Fast Multi-Scale Structural Patch Decomposition for Multi-Exposure Image Fusion

Hui Li, Kede Ma, *Member, IEEE*, Hongwei Yong, and Lei Zhang, *Fellow, IEEE*

***Abstract*— Exposure bracketing is crucial to high dynamic range imaging, but it is prone to halos for static scenes and ghosting artifacts for dynamic scenes. The recently proposed structural patch decomposition for multi-exposure fusion (SPD-MEF) has achieved reliable performance in deghosting, but suffers from visible halo artifacts and is computationally expensive. In addition, its relationship to other MEF methods is unclear. We show that without explicitly performing structural patch decomposition, we arrive at an unnormalized version of SPD-MEF, which enjoys an order of 30× speed-up, and is closely related to pixel-level MEF methods as well as the standard two-layer decomposition method for MEF. Moreover, we develop a fast multi-scale SPD-MEF method, which can effectively reduce halo artifacts. Experimental results demonstrate the effectiveness of the proposed MEF method in terms of speed and quality.**

***Index Terms*— Multi-exposure fusion, high dynamic range imaging, computational photography.**

I. Introduction

FAITHFUL reproduction of natural scenes with high dynamic ranges (HDR) is a challenging task [1]. Due to the low dynamic range (LDR) of current sensors, under-/over-exposure occurs frequently in everyday photo-taking experiences, leading to unpleasing visual appearances. This issue has been addressed computationally by exposure bracketing, which captures and fuses several pictures of the same scene at different exposure levels. An HDR image can be reconstructed if we are able to invert the camera response function [2] and perform fusion in radiance domain (*i.e.*, HDR reconstruction). Tone mapping operators [3]–[7] are needed to render HDR images on a standard display with a limited dynamic range. Multi-exposure fusion (MEF) [8], [9] offers a simpler and more direct alternative by performing fusion in intensity domain, which has been widely employed in mobile devices for HDR imaging [10].

Manuscript received May 2, 2019; revised February 1, 2020 and March 12, 2020; accepted April 6, 2020. Date of publication April 16, 2020; date of current version April 27, 2020. This work was supported in part by NSFC, China, under Grant 61672446, and in part by the RGC RIF, Hong Kong, under Grant R5001-18. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jing-Ming Guo. *(Corresponding author: Lei Zhang.)*

Hui Li is with the Media Engineering Laboratory, Media Technology Institute (MTI), Huawei 2012 Labs, Shenzhen, China, and also with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong (e-mail: cshli@comp.polyu.edu.hk).

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

Hongwei Yong and Lei Zhang are with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong (e-mail: cshyong@comp.polyu.edu.hk; cslzhang@comp.polyu.edu.hk).

Digital Object Identifier 10.1109/TIP.2020.2987133

A common problem of MEF methods is the introduction of ghosting artifacts when dealing with dynamic scenes that contain moving objects (see Fig 1). While many MEF algorithms (also referred to as HDR deghosting methods) are able to produce ghost-free images, they come with their own disadvantages such as substantial computational complexity due to the need of solving a global optimization problem [11]–[13], or suboptimal visual quality due to excessive reliance on the reference exposure for inconsistent motion rejection [14]–[16]. Recently, Ma *et al.* [17] proposed the structural patch decomposition for MEF (SPD-MEF) that demonstrates reliable deghosting performance over a wide range of dynamic scenes. The visual quality improvements of SPD-MEF have been verified by MEF-SSIM [18], [19], a widely used objective quality metric for MEF, and in two independent subjective experiments [20], [21]. Although faster than many HDR deghosting algorithms, SPD-MEF still takes seconds (even minutes) to fuse high-resolution sequences, and therefore is not suitable for real-time mobile applications. Meanwhile, it is likely to generate visible halo artifacts for some natural scenes, where the dynamic range differences between the foreground and the background are large (see Fig. 1).

In this paper, we make an in-depth analysis of SPD-MEF to gain a better understanding of its behavior. As a patch based fusion method, SPD-MEF represents an image patch by its mean intensity, signal strength and signal structure. The desired patch is obtained by fusing the three components separately in SPD-MEF. Our empirical analysis shows that the normalization step can be skipped when fusing signal structures without introducing noticeable differences to the original scheme. This allows us to perform structural patch decomposition implicitly, leading to around 30 times speed-up. The proposed fast SPD-MEF scheme is also closely related to pixel-level MEF methods and the standard two-layer decomposition method for MEF. Moreover, we propose a multi-scale approach by recursively downsampling and processing the mean intensity images, which effectively reduces the halo artifacts. Experiments on a variety of static and dynamic scenes show that our fast multi-scale SPD-MEF algorithm consistently produces visually appealing results, while being one of the fastest among the state-of-the-arts.

II. Related Work

In this section, we provide an overview of existing MEF algorithms with emphasis on how different methods compute perceptual weights for fusion, and how they design exposure-invariant features for motion estimation.

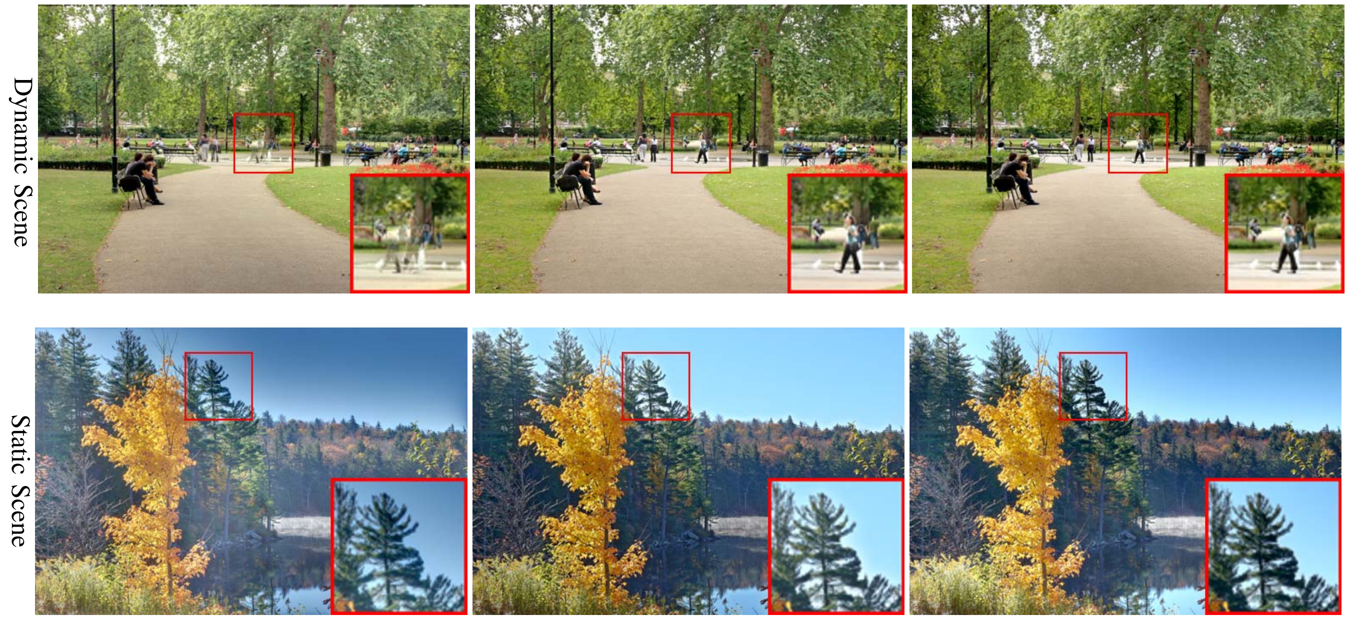


Fig. 1. Examples of MEF results by Mertens09 [8] (left column), SPD-MEF [17] (middle column) and our method (right column) on a dynamic scene (top row) and a static scene (bottom row). One can see that our method successfully suppresses ghost and halo artifacts compared to Mertens09 and SPD-MEF.

A. MEF Methods for Static Scenes

Static MEF methods mainly consist of weight map computation [8] and weighted fusion, followed by post-processing such as detail-enhancement [22]. The weight map smoothing occurs explicitly in pixel-level fusion to keep spatial consistency. Patch-level fusion smooths the weight map implicitly via aggregating overlapping patches [23], [24]. Multi-scale decomposition is widely used in MEF for halo reduction [8], [25]. Post-processing is often adopted to further improve visual quality of fused images.

Specifically, Mertens *et al.* [8] computed the weight maps using contrast, color saturation, and well-exposedness measurements. The fusion is accomplished in a multi-scale framework, where the input images are decomposed into a Laplacian pyramid and the weight maps are smoothed within a Gaussian pyramid. While computationally efficient, this method suffers from possible detail loss. Li *et al.* enhanced the details of Mertens' results by solving a quadratic optimization problem [9], [22]. Shen *et al.* performed MEF in a boosting Laplacian pyramid [26]. Kou *et al.* [25] replaced Gaussian smoothing in [8] with gradient domain guided smoothing to further reduce halos. Ancuti *et al.* [27] provided a fast single-scale approximation to [8] by applying Gaussian filtering to the weight maps and adding back the details extracted using a second-order Laplacian filter.

Li *et al.* [28] decomposed the input sequence into a base layer and a detail layer, and the weight maps were computed by saliency measurements and refined by guided filters [29] with different parameters. Raman and Chaudhuri [30] directly treated the detail layer as the weight map, which results in somewhat dreary appearance. Goshtasby [23] designed the weight maps based on the maximum entropy principle, and smoothed them with a monotonic blending function to reduce blocking artifacts.

Optimization-based methods have also been used in MEF. Ma *et al.* [31] employed a gradient descent-based method to optimize MEF-SSIM [18] in the image space. Despite visual quality improvements, their algorithm is prohibitively slow. Later, feed-forward convolutional networks were trained for MEF [32], [33] by optimizing MEF-SSIM. Cai *et al.* [20] made use of thirteen existing MEF methods to generate a set of fused candidate images, and manually picked the best ones as the ground truths to train a convolutional network for single image contrast enhancement. Since this process requires extensive human intervention, the resulting number of sequences for training is quite limited, which may hinder the generalization ability of the learned network.

B. MEF Methods for Dynamic Scenes

MEF methods that are motion-unaware will inevitably suffer from ghosting artifacts when there is camera or object motion in the scene. Camera motion is relatively mild in practical applications, and it can be easily addressed by means of a tripod or image registration techniques [34]. Most MEF methods that handle object motion rely on specification of a reference image such that inconsistent motion from other exposures can be detected and discarded. A straightforward approach to motion detection is to work in radiance domain and exploit the linear relationship between radiance value and exposure time. The challenge here is how to accurately recover the camera response function using a limited number of exposures. Gallo *et al.* [35] used a simple threshold to reject inconsistent pixels. Sen *et al.* [14] integrated image alignment and reconstruction into joint optimization. Kalantari *et al.* [36] developed a convolutional network for HDR imaging guided by optical flow, which was improved by Wu *et al.* [37], towards end-to-end training and inference. In intensity domain, various exposure-invariant features have been explored for robust motion detection, such as image gradient [38], SIFT [39], deep representation [40], Shannon entropy [41], optical flow [42], normalized structure [16], [17], and intensity mapping [15]. It is worth noting that motion detection in MEF requires dense correspondence across exposures, and typically suffers from heavy computational burdens.

Methods without reliance on the reference image assume that the background dominates the scene; moving objects are modeled as outliers and should be removed. Li and Kang [43] recovered the background by applying median filtering to the histogram-equalized images. Pece and Kautz [44] made use of median threshold bitmap for moving object detection. Lee *et al.* [12] and Oh *et al.* [13] formulated image alignment and motion detection as a rank minimization problem [45]. However, the assumption that background dominates does not hold for the majority of realistic natural scenes, therefore limiting the capability of these methods in HDR deghosting.

III. Fast Multi-Scale SPD-MEF

In this section, we first revisit the algorithm design of SPD-MEF [17], and show that an unnormalized approximation permits a neat acceleration scheme, whose relationship to other MEF approaches is also much clearer. We then develop a fast multi-scale SPD-MEF approach with halo reduction.

A. Background on SPD-MEF

We briefly describe how SPD-MEF [17] computes the fused image. The core idea of SPD-MEF for static scenes is to decompose an image patch $\mathbf{x} \in \mathbb{R}^N$ into three conceptually independent components: mean intensity, signal strength, and signal structure:

$$\begin{aligned} \mathbf{x} &= l \cdot \mathbf{1} + \|\mathbf{x} - l\| \cdot \frac{\mathbf{x} - l}{\|\mathbf{x} - l\|} \\ &= l \cdot \mathbf{1} + \|\tilde{\mathbf{x}}\| \cdot \frac{\tilde{\mathbf{x}}}{\|\tilde{\mathbf{x}}\|} \\ &= l \cdot \mathbf{1} + c \cdot \mathbf{s}, \end{aligned} \tag{1}$$

where $\mathbf{1}$ is an N-dimensional vector of all ones, $\tilde{\mathbf{x}}$ is the mean-removed patch of $\mathbf{x}$, and $\|\cdot\|$ denotes the ℓ_2-norm. l and $c = \|\tilde{\mathbf{x}}\|$ are two scalars, representing the mean intensity and the signal strength of $\mathbf{x}$, respectively. $\mathbf{s} = \frac{\tilde{\mathbf{x}}}{\|\tilde{\mathbf{x}}\|}$ is a unit vector, whose direction encodes signal structure. The desired patch of the output image can be obtained by fusing the three components l, c and $\mathbf{s}$ separately and then inverting the decomposition. Specifically, assuming the input sequence has K exposures, the desired local mean intensity is computed by

$$\hat{l} = \sum_{k=1}^{K} \alpha_k l_k, \tag{2}$$

where the weights $\alpha_k \geq 0$ for $k = 1, \dots K$ measure the well-exposedness, and $\sum_k \alpha_k = 1$. In SPD-MEF, a two-dimensional Gaussian profile is adopted to design α_k:

$$\alpha_k = \exp\left(-\frac{(\mu_k - 0.5)^2}{2\sigma_g^2} - \frac{(l_k - 0.5)^2}{2\sigma_l^2} \right), \tag{3}$$

where μ_k is the global mean intensity of the k-th image $\mathbf{X}_k$. σ_g and σ_l are Gaussian spreads. The desired local signal strength is computed as the largest one across exposures:

$$\hat{c} = \max_{1 \leq k \leq K} \|\tilde{\mathbf{x}}_k\| = \max_{1 \leq k \leq K} c_k. \tag{4}$$

The desired local signal structure is determined by

$$\hat{\mathbf{s}} = \frac{\bar{\mathbf{s}}}{\|\bar{\mathbf{s}}\|}, \quad \text{and} \quad \bar{\mathbf{s}} = \sum_{k=1}^{K} \beta_k \mathbf{s}_k, \tag{5}$$

where the weights $\beta_k \geq 0$ for $k = 1, \dots, K$, and $\sum_k \beta_k = 1$. In SPD-MEF, β_k is given by

$$\beta_k = \|\tilde{\mathbf{x}}_k\|^p, \tag{6}$$

which is proportional to the signal strength $\|\tilde{\mathbf{x}}_k\|$. $p \geq 0$ is an exponent parameter. After this, we are able to compute the desired local patch as

$$\hat{\mathbf{x}} = \hat{l} \cdot \mathbf{1} + \hat{c} \cdot \hat{\mathbf{s}}. \tag{7}$$

SPD-MEF performs patch aggregation by simply averaging all overlapping pixel values to obtain the final fused image [17].

B. Fast SPD-MEF

Much of computation in SPD-MEF comes from the structural patch decomposition in Eq. (1), whose complexity is $\mathcal{O}(NMK)$, where N is the patch size, M is the number of pixels in each exposure, and K is the number of exposures. In this subsection, we show that this complexity can be reduced to $\mathcal{O}(MK)$.

We first analyze $\bar{\mathbf{s}}$ (refer to Eq. (5)), which is a convex combination of K unit vectors. The norm of $\bar{\mathbf{s}}$ satisfies

$$\|\bar{\mathbf{s}}\| = \left\| \sum_{k=1}^{K} \beta_k \mathbf{s}_k \right\| \leq \sum_{k=1}^{K} \beta_k \|\mathbf{s}_k\| = 1, \tag{8}$$

which can be easily proved by induction using the triangle inequality and the absolute homogeneity of norm. The equality holds for arbitrarily chosen $\{\beta_k\}$, when all signal structures are identical. If some $\mathbf{s}_k$ points to a different direction, we may still achieve the equality by assigning the corresponding β_k to zero. Empirically, we find that $\|\bar{\mathbf{s}}\|$ computed by SPD-MEF is close to one (see the histogram in Fig. 2). This is expected because as long as the set of $\{\mathbf{x}_k\}$ are not under-/over-exposed, the corresponding exposure-invariant $\{\mathbf{s}_k\}$ have very similar structures, leading to $\|\bar{\mathbf{s}}\| \approx 1$. For under-exposed regions, $\mathbf{s}_k$ mainly contains amplified noise structure; for over-exposed regions, $\mathbf{s}_k$ is nearly flat, *i.e.*, $\frac{1}{\sqrt{N}}\mathbf{1}$. In either case, $\mathbf{s}_k$ may point to a different direction from the true signal structure. Fortunately, the corresponding β_k of those regions computed by SPD-MEF will be close to zero, giving rise to $\|\bar{\mathbf{s}}\| \approx 1$. This implies that whether the signal structure is normalized or not has little impact on the final fusion performance. In Fig. 3, we show the fused images by SPD-MEF with and without normalization. We can see that the results are very similar, as evidenced by a high SSIM score [46] of 0.999 between them. Similar observations can be found on other sequences in [17].

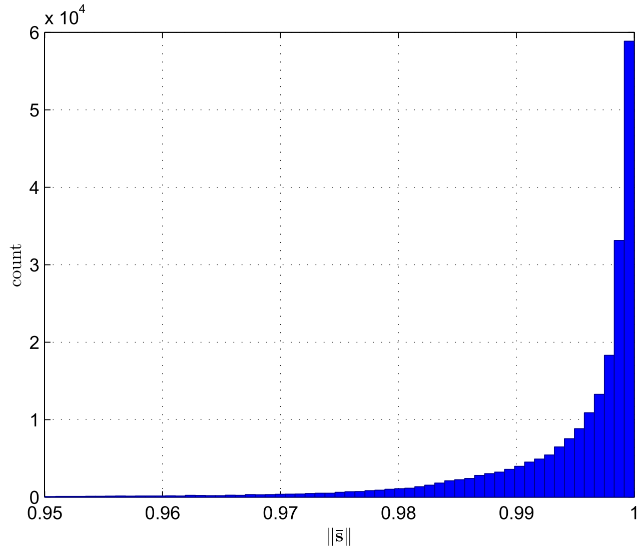
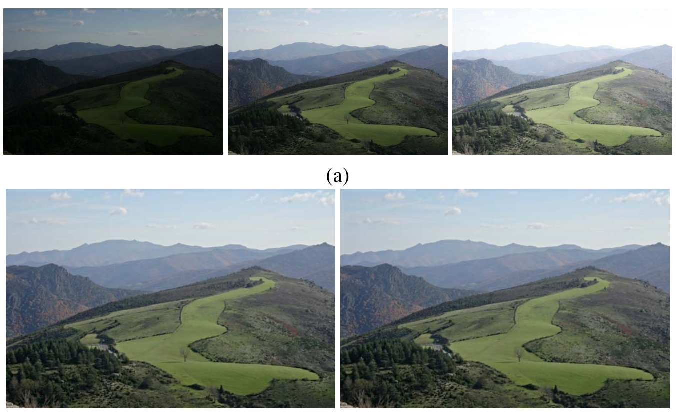


Fig. 2. The histogram of $\|\bar{\mathbf{s}}\|$ computed from six pre-registered static scenes from the image dataset in [17].



(a)

(b) MEF-SSIM = 0.991 (c) MEF-SSIM = 0.991

Fig. 3. SPD-MEF with and without normalization. (a) Image sequence "Landscape" (courtesy of HDRsoft). (b) Result with normalization. (c) Result without normalization. The visual similarity between the two fused images is verified by a high SSIM [46] value of 0.999. Similar observations can be found on other sequences in [17].

We proceed by substituting $\hat{l}$ in Eq. (2) and $\bar{\mathbf{s}}$ in Eq. (5) into Eq. (7), where we have approximated $\hat{\mathbf{s}}$ by $\bar{\mathbf{s}}$. Hence,

$$\hat{\mathbf{x}} \approx \sum_{k=1}^{K} \left(\alpha_k l_k \cdot \mathbf{1} + \hat{c}\beta_k \cdot \mathbf{s}_k\right) \tag{9}$$

$$= \sum_{k=1}^{K} \left(\alpha_k l_k \cdot \mathbf{1} + \frac{\hat{c}\beta_k}{\|\tilde{\mathbf{x}}_k\|} \cdot \tilde{\mathbf{x}}_k\right)$$

$$= \sum_{k=1}^{K} \left(\alpha_k l_k \cdot \mathbf{1} + \gamma_k \cdot (\mathbf{x}_k - l_k)\right), \tag{10}$$

where $\gamma_k = \frac{\hat{c}\beta_k}{\|\tilde{\mathbf{x}}_k\|}$. By incorporating $\|\tilde{\mathbf{x}}_k\|$ into γ_k, we are able to perform structural patch decomposition implicitly, and

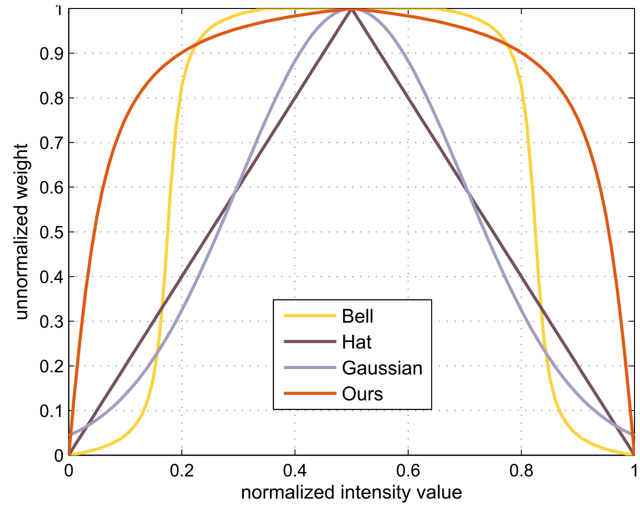


Fig. 4. Comparison of different weight functions based on well-exposedness.

compute the final image $\hat{\mathbf{X}}$ by

$$\hat{\mathbf{X}} = \sum_{k=1}^{K} \Big(\mathbf{L}\big(\boldsymbol{\alpha}_k \odot \mathbf{L}(\mathbf{X}_k)\big) + \mathbf{L}(\boldsymbol{\gamma}_k) \odot \mathbf{X}_k - \mathbf{L}\big(\boldsymbol{\gamma}_k \odot \mathbf{L}(\mathbf{X}_k)\big)\Big), \tag{11}$$

where $\mathbf{L}(\cdot)$ denotes the mean filter with kernel size N, and $\odot$ denotes the Hadamard product. That is, we apply $\mathbf{L}(\cdot)$ to $\mathbf{X}_k$ for computing the mean intensity map $\mathbf{l}_k$, and apply the same filter to the weight maps (*i.e.*, $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_k$) as an equivalent operation of averaging all overlapping pixel values during patch aggregation. The mean filtering process can be implemented in linear time via box filter [29]. As a result, the computational complexity of SPD-MEF is reduced from $\mathcal{O}(NMK)$ to $\mathcal{O}(MK)$, independent of patch size N.

We now take a closer look at Eq. (11). Choosing $\boldsymbol{\alpha}_k = \boldsymbol{\gamma}_k$ yields the classic form of pixel-level MEF with a smoothed weight map $\mathbf{L}(\boldsymbol{\gamma}_k)$. If each pixel computes a separate mean intensity from the patch centered at it, Eq. (11) becomes

$$\hat{\mathbf{X}} = \sum_{k=1}^{K} \Big(\mathbf{L}(\boldsymbol{\alpha}_k) \odot \mathbf{L}\left(\mathbf{X}_k\right) + \mathbf{L}(\boldsymbol{\gamma}_k) \odot \big(\mathbf{X}_k - \mathbf{L}\left(\mathbf{X}_k\right)\big)\Big), \tag{12}$$

which is essentially the two-layer decomposition framework for MEF. The weight maps for the base layer and the detail layer are $\mathbf{L}(\boldsymbol{\alpha}_k)$ and $\mathbf{L}(\boldsymbol{\gamma}_k)$, respectively. In the original development of SPD-MEF [17], the authors speed up the algorithm by sampling patches with a stride larger than one. This can also be incorporated into Eq. (11):

$$\hat{\mathbf{X}} = \sum_{k=1}^{K} \Big(\mathbf{L}\big(\mathbf{M}_k \odot \boldsymbol{\alpha}_k \odot \mathbf{L}(\mathbf{X}_k)\big) + \mathbf{L}(\mathbf{M}_k \odot \boldsymbol{\gamma}_k) \odot \mathbf{X}_k - \mathbf{L}\big(\mathbf{M}_k \odot \boldsymbol{\gamma}_k \odot \mathbf{L}(\mathbf{X}_k)\big)\Big), \tag{13}$$

where $\mathbf{M}_k$ is a binary mask with ones indicating patches that have been sampled.

(a) (b)

Fig. 5. Visual demonstration of the proposed multi-scale SPD-MEF approach on the image sequence "Arno" (courtesy of Bartlomiej Okonek). (a) Desired base and detail layers at four scales. (b) Final fused image.

C. Fast Multi-Scale SPD-MEF

The kernel size of the mean filter $\mathbf{L}$, or equivalently the patch size, has a significant impact on the final fusion performance. A small-size kernel generally recovers more details, but tends to produce noisy weight maps, causing spatial inconsistency of the fused image. A large-size kernel would resolve this problem at the cost of fine detail loss. A medium-size kernel keeps a good balance between spatial consistency and detail preservation, but may encourage halo artifacts near strong edges due to unwanted smoothing [29].

Here we present a fast multi-scale SPD-MEF approach to reduce halos, while preserving the details at different scales. We index the original sequence as scale 1. In Eq. (11), we notice that a desired detail layer that contains rich high-frequency information is computed as

$$\hat{\mathbf{H}}^{(1)} = \sum_{k=1}^{K} \Big(\mathbf{L}(\boldsymbol{\gamma}_k^{(1)}) \odot \mathbf{X}_k^{(1)} - \mathbf{L}\big(\boldsymbol{\gamma}_k^{(1)} \odot \mathbf{L}(\mathbf{X}_k^{(1)})\big) \Big). \quad (14)$$

To make SPD-MEF multi-scale, we do not fuse $\mathbf{L}(\mathbf{X}_k^{(1)})$ directly, but downsample it to obtain

$$\mathbf{X}_k^{(2)} = \mathbf{D}\big(\mathbf{L}(\mathbf{X}_k^{(1)})\big), \quad (15)$$

where $\mathbf{D}(\cdot)$ denotes downsampling by a factor of two. Then, the desired detail layer at scale 2 is computed by

$$\hat{\mathbf{H}}^{(2)} = \sum_{k=1}^{K} \Big(\mathbf{L}(\boldsymbol{\gamma}_k^{(2)}) \odot \mathbf{X}_k^{(2)} - \mathbf{L}\big(\boldsymbol{\gamma}_k^{(2)} \odot \mathbf{L}(\mathbf{X}_k^{(2)})\big) \Big). \quad (16)$$

The above process is applied recursively to $\mathbf{X}_k^{(j)}$ until the coarsest scale

$$J = \lfloor \log_2 \min(H, W) \rfloor - 3 \quad (17)$$

is reached, where H and W represent the height and width of the sequence, respectively. The constant three is subtracted to ensure that the short side of the sequence at scale J has a minimum of eight pixels, as a way of preserving low-frequency intensity information. Finally, we compute the desired base layer as follows:

$$\hat{\mathbf{B}}^{(J)} = \sum_{k=1}^{K} \mathbf{L}\big(\boldsymbol{\alpha}_k^{(J)} \odot \mathbf{L}(\mathbf{X}_k^{(J)})\big). \quad (18)$$

The fused image is obtained by progressively upsampling and adding the detail layers back to intermediate base layers:

$$\hat{\mathbf{B}}^{(j)}\mathbf{L}\big(\mathbf{U}(\hat{\mathbf{B}}^{(j+1)} + \hat{\mathbf{H}}^{(j+1)})\big), \quad j = 1, \ldots, J-1 \quad (19)$$

and

$$\hat{\mathbf{X}} = \hat{\mathbf{B}}^{(1)} + \hat{\mathbf{H}}^{(1)}, \quad (20)$$

where $\mathbf{U}(\cdot)$ denotes upsampling by a factor of two.

D. Weight Calculation

We compute $\boldsymbol{\alpha}_k^{(J)}$ according to the well-exposedness of $\mathbf{X}_k^{(J)}$ at the coarsest scale. Instead of adopting Gaussian curves [8], [17], we use a modified arctan function [6]:

$$\boldsymbol{\alpha}_k^{(J)} = \frac{\arctan\left(0.5\lambda - \left|0.5 - \mathbf{X}_k^{(J)}\right| \lambda\right)}{\sum_{k=1}^{K} \arctan\left(0.5\lambda - \left|0.5 - \mathbf{X}_k^{(J)}\right| \lambda\right)}, \quad (21)$$

where λ is a fixed parameter. We compare four weight functions, including bell-shaped, hat-shaped, Gaussian and the proposed curves in Fig. 4. It can be observed that our measure gives less penalty to slightly under-/over-exposed intensities. This is helpful for better preserving global brightness. The weight map $\boldsymbol{\gamma}_k^{(j)}$ for $\mathbf{X}_k^{(j)}$ is the same as the original SPD-MEF [17], except that it is computed at scale j. Fig. 5 shows the intermediate results of the proposed multi-scale SPD-MEF along with the final output.

E. Handling Dynamic Scenes

When dealing with dynamic scenes that contain noticeable object motion, most MEF algorithms rely on a pre-selected exposure as reference to detect inconsistent motion. Following SPD-MEF, we select the one that has the least number of under- and over-exposed patches [17]. Given a reference patch $\mathbf{s}_r$ and a co-located patch $\mathbf{s}_k$, we make the structural consistency measurement as follows:

$$\rho_k = \mathbf{s}_r^T \mathbf{s}_k \approx \frac{\tilde{\mathbf{x}}_r^T \tilde{\mathbf{x}}_k + \epsilon}{\|\tilde{\mathbf{x}}_r\| \, \|\tilde{\mathbf{x}}_k\| + \epsilon}, \quad (22)$$

where ϵ is a small positive constant to ensure the robustness to sensor noise. We also use the box filter [29] to implement Eq. (22), which has a complexity of $\mathcal{O}(MK)$. With

(a)

(b) Mertens09 [8], MEF-SSIM = 0.989 (c) Shen11 [47], MEF-SSIM = 0.964 (d) Gu12 [48], MEF-SSIM = 0.927

(e) Li13 [28], MEF-SSIM = 0.984 (f) Shen14 [26], MEF-SSIM = 0.767 (g) SPD-MEF [17], MEF-SSIM = 0.991

(h) Nejati17 [49], MEF-SSIM = 0.991 (i) Ancuti17 [27], MEF-SSIM = 0.974 (j) Ours, MEF-SSIM = 0.994

Fig. 6. Visual results and the MEF-SSIM scores of different static MEF algorithms on the image sequence "Chinese garden" (courtesy of Bartlomiej Okonek).

a pre-defined threshold T, a total of K binary maps can be created to identify static and dynamic regions

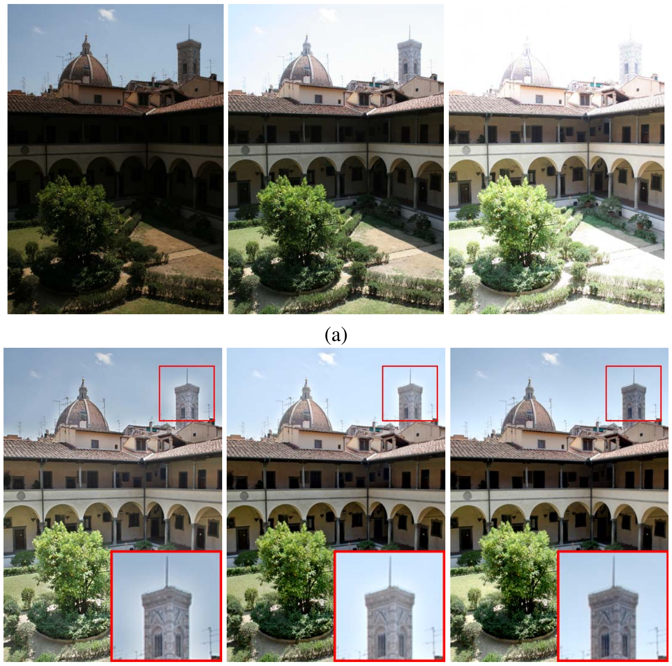
$$\kappa_k(i) = \begin{cases} 1 & \text{if } \rho_k(i) \geq T \\ 0 & \text{if } \rho_k(i) < T, \end{cases} \tag{23}$$

where i denotes the spatial index. κ_k is further refined with the help of the intensity mapping function (IMF) [17]. Although it is straightforward to make and combine structural consistency measurements at multiple scales. For simplicity, in this paper, we perform motion detection at the original scale only, and apply the corresponding binary maps for multi-scale fusion. We summarize the proposed fast multi-scale SPD-MEF method in Algorithm 1.

IV. Experiments

In this section, we first present the implementation details of the proposed fast multi-scale SPD-MEF approach. Then we provide qualitative and quantitative results of our method against the state-of-the-art MEF methods. Last, we conduct theoretical and empirical computational complexity analysis.

Our method does not introduce any new parameter; all are inherited from previous publications [6], [17], [24], including the patch size (*i.e.*, the kernel size of the mean filter)



(a)

(b) MEF-SSIM $= 0.957$ (c) MEF-SSIM $= 0.987$ (d) MEF-SSIM $= 0.989$

Fig. 7. Emergence of halo artifacts. (a) Image sequence "Laurenziana" (courtesy of Bartlomiej Okonek). (b) Ancuti17 [27]. (c) SPD-MEF [17]. (d) Ours.

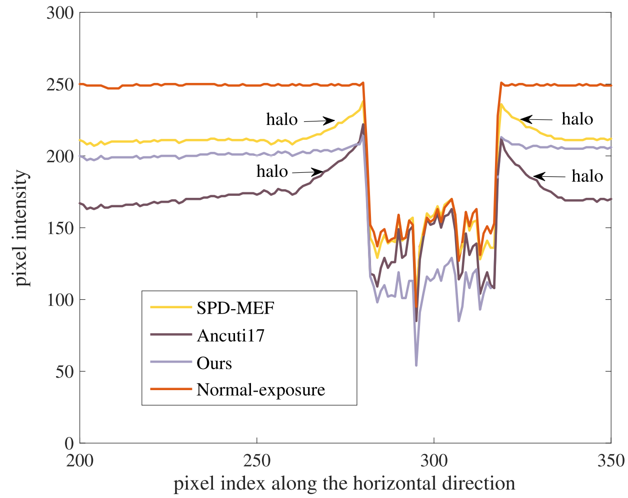


Fig. 8. Pixel intensity analysis of the zoom-in patches in Fig. 7 along the horizontal direction. The patch from the normal-exposure image is used as reference. The halos generated by SPD-MEF [17] and Ancuti17 [27] can be clearly seen due to unwanted smoothing near the boundaries. Our method closely approximates the boundaries of the reference patch as expected.

Algorithm 1 Fast multi-scale SPD-MEF

Input: Aligned image sequence $\{\mathbf{X}_k\}$
Output: Fused image $\hat{\mathbf{X}}$

```
Select a reference image, detect inconsistent motion via structural consistency check, and compensate moving regions using IMF
for scale j ∈ [1, J − 1] do
    Compute the detail layer Ĥ^(j) (refer to Eq. (14))
    Downsample X_k^(j) (refer to Eq. (15))
end for
Compute the detail layer Ĥ^(J)
Generate the base layer B̂^(J) using Eq. (18)
Reconstruct the fused image X̂ with Eq. (19) and Eq. (20)
```

$N = 8 \times 8 \times 3$ from [24] for the finest scale and $N = 8 \times 8$ for other scales, $\lambda = 20$ that determines the arctan curve from [6], and $\epsilon = 0.03^2$ in Eq. (22) and $T = 0.8$ in Eq. (23) from [17].

A. Static Scene Comparison

We compare our method with nine MEF algorithms on 21 static scenes, including Mertens09 [8], Shen11 [47], Gu12 [48], Li13 [28], Shen14 [26], SPD-MEF [17], Nejati17 [49], GGIF [25], and Ancuti17 [27]. All fused images are either from the original authors or generated by the publicly available implementations with default settings.

Fig. 6 visually compares our method with existing MEF algorithms on the image sequence "Chinese garden." Although built upon Mertens09 [8], Shen14 [26] generates an unnatural appearance with color and structure distortions since it nonlinearly enhances the detail layer by a simple sigmoid function. Relying on gradients, Gu12 [48] makes little use of color information, and over-shoots the details by solving the Poisson equation in gradient domain [4]. The color appearance produced by Shen11 [47] is slightly better, but the overall contrast is somewhat reduced. In addition, ringing artifacts appear near strong edges because of excessive nonlinear manipulation of subbands. The above three methods equate detail enhancement with visual quality improvement, which is not always true, especially in the case of over-enhancement. Li13 [28], SPD-MEF [17], Nejati17 [49], and Ancuti17 [27] exhibit different degrees of halo artifacts in the sky regions (zoom in for improved visibility). Compared to Li13, Nejati17 reduces the halos by replacing Gaussian filtering with guided filtering [29] in the two-layer decomposition. Mertens09 [8] and our method produce similar results on this sequence with little artifacts.

To better understand the emergence of halo artifacts in MEF, we show another visual example in Fig. 7, where we compare our method with SPD-MEF [17] and Ancuti17 [27] on the image sequence "Laurenziana." The boundaries (*e.g.*, zoom-in patches) between the foreground and the background with large dynamic range differences are the main sources of halo artifacts. To faithfully reproduce fine details across exposures, single-scale methods such as SPD-MEF and Ancuti17 often choose medium-size kernels, which may lead to unwanted smoothing along strong edges, resulting in visually unpleasant "halos" (see Fig. 8). The proposed multi-scale SPD-MEF approach resolves this issue by diluting the halos more globally, making the appearance more natural (see also Fig. 10).

We objectively evaluate the quality of fused images generated by different MEF algorithms using MEF-SSIM [18], which has been verified by comparing to human data [50] and through perceptual optimization [31]. MEF-SSIM [18] summarizes local structure preservation and global luminance consistency into an overall score between 0 and 1, with a higher value indicating better perceptual quality. From Table I,

TABLE I

QUANTITATIVE COMPARISON OF OUR METHOD WITH EXISTING MEF ALGORITHMS USING MEF-SSIM [18] ON THE SEQUENCES IN [17]. THE SCORE RANGES FROM 0 TO 1 WITH A HIGHER VALUE INDICATING BETTER PERFORMANCE. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Image sequence	Mertens09 [8]	Shen11 [47]	Gu12 [48]	Li13 [28]	Shen14 [26]	SPD-MEF [17]	Nejati17 [49]	GGIF [25]	Ancuti17 [27]	Ours
Arno	**0.991**	0.955	0.890	0.969	0.846	0.984	0.985	0.970	0.915	0.990
Balloons	0.969	0.940	0.913	0.948	0.776	0.969	**0.971**	0.951	0.929	0.963
Belgium house	0.971	0.935	0.896	0.964	0.709	0.973	0.972	0.968	0.938	**0.977**
Cave	0.975	0.946	0.934	0.978	0.788	**0.985**	0.979	0.979	0.958	0.984
Chinese garden	0.989	0.964	0.927	0.984	0.767	0.991	0.991	0.983	0.974	**0.994**
Church	0.989	0.959	0.866	0.992	0.878	**0.993**	0.991	0.992	0.980	0.991
Farmhouse	0.981	0.966	0.932	0.985	0.944	0.984	0.983	0.982	0.976	**0.986**
House	0.964	0.925	0.876	0.957	0.396	0.960	0.949	0.961	0.893	**0.973**
Lamp	**0.969**	0.917	0.875	0.929	0.539	0.956	0.960	0.945	0.877	0.967
Landscape	0.976	0.955	0.941	0.942	0.880	**0.993**	0.992	0.947	0.939	0.989
Laurenziana	0.988	0.956	0.873	0.987	0.881	0.987	0.986	0.985	0.957	**0.989**
Madison capitol	0.977	0.940	0.864	0.968	0.542	0.983	0.978	0.969	0.907	**0.990**
Mask	0.987	0.964	0.879	0.979	0.827	0.988	0.988	0.977	0.948	**0.991**
Office	0.985	0.958	0.900	0.967	0.756	**0.990**	0.988	0.984	0.957	0.989
Ostrow	0.974	0.950	0.877	0.967	0.786	0.978	0.978	0.977	0.925	**0.979**
Room	0.974	0.945	0.853	0.986	0.729	0.978	0.976	**0.983**	0.958	0.980
Set	0.986	0.974	0.911	0.960	0.873	0.988	0.988	0.966	0.905	**0.992**
Tower	0.986	0.946	0.932	0.986	0.779	0.986	0.986	0.986	0.962	**0.988**
Venice	0.966	0.930	0.889	0.954	0.765	**0.984**	0.976	0.952	0.932	**0.984**
Window	**0.982**	0.959	0.876	0.971	0.879	**0.982**	0.981	0.972	0.936	**0.982**
Yellow hall	0.995	0.983	0.869	0.990	0.866	0.995	0.996	0.987	0.966	**0.997**
Average	0.980	0.951	0.894	0.970	0.772	0.982	0.981	0.972	0.940	**0.985**

TABLE II

COMPUTATIONAL COMPLEXITY COMPARISON OF OUR METHOD AGAINST STATE-OF-THE-ART DEGHOSTING SCHEMES

Algorithm	Complexity
Sen12 [14]	$\mathcal{O}(I_i NMK^2)$
Hu13 [11]	$\mathcal{O}(I_i N(M \log M)K)$
Lee14 [12]	$\mathcal{O}(I_o I_i MK^2)$
Li14 [15]	$\mathcal{O}(MK)$
Qin15 [16]	$\mathcal{O}(I_i NM^2 K)$
Oh15 [13]	$\mathcal{O}(I_o I_i MK^2)$
SPD-MEF [17]	$\mathcal{O}(NMK)$
Ours	$\mathcal{O}(MK)$

we observe that our method achieves the best performance on average. Specifically, it outperforms the competing algorithms on 13 out of 21 natural scenes. It should be noted that though MEF-SSIM has been a de facto measure for MEF outputs, it is not good at capturing halos and may even prefer such artifacts [31]. Therefore, the perceptual gains of our method may not be fully reflected in terms of MEF-SSIM.

B. Dynamic Scene Comparison

We compare our method with eight state-of-the-art HDR deghosting algorithms that cover a wide range of design philosophies, including low rank-based methods Lee14 [12] and Oh15 [13], energy-based methods Sen12 [14], Hu13 [11] and Qin15 [16], and feature-based methods Li14 [15], Liu15 and Wang [39] and SPD-MEF [17]. For HDR reconstruction algorithms (*i.e.*, fusion in radiance domain), the Debevec and Malik's method [51] is used to estimate the camera response function. In order to generate LDR images for visual comparison, Lee14 makes use of the MATLAB function tonemap(), and Sen12 and Hu13 fuse aligned sequences using Photomatix [52] and Mertens09, respectively.

Fig. 9 shows the fusion results on the image sequence "Girl." Sen12 [14] produces an over-enhanced image that looks unnatural. This is largely attributed to the aggressive settings of Photomatix [52] to enhance HDR details. In general, it is delicate for HDR reconstruction algorithms to select proper tone mapping operators for dynamic range compression. Lee14 [12] and Oh15 [13] suffer from ghosting artifacts, which is expected because small motion does not satisfy the low rank assumption. In addition, solving such an optimization problem with a limited number of exposures is relatively unstable, and may result in other forms of distortions. Liu15 and Wang [39] relies on dense SIFT features, which may not be robust to exposure change, making deghosting unsuccessful. Some halos around the girl's legs are visible in the fused image generated by SPD-MEF [17]. Hu13 [11] and Qin15 [16] may generate shifted colors and deformed structures due to inaccurate patch matching during energy minimization. The results produced by Li14 [15] and our method are visually similar on this sequence.

C. Computational Complexity Comparison

We conduct a concise computational complexity analysis of HDR deghosting schemes in terms of the number of floating-point operations. We refer the interested readers to [12], [17] for a more detailed complexity analysis. Assume the input sequence has K exposures, each of which contains M pixels ($K \ll M$); for patch-wise methods, the patch size

(a)

(b) Sen12 [14] (c) Hu13 [11] (d) Lee14 [12]

(e) Li14 [15] (f) Liu15 [39] (g) Qin15 [16]

(h) Oh15 [13] (i) SPD-MEF [17] (j) Ours

Fig. 9. Visual comparison of our method with state-of-the-art dynamic MEF algorithms on the image sequence "Girl" (courtesy of Zhengguo Li).

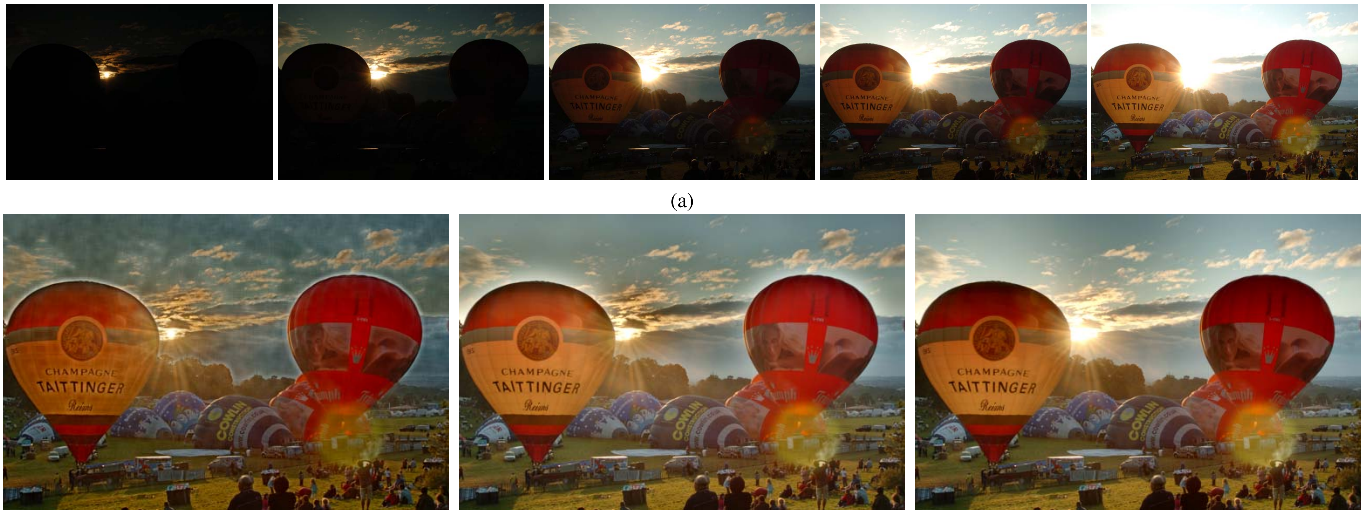
TABLE III

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

Alg	Sen12 [14]	Hu13 [11]	Lee14 [12]	Qin15 [16]	Oh15 [13]	SPD-MEF [17]	Ours
Env	MATLAB+Mex	MATLAB+Mex	MATLAB+Mex	MATLAB+Mex	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

is assumed to be N; for iterative algorithms, the iteration numbers used in the inner and outer loops are I_i and I_o, respectively. The complexities of different methods are listed in Table II, where we find that the proposed method and Li14 [15] enjoy the lowest computational complexity, which is linear with the number of pixels in the sequence.

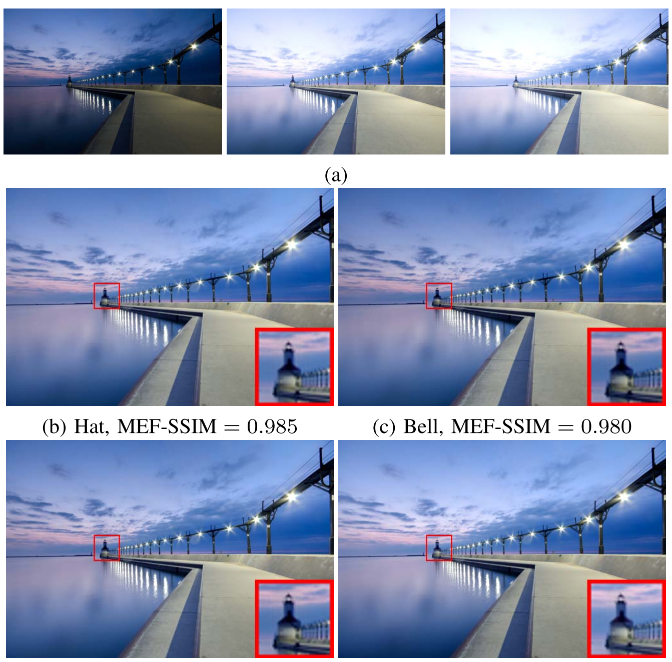
The average running time of different algorithms on 12 natural scenes is listed in Table III. Note that since the implementations of some algorithms in Table II are not publicly available, we are not able to report the running time of those algorithms in Table III. The experiment is conducted on a computer with 4GHz CPU and 32GB RAM. To make a fair comparison, the stride of SPD-MEF is set to one. Our MATLAB code runs the fastest among the competing algorithms, accelerating the original SPD-MEF more than 30 times. When compared to Mertens09 [8] that is widely adopted in mobile devices as a core module to capture HDR-like pictures (*i.e.*, the HDR mode) [10], our method



(a)

(b) Single-scale, MEF-SSIM = 0.851 (c) Three-scale, MEF-SSIM = 0.926 (d) Five-scale, MEF-SSIM = 0.963

Fig. 10. The number of scales in our method plays an important role in fusion quality. (a) Image sequence "Balloons" (courtesy of Erik Reinhard). (b) Single-scale result. (c) Three-scale result. (d) Five-scale result, whose scale is computed adaptively using Eq. (17).



(a)

(b) Hat, MEF-SSIM = 0.985 (c) Bell, MEF-SSIM = 0.980

(d) Gaussian, MEF-SSIM = 0.983 (e) Ours, MEF-SSIM = 0.992

Fig. 11. Visual comparison of different intensity weight functions. (a) Image sequence "Set" (courtesy of Jianbing Shen).

TABLE IV

IMPACT OF THE NUMBER OF SCALES IN TERMS OF MEF-SSIM AVERAGED OVER THE STATIC DATASET IN [17]

# of scale	Single	Three	Adaptive (Ours)
MEF-SSIM	0.849	0.961	0.985

TABLE V

IMPACT OF WEIGHT FUNCTIONS IN TERMS OF MEF-SSIM AVERAGED OVER THE STATIC DATASET IN [17]

Function	Hat	Gaussian	Bell	Ours
MEF-SSIM	0.983	0.982	0.981	0.985

shares the same computational complexity, and therefore has great potentials in enabling real-time mobile applications for challenging dynamic scenes.

D. Ablation Experiments

1) Impact of the Number of Scales: We first visualize the impact of the number of scales J on the final fusion performance using the image sequence "Balloons". When the number of scales increases, our method gradually spreads the halos around the two balloons over the background, making the sky brighter and perceptually more appealing (see Fig. 10). The spatial inconsistency is also effectively reduced at the price of some detail loss (*e.g.*, around the sun). Our adaptive strategy of determining J according to Eq. (17) achieves a satisfactory trade-off among spatial consistency, detail preservation, and halo suppression.

This observation is consistent across many static scenes. In Table IV, we list the average MEF-SSIM scores of the proposed method w.r.t. different scales on the static dataset [17]. We can see that the proposed adaptive scale-selection method achieves the best results.

2) Impact of the Weight Functions: We have drawn four weight functions in Fig. 4. Here we visually compare the fusion results in Fig. 11, where we find that the hat-shaped and Gaussian curves generate visually close results because both weight intensities in a similar fashion. Compared to the bell-shaped curve, the proposed weight function is more friendly to less well-exposed intensities, resulting in a slightly brighter overall appearance with a higher MEF-SSIM value.

In Table V, the average MEF-SSIM scores of the proposed method w.r.t. different weight functions on the static dataset [17] are listed. We can see that the proposed weight function achieves the best results.

V. Conclusion

We studied in-depth the structural patch decomposition (SPD) method for MEF, and presented an unnormalized approximation of it, which speeds up SPD-MEF more than 30 times without sacrificing the MEF performance. We then presented a multi-scale extension of SPD-MEF to effectively reduce halo artifacts near strong edges. Quantitative and qualitative experiments on static and dynamic scenes validated the advantages of the proposed fast multi-scale SPD-MEF method, which provides a practical solution to fusing high-resolution dynamic sequences on mobile devices.

References

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

[2] M. D. Grossberg and S. K. Nayar, "Determining the camera response from images: What is knowable?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1455–1467, Nov. 2003.

[3] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, Jul. 2002.

[4] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 249–256, Jul. 2002.

[5] 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.

[6] H. Li, X. Jia, and L. Zhang, "Clustering based content and color adaptive tone mapping," *Comput. Vis. Image Understand.*, vol. 168, pp. 37–49, Mar. 2018.

[7] V. Laparra, A. Berardino, J. Ballé, and E. P. Simoncelli, "Perceptually optimized image rendering," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 34, no. 9, p. 1511, Sep. 2017.

[8] 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.

[9] Z. Li, Z. Wei, C. Wen, and J. Zheng, "Detail-enhanced multi-scale exposure fusion," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1243–1252, Mar. 2017.

[10] S. W. Hasinoff *et al.*, "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Trans. Graph.*, vol. 35, no. 6, pp. 1–12, Nov. 2016.

[11] J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR deghosting: How to deal with saturation?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1163–1170.

[12] C. Lee, Y. Li, and V. Monga, "Ghost-free high dynamic range imaging via rank minimization," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1045–1049, Sep. 2014.

[13] T.-H. Oh, J.-Y. Lee, Y.-W. Tai, and I. S. Kweon, "Robust high dynamic range imaging by rank minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1219–1232, Jun. 2015.

[14] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based hdr reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–11, Nov. 2012.

[15] 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.

[16] 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.

[17] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, "Robust multi-exposure image fusion: A structural patch decomposition approach," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2519–2532, May 2017.

[18] 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.

[19] Y. Fang, H. Zhu, K. Ma, Z. Wang, and S. Li, "Perceptual evaluation for multi-exposure image fusion of dynamic scenes," *IEEE Trans. Image Process.*, vol. 29, pp. 1127–1138, Dec. 2020.

[20] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.

[21] Y. Fang, H. Zhu, K. Ma, and Z. Wang, "Perceptual quality assessment of HDR deghosting algorithms," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3165–3169.

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

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

[24] K. Ma and Z. Wang, "Multi-exposure image fusion: A patch-wise approach," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1717–1721.

[25] F. Kou, Z. Li, C. Wen, and W. Chen, "Edge-preserving smoothing pyramid based multi-scale exposure fusion," *J. Vis. Commun. Image Represent.*, vol. 53, pp. 235–244, May 2018.

[26] 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.

[27] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and A. C. Bovik, "Single-scale fusion: An effective approach to merging images," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 65–78, Jan. 2017.

[28] 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.

[29] 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.

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

[31] K. Ma, Z. Duanmu, H. Yeganeh, and Z. Wang, "Multi-exposure image fusion by optimizing a structural similarity index," *IEEE Trans. Comput. Imag.*, vol. 4, no. 1, pp. 60–72, Mar. 2018.

[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 *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4724–4732.

[33] K. Ma, Z. Duanmu, H. Zhu, Y. Fang, and Z. Wang, "Deep guided learning for fast multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 2808–2819, 2020.

[34] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[35] O. Gallo, N. Gelfandz, W.-C. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Apr. 2009, pp. 1–7.

[36] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–12, Jul. 2017.

[37] S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang, "Deep high dynamic range imaging with large foreground motions," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 117–132.

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

[39] Y. Liu and Z. Wang, "Dense SIFT for ghost-free multi-exposure fusion," *J. Vis. Commun. Image Represent.*, vol. 31, pp. 208–224, Aug. 2015.

[40] H. Li and L. Zhang, "Multi-exposure fusion with CNN features," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 1723–1727.

[41] K. Jacobs, C. Loscos, and G. Ward, "Automatic high-dynamic range image generation for dynamic scenes," *IEEE Comput. Graph. Appl.*, vol. 28, no. 2, pp. 84–93, Mar. 2008.

[42] H. Zimmer, A. Bruhn, and J. Weickert, "Freehand HDR imaging of moving scenes with simultaneous resolution enhancement," *Comput. Graph. Forum*, vol. 30, no. 2, pp. 405–414, Apr. 2011.

[43] 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.

[44] F. Pece and J. Kautz, "Bitmap movement detection: HDR for dynamic scenes," in *Proc. Conf. Vis. Media Prod.*, Nov. 2010, pp. 1–8.

[45] H. Yong, D. Meng, W. Zuo, and L. Zhang, "Robust online matrix factorization for dynamic background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 7, pp. 1726–1740, Jul. 2018.

[46] 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.

[47] 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.

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

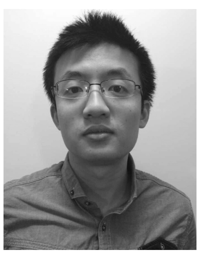
[49] M. Nejati, M. Karimi, S. M. R. Soroushmehr, N. Karimi, S. Samavi, and K. Najarian, "Fast exposure fusion using exposedness function," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2234–2238.

[50] K. Zeng, K. Ma, R. Hassen, and Z. Wang, "Perceptual evaluation of multi-exposure image fusion algorithms," in *Proc. 6th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, Sep. 2014, pp. 27–28.

[51] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from images," in *Proc. Conf. Comput. Graph. Interact. Techn.*, Aug. 1997, pp. 369–378.

[52] Photomatix. (2017). *Commercially-Available HDR Processing Software*. [Online]. Available: http://www.hdrsoft.com

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



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

Hongwei Yong received the B.Sc. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. His research interests include image modeling and deep learning.

Lei Zhang (Fellow, IEEE) received the B.Sc. degree from the Shenyang Institute of Aeronautical Engineering, Shenyang, China, in 1995, and the M.Sc. and Ph.D. degrees in control theory and engineering from Northwestern Polytechnical University, Xi'an, China, in 1998 and 2001, respectively. From 2001 to 2002, he was a Research Associate with the Department of Computing, The Hong Kong Polytechnic University. From January 2003 to January 2006, he worked as a Postdoctoral Fellow of the Department of Electrical and Computer Engineering, McMaster University, Canada. In 2006, he joined the Department of Computing, The Hong Kong Polytechnic University, as an Assistant Professor, where he has been a Chair Professor with the Department of Computing since July 2017. He has published more than 200 articles in his research areas. As of 2020, his publications have been cited more than 49 000 times in literature. He was a Clarivate Analytics Highly-Cited Researcher from 2015 to 2019. His research interests include computer vision, image and video analysis, pattern recognition, and biometrics. He is a Senior Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, and is/was an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the *SIAM Journal of Imaging Sciences*, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and *Image and Vision Computing*.