

Detail-Preserving Multi-Exposure Fusion with Edge-Preserving Structural Patch Decomposition

Hui Li, Tsz Nam Chan, *Member, IEEE*, Xianbiao Qi, and Wuyuan Xie

Abstract—The multi-exposure fusion (MEF) methods have received much attention in recent years due to the importance of constructing high dynamic range images. Among most of the existing studies, multi-scale structural-patch-decomposition-based MEF (MSPD-MEF) has achieved state-of-the-art fusion quality and the fastest running time. However, this method still suffers from detail loss in the fused images. To tackle this issue, we first incorporate the edge-preserving factors into this method to preserve the details in the fused images in a single-scale setting. Then, we develop the novel and flexible bell curve function, which can further preserve the details in both bright and dark regions. After that, we also show that our method can seamlessly plug in to this multi-scale framework. Extensive experimental results indicate that the proposed method can produce pleasing fusion results with little artifacts and low computational cost in both static and dynamic scenes.

Index Terms—Multi-exposure fusion, high dynamic range imaging, edge-preserving, structural patch decomposition

I. INTRODUCTION

THE dynamic range of a natural scene is much higher than that of the image which generic digital cameras can capture within one single shot. As a consequence, over-exposure or under-exposure effect occurs frequently in the general photo-taking experience, leading to unpleasing imaging results with severe information loss. High dynamic range (HDR) imaging can effectively address this issue with a sequence of differently exposed low dynamic range (LDR) images. There are usually two strategies to get an HDR-like image: the multi-exposure fusion (MEF) [1]–[4] in image domain, and HDR content reconstruction via camera response function (CRF) and then tone reproduction [5] in radiance domain. Due to the complexity of recovering CRF and designing tone mappers [5], [6], MEF has been widely applied to HDR imaging in mobile devices, as a direct and simple approach. Besides, MEF has also been used for low-light image enhancement [7], [8], dehazing [9] and saliency detection [10] by fusing generative pseudo exposure sequences.

A good MEF method should work robustly in both static and dynamic scenes with good visual quality, and consume

Hui Li is with the Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail:xiaohui102788@126.com).

Tsz Nam Chan is with the Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR (e-mail: edisonchan@comp.hkbu.edu.hk).

Xianbiao Qi is with the Visual Computing Group, Ping An Property & Casualty Insurance Company, Shenzhen, China (e-mail: qixianbiao@gmail.com).

Wuyuan Xie is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China (e-mail: wuyuan.xie@gmail.com).

Copyright © 20xx IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

low computational resources especially when high-resolution images are captured via cameras or mobile phones. Recently, Zhang *et al.* [11] and Ma *et al.* [3] successfully applied the structure path decomposition (SPD) method to MEF (we term this MEF solution as SPD-MEF), which can achieve the state-of-the-art results in both static and dynamic scenes. Due to the superior performance of SPD-MEF in practice, this method has also been applied to other applications in the past few years, e.g., medical image fusion [12], and infrared and visible image fusion [13]. The general idea of SPD-MEF is to first decompose each overlapped patch into three components: signal strength, signal structure and mean intensity. Then, it produces each fused patch by setting different rules on the three components, and the final result is acquired via aggregation.

However, SPD-MEF suffers from high computational cost, halo effect and detail loss. As such, we proposed the method, called multi-scale structure path decomposition (MSPD), in our preliminary work [4] to significantly alleviate the halo effect for MEF (we call this MEF solution as MSPD-MEF). With a fast approximation algorithm, we illustrated that MSPD-MEF can significantly improve the efficiency over the SPD-MEF method. Nevertheless, since the method MSPD-MEF adopted the multi-scale approach, the fused images can suffer from apparent detail loss, especially for the increased scale, which is a common issue for this type of methods [14].

To tackle this issue, one possible idea is to integrate the edge-preserving-based methods [15]–[17] into our preliminary solution [4], which can simultaneously preserve the detail information, suppress the halo effect and achieve the low computational cost. Based on this idea, we develop the method, called multi-scale edge-preserving structural patch decomposition (MESPD), for MEF, which we term this MEF solution as MESPD-MEF. Fig. 1 shows the results of different solutions, including SPD-MEF [3], MSPD-MEF [4], and our proposed method MESPD-MEF, in the image sequences “Tower” on a static scene. Observe that our method MESPD-MEF can retain more detail information in bright and dark regions than SPD-MEF and MSPD-MEF. Our main contributions are summarized as follows:

- This is the first work that incorporates the edge-preserving factors into the currently state-of-the-art method MSPD [4], which can simultaneously reduce the halo effect and preserve the details in the fused images.
- We further develop the novel and flexible bell curve function for accurately estimate the weight function for the mean intensity component, which can preserve the details in both bright and dark regions.

- Extensive experiments demonstrate that our solution can achieve significant improvement for visual quality in the fused images over different types of state-of-the-art methods in both static scene and dynamic scene settings for a wide range of datasets, without incurring high time overhead, compared with the fastest method MSPD-MEF [4].

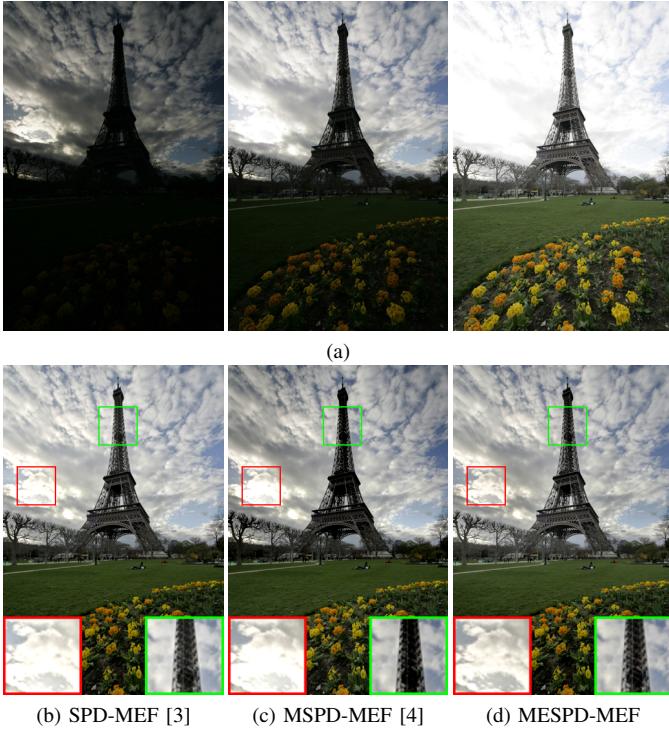


Fig. 1. Fused results by three methods based on SPD. (a) Image sequence “Tower”. (b) SPD-MEF [3]. (c) Fast multi-scale SPD-MEF [4]. (d) The proposed method.

The rest of the paper is organized as follows. Section II shows a brief overview of MEF algorithms in static and dynamic scenes. Section III illustrates the state-of-the-art approach for this task. Section IV presents our solution, which includes ESPD-MEF and its fast multi-scale version. Section V provides the experimental results and discussions for comparing different types of MEF methods and our proposed solution. Section VI concludes this paper.

II. RELATED WORK

Multi-exposure fusion (MEF) is a branch of image fusion [18]–[22], which aims to blend a set of multiple exposure images captured by bracketing into a single composite image. Various kinds of MEF methods, which focus on weight calculation, weight map smoothing, multi-scale implementation, and detail enhancement, have been proposed in the past decades, which can be divided into two main categories, namely static MEF methods and dynamic MEF methods.

A. Static MEF methods

In this section, we review three commonly used approaches for static MEF, which are multi-scale-fusion-framework-based

approach, optimization-based approach, and deep-learning-based approach.

Multi-scale-fusion-framework-based approach: Mertens *et al.* [1] proposed the pioneer multi-scale fusion framework, which first computes the (normalized) weight map using contrast, color saturation and well-exposedness, and then multiplies the Gaussian pyramid of this weight map with the Laplacian pyramid of the multi-exposure images in order to provide the fused image. Based on this framework, many research studies have been proposed to further improve the fusion performance. Shen *et al.* [23] adopted a sigmoid shape curve to enhance the texture details of fused images. Kou *et al.* [15], [16] replaced the Gaussian smoothing in the multi-scale fusion framework by the gradient domain edge-preserving smoothing to suppress the halo effect. Li *et al.* [24], [25] enhanced the details by incorporating the extracted gradient into the multi-scale fusion framework. Li *et al.* [26] utilized the guided filtering method [27] to construct the saliency-based weight map in order to further enhance the details. Even though the above methods [15], [16], [23]–[26] can either enhance the details or suppress the halo effect, they suffer from high computational cost. As such, many methods have been developed in the literature to further boost the efficiency by sacrificing the fusion quality. Ancuti *et al.* [14] only adopted the single-scale in this fusion framework, by filtering the weight map with a larger Gaussian kernel size and retaining the details with the second-order Laplacian filter. Wang *et al.* [28] further incorporated the edge details with this method [14] in the YUV color space to improve the fusion results without incurring high computational cost.

Optimization-based approach: Some optimization-based methods for MEF have been also developed in the literature. Gu *et al.* [29] obtained the fused gradient field based on Riemannian geometry metric, then attained the fused image via Poisson solver. Song *et al.* [30] approximated the ideal luminance image to max-contrast image with gradient constraint under a maximum a posteriori framework. Ma *et al.* [31] obtained the fusion results by globally optimizing the structural similarity index. However, all these methods suffer from obvious artifacts, e.g., ringing effect, detail loss and distortion, which provide inferior fusion results.

Deep-learning-based approach: Prabhakar *et al.* [32] developed the first end-to-end deep learning solution for MEF, in which they defined the ground truth, based on optimizing a non-reference image quality metric [33]. Xu *et al.* [34] adopted the generative adversarial network (GAN) to support MEF. However, instead of fusing multiple images, both methods [32], [34] can only fuse two images. To tackle this issue, Ma *et al.* [35] developed the MEF-NET, which can support fusing multiple images with different sizes. Zhang *et al.* [36] further proposed a general deep learning framework for all kinds of image fusion tasks. However, there are two main drawbacks for this approach. First, the lack of large amount of available data restricts the generalization ability of these deep-learning-based methods. Second, it is hard to define the meaningful ground truth for this task.

B. Dynamic MEF methods

Most of the static MEF methods assume that the objects are static, which may not be true in practice. As such, these methods can result in ghosting effect, once there is a dynamic object in the screen. To tackle this issue, many studies have been proposed in the literature, which can be classified into two domains, namely radiance domain and intensity domain.

Radiance domain: This type of methods [37]–[41] first transforms (using the camera response function) the multi-exposure images to the radiance domain, which assumes that the irradiance is linearly proportional to the exposure time of each image [42]. Based on this property, these methods can then detect and remove those inconsistent pixels (e.g., with motion) in the radiance domain and then perform image fusion. After that, they can utilize the tone mapping methods to obtain those fused images. However, these methods suffer from high computational costs, as evaluating the camera response functions [42]–[44] and the tone mapping methods [5], [45], [46] are time-consuming.

Intensity domain: Different types of feature extraction methods, e.g., gradient direction features [47], SIFT features [48], and CNN features [49], have been proposed to detect the motion, and then remove the inconsistent pixels, in the intensity domain. Moreover, some other motion detection methods [50]–[52] have been also adapted to support for this task. However, all these methods can either provide inaccurate fusion results [47]–[49] or suffer from high computational cost [50]–[52].

III. STATE-OF-THE-ART APPROACH: SPD-MEF

In this section, we first illustrate the state-of-the-art method [3], namely *structural-patch-decomposition-based multi-exposure fusion (SPD-MEF)*, in Section III-A. Then, we discuss the fast algorithm for SPD-MEF [4] in Section III-B. Lastly, we describe the multi-scale version for SPD-MEF [4] in Section III-C.

A. SPD-MEF

The core idea of SPD-MEF method [3] is to decompose the size $r \times r$ image patch \mathbf{x} , which is denoted as an $\mathbb{R}^{r^2 \times 1}$ vector, into three conceptually independent components, namely signal strength $c = \|\mathbf{x} - \mathbf{l}\|$, signal structure $\mathbf{s} = \frac{\mathbf{x} - \mathbf{l}}{\|\mathbf{x} - \mathbf{l}\|}$ and mean intensity \mathbf{l} , i.e.,

$$\begin{aligned} \mathbf{x} &= \|\mathbf{x} - \mathbf{l}\| \cdot \frac{\mathbf{x} - \mathbf{l}}{\|\mathbf{x} - \mathbf{l}\|} + \mathbf{l} \\ &= c \cdot \mathbf{s} + \mathbf{l} \end{aligned} \quad (1)$$

where $\mathbf{l} = [m_{\mathbf{x}}, m_{\mathbf{x}}, \dots, m_{\mathbf{x}}]^T \in \mathbb{R}^{r^2 \times 1}$ and $m_{\mathbf{x}}$ is the mean pixel value of this image patch \mathbf{x} .

Suppose that there are K multi-exposure image patches (with the same position) and \mathbf{x}_k denotes the image patch in the k^{th} image. Ma *et al.* [3] first decomposed \mathbf{x}_k into these three components c_k , \mathbf{s}_k and \mathbf{l}_k . Then, they estimated the signal strength \hat{c} , signal structure $\hat{\mathbf{s}}$, and mean intensity $\hat{\mathbf{l}}$ of the fused image patch $\hat{\mathbf{x}}$ by the following equations.

$$\hat{c} = \max_{1 \leq k \leq K} c_k \text{ and } \hat{\mathbf{s}} = \sum_{k=1}^K \alpha_k \mathbf{s}_k \text{ and } \hat{\mathbf{l}} = \sum_{k=1}^K \beta_k \mathbf{l}_k \quad (2)$$

where they utilized the norm and Gaussian functions to estimate the weightings α_k and β_k respectively. More details can be found in [3], [4].

Once they have obtained these three components \hat{c} , $\hat{\mathbf{s}}$ and $\hat{\mathbf{l}}$, they reconstructed the fused image patch $\hat{\mathbf{x}}$, based on Equation 1.

B. Fast SPD-MEF

The basic algorithm for SPD-MEF method [3] takes $O(WHKr^2)$ time, where W and H denote the width and height of each image. However, this method can be time-consuming. As such, Li *et al.* [4] further observed that each fused image patch $\hat{\mathbf{x}}$ is equivalent to:

$$\hat{\mathbf{x}} = \hat{c} \cdot \hat{\mathbf{s}} + \hat{\mathbf{l}} = \sum_{k=1}^K \gamma_k (\mathbf{x}_k - \mathbf{l}_k) + \beta_k \mathbf{l}_k \quad (3)$$

where: $\gamma_k = \frac{\alpha_k \hat{c}}{\|\mathbf{x}_k - \mathbf{l}_k\|}$. Based on this equation, they can further showed that the fused image $\hat{\mathbf{X}}$, with size $W \times H$, can be obtained using five mean filtering operations, where:

$$\hat{\mathbf{X}} = \sum_{k=1}^K \left(\mathbf{M}(\alpha_k \odot \mathbf{M}(\mathbf{X}_k)) + \mathbf{M}(\gamma_k) \odot \mathbf{X}_k - \mathbf{M}(\gamma_k \odot \mathbf{M}(\mathbf{X}_k)) \right) \quad (4)$$

where the operator \odot is the Hadamard product, i.e., element-wise multiplication, and \mathbf{M} denotes the $r \times r$ meaning filtering operation.

Since mean filtering operations can be efficiently implemented in linear time using box filter [27]. This method can significantly reduce the time complexity to $O(WHK)$ time.

C. Multi-scale SPD-MEF (MSPD-MEF)

The image patch size $r \times r$, a.k.a. kernel size, of the SPD-MEF method can strongly affect the fusion performance. Suppose that we choose the small kernel size (e.g., 3×3), the fused images suffer from severe spatial inconsistency. However, once we adopt the large kernel size, the fused images suffer from the significant detail loss. Even though we can achieve the balance between the spatial consistency and the detail preservation by choosing the medium kernel size, this approach can still cause the halo artifacts [27] near strong edges of object boundaries. To resolve this issue, Li *et al.* [4] further utilized the multi-scale framework (cf. Fig. 2) to support SPD-MEF for fusing K multi-exposure images $\mathbf{X}_1^{(1)}$, $\mathbf{X}_2^{(1)}, \dots, \mathbf{X}_K^{(1)}$.

The core idea of this multi-scale framework [4] is that all K multi-exposure images are first downsampled by a factor of 2 (with operator \mathbf{D}) into J scales (cf. black arrows in Fig. 2), i.e.,

$$\mathbf{X}_k^{(j)} = \mathbf{D}(\mathbf{M}(\mathbf{X}_k^{(j-1)})) \quad (5)$$

where they adopted $J = \lfloor \log_2 \min(H, W) \rfloor - 3$ in order to ensure the image size is not too small, i.e., at least 8×8 .

Then, they only fused the high frequency component in Equation 4 for each scale j (cf. red arrows in Fig. 2), i.e.,

$$\hat{\mathbf{H}}^{(j)} = \sum_{k=1}^K \left(\mathbf{M}(\gamma_k^{(j)}) \odot \mathbf{X}_k^{(j)} - \mathbf{M}(\gamma_k^{(j)} \odot \mathbf{X}_k^{(j)}) \right) \quad (6)$$

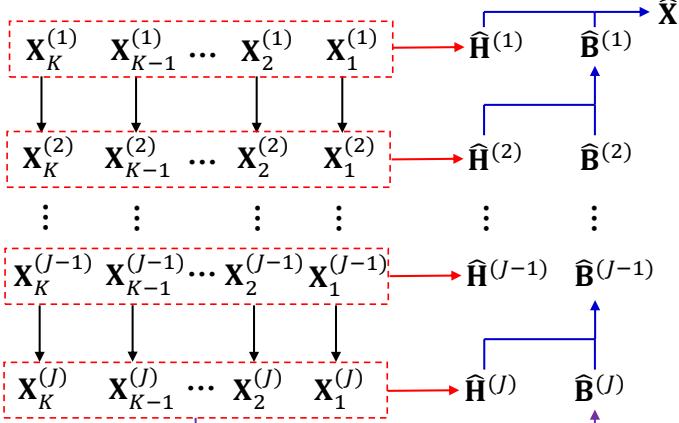


Fig. 2. Multi-scale framework for SPD-MEF.

After that, they computed the low frequency component in Equation 4 for the scale J (cf. purple arrow in Fig. 2), i.e.,

$$\hat{\mathbf{B}}^{(J)} = \sum_{k=1}^K \mathbf{M}(\alpha_k^{(J)} \odot \mathbf{M}(\mathbf{X}_k^{(J)})) \quad (7)$$

By iteratively adding the low and high frequency components and performing upsampling, termed as \mathbf{U} , and mean filtering operations (cf. blue arrows in Fig. 2), they reconstructed the low frequency components in different scales (cf. Equation 8), and lastly obtained the fused image $\hat{\mathbf{X}}$ (cf. Equation 9).

$$\hat{\mathbf{B}}^{(j)} = \mathbf{M}(\mathbf{U}(\hat{\mathbf{B}}^{(j+1)} + \hat{\mathbf{H}}^{(j+1)})) \quad (8)$$

$$\hat{\mathbf{X}} = \hat{\mathbf{B}}^{(1)} + \hat{\mathbf{H}}^{(1)} \quad (9)$$

IV. OUR SOLUTION

In this section, we first propose the new method, called *edge-preserving structural path decomposition (ESPD)* for MEF, i.e., ESPD-MEF (cf. Section IV-A). Then, we further extend this method to multi-scale version (cf. Section IV-B) to suppress the halo effect of the fused images.

A. Edge-preserving SPD-MEF (ESPD-MEF)

Even though the SPD-MEF method [3] and its variant [4] can achieve the state-of-the-art accuracy and efficiency performance for MEF, these methods still suffer from the detail loss (e.g., blurred edges in Fig. 1b and c). To tackle this issue, we aim to integrate the edge information into the mean intensity \mathbf{l} , instead of using the constant vector to represent this component. In particular, we define \mathbf{l} as:

$$\mathbf{l} = a\mathbf{x} + b\mathbf{1} \quad (10)$$

where $\mathbf{1}$ is the $r^2 \times 1$ -dimensional vector with value one in all entries.

Observe from Equation 10, the component $a\mathbf{x}$ incorporates the edge/contour information of each image patch. As such, this representation (cf. Equation 10) can capture more details for each multi-exposure image, compared with SPD, which only utilizes $b\mathbf{1}$ (with $b = m_{\mathbf{x}}$) as \mathbf{l} (cf. Equation 1). To determine the coefficients a and b for \mathbf{l} , we aim to minimize the difference between \mathbf{l} and \mathbf{x} , i.e., $\|\mathbf{l} - \mathbf{x}\|$, to retain more details

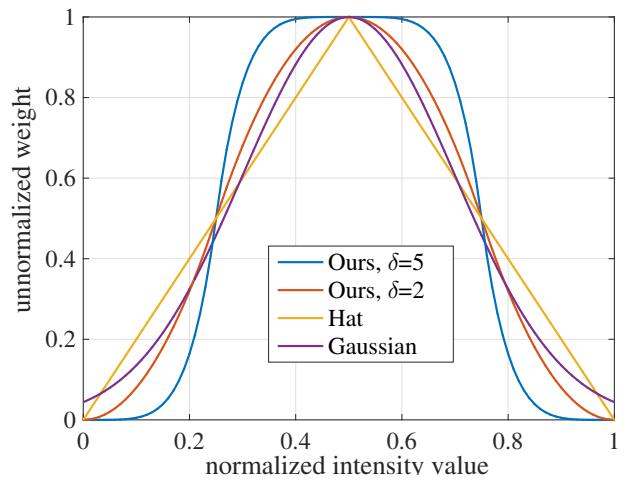


Fig. 3. The weighting method for mean intensity component via different bell curve functions.

of the image patch. The solution of this problem is trivial, by setting $a = 1$ and $b = 0$. However, this approach can retain all details (all high frequency components) for each image patch, which creates noise for the fused image. Therefore, we impose the regularization constant λ to avoid the large value for the parameter a in the optimization problem (cf. Equation 11).

$$\min_{a,b} \|a\mathbf{x} + b\mathbf{1} - \mathbf{x}\|^2 + \lambda a^2 \quad (11)$$

In this optimization problem, we can obtain the close-form solutions for both the parameters a and b (cf. Equation 12).

$$a = \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}}^2 + \lambda} \text{ and } b = \frac{\lambda}{\sigma_{\mathbf{x}}^2 + \lambda} m_{\mathbf{x}} \quad (12)$$

where $m_{\mathbf{x}}$ and $\sigma_{\mathbf{x}}^2$ are the mean and variance of the image patch \mathbf{x} respectively.

Recall that the signal strength $c = \|\mathbf{x} - \mathbf{l}\|$ and the signal structure $s = \frac{\mathbf{x} - \mathbf{l}}{\|\mathbf{x} - \mathbf{l}\|}$. After we substitute a and b to \mathbf{l} , and then substitute \mathbf{l} to c and s , we can obtain the new signal strength c , signal structure s and mean intensity \mathbf{l} as follows.

$$\mathbf{l} = \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}}^2 + \lambda} \mathbf{x} + \frac{\lambda}{\sigma_{\mathbf{x}}^2 + \lambda} m_{\mathbf{x}} \mathbf{1} \quad (13)$$

$$c = \frac{\lambda}{\sigma_{\mathbf{x}}^2 + \lambda} \|\mathbf{x} - m_{\mathbf{x}} \mathbf{1}\| \quad (14)$$

$$s = \frac{\mathbf{x} - m_{\mathbf{x}} \mathbf{1}}{\|\mathbf{x} - m_{\mathbf{x}} \mathbf{1}\|} \quad (15)$$

Observe from Equation 13, once the image patch \mathbf{x} represents the flat region, the variance of this patch $\sigma_{\mathbf{x}}^2$ is close to zero. As such, the mean intensity \mathbf{l} degenerates to $m_{\mathbf{x}} \mathbf{1}$, which is the same as SPD. However, suppose that this image patch contains many high frequency components, including edges and contours, the variance $\sigma_{\mathbf{x}}^2$ can be higher and the first term of \mathbf{l} (cf. Equation 13) can automatically capture more high frequency components, in order to improve the quality of the fused image patch.

To fuse K multi-exposure image patches, we can reuse the Equation 2 to fuse different components and then reuse the Equation 3 for obtaining the fused image patch $\hat{\mathbf{x}}$. However, with the integrated edge information for \mathbf{l} (cf. Equation 10),

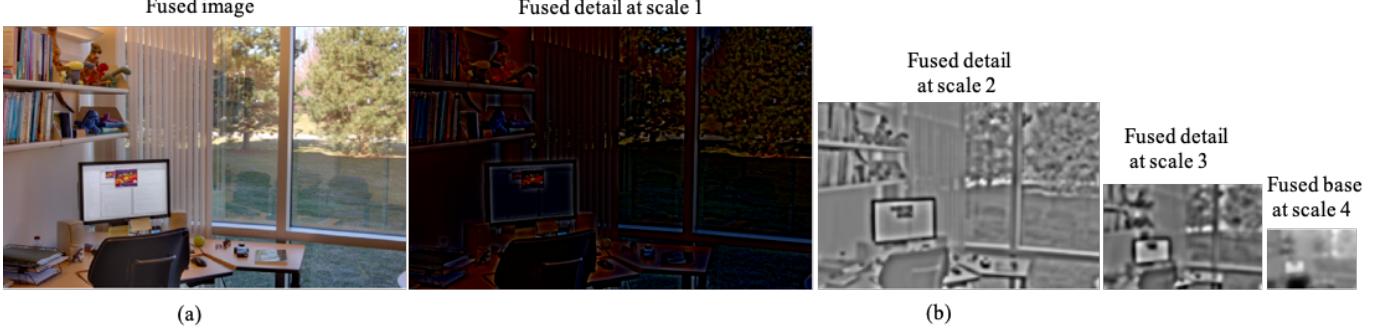


Fig. 4. Visual demonstration of the proposed MESPD approach with four scales on the image sequence “Office”. (a) Final fused image. (b) Fused base and detail layers at different scales.

we also need to modify the expressions of both γ_k and β_k in Equation 3. Here, we follow [4] and utilize the p -norm function to estimate the coefficients α_k . Based on these coefficients, we can obtain the new γ_k as:

$$\gamma_k = \frac{\lambda}{\sigma_{\mathbf{x}_k}^2 + \lambda} \max_{j=1}^K \{\|\mathbf{x}_j - \mathbf{l}_k\|\} \cdot \frac{\|\mathbf{x}_k - \mathbf{l}_k\|^{p-1}}{\sum_{j=1}^K \|\mathbf{x}_j - \mathbf{l}_k\|^p} \quad (16)$$

Since we have incorporated the edge information for \mathbf{l} , we also utilize the new estimation to determine the coefficients β_k in order to further improve the fusion quality. We propose the new bell curve function between the coefficient β_k and the mean intensity \mathbf{l}_k (cf. Equation 13) in Equation 17. Here, we let $C(\mathbf{l}_k)$ denote the mean intensity value in the center pixel of the image patch \mathbf{x}_k .

$$\beta_k = \begin{cases} C(\mathbf{l}_k)^\delta \cdot 0.25^{1-\delta} & \text{if } 0 \leq C(\mathbf{l}_k) \leq 0.25 \\ 0.5 - (0.5 - C(\mathbf{l}_k))^\delta \cdot 0.25^{1-\delta} & \text{if } 0.25 \leq C(\mathbf{l}_k) \leq 0.5 \\ 0.5 - (C(\mathbf{l}_k) - 0.5)^\delta \cdot 0.25^{1-\delta} & \text{if } 0.5 \leq C(\mathbf{l}_k) \leq 0.75 \\ (1 - C(\mathbf{l}_k))^\delta \cdot 0.25^{1-\delta} & \text{if } 0.75 \leq C(\mathbf{l}_k) \leq 1 \end{cases} \quad (17)$$

Fig. 3 illustrates different functions to estimate β_k , including typical Hat function, Gaussian function, and our method with different δ values. Observe that these curves can normally provide higher β_k , once $C(\mathbf{l}_k)$ is close to the middle region (in x-axis), e.g., 0.4 to 0.6, which indicates the good exposure of this image patch. Compared with the Gaussian function, our new bell curve function can be more flexible for controlling the size of this middle region, in order to determine how many details should be preserved in the image patch, by tuning the parameter δ . As an example, once we use $\delta = 5$, the middle region is larger than the one with $\delta = 2$. As a remark, we also need to normalize all β_k in order to ensure each pixel value in the fused image is within the domain $[0, 255]$ ¹.

After we have obtained both the parameters γ_k and β_k , we can also extend the fast implementation for SPD-MEF (cf. Section III-B) to ESPD-MEF method, which can produce the fused image $\widehat{\mathbf{X}}$ (cf. Equation 18).

$$\widehat{\mathbf{X}} = \sum_{k=1}^K \left(\mathbf{M}(\boldsymbol{\alpha}_k \odot \mathbf{F}(\mathbf{X}_k)) + \mathbf{M}(\boldsymbol{\gamma}_k) \odot \mathbf{X}_k - \mathbf{M}(\boldsymbol{\gamma}_k \odot \mathbf{F}(\mathbf{X}_k)) \right) \quad (18)$$

¹The normalization process is to use $\frac{\beta_k}{\sum_{k=1}^K \beta_k}$ as the new β_k .

where \mathbf{M} and \mathbf{F} denote the $r \times r$ mean filter and the edge-preserving filter respectively. Since the derivation of this equation is similar with Equation 4 and \mathbf{F} is the simple extension of the mean filter \mathbf{M} , we omit the details in this paper. Like [4], both \mathbf{M} and \mathbf{F} can be efficiently implemented using box filter [27]. Therefore, we can utilize $O(WHK)$ time to obtain this fused image $\widehat{\mathbf{X}}$.

As a remark, even though we only focus on the static scene fusion in this paper, our method can be easily extended to support dynamic scene fusion, by using the same method in [3], [4], i.e., utilizing the temporal similarity to filter the inconsistent pixels. For more details, please refer to Section III-E in [4].

B. Extension to Multi-scale ESPD-MEF (MESPD-MEF)

In order to suppress the halo effect of the fused images, we extend the multi-scale framework (cf. Fig. 2) to support our ESPD-MEF method. Since we have utilized the edge-preserving filter \mathbf{F} for each multi-exposure image \mathbf{X}_k (cf. Equation 18), instead of using the mean filter (cf. Equation 4), we need to replace the Equation 5 by Equation 19 in order to obtain the images with different scales.

$$\mathbf{X}_k^{(j)} = \mathbf{D}(\mathbf{F}(\mathbf{X}_k^{(j-1)})) \quad (19)$$

Based on Equation 18, we can also modify the expressions of both the high frequency component in each j^{th} scale (cf. Equation 6) and the low frequency component in the J^{th} scale (cf. Equation 7) by the Equations 20 and 21 respectively.

$$\widehat{\mathbf{H}}^{(j)} = \sum_{k=1}^K \left(\mathbf{M}(\boldsymbol{\gamma}_k^{(j)}) \odot \mathbf{X}_k^{(j)} - \mathbf{M}(\boldsymbol{\gamma}_k^{(j)} \odot \mathbf{F}(\mathbf{X}_k^{(j)})) \right) \quad (20)$$

$$\widehat{\mathbf{B}}^{(J)} = \sum_{k=1}^K \mathbf{M}(\boldsymbol{\alpha}_k^{(J)} \odot \mathbf{M}(\mathbf{F}(\mathbf{X}_k^{(J)}))) \quad (21)$$

By adopting the above modifications, the multi-scale framework can also support this method.

There is still a remaining question. Recall that our method utilizes the regularization constant λ (cf. Equation 11) to control the value of a in Equation 10. We notice that the higher the scale level j , the more the detail loss in the image $\mathbf{X}_k^{(j)}$. As such, we need to utilize the higher value of a , i.e.,

TABLE I
IMPACT OF $\lambda^{(1)}$ AND T ON AVERAGE MEF-SSIM VALUES OVER THE STATIC DATASETS FROM [3]

Value of	$\lambda^{(1)} = 0.25$ $T = 0.1$	$\lambda^{(1)} = 0.25$ $T = 0.2$	$\lambda^{(1)} = 0.25$ $T = 0.3$	$\lambda^{(1)} = 0.25$ $T = 0.4$	$\lambda^{(1)} = 0.25$ $T = 0.5$	$\lambda^{(1)} = 0.25$ $T = 0.6$	$\lambda^{(1)} = 0.25$ $T = 0.7$	$\lambda^{(1)} = 0.25$ $T = 0.8$	$\lambda^{(1)} = 0.25$ $T = 0.9$	$\lambda^{(1)} = 0.25$ $T = 1$
MEF-SSIM	0.910	0.940	0.959	0.971	0.978	0.982	0.983	0.984	0.985	0.985
Value of	$\lambda^{(1)} = 0.05$ $T = 0.5$	$\lambda^{(1)} = 0.10$ $T = 0.5$	$\lambda^{(1)} = 0.15$ $T = 0.5$	$\lambda^{(1)} = 0.20$ $T = 0.5$	$\lambda^{(1)} = 0.25$ $T = 0.5$	$\lambda^{(1)} = 0.30$ $T = 0.5$	$\lambda^{(1)} = 0.35$ $T = 0.5$	$\lambda^{(1)} = 0.40$ $T = 0.5$	$\lambda^{(1)} = 0.45$ $T = 0.5$	$\lambda^{(1)} = 0.50$ $T = 0.5$
MEF-SSIM	0.923	0.958	0.969	0.975	0.978	0.980	0.981	0.982	0.983	0.984



(a)

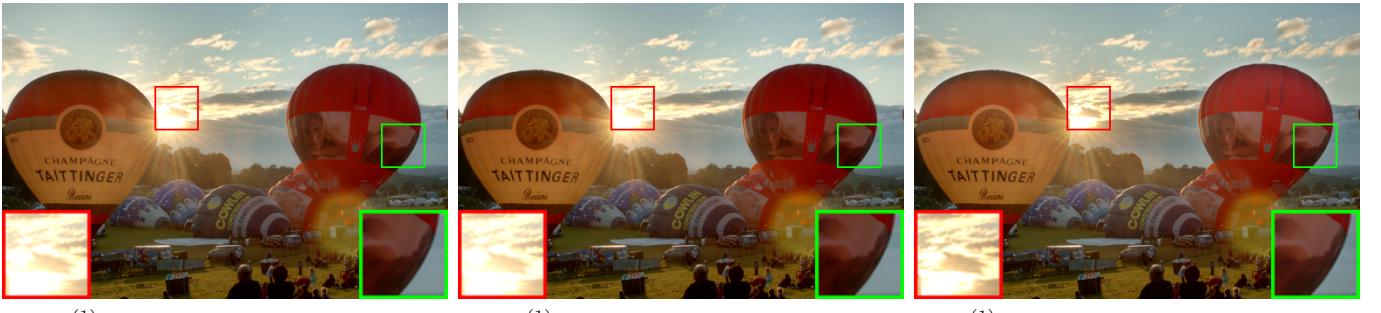
(b) $\lambda^{(1)} = 0.2, T = 0.5$, MEF-SSIM = 0.955(c) $\lambda^{(1)} = 0.2, T = 0.6$, MEF-SSIM = 0.963(d) $\lambda^{(1)} = 0.2, T = 0.8$, MEF-SSIM = 0.972(e) $\lambda^{(1)} = 0.2, T = 1$, MEF-SSIM = 0.974(f) $\lambda^{(1)} = 0.3, T = 1$, MEF-SSIM = 0.975(g) $\lambda^{(1)} = 0.25, T = 0.5$, MEF-SSIM = 0.959

Fig. 5. The results of MESPD by different $\lambda^{(1)}$ and T . (a) Image sequence “Balloons”. (b) Result by $\lambda^{(1)} = 0.2$ and $T = 0.5$. (c) Result by $\lambda^{(1)} = 0.2$ and $T = 0.6$. (d) Result by $\lambda^{(1)} = 0.2$ and $T = 0.8$. (e) Result by $\lambda^{(1)} = 0.2$ and $T = 1$. (f) Result by $\lambda^{(1)} = 0.3$ and $T = 1$. (g) Result by $\lambda^{(1)} = 0.25$ and $T = 0.5$.

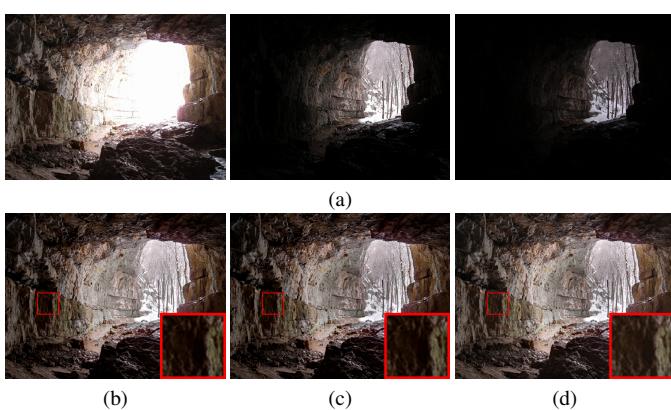


Fig. 6. The fused results of MESPD by different weight functions for mean intensities. (a) Image sequence “Cave”. (b) Result by Hat. (c) Result by Gaussian. (d) Result by Ours.

the smaller value of λ , in order to achieve edge preservation for the image with higher scale level j . Therefore, we define the regularization constant $\lambda^{(j)}$, where $1 < j \leq J$, using the Equation 22. Here, T is the constant such that $0 < T \leq 1$.

$$\lambda^{(j)} = \lambda^{(j-1)} \cdot T \quad (22)$$

Empirically, we find that $\lambda^{(1)} = 0.25$ and $T = 0.5$ can achieve the satisfactory performance. Fig. 4² shows the intermediate results of our method at four scales along with the final output.

V. EXPERIMENTAL EVALUATION

In our experiments, we first conduct ablation studies in our method MESPD-MEF (cf. Section V-A). Then, we provide

²As a remark, the high frequency component at scale 1, i.e., $\hat{\mathbf{H}}^{(1)}$, is computed based on stacking the RGB channels together, which contains both the finest details and the rich color information of the sequence. For other scales j , the high frequency components $\hat{\mathbf{H}}^{(j)}$ are computed in the grayscale domain. The similar idea has been also adopted in [3], [4].

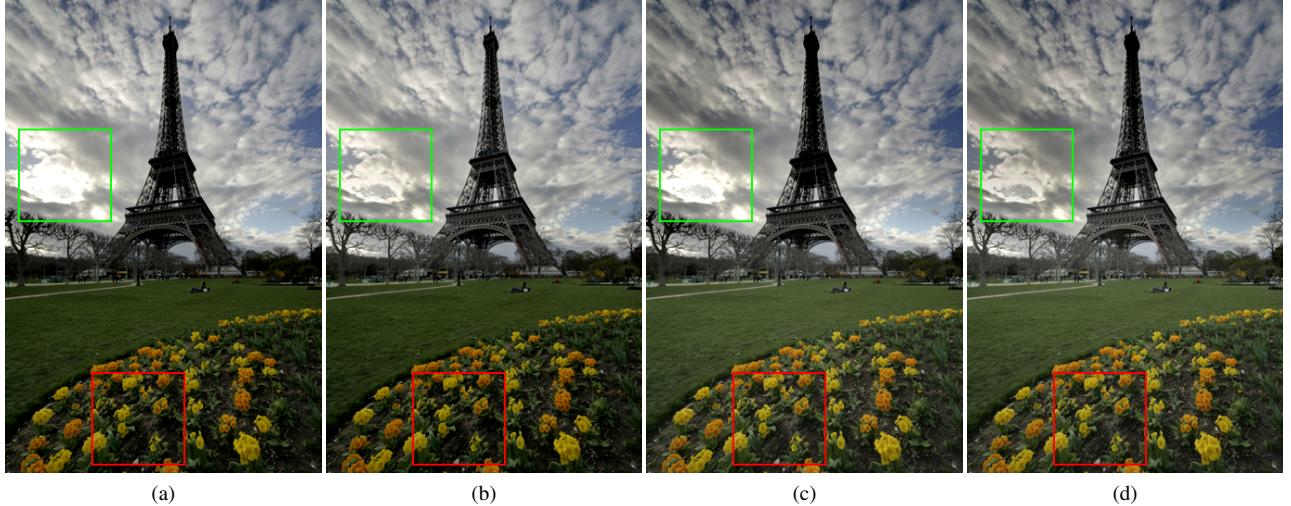


Fig. 7. The fused results of MESPD by different weight functions for mean intensities. (a) MSPD. (b) MESPD with the curve for β_k (Used in MSPD). (c) MSPD with our bell curve in β_k . (d) MESPD.

qualitative and quantitative results of our method against the state-of-the-art MEF methods in both static and dynamic scenes in Section V-B. Lastly, we compare the running time of our method with the state-of-the-art MEF methods in Section V-C. For the sake of simplicity, we abbreviate the methods SPD-MEF (cf. Section III-A), MSPD-MEF (cf. Section III-C) and MESPD-MEF (cf. Section IV-B) as SPD, MSPD and MESPD respectively. As a remark, most parameters of MESPD are inherited from the state-of-the-art method MSPD [4] (cf. Section III-C). Therefore, we adopt the default values in [4] for those parameters in our experiments, except that we change the kernel size from 8×8 in [4] to 9×9 (the smallest and middle scale levels) and 7×7 (the highest scale level), as we need to utilize the center pixel for the image patch in our method (cf. Equation 17).

A. Ablation Studies of Our Method MESPD

In this section, we first analyze the impact of edge preserving factors, including the regularization constant λ (cf. Equation 11) and the parameter T (cf. Equation 22), for the image fusion quality. Then, we investigate how the bell curves (cf. Fig. 3) affect the image fusion quality of our method. Lastly, we investigate how each component of our MESPD method improves the details for the fused image, compared with the state-of-the-art MSPD method [4] (cf. Section III-C).

Impact of edge-preserving factors: In the first experiment, we measure the average MEF-SSIM values of our method MESPD on 21 static scenes (used in [3]), by varying the parameters $\lambda^{(1)}$ and T . Table I shows the results for two sets of parameters (fixing $\lambda^{(1)} = 0.25$, varying T and fixing $T = 0.5$, varying $\lambda^{(1)}$). Observe that once we decrease the value $\lambda^{(1)}$, more high frequency components can be added for the fused image (based on Equations 10 and 11). As such, the MEF-SSIM value can decrease. On the other hand, the smaller the value T , the smaller the $\lambda^{(j)}$ value for each scale level j (cf. Equation 22), which also leads to the smaller MEF-SSIM value. However, even though the high $\lambda^{(1)}$ and T of MESPD method can lead to the high MEF-SSIM value, the

fused images can suffer from detail loss, as pointed out in the next experiment, which can also degrade the visual quality.

In the second experiment, we adopt the image sequence ‘‘Balloons’’ to test the fusion performance (cf. Fig. 5(a)). Observe that once we fix $\lambda^{(1)} = 0.2$ and vary the parameter T from 0.5 to 1 (cf. Fig. 5(b)-(e)), the visual quality of the images is lower (e.g., image patches (red boxes) in Fig. 5(e)). The main reason is that the higher the value T , the larger the regularization constant $\lambda^{(j)}$, where $1 < j \leq J$. However, the large regularization constant can result in smaller value of $a^{(j)}$ (cf. Equation 11), in which the fused image can suffer from detail loss (cf. Equation 10). We also vary different sets of $\lambda^{(1)}$ and T in this experiment (e.g., Fig. 5(f) and (g)). Observe that our method MESPD with $\lambda^{(1)} = 0.3$, $T = 1$ and $\lambda^{(1)} = 0.25$, $T = 0.5$ can achieve the highest MEF-SSIM value and the best visual quality respectively, which are adopted as the default values of our method in later experiments.

Impact of bell curve functions: Fig. 6 illustrates the fused results for the image sequence ‘‘Cave’’, using different bell curve functions for estimating the parameter β_k for each image batch (cf. Equation 2). In this experiment, we adopt $\delta = 5$ for our bell curve function (cf. Equation 17). Observe that both Hat and Gaussian functions suffer from visible detail loss in dark regions (i.e., red boxes in Fig. 6), compared with our bell curve function. The main reason is that our function can be more flexible to tune the shape (cf. Fig. 3) in order to provide less weight for β_k , once the image patch x_k is either under-exposed or over-exposed.

Compared with MSPD: MSPD (cf. Section III-C) can be regarded as the special case of our method MESPD, by setting $a = 0$ for l in Equation 10. On the other hand, MESPD further adopts the flexible bell curve function (cf. Equation 17) for estimating β_k , instead of using the modified arctan function in MSPD. Here, we ask a question, how can each component of our method MESPD improve the details for the fused image, compared with the MSPD method? Fig. 7 shows the results of different combinations for each component in the image sequence ‘‘Tower’’. Observe that once we use the modified



Fig. 8. Fused results by different methods in one static scene. (a) Image sequence “Arto”. (b) Mertens [1]. (c) GFF [26]. (d) GGIF [16]. (e) SPD [3]. (f) MSPD [4]. (g) MEF-NET [35]. (h) IFCNN [36]. (i) MEF-GAN [34]. (j) MESPD.

arctan curve, i.e., omit our bell curve function (cf. Fig. 7(a) and (b)), the fused images are dark as this modified arctan function curve can fuse more under-exposure images, which reduce the quality of the results (red boxes in Fig. 7). Suppose that we set $a = 0$ for \mathbf{l} (cf. Fig. 7(a) and (c)), the fused images suffer from detail loss (green boxes in Fig. 7). By incorporating both components, our method MESPD (cf. Fig. 7(d)) achieves the best fusion result.

B. Quality Comparisons with the State-of-the-art MEF Methods

We proceed to compare the fusion quality between our method and the state-of-the-art methods in both static and dynamic scenes.

Static scene comparisons: To verify the effectiveness of the proposed method, we compare our method with eight state-

of-the-art MEF methods, including Mertens [1], GGIF [16], GFF [26], MEF-NET [35], MEF-GAN [34], IFCNN [36], SPD [3], and MSPD [4], on 21 static scenes. Mertens [1] is the most representative baseline MEF method. GGIF is an improved method based on Mertens [1]. GFF [26] is a simple yet effective two-layer decomposition based method. Both MEF-NET [35], MEF-GAN [34] and IFCNN [36] are the recently proposed deep-learning-based methods. MEF-NET [35] is the first deep-learning-based approach, which extends Deep-Fuse [32], to support MEF with arbitrary numbers of multi-exposure images with unrestricted image resolution. MEF-GAN [34] uses GAN to achieve better detail and photorealistic performance. IFCNN [36] is a representative fusion framework under which multiple fusion issues can be handled. The fused results are produced by the available implementations on the

TABLE II

QUANTITATIVE COMPARISON OF OUR METHOD WITH EXISTING MEF ALGORITHMS USING MEF-SSIM [33] ON THE SEQUENCES IN [3]. THE SCORE RANGES FROM 0 TO 1 WITH A HIGHER VALUE INDICATING BETTER PERFORMANCE. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Image sequence	Mertens09 [1]	GFF [26]	SPD-MEF [3]	GGIF [16]	Ancuti17 [14]	MSPD [4]	MEF-NET [35]	IFCNN [36]	MEF-GAN [34]	MESPD $\lambda = 0.25$ $T = 0.5$	MESPD $\lambda = 0.3$ $T = 1$
	0.991	0.969	0.984	0.970	0.915	0.990	0.958	0.900	0.834	0.975	0.991
Arno	0.969	0.948	0.969	0.951	0.929	0.963	0.960	0.931	0.716	0.959	0.975
Balloons	0.971	0.964	0.973	0.968	0.938	0.977	0.962	0.942	0.896	0.969	0.975
Belgium house	0.975	0.978	0.985	0.979	0.958	0.984	0.958	0.955	0.804	0.984	0.983
Cave	0.989	0.984	0.991	0.983	0.974	0.994	0.981	0.948	0.635	0.987	0.995
Chinese garden	0.989	0.992	0.993	0.992	0.980	0.991	0.983	0.939	0.435	0.985	0.991
Church	0.981	0.985	0.984	0.982	0.976	0.986	0.977	0.948	0.715	0.970	0.980
Farmhouse	0.964	0.957	0.960	0.961	0.893	0.973	0.902	0.934	0.615	0.969	0.971
Lamp	0.969	0.929	0.956	0.945	0.877	0.967	0.938	0.915	0.648	0.968	0.970
Landscape	0.976	0.942	0.993	0.947	0.939	0.989	0.974	0.889	0.662	0.984	0.991
Laurenziana	0.988	0.987	0.987	0.985	0.957	0.989	0.977	0.956	0.791	0.980	0.993
Madison capitol	0.977	0.968	0.983	0.969	0.907	0.990	0.967	0.957	0.719	0.981	0.985
Mask	0.987	0.979	0.988	0.977	0.948	0.991	0.981	0.933	0.734	0.989	0.993
Office	0.985	0.967	0.990	0.984	0.957	0.989	0.984	0.913	0.747	0.986	0.991
Ostrow	0.974	0.967	0.978	0.977	0.925	0.979	0.964	0.916	0.750	0.965	0.983
Room	0.974	0.986	0.978	0.983	0.958	0.980	0.976	0.934	0.698	0.976	0.979
Set	0.986	0.960	0.988	0.966	0.905	0.992	0.950	0.914	0.820	0.984	0.990
Tower	0.986	0.986	0.986	0.986	0.962	0.988	0.980	0.927	0.697	0.990	0.988
Venice	0.966	0.954	0.984	0.952	0.932	0.984	0.953	0.924	0.650	0.980	0.985
Window	0.995	0.990	0.995	0.987	0.966	0.982	0.967	0.925	0.916	0.969	0.981
Yellow hall	0.980	0.970	0.982	0.972	0.940	0.985	0.966	0.927	0.845	0.991	0.994
Average										0.978	0.985

TABLE III

AVERAGE RUNNING TIME COMPARISON ON 12 DYNAMIC SCENES OF APPROXIMATELY THE SAME SIZE ($683 \times 1024 \times 3 \times 3$)

Alg	Sen12 [38]	Hu13 [53]	Lee14 [39]	Qin15 [50]	Oh15 [40]	SPD [3]	MSPD [4]	MESPD
Env	MATLAB+Mex	MATLAB+Mex	MATLAB+Mex	MATLAB+Mex	MATLAB	MATLAB	MATLAB	MATLAB
Time (s)	75.28 ± 20.48	114.96 ± 45.29	36.91 ± 11.55	465.06 ± 298.87	40.93 ± 9.93	57.48 ± 3.21	1.92 ± 0.20	2.12 ± 0.22

public with default optimized settings. The deep-learning-based methods are tested by GPU GeForce GTX 1080 with 16 GB memory with the released pre-trained models. The visual comparison results in the image sequence “Arto” can be found in Fig. 8.

In Fig. 8, the result by Mertens cannot effectively retain the details, as marked in red boxes. GFF recovers more details than Mertens, but information in the dark region is still not sufficient. GGIF is an improved method based on Mertens with edge-preserving factor and decreased decomposition level. However, the detail in the green box is still lost. MEF-GAN severely loses the details in the whole picture due to limited number of multi-exposure images. Even though IFCNN can present details in dark region, the result is over-enhanced with visible noise. Compared with SPD, MSPD loses more details in dark region. Both MEF-NET and MESPD can achieve similar visual performance, which can preserve the details on the building and bridge.

Moreover, we use MEF-SSIM [33] to quantitatively assess the quality of results generated by different MEF algorithms (cf. Table II). MEF-SSIM is specifically designed for multi-exposure fusion task, which has been the most widely used metric recently. A higher score between 0 and 1 indicates a better quality. Observe that our proposed method MESPD ($\lambda^{(1)} = 0.3$ and $T = 1$) can obtain the best performance with 10 highest scores out of 21 scenarios and highest average score. It should be noted that MEF-SSIM is not accurately consistent with human visual system. It prefers halo and detail loss to some extent. Similar observations and statements can be also found in recent studies [4], [31], [34]. Therefore, it may

not be important to blindly pursue a high MEF-SSIM value for developing a new MEF method. However, MEF-SSIM is still an important reference for objective evaluation of MEF task. A good MEF method should keep a good balance between dynamic range recovery and decent MEF-SSIM value.

Furthermore, we conduct the subjective experiment in an indoor environment for these nine MEF methods. In this experiment, we have found six volunteer subjects, including three males and three females, to provide the score from 1 (the worst) to 10 (the best) for each fused image. Based on their scores, we measure the mean opinion score (MOS) for each MEF method. Fig. 9 shows the MOS of all methods. Observe that our method MESPD can achieve the highest MOS, i.e., 7.33 ± 1.03 , which is higher than the second best method MEF-NET with MOS 7.17 ± 1.17 .

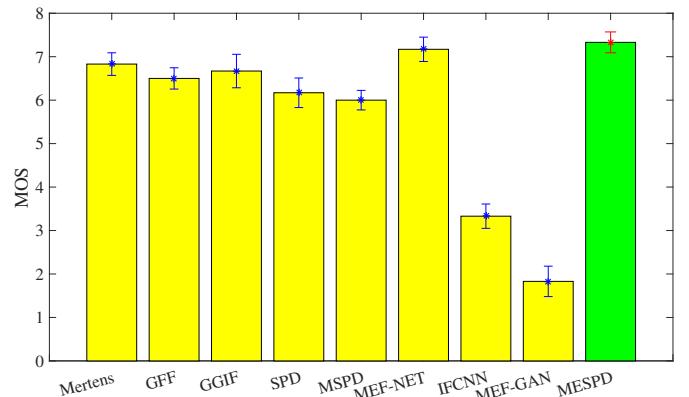


Fig. 9. MOS of all methods, using the 21 static scenes (from [3]) for testing.

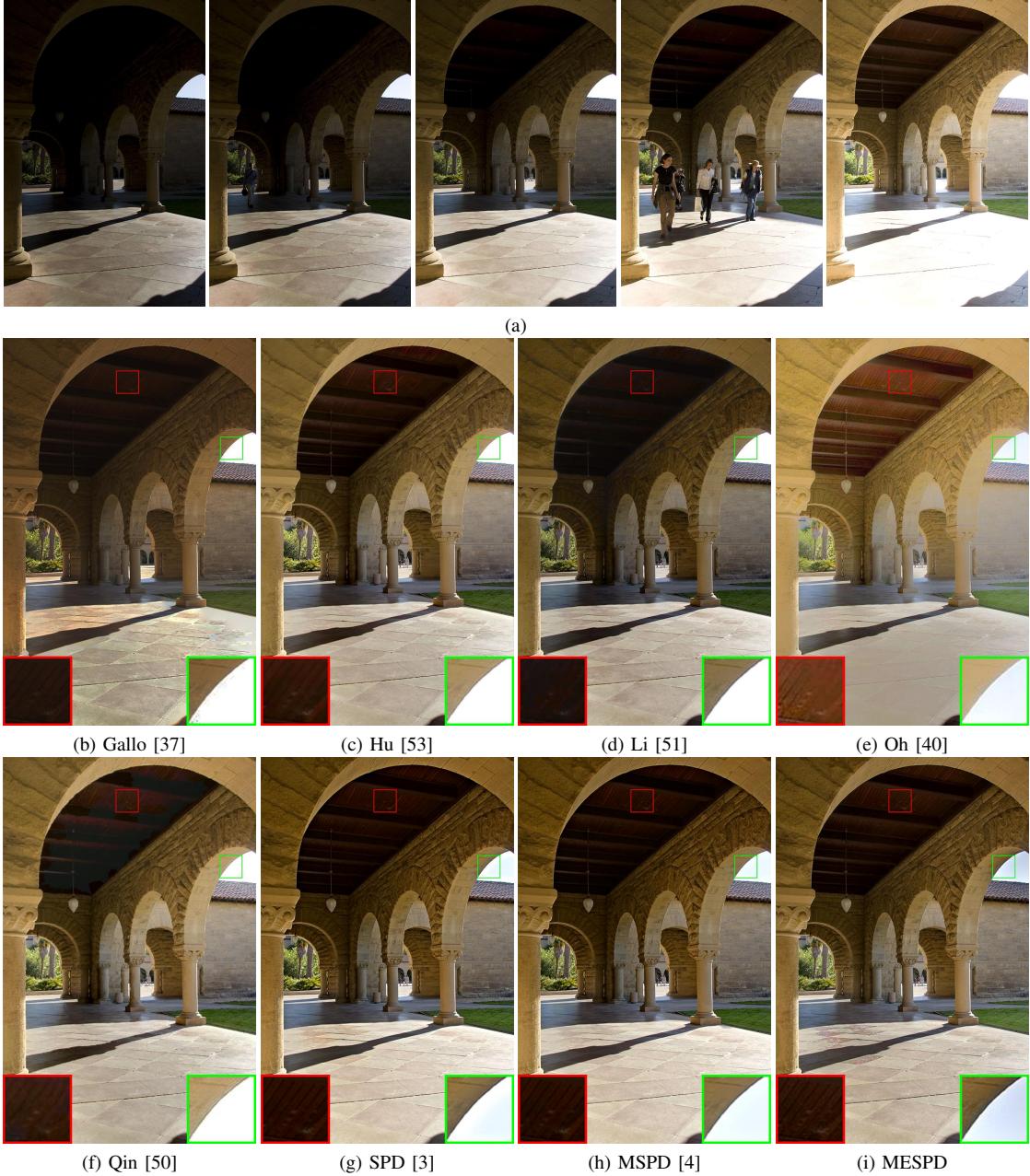


Fig. 10. Fused results by different methods in one dynamic scene. (a) Image sequence “ArchSequence”. (b) Gallo [37]. (c) Hu [53]. (d) Li [51]. (e) Oh [40]. (f) Qin [50]. (g) SPD [3]. (h) MSPD [4]. (i) MESPD.

Dynamic scene comparisons: We proceed to compare the performance between our method and the state-of-the-art methods, including Gallo [37], Hu [53], Li [51], Oh [40], Qin [50], SPD [3], and MSPD [4], under dynamic scene setting. Fig. 10 shows the results of all methods in the image sequence “ArchSequence”. Gallo fails to present the details in dark arch and bright sky. Hu can recover a few details in dark region, but the sky in the fused image is over-exposed. Li shares similar performance with Gallo. Oh can show the blue sky, but it suffers from evident artifacts on the arch. Qin suffers from the color distortions on the dark region. Even though SPD can recover the detail in both dark and bright regions, it suffers from halo effect, as shown in the green box. MSPD can suppress the halo effect. However, it loses

the details in the dark region. Our presented method MESPD can simultaneously recover the blue sky and visible details in the dark region.

C. Running Time Comparisons with the State-of-the-art MEF Methods

In this section, we compare the running time between our method and the state-of-the-art methods, under dynamic scene setting, using a computer with 4GHz CPU and 32GB RAM. We report the average running time of different methods on 12 natural scenes in Table III. In our preliminary work [4], we have already shown that MSPD can achieve the smallest time complexity, compared with other methods. Therefore, the running time of this method is the smallest (cf. Table III). Since our MESPD method shares the same theoretical computational

cost as MSPD, our method MESPD, which can significantly preserve the details (cf. Fig. 10), only incurs a slightly higher computational time, compared with the fastest method MSPD.

VI. CONCLUSION

In this paper, we proposed the multi-scale edge-preserving structural patch decomposition for supporting multi-exposure fusion, i.e., MESPD-MEF, which can preserve more details in the fused images, based on incorporating the edge-preserving factors. Compared with the state-of-the-art methods, SPD-MEF [3] and MSPD-MEF [4], our method can inherit all their advantages, e.g., suppress halo artifacts [4], avoid ghosting effect [3], [4], and achieve the fastest running time [4], and overcome their weakness, a.k.a. detail loss. As such, the extensive experiments validate that our proposed method MESPD-MEF can achieve the state-of-the-art visual quality for both static scene and dynamic scene settings, compared with different types of MEF methods. In addition, MESPD-MEF only incurs a slight time overhead (10-35% in Table III), compared with the fastest method [4].

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China project 61902251.

REFERENCES

- [1] T. Mertens, J. Kautz, and F. Van Reeth, “Exposure fusion: A simple and practical alternative to high dynamic range photography,” *Computer Graphics Forum*, vol. 28, no. 1, pp. 161–171, Mar. 2009.
- [2] K. Ma and Z. Wang, “Multi-exposure image fusion: A patch-wise approach,” in *IEEE International Conference on Image Processing*, 2015, pp. 1717–1721.
- [3] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, “Robust multi-exposure image fusion: A structural patch decomposition approach,” *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2519–2532, May 2017.
- [4] H. Li, K. Ma, H. Yong, and L. Zhang, “Fast multi-scale structural patch decomposition for multi-exposure image fusion,” *IEEE Transactions on Image Processing*, vol. 29, pp. 5805–5816, 2020.
- [5] H. Li, X. Jia, and L. Zhang, “Clustering based content and color adaptive tone mapping,” *Computer Vision and Image Understanding*, vol. 168, pp. 37–49, Mar. 2018.
- [6] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, “High dynamic range image compression by optimizing tone mapped image quality index,” *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3086–3097, Oct. 2015.
- [7] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, “A fusion-based enhancing method for weakly illuminated images,” *Signal Processing*, vol. 129, pp. 82–96, 2016.
- [8] C. Zheng, Z. Li, Y. Yang, and S. Wu, “Single image brightening via multi-scale exposure fusion with hybrid learning,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.
- [9] C. O. Ancuti and C. Ancuti, “Single image dehazing by multi-scale fusion,” *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3271–3282, 2013.
- [10] X. Wang, Z. Sun, Q. Zhang, Y. Fang, L. Ma, S. Wang, and S. Kwong, “Multi-exposure decomposition-fusion model for high dynamic range image saliency detection,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.
- [11] W. Zhang, S. Hu, and K. Liu, “Patch-based correlation for deghosting in exposure fusion,” *Information Sciences*, vol. 415, pp. 19–27, 2017.
- [12] Y. Yang, J. Wu, S. Huang, Y. Fang, P. Lin, and Y. Que, “Multimodal medical image fusion based on fuzzy discrimination with structural patch decomposition,” *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1647–1660, 2018.
- [13] H. Li, X. Qi, and W. Xie, “Fast infrared and visible image fusion with structural decomposition,” *Knowledge-Based Systems*, vol. 204, p. 106182, 2020.
- [14] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and A. C. Bovik, “Single-scale fusion: An effective approach to merging images,” *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 65–78, Jan. 2017.
- [15] F. Kou, Z. Li, C. Wen, and W. Chen, “Multi-scale exposure fusion via gradient domain guided image filtering,” in *IEEE International Conference on Multimedia and Expo*, 2017, pp. 1105–1110.
- [16] F. Kou, W. Chen, C. Wen, and Z. Li, “Edge-preserving smoothing pyramid based multi-scale exposure fusion,” *Journal of Visual Communication and Image Representation*, vol. 53, pp. 235–244, May 2018.
- [17] Y. Que, Y. Yang, and H. J. Lee, “Exposure measurement and fusion via adaptive multiscale edge-preserving smoothing,” *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 12, pp. 4663–4674, 2019.
- [18] H. Li, L. Li, and J. Zhang, “Multi-focus image fusion based on sparse feature matrix decomposition and morphological filtering,” *Optics Communications*, vol. 342, pp. 1–11, 2015.
- [19] M. Yin, X. Liu, Y. Liu, and X. Chen, “Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain,” *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 1, pp. 49–64, 2018.
- [20] W. Zhao, D. Wang, and H. Lu, “Multi-focus image fusion with a natural enhancement via a joint multi-level deeply supervised convolutional neural network,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 4, pp. 1102–1115, 2018.
- [21] S. Liu, J. Chen, and S. Rahardja, “A new multi-focus image fusion algorithm and its efficient implementation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 5, pp. 1374–1384, 2019.
- [22] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, “U2fusion: A unified unsupervised image fusion network,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [23] J. Shen, Y. Zhao, S. Yan, and X. Li, “Exposure fusion using boosting Laplacian pyramid,” *IEEE Transactions on Cybernetics*, vol. 44, no. 9, pp. 1579–1590, Sep. 2014.
- [24] Z. Li, J. Zheng, and S. Rahardja, “Detail-enhanced exposure fusion,” *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4672–4676, Nov. 2012.
- [25] Z. Li, Z. Wei, C. Wen, and J. Zheng, “Detail-enhanced multi-scale exposure fusion,” *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1243–1252, Mar. 2017.
- [26] S. Li, X. Kang, and J. Hu, “Image fusion with guided filtering,” *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.
- [27] K. He, J. Sun, and X. Tang, “Guided image filtering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [28] Q. Wang, W. Chen, X. Wu, and Z. Li, “Detail-enhanced multi-scale exposure fusion in yuv color space,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.
- [29] B. Gu, W. Li, J. Wong, M. Zhu, and M. Wang, “Gradient field multi-exposure images fusion for high dynamic range image visualization,” *Journal of Visual Communication and Image Representation*, vol. 23, no. 4, pp. 604–610, May 2012.
- [30] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, “Probabilistic exposure fusion,” *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 341–357, Jan. 2012.
- [31] K. Ma, Z. Duanmu, H. Yeganeh, and Z. Wang, “Multi-exposure image fusion by optimizing a structural similarity index,” *IEEE Transactions on Computational Imaging*, vol. 4, no. 1, pp. 60–72, Mar. 2018.
- [32] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, “DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs,” in *IEEE International Conference on Computer Vision*, 2017, pp. 4724–4732.
- [33] K. Ma, K. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
- [34] H. Xu, J. Ma, and X.-P. Zhang, “Mef-gan: Multi-exposure image fusion via generative adversarial networks,” *IEEE Transactions on Image Processing*, 2020.
- [35] K. Ma, Z. Duanmu, H. Zhu, Y. Fang, and Z. Wang, “Deep guided learning for fast multi-exposure image fusion,” *IEEE Transactions on Image Processing*, vol. 29, pp. 2808–2819, 2019.

- [36] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, “Ifcnn: A general image fusion framework based on convolutional neural network,” *Information Fusion*, vol. 54, pp. 99–118, 2020.
- [37] O. Gallo, N. Gelfandz, W.-C. Chen, M. Tico, and K. Pulli, “Artifact-free high dynamic range imaging,” in *IEEE International Conference on Computational Photography*, 2009, pp. 1–7.
- [38] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, “Robust patch-based HDR reconstruction of dynamic scenes,” *ACM Transactions on Graphics*, vol. 31, no. 6, pp. 203:1–203:11, Nov. 2012.
- [39] C. Lee, Y. Li, and V. Monga, “Ghost-free high dynamic range imaging via rank minimization,” *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1045–1049, Sep. 2014.
- [40] T.-H. Oh, J.-Y. Lee, Y.-W. Tai, and I. S. Kweon, “Robust high dynamic range imaging by rank minimization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1219–1232, Jun. 2015.
- [41] N. K. Kalantari and R. Ramamoorthi, “Deep high dynamic range imaging of dynamic scenes,” *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 144:1–144:12, Jul. 2017.
- [42] P. E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in *ACM SIGGRAPH 2008 classes*, 2008, pp. 1–10.
- [43] A. Chakrabarti, Y. Xiong, B. Sun, T. Darrell, D. Scharstein, T. Zickler, and K. Saenko, “Modeling radiometric uncertainty for vision with tone-mapped color images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2185–2198, 2014.
- [44] A. Badki, N. K. Kalantari, and P. Sen, “Robust radiometric calibration for dynamic scenes in the wild,” in *2015 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2015, pp. 1–10.
- [45] R. Fattal, D. Lischinski, and M. Werman, “Gradient domain high dynamic range compression,” *ACM transactions on graphics*, vol. 21, no. 3, pp. 249–256, Jul. 2002.
- [46] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, “Edge-preserving decompositions for multi-scale tone and detail manipulation,” *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 1–10, 2008.
- [47] W. Zhang and W.-K. Cham, “Gradient-directed multiexposure composition,” *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2318–2323, Apr. 2012.
- [48] Y. Liu and Z. Wang, “Dense SIFT for ghost-free multi-exposure fusion,” *Journal of Visual Communication and Image Representation*, vol. 31, pp. 208–224, Aug. 2015.
- [49] H. Li and L. Zhang, “Multi-exposure fusion with CNN features,” in *IEEE International Conference on Image Processing*, 2018, pp. 1723–1727.
- [50] X. Qin, J. Shen, X. Mao, X. Li, and Y. Jia, “Robust match fusion using optimization,” *IEEE Transactions on Cybernetics*, vol. 45, no. 8, pp. 1549–1560, Aug. 2015.
- [51] Z. Li, J. Zheng, Z. Zhu, and S. Wu, “Selectively detail-enhanced fusion of differently exposed images with moving objects,” *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4372–4382, Oct. 2014.
- [52] H. Zimmer, A. Bruhn, and J. Weickert, “Freehand HDR imaging of moving scenes with simultaneous resolution enhancement,” *Computer Graphics Forum*, vol. 30, no. 2, pp. 405–414, Apr. 2011.
- [53] J. Hu, O. Gallo, K. Pulli, and X. Sun, “HDR deghosting: How to deal with saturation?” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1163–1170.



Hui Li received his Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, in 2019. Currently, he is working as a Senior Engineer in Media Engineering Laboratory, Huawei 2012 Labs, Shenzhen, China. His research interests include image processing, computational photography, and computer vision.



Tsz Nam Chan (M’20) received the bachelor’s degree in electronic and information engineering and the PhD degree in computer science from the Hong Kong Polytechnic University in 2014 and 2019 respectively. He worked as the postdoctoral researcher in The University of Hong Kong from Sep 2018 to Aug 2020. He is currently a research assistant professor in the Hong Kong Baptist University. His research interests include multidimensional similarity search, pattern matching and kernel methods for machine learning. He is a member of IEEE.



Xianbiao Qi received the B.E. and Ph.D. degrees from the Beijing University of Posts and Telecommunications, Beijing, China, in 2008 and 2015, respectively. He visited the Web Search and Mining Group in Microsoft Research Asia from January 2011 to May 2012. He was a researcher with the Center of Machine Vision Group from January 2014 to December 2015, University of Oulu, Finland. He spent two years as a post-doctor in Hong Kong Polytechnic University from 2016 to 2018. Currently, he is with the Ping An Property and Casualty Insurance Company, Shenzhen, China. His research interests include computer vision, and deep learning.



Wuyuan Xie is currently an assistant professor in Shenzhen University, Shenzhen, China. Before that, she was a research associate in the Department of Computing, The Hong Kong Polytechnic University. She gained Ph.D degree in Mechanical and Automation Engineering from The Chinese University of Hong Kong in 2015. From 2008 to 2011, she was with Shenzhen Institutes of Advanced Technology, Chinese Academy of Science. Her research interests include shape from shading, motion learning, photometric stereo, and 3D surface details enhancement.