Multi-Exposure Image Fusion by Optimizing A Structural Similarity Index

Kede Ma, *Student Member, IEEE*, Zhengfang Duanmu, *Student Member, IEEE*, Hojatollah Yeganeh, *Member, IEEE*, and Zhou Wang, *Fellow, IEEE*

***Abstract*—We propose a multi-exposure image fusion (MEF) algorithm by optimizing a novel objective quality measure, namely the color MEF structural similarity (MEF-SSIM$_c$) index. The design philosophy we introduce here is substantially different from existing ones. Instead of pre-defining a systematic computational structure for MEF (e.g., multiresolution transformation and transform domain fusion followed by image reconstruction), we directly operate in the space of all images, searching for the image that optimizes MEF-SSIM$_c$. Specifically, we first construct the MEF-SSIM$_c$ index by improving upon and expanding the application scope of the existing MEF-SSIM algorithm. We then describe a gradient ascent-based algorithm, which starts from any initial point in the space of all possible images and iteratively moves towards the direction that improves MEF-SSIM$_c$ until convergence. Numerical and subjective experiments demonstrate that the proposed algorithm consistently produces better quality fused images both visually and in terms of MEF-SSIM$_c$. The final high quality fused image appears to have little dependence on the initial image. The proposed optimization framework is readily extensible to construct better MEF algorithms when better objective quality models for MEF are available.**

***Index Terms*—Multi-exposure image fusion (MEF), gradient ascent, structural similarity (SSIM), perceptual optimization.**

I. Introduction

MULTI-EXPOSURE image fusion (MEF) is a cost-effective technique that bridges the gap between the high dynamic range (HDR) of luminance levels in natural scenes and the low dynamic range (LDR) of standard display devices [1]. The input sequence of an MEF algorithm consists of multiple pictures of the same scene taken at different exposure levels, each of which captures partial information of the scene. An excellent MEF algorithm is expected to fuse perceptually important information of all input images into a single LDR image that is visually appealing and has minimal visual artifacts [2].

Manuscript received November 15, 2015; revised June 13, 2017; accepted December 14, 2017. Date of publication December 21, 2017; date of current version February 8, 2018. This work was supported by the Natural Sciences and Engineering Research Council of Canada. The work of K. Ma was partially supported by the CSC. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sergio Goma. *(Corresponding author: Kede Ma.)*

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada (e-mail: k29ma@uwaterloo.ca; zduanmu@uwaterloo.ca; hyeganeh@uwaterloo.ca; zhou.wang@uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCI.2017.2786138

In recent years, many MEF algorithms have been proposed [3]–[21]. However, none of them has been designed to optimize a proven quality measure that correlates well with human visual perception of image quality. Instead, they make certain explicit or implicit hypotheses about what good visual quality means, with little effort justifying the validity of such hypotheses. For example, a commonly used approach is to maximize the fine details in fused images as a way to create vivid appearance [9], [10]. However, detail enhancement does not necessarily lead to consistent perceptual quality improvement and may sometimes create unwanted visual artifacts [22]. Moreover, all existing algorithms start by pre-defining a systematic computational structure for MEF (*e.g.*, multi-resolution transformation and transform domain fusion followed by image reconstruction), with weak and indirect support of the validity and optimality of such a structure. In addition, most existing MEF algorithms are demonstrated using a limited number of hand-picked examples, without subjective verifications on databases that contain sufficient variations of image content or objective assessment by well-established and subject-validated quality models [23].

A subjective database that is dedicated to MEF was created recently [22], based on which a somewhat surprising result is that previously published objective quality models [23]–[32] for general purpose image fusion are very limited in predicting the perceived quality of multi-exposure fused images. Motivated by the lack of proper objective quality models for MEF, Ma *et al.* [2] proposed one of the first objective measures specifically for MEF by combining the design principle of the well-known structural similarity (SSIM) index [33] with a patch consistency measure. Here we call this measure the multi-exposure fusion structural similarity (MEF-SSIM) index, which has been shown to well correlate with human perception of image quality and meanwhile inherit some nice mathematical properties from SSIM [34] for optimization purposes [35]–[37].

In this paper, we propose a substantially different framework to design MEF. Unlike existing MEF methods that employ a pre-defined computational structure, we directly explore in the space of all images, searching for the image that optimizes MEF-SSIM$_c$, which is a more advanced model built upon MEF-SSIM. More specifically, we first construct the MEF-SSIM$_c$ model by expanding the application scope of MEF-SSIM from grayscale to color images and by better accounting for the impact of luminance changes on image quality. We then derive an analytic form of the gradient of MEF-SSIM$_c$ in the space of all images and use it to iteratively search for the optimal MEF-SSIM$_c$ image.

Extensive numerical and subjective experiments demonstrate that the proposed algorithm consistently produces better quality fused images, regardless of the starting point of the iteration, *i.e.*, the initial image, which could be an image produced by an existing MEF algorithm, a uniform gray image, a random noise image, or even another completely irrelevant natural image. The novel optimization framework is more general than the proposed MEF algorithm. Whenever a better objective quality model for MEF is available, it may be incorporated into the framework to produce better MEF images.

II. Related Work

A. Existing MEF Algorithms

Many existing MEF algorithms follow a weighted summation framework

$$\mathbf{y} = \sum_{k=1}^{K} w_k \mathbf{x}_k \,, \tag{1}$$

where K is the number of exposure levels in the source image sequence. $\mathbf{x}_k$ represents a co-located pixel or patch in the k-th exposure image $\mathbf{X}_k$, depending on whether the algorithm is a pixel-wise or patch-wise method. $\mathbf{y}$ denotes the corresponding pixel or patch in the fused image $\mathbf{Y}$. The weight w_k carries the information about the perceptual importance of $\mathbf{x}_k$ in the fusion process. In transform domain approaches, $\mathbf{x}_k$ and $\mathbf{y}$ may also be co-located transform coefficients or a group of neighboring coefficients. Most existing algorithms differ in the computation of w_k and how it may adapt over space or scale based on image content. Equation (1) has been taken for granted by a majority of MEF algorithms, but there has been very little discussion about why weighted summation is a good way of fusion and how far it is from optimality.

As one of the first attempts, an efficient Laplacian pyramid decomposition for binocular image fusion was proposed in [38]. This decomposition scheme was later applied to MEF [3], [11]. A similar decomposition scheme, namely boosting Laplacian pyramid [19], was proposed for MEF with the weights determined by well exposedness, gradient direction, and just noticeable distortion-based saliency measures. Goshtasby [4] proposed one of the first patch-wise MEF approaches by directly choosing the patch with the highest entropy to construct the fused image. Later on, Ma and Wang developed a structural patch decomposition for MEF [39]. Raman *et al.* adopted a bilateral filter [40] to extract edge information, which is subsequently added to the base layer image for detail enhancement. Song *et al.* [7] proposed a probabilistic fusion scheme by first estimating the initial image with the maximum visual contrast and scene gradient, and then generating the final image with reversals in image gradients suppressed. Another conditional random field based MEF method was proposed in [14], where the weights were determined by local contrast and color saturation. A gradient-based detail-enhanced MEF method was proposed in [12]. Li *et al.* [9] also enhanced the details of a given fused image in a quadratic optimization framework. A guided filter [41] was adopted in [10] to control the roles of pixel saliency and spatial consistency when constructing the fused image. Motivated by the fact that traditional edge-preserving smoothing techniques suffer from halo artifacts, a weighted guided image filter was introduced in [20] and used for MEF [20]. A variational approach for MEF was proposed in [15] by combining color matching and gradient direction information. Hara *et al.* determined the global and local weights of their MEF algorithm via a gradient-based contrast maximization and an image saliency detection method, respectively [16].

To overcome the misalignment problem caused by camera and object motion, several algorithms have been proposed. Zhang *et al.* [13] used gradient direction to differentiate the dominant background from the moving object. A median filter was used to filter out the moving object in [8]. Li *et al.* [17] enabled their two level detail enhancing image fusion scheme to account for dynamic scenes by explicitly detecting and correcting inconsistent pixels with respect to a chosen reference image. Qin *et al.* tackled camera and object motions in the source sequence by a patch-wise matching algorithm. The weight for each patch was computed using a random walker method [18].

B. MEF-SSIM

Since the first step of our approach is to improve upon the objective quality model, MEF-SSIM [2], we provide a brief introduction here. Similar to the design philosophy of SSIM [33], MEF-SSIM looks at an image patch from three conceptually different aspects: luminance, contrast and structure. In particular, it explicitly performs the following patch decomposition

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

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 the mean-removed 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 [2].

The desired contrast of the fused image patch is determined by the highest contrast of all source image patches

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

The desired structure of the fused image patch is computed by a weighted average of the input structure vectors

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

and $w(\cdot)$ is a power weighting function given by

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

Here $p \ge 0$ is an exponential parameter that is determined adaptively based on a patch consistency measure [2]. Once $\hat{c}$ and $\hat{\mathbf{s}}$

are determined at each spatial location, they are combined using (2) to yield a new vector

$$\hat{\mathbf{x}} = \hat{c} \cdot \hat{\mathbf{s}} \,. \tag{6}$$

A simplified formation of the SSIM index is then used to evaluate the local image quality

$$S(\{\mathbf{x}_k\}, \mathbf{y}) = \frac{2\sigma_{\hat{\mathbf{x}}\mathbf{y}} + C_2}{\sigma_{\hat{\mathbf{x}}}^2 + \sigma_{\mathbf{y}}^2 + C_2} \,, \tag{7}$$

where $\{\mathbf{x}_k\} = \{\mathbf{x}_k | 1 \leq k \leq K\}$ denotes the set of co-located image patches in the source image sequence. $\sigma_{\hat{\mathbf{x}}}^2$, $\sigma_{\mathbf{y}}^2$, 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_2 is a small positive stabilizing constant that accounts for the saturation effects of the visual system at low contrast [33].

The local MEF-SSIM comparison 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 (some examples will be given later in Fig. 3). The local MEF-SSIM values are averaged to obtain an overall quality measure of the fused image

$$Q(\{\mathbf{X}_k\}, \mathbf{Y}) = \frac{1}{M} \sum_{j=1}^{M} S(\{\mathbf{x}_k(j)\}, \mathbf{y}(j)) \,, \tag{8}$$

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

III. MEF By MEF-SSIM$_c$ Optimization

In this section, we first introduce the general framework to design MEF algorithm, which uses a gradient ascent-based method to optimize an improved MEF-SSIM quality measure named MEF-SSIM$_c$. We then provide detailed descriptions on each component of the proposed approach, which includes the derivation of the MEF-SSIM$_c$ model, the computation of the gradient of MEF-SSIM$_c$ in the space of all images, and additional implementation details.

A. Framework

Considering MEF-SSIM$_c$ as the quality criterion of the fused image, the problem of MEF can be formulated as

$$\mathbf{Y}_{\text{opt}} = \arg\max_{\mathbf{Y}} Q(\{\mathbf{X}_k\}, \mathbf{Y}) \,. \tag{9}$$

Due to the nonconvexity of MEF-SSIM$_c$ and the high dimensionality of the optimization problem (which equals to the number of pixels in the image), the analytic solution of (9) is implausible. A practical alternative is to employ a gradient-based iterative optimization procedure [34], [35], [42]. Specifically, we consider the fused image as a point or a vector starting from the origin in the space of all possible images, which has the same dimension as the number of pixels in the image, and each dimension is the intensity value of a pixel. At the i-th iteration, we use a gradient ascent algorithm to improve the MEF-SSIM$_c$ index of the resulting image $\mathbf{Y}_i$ from the last iteration by

$$\mathbf{Y}_{i+1} = \mathbf{Y}_i + \lambda \, \nabla_{\mathbf{Y}} Q(\{\mathbf{X}_k\}, \mathbf{Y})\Big|_{\mathbf{Y}=\mathbf{Y}_i} \,, \tag{10}$$

where $\nabla_{\mathbf{Y}} Q(\{\mathbf{X}_k\}, \mathbf{Y})$ is the gradient of $Q(\{\mathbf{X}_k\}, \mathbf{Y})$ with respect to $\mathbf{Y}$ and constitutes another vector in the image space. λ is a step parameter controlling the speed of movement in the image space. The iterations continue until a pre-defined convergence criterion is satisfied.

B. MEF-SSIM$_c$ Model

The original MEF-SSIM index works with the luminance component only. This is sufficient for many quality assessment tasks when the main distortions introduced by MEF algorithms are captured by luminance contrast and structural changes [2]. However, in practical MEF applications, recovering a vivid color appearance is pivotal. Therefore, we extend the MEF-SSIM index to work with color images. Recall the patch decomposition used in MEF-SSIM [2] given in (2). To extend such a patch representation for color patches, we concatenate the patch vectors of all color channels into one. Thus, the vector length of $\mathbf{x}_k$ and $\mathbf{s}_k$ becomes CN^2, where N is the patch size and C is the number of color channels. Correspondingly, the physical meanings of c_k, $\mathbf{s}_k$ and l_k of such color patches are generalized to patch strength, patch structure and mean patch intensity, respectively. This patch decomposition inherently takes into account color information as part of the overall contrast and structure. For example, in order to preserve color information, the relative signal strength between multiple color channels in a color space such as RGB is contained in the structural component of the proposed color patch representation. As such, preserving patch structure will also implicitly preserve color information. By contrast, existing MEF algorithms that treat RGB channels separately do not have an appropriate mechanism to enforce color preservation and thus often produce unwanted color or luminance changes.

The original MEF-SSIM index excludes the luminance comparison. When it comes to constructing MEF algorithms, the mean intensity of each color patch needs to be explicitly specified. Inspired by the method used in [11], [39], we estimate the desired mean intensity of the fused image patch by

$$\hat{l} = \frac{\sum_{k=1}^{K} u\left(\mu_k, l_k\right) l_k}{\sum_{k=1}^{K} u\left(\mu_k, l_k\right)} \,, \tag{11}$$

where $u(\cdot)$ is a weighting function that takes the global mean intensity μ_k of the color image $\mathbf{X}_k$ and the local mean intensity of the current patch $\mathbf{x}_k$ as inputs. $u(\cdot)$ quantifies the well exposedness of $\mathbf{x}_k$ in $\mathbf{X}_k$ so that a large penalty is given when $\mathbf{X}_k$ and/or $\mathbf{x}_k$ are under/over-exposed. A two dimensional Gaussian profile is adopted

$$u\left(\mu_k, l_k\right) = \exp\left(-\frac{(\mu_k - \mu_c)^2}{2\sigma_g^2} - \frac{(l_k - l_c)^2}{2\sigma_l^2}\right) \,, \tag{12}$$

where σ_g and σ_l control the spreads of the profile along μ_k and l_k dimensions, respectively. μ_c and l_c are constants for the mid-intensity values. For example, for source image sequences normalized to $[0, 1]$, both μ_c and l_c are 0.5. The first term encourages the spatial intensity consistency across the image while the second term preserves the local structures that are well exposed. By imposing both factors, we are able to produce more visually

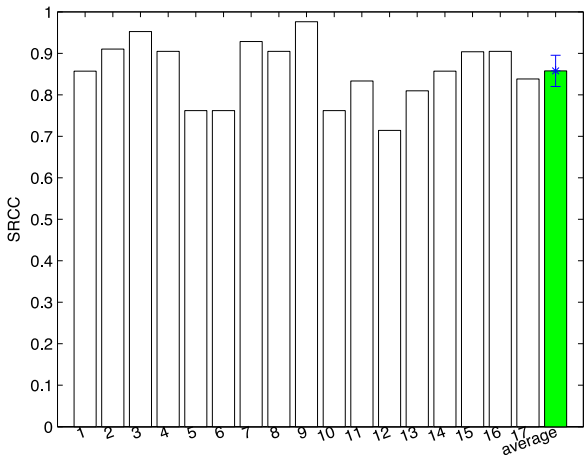


Fig. 1. SRCC between MEF-SSIM$_c$ and MOS on the subjective database [2]. Rightmost column: average performance over image content.

appealing fused images as will be shown in Section IV. Once $\hat{l}$ is computed, we update (6) by

$$\hat{\mathbf{x}} = \hat{c} \cdot \hat{\mathbf{s}} + \hat{l}\,, \tag{13}$$

which now conforms to the patch decomposition of Eq. (2).

The construction of MEF-SSIM$_c$ follows the definition of the SSIM index [33]

$$S(\{\mathbf{x}_k\}, \mathbf{y}) = \frac{(2\mu_{\hat{\mathbf{x}}}\mu_{\mathbf{y}} + C_1)(2\sigma_{\hat{\mathbf{x}}\mathbf{y}} + C_2)}{(\mu_{\hat{\mathbf{x}}}^2 + \mu_{\mathbf{y}}^2 + C_1)(\sigma_{\hat{\mathbf{x}}}^2 + \sigma_{\mathbf{y}}^2 + C_2)}\,, \tag{14}$$

where $\mu_{\hat{\mathbf{x}}}$ and $\mu_{\mathbf{y}}$ denote the mean intensities of the desired color image patch and a given color image patch, respectively. C_1 and C_2 are small positive constants to avoid instability when the denominator is close to 0.

The local quality values of the MEF-SSIM$_c$ index are averaged to obtain an overall quality measure of the fused image

$$Q(\{\mathbf{X}_k\}, \mathbf{Y}) = \frac{1}{M}\sum_{j=1}^{M} S(\{\mathbf{R}_j\mathbf{X}_k\}, \mathbf{R}_j\mathbf{Y})\,, \tag{15}$$

where $\mathbf{R}_j$ is a binary matrix whose number of columns equals the image dimension and number of rows equals to patch size CN^2. It serves as an operator that extracts the j-th local patch from the image. For example, at the j-th location, $\mathbf{x}_k = \mathbf{R}_j\mathbf{X}_k$ and $\mathbf{y} = \mathbf{R}_j\mathbf{Y}$.

In order to demonstrate that MEF-SSIM$_c$ well correlates with human perception, we compute Spearman's rank correlation coefficient (SRCC) between its quality predictions and mean opinion scores (MOSs) on the subjective database in [2], which contains 17 source sequences and 8 state-of-the-art MEF algorithms, resulting in a total of 136 fused images. The SRCC results across different content are shown in Fig. 1 and we achieve 0.858 on average. This verifies the prediction performance of MEF-SSIM$_c$ and its suitability for perceptual optimization.

C. Gradient of MEF-SSIM$_c$

To compute the gradient $\nabla_{\mathbf{Y}} Q(\{\mathbf{X}_k\}, \mathbf{Y})$, we start from the local MEF-SSIM$_c$ index and rewrite (14) as

$$S(\{\mathbf{x}_k\}, \mathbf{y}) = \frac{A_1\,A_2}{B_1\,B_2}\,, \tag{16}$$

where

$$A_1 = 2\mu_{\hat{\mathbf{x}}}\mu_{\mathbf{y}} + C_1 \tag{17}$$

$$B_1 = \mu_{\hat{\mathbf{x}}}^2 + \mu_{\mathbf{y}}^2 + C_1 \tag{18}$$

$$A_2 = 2\sigma_{\hat{\mathbf{x}}\mathbf{y}} + C_2 \tag{19}$$

$$B_2 = \sigma_{\hat{\mathbf{x}}}^2 + \sigma_{\mathbf{y}}^2 + C_2\,. \tag{20}$$

Recalling that both image patches are treated as column vectors of length CN^2, we have the sample statistics given by

$$\mu_{\hat{\mathbf{x}}} = \frac{1}{CN^2}\mathbf{1}^T\,\hat{\mathbf{x}} \tag{21}$$

$$\sigma_{\hat{\mathbf{x}}}^2 = \frac{1}{CN^2}(\hat{\mathbf{x}} - \mu_{\hat{\mathbf{x}}})^T(\hat{\mathbf{x}} - \mu_{\hat{\mathbf{x}}}) \tag{22}$$

$$\sigma_{\hat{\mathbf{x}}\mathbf{y}} = \frac{1}{CN^2}(\hat{\mathbf{x}} - \mu_{\hat{\mathbf{x}}})^T(\mathbf{y} - \mu_{\mathbf{y}})\,, \tag{23}$$

where $\mathbf{1}$ is a CN^2-vector with all entries equaling to 1. $\mu_{\mathbf{y}}$ and $\sigma_{\mathbf{y}}^2$ are computed similarly. The gradient of the local MEF-SSIM$_c$ measure with respect to $\mathbf{y}$ can then be expressed as

$$\nabla_{\mathbf{y}} S(\{\mathbf{x}_k\}, \mathbf{y}) = \frac{(A_1'A_2 + A_1A_2')}{B_1B_2} - \frac{(B_1'B_2 + B_1B_2')A_1A_2}{(B_1B_2)^2}\,, \tag{24}$$

where

$$A_1' = \nabla_{\mathbf{y}} A_1 = \frac{2\mu_{\hat{\mathbf{x}}}}{CN^2}\mathbf{1} \tag{25}$$

$$B_1' = \nabla_{\mathbf{y}} B_1 = \frac{2\mu_{\mathbf{y}}}{CN^2}\mathbf{1} \tag{26}$$

$$A_2' = \nabla_{\mathbf{y}} A_2 = \frac{2}{CN^2}\,(\hat{\mathbf{x}} - \mu_{\hat{\mathbf{x}}}) \tag{27}$$

$$B_2' = \nabla_{\mathbf{y}} B_2 = \frac{2}{CN^2}\,(\mathbf{y} - \mu_{\mathbf{y}})\,. \tag{28}$$

Plugging (17), (18), (19), (20), (25), (26), (27), and (28) into (24), we obtain the gradient of the local MEF-SSIM$_c$ index. Finally, we compute the gradient of the overall MEF-SSIM$_c$ index with respect to the fused image $\mathbf{Y}$. Since gradient is a linear operator, this can be done by summing over all the local gradients

$$\nabla_{\mathbf{Y}} Q(\{\mathbf{X}_k\}, \mathbf{Y}) = \frac{1}{M}\sum_{j=1}^{M} \mathbf{R}_j^T \nabla_{\mathbf{y}} S(\{\mathbf{R}_j\mathbf{X}_k\}, \mathbf{R}_j\mathbf{Y})\,, \tag{29}$$

where $\mathbf{R}_j^T$ inverts the process of $\mathbf{R}_j$ by placing the local gradient patch back into the corresponding location in the image.

TABLE I
INFORMATION OF SOURCE INPUT IMAGE SEQUENCES

Source sequence	Size	Image credit
Arno	$339 \times 512 \times 3$	Bartlomiej Okonek
Balloons	$339 \times 512 \times 9$	Erik Reinhard
Belgium house	$384 \times 512 \times 9$	Dani Lischinski
Cave	$384 \times 512 \times 4$	Bartlomiej Okonek
Chinese garden	$340 \times 512 \times 3$	Bartlomiej Okonek
Church	$512 \times 335 \times 3$	Jianbing Shen
Farmhouse	$341 \times 512 \times 3$	HDR projects
House	$340 \times 512 \times 4$	Tom Mertens
Kluki	$341 \times 512 \times 3$	Bartlomiej Okonek
Landscape	$341 \times 512 \times 3$	HDRsoft
Lamp	$384 \times 512 \times 15$	Martin Ĉadík
Laurenziana	$512 \times 356 \times 3$	Bartlomiej Okonek
Lighthouse	$340 \times 512 \times 3$	HDRsoft
Madison capitol	$384 \times 512 \times 30$	Chaman Singh Verma
Mask	$341 \times 512 \times 3$	HDRsoft
Office	$340 \times 512 \times 6$	MATLAB
Ostrow	$341 \times 512 \times 3$	Bartlomiej Okonek
Room	$341 \times 512 \times 3$	Pangeasoft
Set	$341 \times 512 \times 3$	Jianbing Shen
Studio	$341 \times 512 \times 5$	HDRsoft
Tower	$512 \times 341 \times 3$	Jacques Joffre
Venice	$341 \times 512 \times 3$	HDRsoft
Window	$384 \times 512 \times 3$	Hvdwolf
Yellow hall	$339 \times 512 \times 3$	Jianbing Shen

Fig. 2. Source image sequences used in the subjective experiment. Each sequence is represented by one fused image that has the best quality in the subjective experiment.

D. Implementation Details

The parameters of the proposed approach come from two parts: the model parameters in the MEF-SSIM$_c$ index and the parameters that control the optimization process. Most of the model parameters are inherited from previous publications. These include the adaptively determined p in (5) based on the patch consistency measure in [2]; two spread parameters $\sigma_g = 0.2$ and $\sigma_l = 0.2$ in (12) from [11]; two stabling constants $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$, where $K_1 = 0.01$ and $K_2 = 0.03$ from [33], and L is the dynamic range of the source image sequence ($L = 255$ for an 8-bit sequence). To reduce the computation in each iteration, we use an 8×8 square window to compute the local statistics. Our optimization procedure stops when the difference between the MEF-SSIM$_c$ values of two consecutive iterations is smaller than a threshold, that is, $|Q_{k+1} - Q_k| < \epsilon = 10^{-6}$. The step size λ is empirically set to 150 throughout the paper to achieve a balance between convergence speed and stability.

IV. EXPERIMENTAL RESULTS

Twenty four source image sequences are selected in our experiment, which span diverse scenes containing both light and dark regions with different color appearances, as shown in Table I and Fig. 2. On the other hand, the proposed algorithm is initialized with the fused images created by 14 existing MEF algorithms. These include two simple operators that linearly combine the input images using local and global energy as weighting factors denoted by LE and GE, respectively, and sophisticated ones with different perceptual emphasis such as Mertens09 [11], Raman09 [5], Shen11 [6], Zhang12 [13], Song12 [7], Li12 [9], Shutao12 [8], Gu12 [12], Li13 [10], Bruce14 [43], Shen14 [19], and Ma15 [39]. All fusion results of existing MEF algorithms are either obtained from the authors or generated by publicly available implementations with default settings.

A. A Visual Demonstration

To better understand and to demonstrate the proposed MEF-SSIM$_c$ optimization algorithm, we visually exam how an arbitrary image evolves to a fused image of high quality. Specifically, we first apply the proposed algorithm to initial images that contain no specific information about the source image sequences. Three types of such initial images are chosen: 1) a flat uniform gray image with a constant mid-level pixel intensity (127 for an 8-bit image); 2) a white Gaussian random noise image; 3) a natural image that is completely irrelevant to the source image sequence. The demonstration is shown in Fig. 3, where it can be observed that regardless of the arbitrary initializations, the proposed algorithm is able to explore the image space over iterations and finally find (converge to) a high quality fused image of faithful structures and vivid color appearance with respect to the given source sequence. The quality improvement during the MEF-SSIM$_c$ optimization is also clearly reflected in the corresponding MEF-SSIM$_c$ quality maps, where brighter indicates better quality.

B. Improvement Upon Existing MEF Algorithms

In practice, to explore the image space more efficiently, it is useful to employ an image created by an existing MEF algorithm

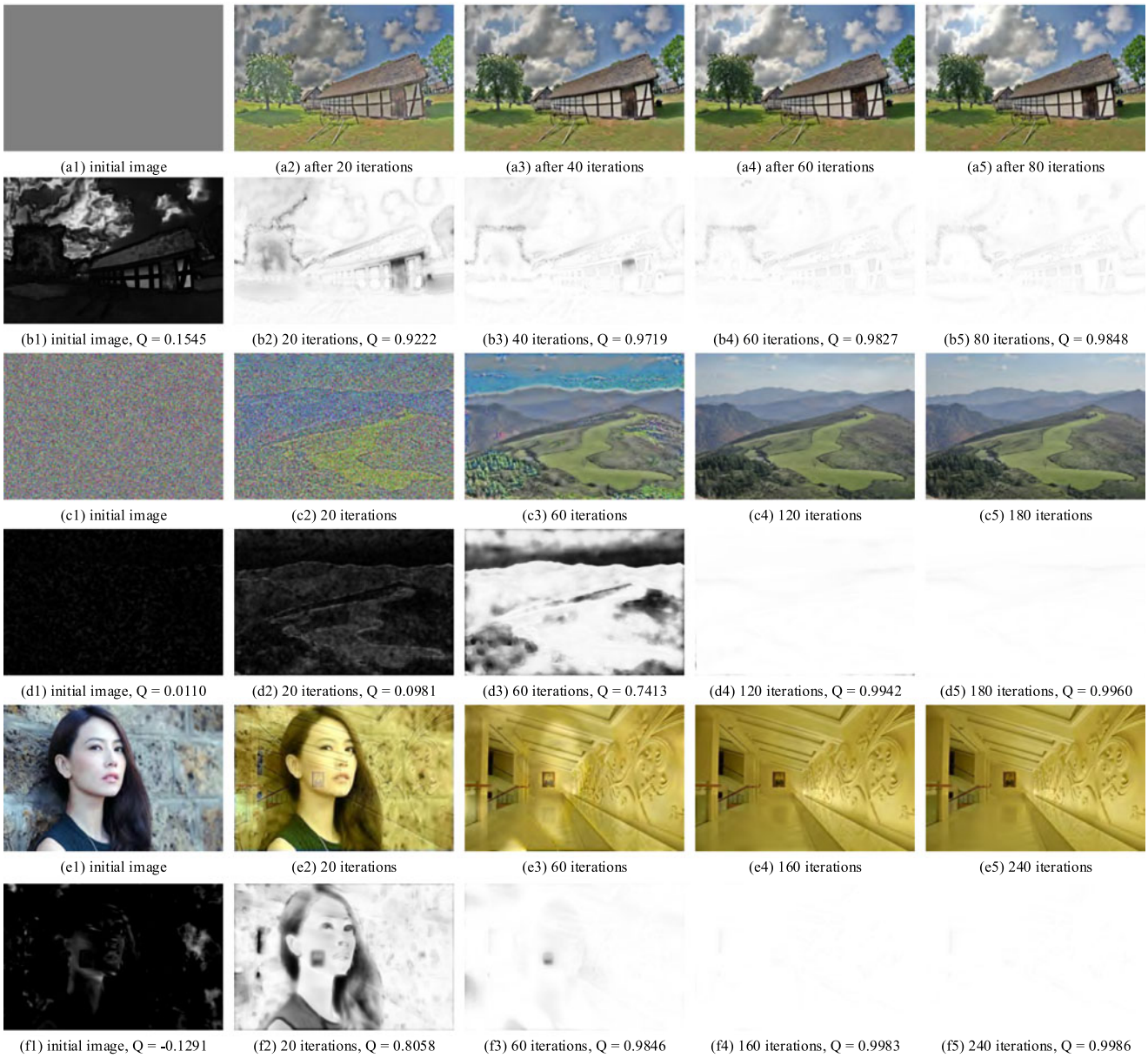


Fig. 3. Visual demonstration of the proposed MEF-SSIM$_c$ optimization algorithm. (a1), (c1), and (e1) are initial images that contain no information about the source image sequences. (a2)–(a5), (c2)–(c5), and (e2)–(e5) are the fused images "Kluki", "Landscape", and "Yellow hall" created by the proposed iterative method, respectively. (b1)–(b5), (d1)–(d5), and (f1)–(f5) are the MEF-SSIM$_c$ quality maps of (a1)–(a5), (c1)–(c5), and (e1)–(e5), respectively, where brighter indicates better quality. Image (e1) by courtesy of Yuanyuan Gao's Studio.

as the initial image for the iterative MEF-SSIM$_c$ optimization. We have carried out extensive experiments this way to improve upon 10 existing MEF algorithms. Here, we first provide visual comparisons of MEF images before and after MEF-SSIM$_c$ optimization. We then provide numerical analysis to confirm the quality gains obtained through the process. Further verifications by subjective testing are presented in Section IV-C.

Fig. 4 provides the visual comparison where we use Song12 [7] to generate the initial image of the "Farmhouse" sequence. As can be observed, Song12 [7] introduces severe color distortions in the fused image with a noisy overall appearance. These are typical distortions generated by pixel-wise MEF algorithms if their weighting maps are not heavily smoothed before fusion. The proposed approach, without specific detection and treatment of the artifacts, successfully corrects the problem and makes the image much sharper and cleaner.

Fig. 5 compares the case when using Gu12 [12] to create the initial image of the "House" sequence. Based on gradient information, Gu12 [12] focuses on detail enhancement only and fails to preserve the color appearance in the source sequence. As a result, the fused image appears to be pale and

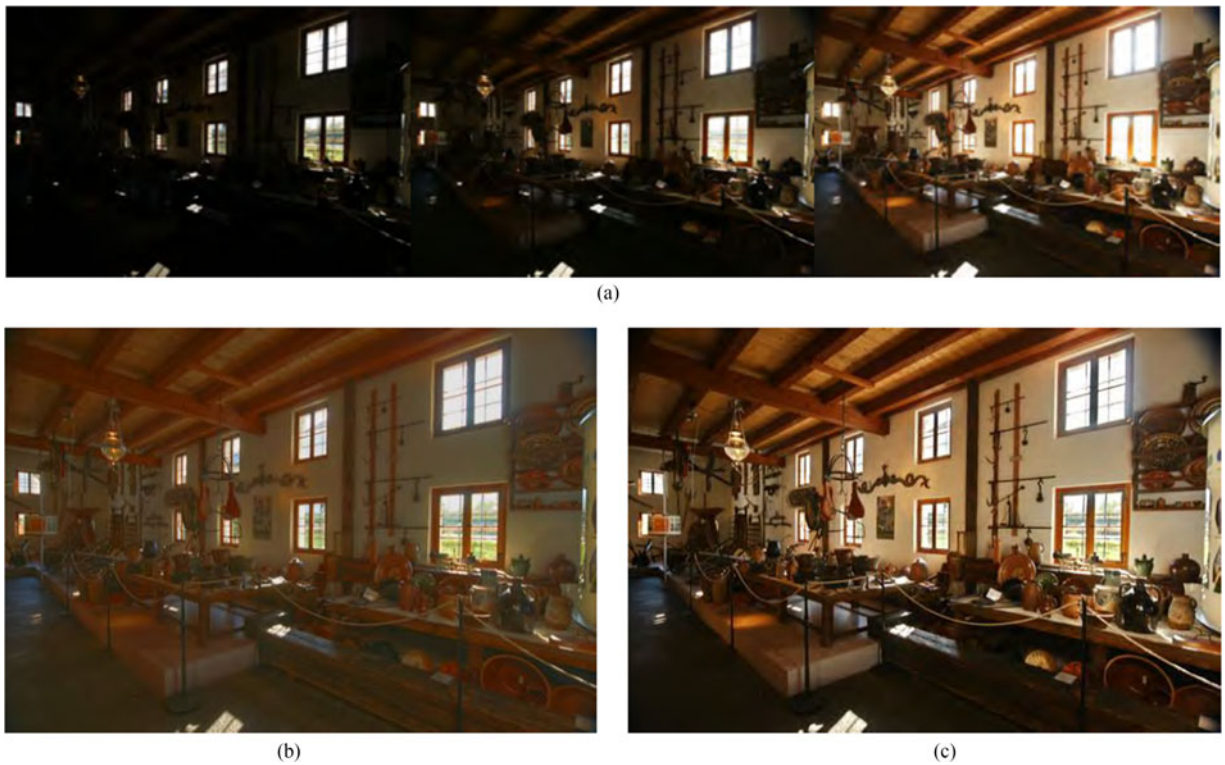


Fig. 4. Visual comparison of MEF images. (a) Source image sequence "Farmhouse" by courtesy of HDR projects. (b) Initial image generated by Song12 [7]. (c) Final image after MEF-SSIM$_c$ optimization.

Fig. 5. Visual comparison of MEF images. (a) Source image sequence "House" by courtesy of Tom Mertens. (b) Initial image generated by Gu12 [12]. (c) Final image after MEF-SSIM$_c$ optimization.

Fig. 6. Visual comparison of MEF images. (a) Source image sequence "Arno" by courtesy of Bartlomiej Okonek. (b) Initial image generated by Shen14 [19]. (c) Final image after MEF-SSIM$_c$ optimization.

unnatural. By contrast, the MEF-SSIM$_c$ optimized image recovers the vivid color appearance and looks much more natural and warmer.

Fig. 6 demonstrates the case when using Shen14 [19] for the initialization of the "Arno" sequence. It is apparent that the fused image by Shen14 [19] suffers from heavy halo artifacts near edges and drastic luminance changes between the sun and cloud regions. Through MEF-SSIM$_c$ optimization, these artifacts are fully repaired. As a result, the MEF-SSIM$_c$ optimized image looks much more natural and visually appealing.

Finally, Fig. 7 shows the case when the initial image is given by Li12 [9], which is a detail enhancement algorithm based on Mertens09 [11]. It can be seen that it fails to preserve some salient details such as the top of the tower and the brightest region of the cloud at the middle-left part of the image. By contrast, those details are faithfully recovered in the MEF-SSIM$_c$ optimized image. Moreover, the overall brightness of the image is much more consistent.

To provide a comprehensive quantitative analysis regarding the improvement of the proposed MEF-SSIM$_c$ optimization approach upon existing MEF algorithms, we list the MEF-SSIM$_c$ values between the initial images and the corresponding MEF-SSIM$_c$ optimized images in Table II. The initial images of 24 source image sequences are created by 10 MEF algorithms including 2 naïve and 8 state-of-the-art ones. From the table, we observe that consistent improvement in terms of MEF-SSIM$_c$ is achieved even when the initial images are created by the most competitive MEF algorithms [6], [11]. This suggests that the proposed MEF algorithm is highly effective at optimizing MEF-SSIM$_c$.

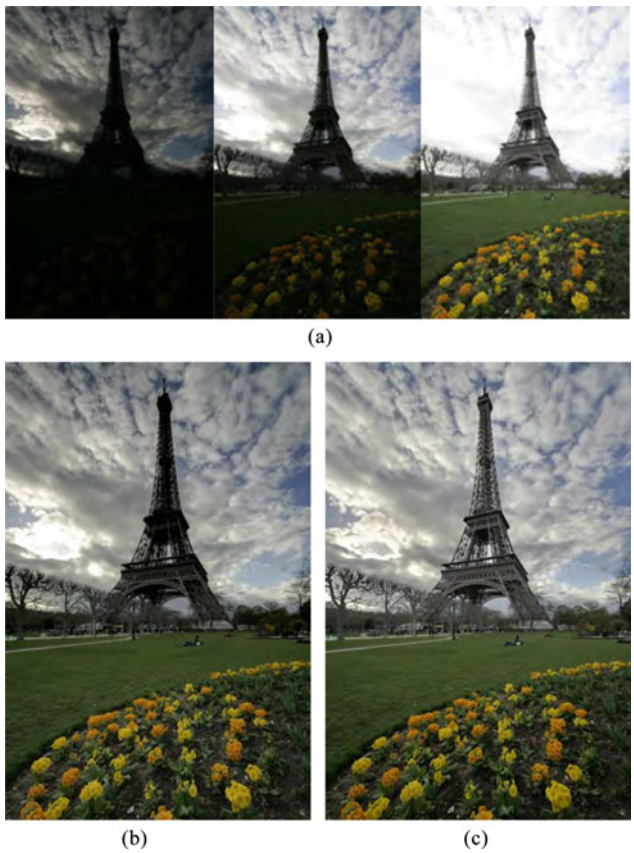


Fig. 7. Visual comparison of MEF images. (a) Source image sequence "Tower" by courtesy of Jacques Joffre. (b) Initial image generated by Li12 [9]. (c) Final image after MEF-SSIM$_c$ optimization.

TABLE II
MEF-SSIM$_c$ COMPARISON BETWEEN INITIAL AND OPTIMIZED IMAGES

Image		LE	GE	Mertens09 [11]	Raman09 [5]	Shen11 [6]	Shutao12 [8]	Zhang12 [13]	Bruce14 [43]	Li13 [10]	Ma15 [39]
Arno	Initial	0.9468	0.9388	0.9474	0.8936	0.9604	0.9681	0.9342	0.9258	0.9002	0.9767
	Optimized	0.9935	0.9935	0.9935	0.9934	0.9936	0.9935	0.9935	0.9935	0.9934	0.9935
Balloons	Initial	0.8425	0.7422	0.9509	0.5408	0.9343	0.9288	0.8478	0.6044	0.9470	0.9501
	Optimized	0.9913	0.9913	0.9913	0.9911	0.9913	0.9913	0.9913	0.9911	0.9913	0.9913
Belgium house	Initial	0.8073	0.7389	0.9495	0.5854	0.9015	0.9207	0.8349	0.6136	0.9490	0.9544
	Optimized	0.9817	0.9815	0.9820	0.9812	0.9819	0.9819	0.9818	0.9813	0.9820	0.9819
Cave	Initial	0.9251	0.7468	0.9526	0.4926	0.9385	0.9689	0.7833	0.8808	0.9813	0.9723
	Optimized	0.9907	0.9906	0.9906	0.9905	0.9907	0.9907	0.9908	0.9907	0.9908	0.9908
Chinese garden	Initial	0.9296	0.8374	0.9577	0.7976	0.9179	0.9734	0.8995	0.8885	0.9819	0.9695
	Optimized	0.9925	0.9925	0.9926	0.9928	0.9925	0.9926	0.9929	0.9928	0.9926	0.9927
Church	Initial	0.9156	0.8596	0.9503	0.7385	0.8763	0.9674	0.8688	0.8320	0.9870	0.9813
	Optimized	0.9927	0.9926	0.9928	0.9922	0.9927	0.9928	0.9905	0.9924	0.9928	0.9928
Farmhouse	Initial	0.9428	0.8280	0.9760	0.7296	0.9450	0.9766	0.9155	0.8618	0.9788	0.9807
	Optimized	0.9930	0.9929	0.9931	0.9928	0.9930	0.9931	0.9930	0.9929	0.9931	0.9931
House	Initial	0.8049	0.8003	0.9196	0.7198	0.8867	0.8578	0.9046	0.8377	0.9005	0.9134
	Optimized	0.9690	0.9690	0.9690	0.9693	0.9691	0.9688	0.9692	0.9692	0.9689	0.9692
Kluki	Initial	0.9319	0.8906	0.9323	0.8807	0.9341	0.9484	0.9223	0.9282	0.9391	0.9428
	Optimized	0.9853	0.9851	0.9852	0.9854	0.9852	0.9852	0.9855	0.9854	0.9850	0.9852
Lamp	Initial	0.8043	0.7269	0.9451	0.5733	0.9054	0.9212	0.9062	0.5756	0.9376	0.9309
	Optimized	0.9806	0.9803	0.9806	0.9805	0.9806	0.9806	0.9807	0.9804	0.9805	0.9807
Landscape	Initial	0.9624	0.9252	0.9914	0.9045	0.9423	0.9837	0.9618	0.9620	0.9764	0.9747
	Optimized	0.9963	0.9963	0.9963	0.9963	0.9963	0.9962	0.9963	0.9963	0.9961	0.9963
Laurenziana	Initial	0.9302	0.8654	0.9418	0.8481	0.9128	0.9647	0.9125	0.9214	0.9615	0.9490
	Optimized	0.9880	0.9880	0.9880	0.9880	0.9880	0.9880	0.9880	0.9879	0.9880	0.9880
Lighthouse	Initial	0.9521	0.8807	0.9706	0.8820	0.9472	0.9185	0.9582	0.9571	0.9261	0.9599
	Optimized	0.9953	0.9953	0.9953	0.9953	0.9953	0.9951	0.9953	0.9953	0.9950	0.9953
Madison capitol	Initial	0.8534	0.7594	0.9298	0.5747	0.8659	0.8476	0.8520	0.5999	0.9280	0.9471
	Optimized	0.9801	0.9802	0.9802	0.9801	0.9801	0.9799	0.9802	0.9801	0.9799	0.9802
Mask	Initial	0.9352	0.8581	0.9649	0.8055	0.9188	0.9797	0.9220	0.8816	0.9756	0.9673
	Optimized	0.9927	0.9928	0.9928	0.9924	0.9928	0.9928	0.9928	0.9928	0.9928	0.9928
Office	Initial	0.9178	0.9457	0.9627	0.8696	0.9439	0.9640	0.9442	0.9224	0.9450	0.9773
	Optimized	0.9905	0.9912	0.9906	0.9914	0.9898	0.9913	0.9916	0.9915	0.9910	0.9906
Ostrow	Initial	0.9052	0.9271	0.8941	0.9186	0.9280	0.9474	0.9256	0.9434	0.9532	0.9662
	Optimized	0.9932	0.9932	0.9932	0.9906	0.9932	0.9932	0.9896	0.9896	0.9932	0.9932
Room	Initial	0.9366	0.8851	0.9084	0.8647	0.9007	0.9438	0.8936	0.9048	0.9646	0.9673
	Optimized	0.9878	0.9877	0.9878	0.9709	0.9878	0.9877	0.9718	0.9722	0.9878	0.9878
Set	Initial	0.9527	0.9290	0.9739	0.9303	0.9475	0.9615	0.9524	0.9649	0.9656	0.9821
	Optimized	0.9954	0.9953	0.9953	0.9953	0.9954	0.9953	0.9954	0.9954	0.9953	0.9954
Studio	Initial	0.8839	0.6901	0.8993	0.6243	0.8957	0.9255	0.8352	0.7524	0.9347	0.9170
	Optimized	0.9799	0.9797	0.9799	0.9796	0.9799	0.9799	0.9798	0.9796	0.9798	0.9799
Tower	Initial	0.9405	0.8678	0.9541	0.8183	0.9201	0.9766	0.9265	0.9223	0.9780	0.9563
	Optimized	0.9925	0.9924	0.9925	0.9925	0.9925	0.9925	0.9925	0.9925	0.9925	0.9925
Venice	Initial	0.9020	0.8372	0.8848	0.7970	0.8952	0.9321	0.8445	0.9016	0.9500	0.9360
	Optimized	0.9757	0.9756	0.9756	0.9756	0.9756	0.9756	0.9756	0.9756	0.9757	0.9757
Window	Initial	0.9281	0.8821	0.9381	0.7960	0.9236	0.9647	0.8881	0.8849	0.9705	0.9713
	Optimized	0.9866	0.9863	0.9847	0.9887	0.9868	0.9865	0.9889	0.9888	0.9861	0.9867
Yellow hall	Initial	0.9742	0.9591	0.9830	0.9577	0.9677	0.9600	0.9668	0.9633	0.9477	0.9902
	Optimized	0.9986	0.9986	0.9986	0.9985	0.9986	0.9986	0.9986	0.9985	0.9985	0.9986
Average	Initial	**0.9094**	**0.8467**	**0.9449**	**0.7726**	**0.9212**	**0.9459**	**0.9000**	**0.8513**	**0.9533**	**0.9597**
	Optimized	**0.9885**	**0.9884**	**0.9884**	**0.9877**	**0.9884**	**0.9885**	**0.9877**	**0.9877**	**0.9884**	**0.9885**

LE and GE stand for local and global energy based weighting, respectively.

TABLE III
MEAN OPINION SCORES OF FUSED IMAGES BEFORE AND AFTER MEF-SSIM$_c$ OPTIMIZATION

Image set	Mean opinion scores (MOS)							
	Mertens09 [11]		Gu12 [12]		Bruce14 [43]		Shen14 [19]	
	Initial	Optimized	Initial	Optimized	Initial	Optimized	Initial	Optimized
Arno	5.71	7.83	3.88	8.13	4.83	8.08	5.21	7.88
Balloons	7.21	8.33	5.21	7.96	3.25	8.83	3.13	7.54
Belgium house	6.46	8.17	4.67	7.96	4.46	8.00	2.63	8.29
Cave	7.08	7.83	5.67	8.38	5.71	8.04	2.83	8.25
Chinese garden	7.38	8.29	4.50	8.38	6.17	7.83	2.67	8.21
Church	6.68	7.71	5.63	7.63	5.83	7.96	4.42	8.25
Farmhouse	6.58	7.29	8.33	7.21	4.50	7.46	5.42	7.33
House	7.92	7.54	3.29	8.13	5.08	8.04	2.21	7.88
Kluki	6.79	8.13	4.50	8.13	3.29	8.00	3.29	8.33
Lamp	7.79	7.04	3.54	7.75	3.92	7.54	1.29	7.33
Landscape	7.33	6.75	3.96	7.67	8.33	7.50	3.13	6.88
Laurenziana	7.46	8.25	3.63	7.92	6.33	8.13	3.38	8.21
Lighthouse	7.79	8.04	6.38	7.38	7.21	7.92	2.38	8.00
Madison capitol	7.04	7.67	5.00	7.13	4.58	7.58	2.00	7.33
Mask	7.25	8.04	4.00	8.25	4.21	7.96	2.29	8.38
Office	8.08	7.75	4.13	8.21	6.25	7.25	2.08	7.88
Ostrow	7.38	7.50	5.25	7.71	6.00	7.88	4.79	7.92
Room	7.00	7.92	4.75	7.71	6.54	7.92	3.50	7.79
Set	7.96	8.25	5.21	8.21	8.08	8.50	3.83	8.29
Studio	6.96	7.17	3.92	7.42	5.50	6.96	2.42	6.80
Tower	6.38	7.75	4.67	8.33	5.50	8.00	2.75	8.08
Venice	7.13	7.96	5.00	7.88	6.29	7.25	2.29	8.08
Window	7.63	7.58	6.04	7.46	7.00	7.83	3.88	8.00
Yellow hall	8.33	7.79	7.25	8.04	7.04	7.67	5.50	8.33
Average	**7.22**	**7.77**	**4.93**	**7.87**	**5.82**	**7.82**	**3.22**	**7.88**

(a) (b)

Fig. 8. A representative image pair for the case of the MOS drop after MEF-SSIM$_c$ optimization. (a) "Farmhouse" by Gu12 [12]. (b) After MEF-SSIM$_c$ optimization.

C. Validation by Subjective Testing

In order to further verify that MEF-SSIM$_c$ optimization indeed results in perceptual gains of the fused images, we carry out a subjective experiment. Specifically, the same 24 source image sequences are included (shown in Fig 2). Four existing MEF algorithms are included, which are LE, Gu12 [12], Bruce14 [43], and Mertens09 [11], respectively. The last method is shown to be the best MEF algorithm on average in a recent subjective experiment [2]. The 96 fused images are then served as initial images of the proposed MEF-SSIM$_c$ optimization algorithm, resulting in 96 corresponding MEF-SSIM$_c$ optimized images. In summary, we have a total of 192 fused images, which can be divided into 24 sets with 8 images per set according to the scene content. In our subjective test, we adopt a multi-stimulus quality scoring strategy without showing the reference source image sequence. More specifically, all 8 fused images in one set are shown to the subject at the same time on one computer screen at actual pixel resolution but in random spatial order. The subject is asked to give an integer score for each images that best reflects their perceptual quality. The score ranges from 0 to 10, where 10 denotes the best quality and 0 the worst. A total of 24 naïve observers, including 11 males and 13 females participate in the subjective experiment. One outlier is detected and removed based on the outlier removal scheme suggested in [44]. The final rating score for each image is computed as the average of the subjective scores, namely the MOS, from all valid subjects.

TABLE IV
STATISTICAL SIGNIFICANCE MATRIX BASED ON T-STATISTICS

	M	G	B	S	MO	GO	BO	SO
M	-	1	1	1	0	0	0	0
G	0	-	0	1	0	0	0	0
B	0	1	-	1	0	0	0	0
S	0	0	0	-	0	0	0	0
MO	1	1	1	1	-	-	-	-
GO	1	1	1	1	-	-	-	-
BO	1	1	1	1	-	-	-	-
SO	1	1	1	1	-	-	-	-

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. M: Mertens09 [11]; G: Gu12 [12]; B: Bruce14 [43]; S: Shen14 [19]. O: After MEF-SSIM$_c$ optimization.

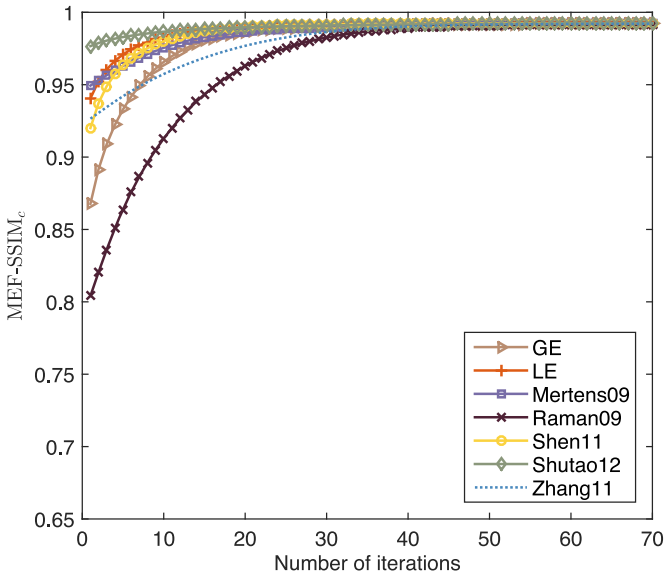


Fig. 9. MEF-SSIM$_c$ as a function of iteration on the "Tower" sequence with initial fused images created by different MEF algorithms.

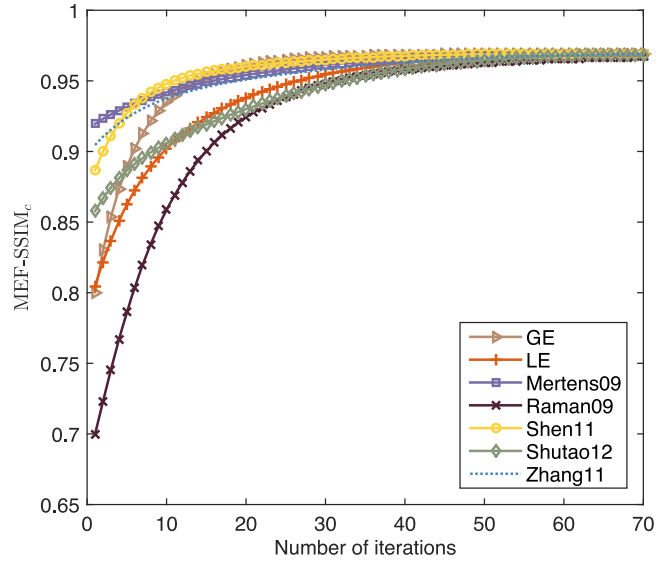


Fig. 10. MEF-SSIM$_c$ as a function of iteration on the "House" sequence with initial fused images created by different MEF algorithms.

The results are listed in Table III, from which we have several interesting observations. First, using MEF-SSIM$_c$ as the optimization goal, the proposed algorithm leads to consistent perceptual gains for all four types of initial images. Second, the proposed algorithm is generally insensitive to initializations in the sense that no matter which initial image is used, the proposed algorithm is able to recover a high quality image of similar perceptual quality. Third, 7 out of 96 images experience a MOS drop after MEF-SSIM$_c$ optimization and a representative image pair is shown in Fig. 8. It can be observed that the proposed optimization process enhances the overall contrast and reduces the ringing artifacts near sharp edges, but meanwhile reduces the overall luminance. This may explain the MOS drop. We have similar observations on the other cases as well.

We use a hypothesis testing approach based on t-statistics [45] to evaluate the statistical significance of the subjective experimental results. Specifically, we treat the MOSs of each column in Table III as one category. The null hypothesis is that the MOSs in one category is statistically indistinguishable (with 95% confidence) from those in another category. The test is carried out for all possible pair combinations of categories and the results are summarized in Table IV, from which we can see that MEF-SSIM$_c$ optimized images have statistically significantly better MOSs in all cases. The MOSs between MEF-SSIM$_c$ optimized images with different initializations are statistically indistinguishable.

In summary, we believe that the proposed approach robustly produces better quality fused images regardless of initial images. Moreover, it reinforces the robustness and usefulness of MEF-SSIM$_c$ as an objective quality model in the area of MEF for not only comparing the performance and tuning the parameters of MEF algorithms but also guiding the design of new MEF algorithms.

D. Convergence and Complexity

Because of the nonconvexity of MEF-SSIM$_c$ and the dimension of the search space, analytical convergence assessment of the proposed algorithm is difficult. Therefore, we observe the convergence performance empirically. Figs. 9 and 10 show the MEF-SSIM$_c$ value as a function of iteration on the "Tower" and "House" sequences, respectively, using different initial images as starting points. There are several useful observations. First, the MEF-SSIM$_c$ values increase monotonically with iterations. Second, the proposed algorithm converges in all cases regardless of the initial images. Third, different initial images may result in slightly different converged images. From these observations, we conclude that the proposed iterative algorithm is well behaved, but the high-dimensional search space is complex and contains many local optima, and the proposed algorithm may be trapped in one of them. Even though finding the absolute global optimum is difficult, all local optima correspond to similar and high quality images; thus a local optimum is sufficient to provide a useful solution in practical applications.

Since the statistics of the source image sequence can be pre-computed and stored before optimization, the computation

complexity of the proposed algorithm does not increase with the number of exposures in the source image sequence and has a computational complexity of $\mathcal{O}(N^2M)$ in each iteration, where M is the number of patches in one exposure and N is the window size. Note that although most MEF algorithms are pixel-wise, they compute the perceptual weights within a local window and therefore have a similar complexity of $\mathcal{O}(N^2MK)$ [4], [5], [8]–[13], [39], where K is the number of exposures. Our unoptimized MATLAB implementation takes around 2.5 seconds per iteration for a $341 \times 512 \times 30$ sequence on a computer with an Intel Quad-Core 2.67 GHz processor.

V. Conclusion and Discussion

In this paper, we propose a substantially different approach to design MEF algorithms. In particular, we directly operate in the space of all images, iteratively searching for an image that optimizes MEF-SSIM$_c$, which is an advanced MEF image quality model built upon the existing MEF-SSIM index. Extensive experimental results show that the proposed approach consistently produces better quality fused images both qualitatively and quantitatively. Moreover, the optimization algorithm is well behaved in the sense that given any initial image, it is able to find a high quality fused image with both sharp structures and vivid color appearance.

The proposed optimization framework is general. Whenever a new and better MEF image quality model is available, it can be integrated into the same framework and creates a better MEF algorithm. The most challenging part of such an algorithm would likely to be the computation of the gradient of the new model in the space of images. On the other hand, the MEF-SSIM$_c$ optimization algorithm proposed in this paper can be improved in many ways. Specifically, since the proposed algorithm is iterative, it may not be suitable for real-time applications. An efficient non-iterative algorithm with the spirit of MEF-SSIM$_c$ in mind is highly desirable. Furthermore, the proposed algorithm can only find local optima due to the nonconvexity of MEF-SSIM$_c$. Relaxing MEF-SSIM$_c$ into a convex metric would benefit from many powerful convex optimization tools [46] to find the global optimum efficiently.

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*. Burlington, MA, USA: Morgan Kaufmann, 2010.

[2] 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.

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

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

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

[6] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3634–3646, Dec. 2011.

[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] 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.

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

[10] 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.

[11] 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, Mar. 2009.

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

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

[14] 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.

[15] M. Bertalmio and S. Levine, "Variational approach for the fusion of exposure bracketed pairs," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 712–723, Feb. 2013.

[16] K. Hara, K. Inoue, and K. Urahama, "A differentiable approximation approach to contrast-aware image fusion," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 742–745, Jun. 2014.

[17] 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.

[18] 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.

[19] J. Shen, Y. Zhao, S. Yan, and X. Li, "Exposure fusion using boosting Laplacian pyramid," *IEEE Trans. Cybern.*, vol. 44, no. 9, pp. 1579–1590, Sep. 2014.

[20] Z. Li, J. Zheng, Z. Zhu, W. Yao, and S. Wu, "Weighted guided image filtering," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 120–129, Jan. 2015.

[21] K. Ma, H. Li, H. Yong, D. Meng, Z. Wang, 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.

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

[23] 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.

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

[25] 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.

[26] Q. Wang, Y. Shen, and J. Jin, "Performance evaluation of image fusion techniques," *Image Fusion: Algorithms Appl.*, pp. 469–492, 2008.

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

[28] 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, Apr. 2007.

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

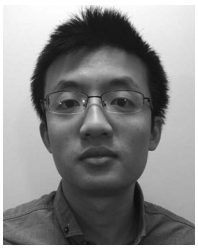
[30] H. Chen and P. K. Varshney, "A human perception inspired quality metric for image fusion based on regional information," *Inf. Fusion*, vol. 8, no. 2, pp. 193–207, Apr. 2007.

[31] 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, Sep. 2009.

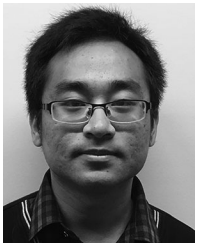
[32] C. Yang, J.-Q. Zhang, X.-R. Wang, and X. Liu, "A novel similarity based quality metric for image fusion," *Inf. Fusion*, vol. 9, no. 2, pp. 156–160, Apr. 2008.

[33] 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.

[34] D. Brunet, E. R. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1488–1499, Apr. 2012.
[35] Z. Wang and E. P. Simoncelli, "Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities," *J. Vis.*, vol. 8, no. 12, pp. 1–13, Sep. 2008.
[36] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.
[37] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Perceptual video coding based on SSIM-inspired divisive normalization," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1418–1429, Apr. 2013.
[38] P. J. Burt, *The Pyramid as A Structure for Efficient Computation*. Berlin, Germany: Springer, 1984.
[39] K. Ma and Z. Wang, "Multi-exposure image fusion: A patch-wise approach," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 1717–1721.
[40] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1998, pp. 839–846.
[41] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[42] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image compression by optimizing tone mapped image quality index," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3086–3097, Oct. 2015.
[43] N. D. Bruce, "Expoblend: Information preserving exposure blending based on normalized log-domain entropy," *Comput. Graph.*, vol. 39, pp. 12–23, Apr. 2014.
[44] ITU-R, "BT.500-13: Methodology for the subjective assessment of the quality of television pictures," Jan. 2012.
[45] D. C. Montgomery, *Applied Statistics and Probability for Engineers*, 6th ed. New York, NY, USA: Wiley, 2013.
[46] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.



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.S. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2014 and 2017, respectively. He is currently a Research Associate with Howard Hughes Medical Institute and Laboratory for Computational Vision, New York University, New York, NY, USA. His research interests include perceptual image processing, computational vision, and computational photography.



Zhengfang Duanmu (S'15) received the B.A.Sc. and the M.A.Sc. degrees in electrical and computer engineering from the University of Waterloo in 2015 and 2017, respectively, where he is currently working toward the Ph.D. degree in electrical and computer engineering. His research interests include perceptual image processing and quality of experience.

Hojatollah Yeganeh (S'10–M'14) received the B.S. degree in electrical engineering from the Isfahan University of Technology, Isfahan, Iran, and the M.Sc. degree in electrical engineering, specializing in speech processing and speech recognition, from the Amirkabir University of Technology, Tehran, Iran, in 2006 and 2009, respectively, and the the Ph.D. degree in electrical and computer engineering from University of Waterloo, Waterloo, ON, Canada, in 2014. where he is currently a visiting scholar with the Department of Electrical and Computer Engineering, University of Waterloo. His main research interests include image and video processing, quality assessment of High Dynamic Range images and videos, Quality of Experience of Internet videos, multimedia communications, and biomedical signal processing.

Zhou Wang (S'99–M'02–SM'12–F'14) received the Ph.D. degree from The University of Texas, Austin, TX, USA, in 2001. He is currently a Professor and University Research Chair with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, 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 more than 200 publications in these fields with more than 40,000 citations (Google Scholar).

Dr. Wang serves as a Senior Area Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (2015–present), and an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2016–present). Previously, he served as a member for the IEEE MULTIMEDIA SIGNAL PROCESSING TECHNICAL COMMITTEE (2013–2015), an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (2009–2014), *Pattern Recognition* (2006–present) and IEEE SIGNAL PROCESSING LETTERS (2006–2010), and a Guest Editor of IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (2013–2014 and 2007–2009). He is a Fellow of Canadian Academy of Engineering, and the recipient of the 2017 Faculty of Engineering Research Excellence Award at 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.