parameters.

Fig. 11 gives example images generated with different  $T_s$  and  $T_m$  values on the image sequence “Horse”, where warmer color in the surface plot indicates better predicted quality of the corresponding fused image. By varying  $T_s$  and  $T_m$ , we obtain fused images of significantly different quality. For example, Fig. 11(d) exhibits strong ghosting artifacts, while Fig 11(a) is clear and vivid. Moreover, MEF-SSIM<sub>d</sub> aligns well with human perception of fused image quality, suggesting that it is a useful tool to automatically select the optimal values of  $T_s$  and  $T_m$ .

## VII. CONCLUSION

We have performed perceptual evaluation for MEF of dynamic scenes. We created an MEF database and conducted a subjective experiment to collect human opinions of fused image quality. No objective quality models are able to accurately predict human data. Therefore, we design the MEF-SSIM<sub>d</sub> index, which successfully captures the ghosting artifacts, resulting in the best quality prediction performance. The use of MEF-SSIM<sub>d</sub> was also demonstrated through automatic parameter tuning of MEF algorithms.

Our subjective database provides a useful, but limited means of assessing the performance of MEF-SSIM<sub>d</sub>. A more direct assessment can arise from perceptual optimization, namely maximizing MEF-SSIM<sub>d</sub> for best fused images in either image domain or through a deep neural network. The images generated along the trajectory would provide a direct visualization of how MEF-SSIM<sub>d</sub> aligns with human judgments, and thus an effective way to identify its advantages and disadvantages.

The baseline MEF-SSIM has been successfully demonstrated to guide the design of better MEF algorithms for static scenes [7], therefore we believe MEF-SSIM<sub>d</sub> holds much promise in perceptual optimization of MEF methods for dynamic scenes as well.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for deeply insightful comments, and the subjects who participate in the subjective user study.

## 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. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, “Robust patch-based HDR reconstruction of dynamic scenes,” *ACM Trans. Graph.*, vol. 31, no. 6, Nov. 2012, Art. no. 203.
- [3] A. Eden, M. Uyttendaele, and R. Szeliski, “Seamless image stitching of scenes with large motions and exposure differences,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 2498–2505.
- [4] A. Srikantha and D. Sidibé, “Ghost detection and removal for high dynamic range images: Recent advances,” *Signal Process., Image Commun.*, vol. 27, no. 6, pp. 650–662, Jul. 2012.
- [5] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem, “The state of the art in HDR deghosting: A survey and evaluation,” *Comput. Graph. Forum*, vol. 34, no. 2, pp. 683–707, May 2015.
- [6] K. Ma, K. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
- [7] K. Ma, Z. Duanmu, H. Yeganeh, and Z. Wang, “Multi-exposure image fusion by optimizing a structural similarity index,” *IEEE Trans. Comput. Imag.*, vol. 4, no. 1, pp. 60–72, Mar. 2018.
- [8] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, “Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4724–4732.
- [9] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem, “An objective deghosting quality metric for HDR images,” *Comput. Graph. Forum*, vol. 35, no. 2, pp. 139–152, May 2016.

- [10] F. Pece and J. Kautz, "Bitmap movement detection: HDR for dynamic scenes," in *Proc. IEEE Conf. Vis. Media Prod.*, Nov. 2010, pp. 1–8.
- [11] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, "Robust multi-exposure image fusion: A structural patch decomposition approach," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2519–2532, May 2017.
- [12] O. Gallo, N. Gelfandz, W.-C. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *Proc. IEEE Int. Conf. Comput. Photography*, Apr. 2009, pp. 1–7.
- [13] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
- [14] C. Lee, Y. Li, and V. Monga, "Ghost-free high dynamic range imaging via rank minimization," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1045–1049, Sep. 2014.
- [15] T.-H. Oh, J.-Y. Lee, Y.-W. Tai, and I. S. Kweon, "Robust high dynamic range imaging by rank minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1219–1232, Jun. 2015.
- [16] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, Jul. 2017, Art. no. 144.
- [17] S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang, "Deep high dynamic range imaging with large foreground motions," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 120–135.
- [18] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures," *J. Graph. Tools*, vol. 8, no. 2, pp. 17–30, 2003.
- [19] 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.
- [20] J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR deghosting: How to deal with saturation?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1163–1170.
- [21] X. Qin, J. Shen, X. Mao, X. Li, and Y. Jia, "Robust match fusion using optimization," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1549–1560, Aug. 2015.
- [22] M. D. Grossberg and S. K. Nayar, "Determining the camera response from images: What is knowable?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1455–1467, Nov. 2003.
- [23] 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.
- [24] K. K. Hadziabdic, J. H. Telalovic, and R. Mantiuk, "Comparison of deghosting algorithms for multi-exposure high dynamic range imaging," in *Proc. 29th Spring Conf. Comput. Graph.*, 2013, pp. 21–28.
- [25] K. Karađuzović-Hadziabdić, J. H. Telalović, and R. K. Mantiuk, "Assessment of multi-exposure HDR image deghosting methods," *Comput. Graph.*, vol. 63, pp. 1–17, Apr. 2017.
- [26] 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.
- [27] 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, May 2006.
- [28] 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.
- [29] C. S. Xydeas and V. S. Petrovic, "Objective pixel-level image fusion performance measure," in *Proc. AeroSense*, 2000, pp. 89–98.
- [30] P.-W. Wang and B. Liu, "A novel image fusion metric based on multi-scale analysis," in *Proc. IEEE Int. Conf. Signal Process.*, Oct. 2008, pp. 965–968.
- [31] 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.
- [32] 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.
- [33] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2003, pp. 169–173.
- [34] 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, Apr. 2007.
- [35] K. Zeng, K. Ma, R. Hassen, and Z. Wang, "Perceptual evaluation of multi-exposure image fusion algorithms," in *Proc. 6th IEEE Int. Workshop Qual. Multimedia Exper.*, Sep. 2014, pp. 7–12.
- [36] Y. Fang, H. Zhu, K. Ma, and Z. Wang, "Perceptual quality assessment of HDR deghosting algorithms," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2017, pp. 3165–3169.
- [37] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [38] Photomatix. (2017). *Commercially-Available HDR Processing Software*. [Online]. Available: <http://www.hdrsoft.com>
- [39] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. Conf. Comput. Graph. Interact. Technol.*, 1997, pp. 369–378.
- [40] 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, Sep. 2008.
- [41] VQEG. (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>
- [42] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," *Biometrika*, vol. 39, nos. 3–4, pp. 324–345, 1952.
- [43] K. Tsukida and M. R. Gupta, "How to analyze paired comparison data," Univ. Washington, Seattle, WA, USA, Tech. Rep. UWEETR-2011-0004, May 2011.
- [44] Q. Wang and Y. Shen, "Performances evaluation of image fusion techniques based on nonlinear correlation measurement," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, May 2004, pp. 472–475.
- [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] P. Hanhart, L. Krasula, P. Le Callet, and T. Ebrahimi, "How to benchmark objective quality metrics from paired comparison data?" in *Proc. IEEE Int. Conf. Qual. Multimedia Exper.*, Jun. 2016, pp. 1–6.
- [47] L. Krasula, P. Le Callet, K. Fliegel, and M. Klíma, "Quality assessment of sharpened images: Challenges, methodology, and objective metrics," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1496–1508, Mar. 2017.
- [48] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, Jul. 2002.
- [49] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, 6th ed. Hoboken, NJ, USA: Wiley, 2013.

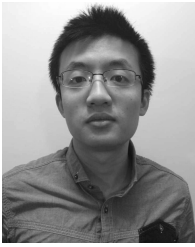


**Yuming Fang** (M'13–SM'17) received the B.E. degree from Sichuan University, Chengdu, China, the M.S. degree from the Beijing University of Technology, Beijing, China, and the Ph.D. degree from Nanyang Technological University, Singapore. He is currently a Professor with the School of Information Management, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include visual attention modeling, visual quality assessment, computer vision, and 3D image/video processing. He serves as an Associate Editor for IEEE ACCESS. He serves on the Editorial Board of *Signal Processing: Image Communication*.



**Hanwei Zhu** (S'17) received the B.E. degree from the Jiangxi University of Finance and Economics, Nanchang, China, in 2017, where he is currently pursuing the M.A.Sc. degree in computer science. His research interest includes perceptual image processing.





**Kede Ma** (S'13–M'18) received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2012, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2014 and 2017, respectively. He was a Research Associate with the Howard Hughes Medical Institute and New York University, New York, NY, USA, in 2018. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. His research interests

include perceptual image processing, computational vision, and computational photography.



**Zhou Wang** (S'99–M'02–SM'12–F'14) received the Ph.D. degree from The University of Texas at Austin in 2001.

He is currently the Canada Research Chair and a Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image and video processing and coding, visual quality assessment and optimization, computational vision and pattern analysis, multimedia communications, and biomedical signal processing. He has over 200 publications in these fields with over 50 000 citations (Google Scholar). He has been serving as a Senior Area Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING since 2015. He has served/has been serving as a member for the IEEE Multimedia Signal Processing Technical Committee from 2013 to 2015, an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2016 to 2018, the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2009 to 2014, *Pattern Recognition* since 2006, and the IEEE SIGNAL PROCESSING LETTERS from 2006 to 2010, and a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING from 2007 to 2009 and from 2013 to 2014. He is also a fellow of the Royal Society of Canada and the Canadian Academy of Engineering. He was a recipient of the 2017 Faculty of Engineering Research Excellence Award from the University of Waterloo, the 2016 IEEE Signal Processing Society Sustained Impact Paper Award, the 2015 Primetime Engineering Emmy Award, the 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, the 2013 IEEE Signal Processing Magazine Best Paper Award, and the 2009 IEEE Signal Processing Society Best Paper Award.

He is currently the Canada Research Chair and a Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image and video processing and coding, visual quality assessment and optimization, computational vision and pattern analysis, multimedia communications, and biomedical signal processing. He has over 200 publications in these fields with over 50 000 citations (Google Scholar). He has been serving as a Senior Area Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING since 2015. He has served/has been serving as a member for the IEEE Multimedia Signal Processing Technical Committee from 2013 to 2015, an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2016 to 2018, the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2009 to 2014, *Pattern Recognition* since 2006, and the IEEE SIGNAL PROCESSING LETTERS from 2006 to 2010, and a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING from 2007 to 2009 and from 2013 to 2014. He is also a fellow of the Royal Society of Canada and the Canadian Academy of Engineering. He was a recipient of the 2017 Faculty of Engineering Research Excellence Award from the University of Waterloo, the 2016 IEEE Signal Processing Society Sustained Impact Paper Award, the 2015 Primetime Engineering Emmy Award, the 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, the 2013 IEEE Signal Processing Magazine Best Paper Award, and the 2009 IEEE Signal Processing Society Best Paper Award.



**Shutao Li** (M'07–SM'15–F'19) received the B.S., M.S., and Ph.D. degrees from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University, where he is currently a Full Professor. From May 2001 to October 2001, he was a Research Associate with the Department of Computer Science, The Hong Kong University of Science and Technology, where he was a Visiting Professor from April 2005 to June 2005.

From November 2002 to November 2003, he was a Post-Doctoral Fellow with the Royal Holloway College, University of London. He has authored or coauthored over 200 refereed articles. His current research interests include image processing, pattern recognition, and artificial intelligence. He is also a member of the Editorial Board of *Information Fusion* and *Sensing and Imaging*. He was a recipient of the 2nd-Grade State Scientific and Technological Progress Awards of China in 2004 and 2006. He is also an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.