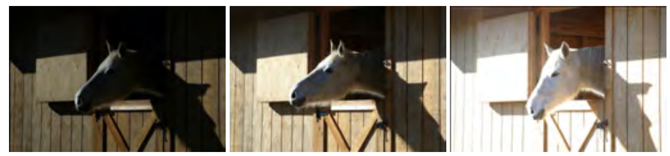
Robust Multi-Exposure Image Fusion: A Structural Patch Decomposition Approach

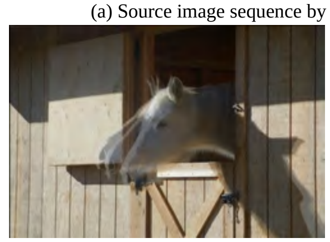
Kede Ma, *Student Member, IEEE,* Hui Li, Hongwei Yong, Zhou Wang, *Fellow, IEEE,* Deyu Meng, *Member, IEEE,* and Lei Zhang, *Senior Member, IEEE*

***Abstract*—We propose a simple yet effective structural patch decomposition (SPD) approach for multi-exposure image fusion (MEF) that is robust to ghosting effect. We decompose an image patch into three conceptually independent components: signal strength, signal structure, and mean intensity. Upon fusing these three components separately, we reconstruct a desired patch and place it back into the fused image. This novel patch decomposition approach benefits MEF in many aspects. First, as opposed to most pixel-wise MEF methods, the proposed algorithm does not require post-processing steps to improve visual quality or to reduce spatial artifacts. Second, it handles RGB color channels jointly and thus produces fused images with more vivid color appearance. Third and most importantly, the direction of the signal structure component in the patch vector space provides ideal information for ghost removal. It allows us to reliably and efficiently reject inconsistent object motions w.r.t. a chosen reference image without performing computationally expensive motion estimation. We compare the proposed algorithm with 12 MEF methods on 21 static scenes and 12 deghosting schemes on 19 dynamic scenes (with camera and object motion). Extensive experimental results demonstrate that the proposed algorithm not only outperforms previous MEF algorithms on static scenes but also consistently produces high quality fused images with little ghosting artifacts for dynamic scenes. Moreover, it maintains a lower computational cost compared with state-of-the-art deghosting schemes.**[1]

***Index Terms*—Multi-exposure image fusion, high dynamic range imaging, structural patch decomposition, deghosting.**



(a) Source image sequence by courtesy of Sing Bing Kang [42]



(b) Mertens09 [9]

(c) SPD-MEF

Fig. 1. Demonstration of MEF in the presence of camera and object motion. Source image sequence is aligned by SIFT feature matching to compute an affine transformation matrix with an l_{21}-norm loss.

I. Introduction

Natural scenes often span greater dynamic ranges of luminance values than those captured by commercial imaging products [2]. High dynamic range (HDR) imaging techniques overcome this limitation by first capturing multiple pictures with different exposure levels and then reconstructing an HDR image through inverting the camera response function (CRF). The main challenge is the estimation of CRF, which is an ill-posed problem. Additional information (e.g., exposure time) and constraints (e.g., assuming some particular parametric forms for CRF) are needed in order to break the self-similar and exponential ambiguities [3], [4]. After acquiring HDR images, a tone mapping process is needed to compress the dynamic range of HDR images for display purpose, since most standard displays currently in use are of low dynamic range (LDR) [5]. Multi-exposure image fusion (MEF) provides a cost-effective alternative to circumvent the gap between HDR imaging and LDR displays. Taking the source image sequence with different exposure levels as input, it directly synthesizes an LDR image that is expected to be more informative and perceptually appealing than any of the input images [6], [7].

Since first introduced in 1980's [6], MEF has been attracting considerable interest from both academia and industry. Most existing MEF algorithms are pixel-wise methods [8]–[13], which however suffer from a main drawback: the weighting map is often too noisy and may result in various artifacts if directly applied to the fusion process. Thus, ad-hoc remedies have been proposed to post-process the weighting map by either smoothing [8]–[10] or edge preserving filtering [11]–[13]. To fairly compare these algorithms, Ma *et al.* [7] made one of the first attempts to develop an objective quality measure for MEF by combining the principle of the structural similarity (SSIM) index [14] with a patch consistency measure. We will refer to this measure as MEF-SSIM in the rest of the paper.

Despite the demonstrated success, typical MEF algorithms require the input source image sequence to be perfectly aligned and there is little object motion in the scene. In practice, however, a small displacement due to hand-held cameras

Kede Ma and Zhou Wang are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada (e-mail: {k29ma, zhou.wang}@uwaterloo.ca).

Hui Li, Hongwei Yong, and Lei Zhang are with the Department of Computing, the Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: {cshli, cshyong, cslzhang}@comp.polyu.edu.hk).

Hongwei Yong and Deyu Meng are with the Institute for Information and System Sciences and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: yonghw@stu.xjtu.edu.cn; dymeng@mail.xjtu.edu.cn).

[1]The MATLAB code of the proposed algorithm will be made available online. Preliminary results of Section III-A [1] were presented at the IEEE International Conference on Image Processing, Canada, 2015.

or object motion (such as ripples and human movement) would neutralize the advantages brought by MEF and cause artifacts referred to as "ghosting", as exemplified in Fig. 1. Similar problem would occur in HDR reconstruction if the displacement is not taken good care of [2]. The difference is that HDR reconstruction works in the radiance domain (where the value is linear w.r.t. the exposure time), while MEF works in the intensity domain (after applying CRF to the radiance value). Compared with object motion, camera motion is relatively easy to tackle via either setting a tripod or employing some registration techniques [15]–[17]. As a result, substantial efforts have been put to develop ghost removal algorithms with an emphasis on object motion. Many existing deghosting algorithms require pixel- or patch-level motion estimation [18]–[21], and their performance is highly dependent on the motion estimation accuracy. Some other methods [22], [23] impose a low rank constraint on the input sequence using a non-convex iterative optimization framework, which typically require more than 3 input images to provide a reasonable low rank structure. One problem shared by these design principles is that they suffer from high computational burden, which may not be affordable by mobile devices.

In this paper, we propose a simple yet robust MEF method, which we name structural patch decomposition (SPD) based MEF (SPD-MEF). Different from the commonly used pixel-wise MEF methods in the literature, we work on image patches. Specifically, we first decompose an image patch into three conceptually independent components: signal strength, signal structure and mean intensity, and process each component based on patch strength, exposedness and structural consistency measures. This novel patch decomposition brings many benefits to the fusion process. First, the weighting maps generated by SPD-MEF are free of noise. As a result, the proposed method does not need post-processing steps to improve the perceived quality or to suppress the spatial artifacts of fused images. Second, it makes use of color information more naturally by treating RGB color channels of an image patch jointly. More importantly, the direction information of the signal structure component enables us to easily check the structural consistency of multi-exposure patches so as to produce a high quality fused image with little ghosting artifacts. We conduct comprehensive experiments by comparing SPD-MEF with 12 MEF methods on 21 static scenes and 12 ghost removal schemes on 19 dynamic scenes (in the presence of camera and object motion). The proposed SPD-MEF method consistently produces better quality fused images for static scenes qualitatively and quantitatively (in terms of the MEF-SSIM index [7]). It also provides significant perceptual gains for dynamic scenes while keeping the computational complexity manageable as verified by our complexity analysis and execution time comparison.

The rest of the paper is organized as follows. Section II reviews existing MEF and ghost removal algorithms. Section III presents in detail the robust SPD-MEF algorithm that is resistant to ghosting effect. Section IV compares the proposed SPD-MEF algorithm with representative MEF and deghosting methods, followed by the computational complexity analysis. We conclude the paper in Section V.

II. Related Work

A. MEF algorithms for static scenes

Most existing MEF algorithms are pixel-wise methods that typically follow a weighted summation framework

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

where K is the number of input images in the multi-exposure source sequence, $\mathbf{W}_k(i)$ and $\mathbf{X}_k(i)$ indicate the weight and intensity values at the i-th pixel in the k-th exposure image, respectively; $\hat{\mathbf{X}}$ represents the fused image. A straightforward extension of this approach in transform domain is to replace $\mathbf{X}_k(i)$ with transform coefficients. The weighting map $\mathbf{W}_k$ often bears information regarding structure preservation and visual importance of the k-th input image at the pixel level. With specific models to quantify this information, existing MEF algorithms differ mainly in the computation of $\mathbf{W}_k$ and how it may adapt over space or scale based on image content.

In 1980's, Burt and Adelson proposed the well known Laplacian pyramid decomposition for binocular image fusion [24]. This decomposition was later adopted by other MEF algorithms [9], [25], [26] to refine $\mathbf{W}_k$ so as to reduce the spatial distortions in the fused image. Edge preserving filters such as bilateral filter [27], guided filter [28] and recursive filter [29] had been applied to retrieve edge information and/or refine $\mathbf{W}_k$ in [11], [30], [12], [31] and [13]. Box and Gaussian filters were also common choices for MEF algorithms to damp $\mathbf{W}_k$ [8]–[10]. Song *et al.* incorporated MEF into a MAP framework [32]. Since no explicit refinement of $\mathbf{W}_k$ was introduced, the fused images tend to be noisy on many test sequences. Another MAP based approach used perceived local contrast and color saturation to construct $\mathbf{W}_k$, whose smoothing was done within a hierarchical multivariate Gaussian framework [33]. A variational approach for MEF was proposed in [34] by combining color matching and gradient direction information. The noise problem of $\mathbf{W}_k$ is less salient because only two input images are required, one serving as the base layer and the other as the detail layer. Hara *et al.* determined the global and local weights via gradient-based contrast maximization and image saliency detection [35]. The work in [36] divided input images into several non-overlapping patches and selected the ones with the highest entropy as the winners. The blocking artifacts were reduced by a blending function.

The above mentioned pixel-wise MEF methods need to deliberately take into account the noisy characteristics of $\mathbf{W}_k$. Post-processing is a must to produce a reasonable fused image, which is a main drawback of this type of methods. Moreover, most existing MEF algorithms are only verified using limited examples, without comprehensive verifications on databases that contain sufficient variations of image content.

B. Ghost removal algorithms for dynamic scenes

As shown in Fig. 1, the MEF methods may produce ghosting artifacts in the presence of camera and object motion. To reduce such artifacts during fusion, a variety of ghost removal algorithms have been proposed. In the radiance domain[2], the linearity between the sensor radiance and the exposure time have been well exploited either directly [37] or through some mathematical models such as energy minimization [18], [38] and low rank minimization [22], [23]. The assumption here is that the linearity should only be broken when the scene changes due to moving objects, provided that the alignment and CRF estimation are perfect. In addition, Eden *et al.* selected one radiance value from one of the input images for each spatial location as an attempt to eliminate ghosting artifacts. However, moving object duplication or deformation may appear [39].

In the intensity domain, the intensity mapping function (IMF) [3] has been heavily used to map between intensity values of any two exposures, which makes the motion estimation [40], [41] and the inconsistent motion detection [19], [42], [43] easier. Exposure invariant features also play an important role in detecting motion pixels, which include image gradient and structure [21], [30], [44], entropy [45] and median threshold bitmap [16], [46]. The proposed ghost removal scheme takes advantage of the above two strategies. We adopt an exposure invariant feature, namely the structural consistency measure, to reject inconsistent motions and refine the procedure with the help of IMF. Some other algorithms also assume that the background dominates the scene and moving objects only appear in one exposure [13], [47] in order to simplify the ghost removal process.

Most existing ghost removal algorithms that deliver state-of-the-art performance require motion estimation in an iterative optimization framework, which suffers from substantial computational complexity and is not suitable for mobile devices. Moreover, deformation of objects often appears as a consequence of inaccurate motion estimation.

III. Structural Patch Decomposition for MEF

In this section, we detail the proposed structural patch decomposition (SPD) approach for MEF. We first describe a baseline version that works for static scenes, and then extend it to dynamic scenes by adding a structural consistency check, resulting in the robust SPD-MEF algorithm.

A. Baseline SPD-MEF

Let $\{\mathbf{x}_k\} = \{\mathbf{x}_k | 1 \leq k \leq K\}$ be a set of color image patches extracted at the same spatial location of the source sequence that contains K multi-exposure images. Here $\mathbf{x}_k$ for all k are column vectors of CN^2 dimensions, where C is the number of color channels of the input images and N is the spatial size of a square patch. Each entry of the vector is one of the three intensity values in RGB channels of a pixel in the patch. Given a color patch, we first decompose it into three components: signal strength, signal structure, and mean intensity

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

where $\|\cdot\|$ denotes the l_2 norm of a vector, $\mu_{\mathbf{x}_k}$ is the mean value of the patch, and $\tilde{\mathbf{x}}_k = \mathbf{x}_k - \mu_{\mathbf{x}_k}$ denotes a mean-removed patch. The scalar $c_k = \|\tilde{\mathbf{x}}_k\|$, the unit-length vector $\mathbf{s}_k = \tilde{\mathbf{x}}_k / \|\tilde{\mathbf{x}}_k\|$, and the scalar $l_k = \mu_{\mathbf{x}_k}$ represent the signal strength, signal structure, and mean intensity components of $\mathbf{x}_k$, respectively. Any patch can be uniquely decomposed into the three components and the process is invertible. As such, the problem of constructing a patch in the fused image is converted to processing the three components separately and then inverting the decomposition.

We first process the component of signal strength. The visibility of the local patch structure largely depends on local contrast, which is directly related to signal strength. Usually, the higher the contrast, the better the visibility. Considering that all input source image patches as realistic capturing of the scene, the patch that has the highest contrast among them would correspond to the best visibility. Therefore, the desired signal strength of the fused image patch is determined by the highest signal strength of all source image patches

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

Different from signal strength, the unit-length structure vector $\mathbf{s}_k$ points to a specific direction in the CN^2 dimensional space. The desired structure of the fused image patch is expected to best represent the structures of all source image patches. A simple implementation of this relationship is given by

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

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

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

where $p \geq 0$ is an exponent parameter. With various choices of the value of p, this general formulation leads to a family of weighting functions with different physical meanings. The larger the p is, the more emphasis is put on the patches that have relatively larger strength.

Due to the construction of $\mathbf{x}_k$ that stacks RGB channels of a patch into one vector, Eq. (3) and Eq. (4) inherently take into account color contrast and structure. An example is shown in Fig. 2. For smooth patches (such as the door frames in the middle of the image) that contain little structure information, SPD-MEF prefers the ones in the 3-rd image that contain strong color information than grayish ones that usually

[2]This type of methods assume the availability of CRF or raw radiance values as in the case of in-camera image processing.

(a) $\underline{c}_1$=0.1, c_1=0.2 (b) $\underline{c}_2$=0.3, c_2=0.3 (c) $\underline{c}_3$=0.3, c_3=7.5 (d) $\underline{c}_4$=0.0, c_4=0.0

(e) Gu12 [8] (f) Shutao12 [13] (g) SPD-MEF

Fig. 2. Making use of color contrast. $\underline{c}_k$ is the average signal strength of the k-th inset patch computed from RGB channels separately. c_k is the corresponding signal strength by treating RGB channels jointly. Source image sequence by courtesy of Tom Mertens [9].

result from under/over-exposure. By contrast, MEF algorithms that treat RGB channels separately may not make proper use of color information and give patches across exposures similar perceptual importance for fusion.

With regard to the mean intensity of the local patch, we take a similar form of Eq. (4)

$$\hat{l} = \frac{\sum_{k=1}^{K} L\left(\mu_k, l_k\right) l_k}{\sum_{k=1}^{K} L\left(\mu_k, l_k\right)}, \tag{6}$$

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

$$L\left(\mu_k, l_k\right) = \exp\left(-\frac{(\mu_k - \mu_c)^2}{2\sigma_g^2} - \frac{(l_k - l_c)^2}{2\sigma_l^2}\right), \tag{7}$$

where σ_g and σ_l control the spreads of the profile along μ_k and l_k dimensions, respectively. μ_c and l_c are constants for the mid-intensity values. For example, both μ_k and l_c are 0.5 for source image sequences normalized to $[0, 1]$.

Once $\hat{c}$, $\hat{\mathbf{s}}$ and $\hat{l}$ are computed, they uniquely define a new vector

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

We extract patches from the source sequence using a moving window with a fixed stride D. The pixels in overlapping patches are averaged to produce the final output. By determining the desired patch using the proposed SPD approach, we make full use of perceptually meaningful information scattered across exposures in the same spatial location.

B. Robust SPD-MEF

We extend the baseline SPD-MEF to account for dynamic scenes in the presence of camera and object motion. We assume that the input source sequence is aligned, for example by setting a tripod or some image registration algorithms [15]–[17], [48]. This assumption is mild because the camera motion is usually small and relatively uniform. In this paper, we implement image registration by first performing SIFT [17] matching and then computing an affine transformation matrix from matched points with an l_{21}-norm loss. It works well on all test sequences that need to be aligned. The use of l_{21}-norm loss is because it is robust to mismatched points and can be efficiently solved using iteratively reweighted least squares. Similar to the methods in [37], [43], we also pick one exposure as the reference to determine the motion appeared in the fused image and reject inconsistent motions in the rest images w.r.t. it. Throughout the paper, we select the one with normal exposure if the source image sequence contains three input images. Otherwise, we choose the one that has the least number of under- or over-exposed patches, as suggested in [19], [20], [37], [43].

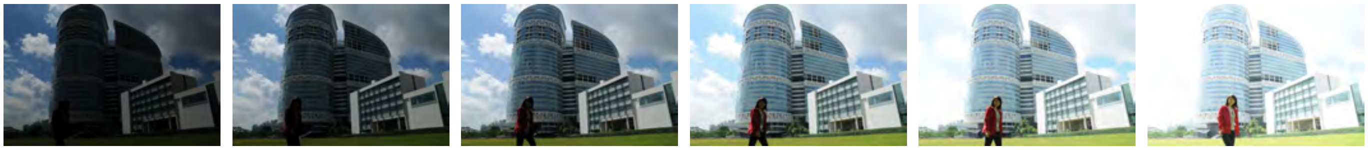
Within the framework of the proposed SPD, it is very convenient to detect inconsistent motions across exposures by making use of the structure vector $\mathbf{s}_k$. To be specific, we compute the inner product between the reference signal structure $\mathbf{s}_r$ and the signal structure $\mathbf{s}_k$ of another exposure

$$\rho_k = \mathbf{s}_r^T \mathbf{s}_k = \frac{(\mathbf{x}_r - l_r)^T (\mathbf{x}_k - l_k)}{\|\mathbf{x}_r - l_r\| \|\mathbf{x}_k - l_k\|}\,. \tag{9}$$

ρ_k lies in $[-1, 1]$ with a larger value indicating higher consistency between $\mathbf{s}_k$ and $\mathbf{s}_r$. Since $\mathbf{s}_k$ is constructed by mean removal and strength normalization, it is robust to exposure and contrast variations. We make an additional modification on Eq. (9) by adding a small constant ϵ to both the numerator and the denominator

$$\rho_k = \frac{(\mathbf{x}_r - l_r)^T (\mathbf{x}_k - l_k) + \epsilon}{\|\mathbf{x}_r - l_r\| \|\mathbf{x}_k - l_k\| + \epsilon}\,. \tag{10}$$

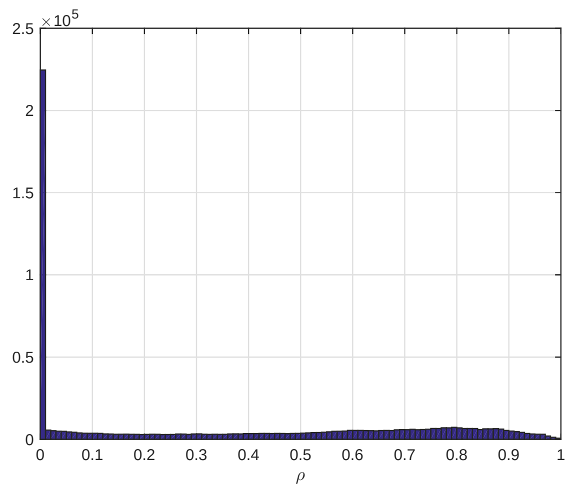
The constant ϵ is to ensure the robustness of the structural consistency to sensor noise. More specifically, in the darkest



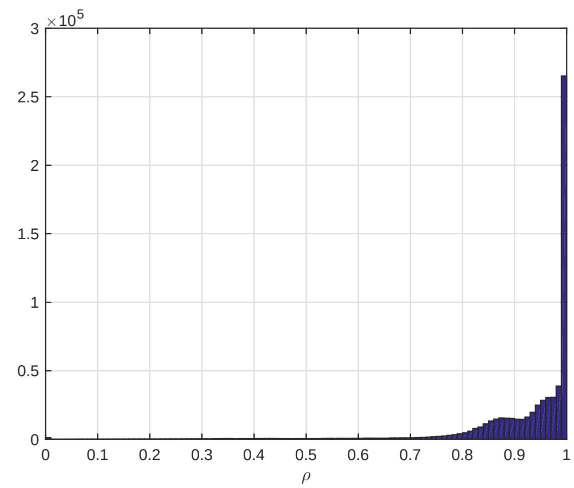
(a) Source image sequence by courtesy of Zhengguo Li [43]

(b) Structural consistency maps corresponding to the above sequence

Fig. 3. Demonstration of the structural consistency map. The girl in the source sequence is walking slowly from left to right. The 4-th exposure image is chosen as the reference.



(a) Using Eq.(9)



(b) Using Eq.(10)

Fig. 4. The histograms of the structural consistency ratio ρ on the darkest regions of 19 test source sequences computed by Eq.(9) and Eq.(10), respectively.

areas where signal strengths are weak, when the structure vector $\mathbf{s}_r$ is scaled to unit length, it will mainly contain amplified noise structures, making the structural consistency check in Eq. (9) unreliable. Fortunately, this issue can be well addressed by adding ϵ to both the denominator and the numerator as in Eq. (10). In those regions, the noise strength is small as compared to ϵ and thus the consistency ratio ρ_k will be close to 1, regardless of the noise structure. To justify our claim, we collect all the darkest regions of the reference images in the 19 test source sequences used in the paper and plot histograms of the structural consistency ratio ρ computed by Eq. (9) and Eq. (10), respectively. As can be observed in Fig. 4, adding ϵ boosts ρ to be close to 1, which allows for retrieving faithful structures from co-located regions in other exposures.

To reject inconsistent patches, we binarize ρ_k with a predefined threshold T_s

$$\tilde{B}_k = \begin{cases} 1 & \text{if } \rho_k \geq T_s \\ 0 & \text{if } \rho_k < T_s \end{cases} . \tag{11}$$

The corresponding binary map generated for each exposure (including the reference which is uniformly one) is referred to as the structural consistency map, as shown in Fig. 3. From the figure, we observe that the inconsistent motions across exposures are reliably identified with minimal false positive detection, and the structure vectors of over-exposed areas in the reference image (e.g., the clouds in the left part of the 4-th image) are consistent with the same regions in other exposures, which verifies our claim of properly handling under- or over-exposed regions.

Although we leave open the possibility of filling in the under- or over-exposed regions of the reference image with structures from other exposures, we add another constraint to check whether those structures are proper for fusion in order to minimize ghosting artifacts by invoking IMF, which is capable of mapping between intensity values of any two exposures. For example, we can easily create a latent image that contains the same motion as the 4-th image of Fig. 3(a) but has an exposure level like the 2-nd image of Fig. 3(a) by mapping the intensity values of the former to the latter using IMF. We first create $K-1$ latent images by mapping the intensity values of the reference image to the rest $K-1$ exposures and compute the absolute mean intensity difference of co-located patches in the k-th exposure and its corresponding latent image. We again threshold the difference

$$\bar{B}_k = \begin{cases} 1 & \text{if } |l_k - l'_k| < T_m \\ 0 & \text{if } |l_k - l'_k| \geq T_m \end{cases} , \tag{12}$$

where l'_k is the mean intensity of the co-located patch in the k-th latent image created from the reference image and T_m is a pre-defined threshold. We define the final structural consistency measure w.r.t. a reference patch by multiplying $\tilde{B}_k$ and $\bar{B}_k$

$$B_k = \tilde{B}_k \cdot \bar{B}_k \, . \tag{13}$$

In general, $\bar{B}_k$ mainly works as a supplement to $\tilde{B}_k$ to more conservatively fill in the under- or over-exposed regions of the reference image.

The remaining work is to incorporate B_k into our baseline SPD-MEF. Specifically, for patches in other exposures that are rejected through B_k, we compensate for them by choosing the corresponding patches in their latent images generated by IMF and the fusion process remains the same. By doing so, we save substantial computation by avoiding explicit motion estimation to find the corresponding patch in the original image that may be in different intensities, orientations and scales. Moreover, we are able to adjust the mean intensity of the moving object in the reference image to adapt it to the neighborhood environment, which avoids abrupt intensity changes in a much cheaper way.

Algorithm 1 SPD-MEF

Input: Source image sequence $\{\mathbf{X}_k\} = \{\mathbf{X}_k | 1 \leq k \leq K\}$

1: Select the reference image $\mathbf{X}_r$ and create $K-1$ latent images $\{\mathbf{X}'_k\} = \{\mathbf{X}'_k | k \neq r\}$ of $\mathbf{X}_r$ using IMF
2: **for** each reference patch $\mathbf{x}_r$ **do**
3: Extract its co-located patches $\{\mathbf{x}_k, \mathbf{x}'_k | k \neq r\}$
4: Check the structural consistency of $\{\mathbf{x}_k\}$ using B_k
5: Reject inconsistent $\mathbf{x}_k$ compensated by $\mathbf{x}'_k$
6: Compute $\hat{c}$, $\hat{\mathbf{s}}$ and $\hat{l}$ separately
7: Reconstruct the fused patch $\hat{\mathbf{x}} = \hat{c} \cdot \hat{\mathbf{s}} + \hat{l}$
8: **end for**
9: Aggregate fused patches into $\hat{\mathbf{X}}$

Output: Fused image $\hat{\mathbf{X}}$

C. Implementation details

We summarize the proposed SPD-MEF approach in Algorithm 1. SPD-MEF has eight parameters in total, including 1) a small positive constant ϵ in Eq. (10), 2) the exponent parameter p to determine the weight of the structure vector component, 3-4) two Gaussian spread parameters σ_g and σ_l to determine the weight of the mean intensity component, 5-6) two thresholds T_s and T_m that binarize the structural consistency map, and 7-8) the patch size N and its associated stride D. The details of how the parameters are set can be found in the supplementary file to this paper. In the following, we briefly present the parameter setting.

The value of ϵ is inherited from the corresponding normalization term of SSIM [14] and is equal to $\frac{1}{2}(0.03L_d)^2$, where

TABLE I
INFORMATION ABOUT STATIC SOURCE SEQUENCES

Source sequence	Size	Image origin
Arno	$339 \times 512 \times 3$	Bartlomiej Okonek [49]
Balloons	$339 \times 512 \times 9$	Erik Reinhard [2]
Belgium house	$384 \times 512 \times 9$	Dani Lischinski [50]
Cave	$384 \times 512 \times 4$	Bartlomiej Okonek [49]
Chinese garden	$340 \times 512 \times 3$	Bartlomiej Okonek [49]
Church	$512 \times 335 \times 3$	Jianbing Shen [26]
Farmhouse	$341 \times 512 \times 3$	HDR projects [51]
House	$340 \times 512 \times 4$	Mertens09 [9]
Lamp	$384 \times 512 \times 15$	Martin Čadík [52]
Landscape	$341 \times 512 \times 3$	HDRsoft [53]
Laurenziana	$512 \times 356 \times 3$	Bartlomiej Okonek [49]
Madison capitol	$384 \times 512 \times 30$	Chaman Singh Verma [54]
Mask	$341 \times 512 \times 3$	HDRsoft [53]
Office	$340 \times 512 \times 6$	MATLAB [55]
Ostrow	$341 \times 512 \times 3$	Bartlomiej Okonek [49]
Room	$341 \times 512 \times 3$	Pangeasoft [56]
Set	$341 \times 512 \times 3$	Jianbing Shen [26]
Tower	$512 \times 341 \times 3$	Jacques Joffre [53]
Venice	$341 \times 512 \times 3$	HDRsoft [53]
Window	$384 \times 512 \times 3$	Hvdwolf [57]
Yellow hall	$339 \times 512 \times 3$	Jianbing Shen [26]

L_d is the maximum intensity value of the source sequence (For a normalized sequence, $L_d = 1$). It turns out that SPD-MEF is insensitive to ϵ. The exponent parameter p and two Gaussian spread parameters σ_g and σ_l in the baseline SPD-MEF algorithm are jointly determined by maximizing MEF-SSIM [7] on 5 held-out static source sequences using a grid search method. The possible values of p, σ_g and σ_l are chosen to be $p \in \{1, 2, \cdots, 10\}$, $\sigma_g \in \{0.1, 0.2, \cdots, 1\}$ and $\sigma_l \in \{0.1, 0.2, \cdots, 1\}$, respectively. In other words, there are $1,000$ possible parameter combinations and the one that achieves the highest MEF-SSIM value on average is selected, which turns out to be $\{p = 4,\ \sigma_g = 0.2,\ \sigma_l = 0.5\}$.

The two thresholds T_s and T_m are crucial for SPD-MEF to work with dynamic scenes in the presence of camera and object motion. Both T_s and T_m have the same range $[0, 1]$. Ideally, the structural consistency map should be able to reject inconsistent motions w.r.t. the reference exposure while incorporating as many consistent patches as possible to make full use of all valid information for fusion. Empirically, we find that $T_s = 0.8$ and $T_m = 0.1$ make a good balance between reliably identifying inconsistent motions across exposures and having a low rate of false positive detection.

We now discuss the impact of patch size N on the fusion performance and computational time. Intuitively, the larger the N is, the more robust the signal structure vector is in terms of structural consistency. However, the computational complexity also increases with N significantly. We find that $N = 21$ provides a good balance between the performance and the complexity. The stride of moving window is determined by $D = \lfloor \frac{N}{10} \rfloor$ accordingly.

Lastly, to generate latent images, we adopt IMF proposed by Grossberg and Nayar [3], which is a non-iterative method, robust to camera motion and can be efficiently implemented by histogram matching. Specifically, the MATLAB function $\mathrm{Imhistmatch}$ with the default settings is used in our implementation.

TABLE II
PERFORMANCE COMPARISON OF SPD-MEF WITH EXISTING MEF ALGORITHMS USING MEF-SSIM [7]. THE QUALITY VALUE RANGES FROM 0 TO 1 WITH A HIGHER VALUE INDICATING BETTER PERCEPTUAL QUALITY. LE AND GE STAND FOR TWO NAÏVE METHODS THAT LINEARLY COMBINE THE INPUT IMAGES USING LOCAL ENERGY AND GLOBAL ENERGY AS WEIGHTING FACTORS, RESPECTIVELY

Source sequence	LE	GE	Raman09 [11]	Gu12 [8]	Shen11 [58]	Shen15 [26]	Bruce14 [59]	Mertens09 [9]	Li13 [12]	SPD-MEF
Arno	0.846	0.960	0.946	0.890	0.955	0.846	0.946	**0.991**	0.969	0.984
Balloons	0.771	0.862	0.768	0.913	0.940	0.776	0.801	**0.969**	0.948	**0.969**
Belgium house	0.732	0.874	0.810	0.896	0.935	0.709	0.832	0.971	0.964	**0.973**
Cave	0.861	0.837	0.694	0.934	0.946	0.788	0.910	0.975	0.978	**0.985**
Chinese garden	0.917	0.928	0.911	0.927	0.964	0.767	0.941	0.989	0.984	**0.991**
Church	0.911	0.948	0.898	0.866	0.959	0.878	0.931	0.989	0.992	**0.993**
Farmhouse	0.942	0.916	0.877	0.932	0.966	0.944	0.930	0.981	**0.985**	0.984
House	0.657	0.836	0.770	0.876	0.925	0.396	0.888	**0.964**	0.957	0.960
Lamp	0.577	0.836	0.729	0.875	0.917	0.539	0.758	**0.969**	0.929	0.956
Landscape	0.901	0.962	0.954	0.941	0.955	0.880	0.979	0.976	0.942	**0.993**
Laurenziana	0.881	0.940	0.934	0.873	0.956	0.881	0.950	**0.988**	0.987	0.987
Madison capitol	0.780	0.886	0.763	0.864	0.940	0.542	0.804	0.977	0.968	**0.983**
Mask	0.876	0.940	0.918	0.879	0.964	0.827	0.939	0.987	0.979	**0.988**
Office	0.831	0.955	0.907	0.900	0.958	0.756	0.943	0.985	0.967	**0.990**
Ostrow	0.793	0.952	0.930	0.877	0.950	0.786	0.931	0.974	0.967	**0.978**
Room	0.904	0.932	0.916	0.853	0.945	0.729	0.938	0.974	**0.986**	0.978
Set	0.856	0.980	0.979	0.911	0.974	0.873	0.975	0.986	0.960	**0.988**
Tower	0.898	0.912	0.895	0.932	0.946	0.779	0.948	**0.986**	**0.986**	**0.986**
Venice	0.846	0.913	0.892	0.889	0.930	0.765	0.933	0.966	0.954	**0.984**
Window	0.884	0.941	0.922	0.876	0.959	0.879	0.937	**0.982**	0.971	**0.982**
Yellow hall	0.963	0.989	0.988	0.869	0.983	0.866	0.990	**0.995**	0.990	**0.995**
Average	0.839	0.919	0.876	0.894	0.951	0.772	0.914	0.980	0.970	**0.982**

IV. EXPERIMENTAL RESULTS

In this section, we conduct comprehensive experiments to verify the performance of SPD-MEF. Throughout the paper, we apply the proposed robust SPD-MEF algorithm to all test sequences (both static and dynamic) with fixed parameter settings. We compare SPD-MEF with state-of-the-art and representative MEF and deghosting algorithms that are specifically designed for static or dynamic scenes. In particular, we first compare SPD-MEF with 12 existing MEF methods on 21 static scenes both visually and in terms of MEF-SSIM [7]. Then, 12 state-of-the-art ghost removal algorithms are compared with SPD-MEF on 19 dynamic scenes. Finally, we perform complexity analysis of state-of-the-art deghosting algorithms and report their average execution time on 12 source sequences. In order to make a fair comparison, all fusion and deghosting results are either generated by the original authors or by the codes that are publicly available with default settings.

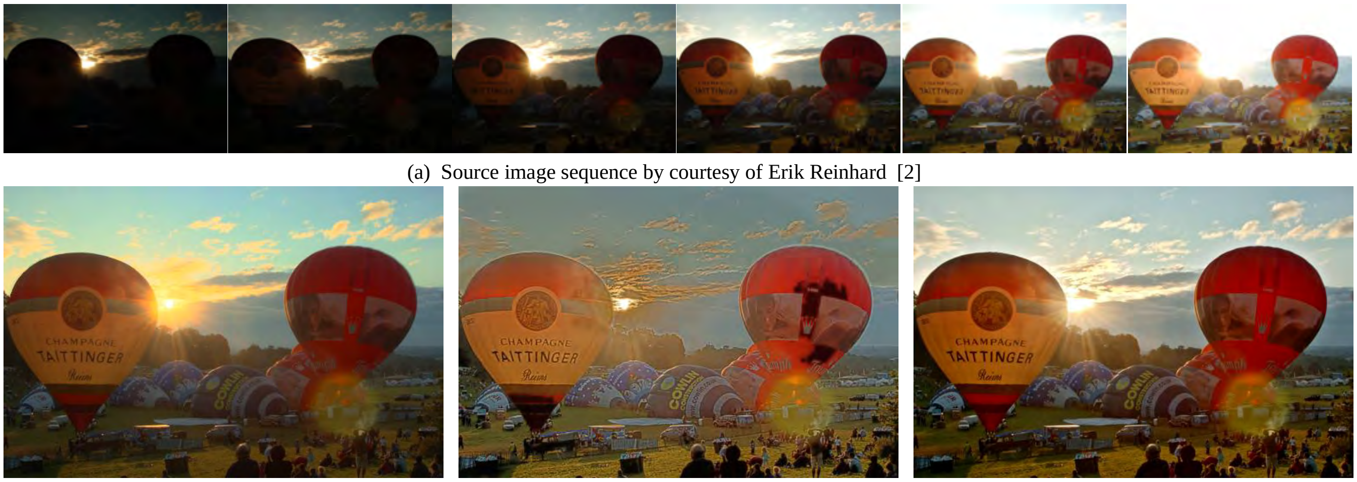
A. Comparison with existing MEF algorithms

We test SPD-MEF on 21 static natural scenes with different exposure levels against 12 existing algorithms. The test source sequences are selected to span a variety of contents including day and light, indoor and outdoor, human and still-life scenes, as listed in Table I. The competing algorithms are chosen to cover a diversity of types, including two simple operators that linearly combine the input images using local and global energy as weighting factors, denoted by LE and GE, respectively, and sophisticated ones with different perceptual emphasis such as Mertens09 [9], Raman09 [11], Shen11 [58], Song12 [32], Li12 [10], Shutao12 [13], Gu12 [8], Li13 [12], Bruce14 [59], and Shen15 [26].

Fig. 2 compares Gu12 [8], Shutao12 [13] with SPD-MEF on the "House" sequence. Based on gradient information, Gu12 [8] focuses on detail enhancement only and ignores color information, therefore it fails to preserve the color appearance in the source sequence. Shutao12 [13] treats RGB channels separately, making it difficult to make proper use of color information. As a result, the color in smooth areas (e.g., the walls and window frames) appears dreary. The global intensity of the fused image also changes drastically, where the left part of the image is much brighter than the right part. By contrast, the proposed method better preserves the color information and the overall appearance of the fused image is more appealing.

Fig. 5 shows the fused images produced by Song12 [32], Shen15 [26] and SPD-MEF on the "Balloons" sequence. Compared with Song12 [32], SPD-MEF produces more natural and vivid color appearance on the sky and the meadow regions. Moreover, it does a better job on structure preservation around the sun area. On the contrary, the fused image produced by Song12 [32] suffers from color distortions and detail loss. Shen15 [26] produces images with sudden intensity changes and uncomfortable colors which are either saturated or pale.

In Fig. 6, we compare Mertens09 [9] and Li12 [10] with SPD-MEF on the "Tower" sequence. In a recent subjective user study [60], Mertens09 [9] performs the best on average among eight MEF algorithms [60]. Li12 [10] is a detail enhanced version of Mertens09 [9]. Compared with Mertens09 [9], we can clearly observe perceptual gains on the fused image produced by SPD-MEF. For example, the structures of the tower at the top and the brightest cloud area are much better preserved. Also, the color appearance of the sky and the meadow regions is more natural and consistent with the source sequence. Li12 [10] is not able to recover such structures lost by Mertens09 [9] but overshoots the details of flowers



(a) Source image sequence by courtesy of Erik Reinhard [2]

(b) Song12 (c) Shen15 (d) SPD-MEF

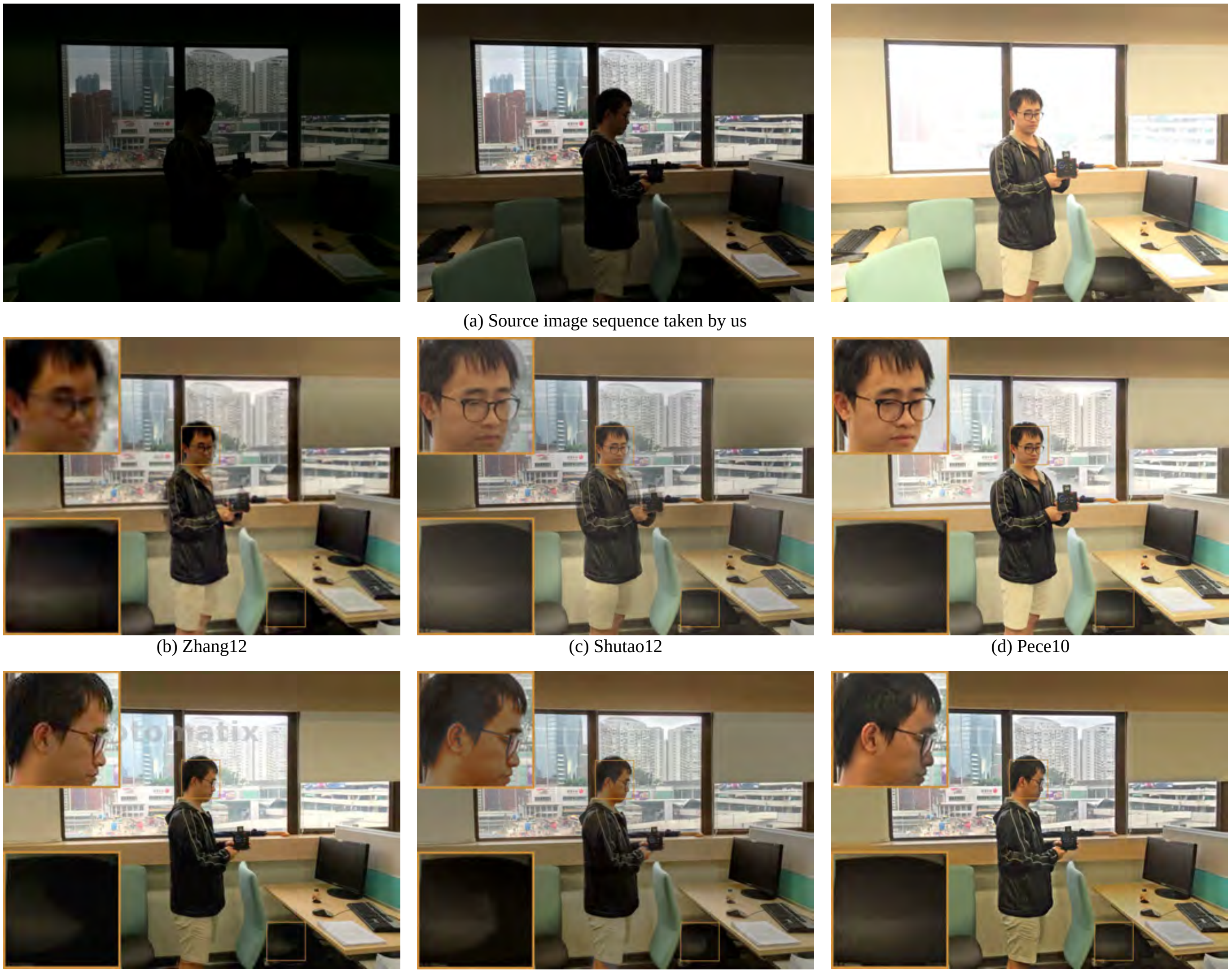
Fig. 5. Comparison of SPD-MEF with Song12 [32] and Shen15 [26].

(a) (b) Mertens09 (c) Li12 (d) SPD-MEF

Fig. 6. Comparison of SPD-MEF with Mertens09 [9] and Li12 [10]. (a) Source image sequence by courtesy of Jacques Joffre [53].

that look artificial. Another comparison of Mertens09 [9] with SPD-MEF can be found in Fig. 1. SPD-MEF better recovers the details inside the stable and makes the overall appearance brighter and warmer.

In order to evaluate the performance of MEF algorithms objectively, we adopt MEF-SSIM (specifically designed for MEF) that well correlates with subjective judgments [7]. MEF-SSIM [7] is based on the multi-scale SSIM framework and a patch consistency measure. It keeps a good balance between local structure preservation and global luminance consistency. The quality value of MEF-SSIM ranges from 0 to 1 with a higher value indicating better quality. The comparison results of SPD-MEF with 9 MEF algorithms on 21 source sequences are listed in Table II, from which we observe that SPD-MEF is comparable to Mertens09 [9], whose quality values are higher than other MEF algorithms by a large margin. Note that MEF-SSIM [7] works with the luminance component only and may underestimate the quality gain of SPD-MEF since producing a natural and vivid color appearance is one of its main advantages. In addition, the differentiability of MEF-SSIM is relatively low at very high image quality levels, a phenomenon similar to SSIM [14]. Nevertheless, in most testing cases, Mertens09 [9] and SPD-MEF give the best performance compared to all other methods (confirmed by visual inspection), but between the two algorithms, the winner is not always clear. So overall the MEF-SSIM results are reasonable.



(a) Source image sequence taken by us

(b) Zhang12 (c) Shutao12 (d) Pece10

(e) Photomatix (f) Li14 (g) SPD-MEF

Fig. 7. Comparison of SPD-MEF with Zhang12 [30], Shutao12 [13], Pece10 [46], Photomatix [61], and Li14 [43].

TABLE III
INFORMATION ABOUT DYNAMIC SOURCE SEQUENCES

Source sequence	Size	Image origin
Arch	$1024 \times 669 \times 5$	Orazio Gallo [37]
Forrest	$683 \times 1024 \times 4$	Orazio Gallo [37]
Noise camera	$480 \times 640 \times 10$	Orazio Gallo [37]
Puppets	$1024 \times 812 \times 5$	Orazio Gallo [37]
Sculpture garden	$754 \times 1024 \times 5$	Orazio Gallo [37]
Lady	$1024 \times 686 \times 3$	Jun Hu [20]
Horse	$690 \times 1024 \times 3$	Sing Bing Kang [42]
Prof. JeonEigth	$681 \times 1024 \times 7$	Zhengguo Li [43]
YWFusionopolis	$681 \times 1024 \times 6$	Zhengguo Li [43]
Corridor	$768 \times 1024 \times 3$	Us
Office	$768 \times 1024 \times 3$	Us
Brunswick	$683 \times 1024 \times 3$	Fabrizio Pece [46]
Cliffs1	$683 \times 1024 \times 3$	Fabrizio Pece [46]
Llandudno	$683 \times 1024 \times 3$	Fabrizio Pece [46]
Russ1	$683 \times 1024 \times 3$	Fabrizio Pece [46]
Square	$683 \times 1024 \times 3$	Fabrizio Pece [46]
Tate3	$683 \times 1024 \times 3$	Fabrizio Pece [46]
Wroclav	$683 \times 1024 \times 3$	Fabrizio Pece [46]
Campus	$648 \times 1011 \times 6$	Wei Zhang [30]

B. Comparison with existing ghost removal algorithms

We then compare SPD-MEF with 12 state-of-the-art ghost removal algorithms, including methods that require motion estimation: Sen12 [18], Hu12 [19], Hu13 [20], and Qin15 [21]; methods that exploit low rank property of input sequences: Lee14 [22] and Oh15 [23]; and methods that make use of other prior information about the sequences: Gallo09 [37], Pece10 [46], Zhang12 [30], Shutao12 [13], Li14 [43], and Photomatix [61]. For deghosting methods that follow the HDR pipeline, Gallo09 [37], Sen12 [18] and Lee14 [22] recover the CRF using Debevec and Malik's method [62]. Oh15 [23] utilizes the method in Lee *et al.* [4]. To generate an LDR image for display, Gallo09 [37] tone maps the HDR image with the algorithm proposed by Lischinski *et al.* [63]. Sen12 [18] fuses the aligned LDR sequence using Mertens09 [9]. Lee14 [22] adopts the MATLAB function `tonemap`. Oh15 [23] applies the tone mapping using the local adaptation method in Photoshop CS6.

Fig. 7 compares Zhang12 [30], Shutao12 [13], Pece10 [46], Photomatix [61], Li14 [43] with SPD-MEF on the source sequence “Office”. The former three algorithms do not need to specify a reference image. The latter three methods including SPD-MEF select the normal exposure one (2-nd) as the reference for fair comparison. From Fig. 7, one can observe that Zhang12 [30] and Shutao12 [13] suffer from visible ghosting artifacts. This is not surprising because both algorithms assume that the background dominates the scene and the moving objects only appear in one exposure, which is typically not the case encountered in the real-world. Some areas (e.g., the wall) of the fused images generated by the two algorithms also exhibit abrupt intensity changes. Although Pece10 [46] successfully avoids ghost, it generates band distortions near the person and fails to preserve details in areas such as the letters of the paper on the table. The results of Photomatix [61] and Li14 [43] are comparable to that of SPD-MEF, except for some detail loss of the black jacket and the chair on the right. SPD-MEF generates a fused image with vivid color appearance and excellent detail preservation but exhibits some ghosting artifacts.

In Fig. 8, we show the comparison results of Gallo09 [37], Hu12 [19], Hu13 [20], Lee14 [22], Photomatix [61] together with SPD-MEF on the source sequence “Forrest”, which provides an ideal test for the robustness of deghosting schemes in the presence of both tiny random motion (tree branches in the wind) and large motion (person). The second image is selected as the reference for all algorithms for fair comparison. As can be seen, Lee14 [22] and Photomatix [61] suffer from ghosting artifacts resulting from the person appearing in the 4-th image. Lee14 [22] also exhibits color speckle noise due to the inaccurate CRF estimation. Hu13 [20] generates blurry tree branches due to errors in motion estimation, which is difficult to avoid in the presence of tiny random motion. Gallo09 [37] has an overall dim appearance while Hu12 [19] has an overall dazzling appearance with detail loss in their respective under- and over-exposed areas. SPD-MEF on the other hand provides a more vivid appearance with sharp edge and little ghost or blur on tree branches and barks.

The comparison of SPD-MEF with Hu13 [20], Qin15 [21], Sen12 [18], Oh15 [23], and Li14 [43] on the source sequence “Noise camera” is shown in Fig. 9. This sequence is captured in a dark room with a high ISO sensitivity and thus contains substantial sensor noise. An excellent MEF algorithm should on one hand reject inconsistent small motions while on the other hand make full use of all available information in the source sequence to denoise the static scene areas by averaging them. Note that the exposure level is set to be constant and thus the same static scene area of each image should be treated equally during fusion. Through this example, we observe that Sen12 [18], Li14 [43], and Qin15 [21] rely much on the reference image to reject inconsistent motions and are reluctant to make use of information from other images. As a result, substantial noise still remains in static areas such as the table and wall. The noise in the fused image produced by Hu13 [20] is less severe but still visible. Oh15 [23] does a good job in noise removal in static scene areas but fails to prevent ghosting artifacts on the person’s head and arm. This may be because low rank schemes are proved to be effective in denoising but small motions that do not follow the sparsity assumption may result in artifacts. Overall, SPD-MEF is successful in denoising static regions by taking advantage of the fusion scheme and in preventing ghost in dynamic areas. However, noise may not be removed in dynamic scene areas for all algorithms.

TABLE IV
COMPUTATIONAL COMPLEXITIES OF SPD-MEF AND STATE-OF-THE-ART DEGHOSTING SCHEMES

Alg	Complexity
Pece10 [46]	$\mathcal{O}(MK)$
Sen12 [18]	$\mathcal{O}(I_iN^2MK^2)$
Hu13 [20]	$\mathcal{O}(I_iN^2(M\log M)K)$
Lee14 [22]	$\mathcal{O}(I_oI_iMK^2)$
Li14 [43]	$\mathcal{O}(MK)$
Oh15 [23]	$\mathcal{O}(I_oI_iMK^2)$
Qin15 [21]	$\mathcal{O}(I_iN^2M^2K)$
SPD-MEF	$\mathcal{O}(N^2MK)$

C. Computational complexity comparison

We conduct a comprehensive computational complexity analysis of state-of-the-art deghosting schemes together with SPD-MEF in terms of the number of floating point operations. Here, we only consider the dominant computation for all algorithms and make conservative estimates because the details of some algorithms are not precisely clear. Suppose we have K exposures, each of which contains M pixels, where $K \ll M$; the patch size used in patch-based methods is N^2; the iteration numbers used in the inner and outer loops for iterative methods are I_i and I_o, respectively. For Pece10 [46] and Li14 [43], all computation is point-wise operations with a complexity of $\mathcal{O}(MK)$. For Sen12 [18], the heaviest computation lies in computing the cost function through the multi-source bidirectional similarity measure, which has a complexity of $\mathcal{O}(I_iN^2MK^2)$. The K^2 term results from the nested summation over K exposures to compute the cost function. For Hu13 [20], motion estimation using generalized PatchMatch contributes to the main computation compared with updating the latent image and refining the IMF, which has a complexity of $\mathcal{O}(I_iN^2(M\log M)K)$. For low rank minimization based algorithms Lee14 [22] and Oh15 [23], the most costly operation is singular value decomposition with a complexity of $\mathcal{O}(I_oI_i(MK^2+K^3)) = \mathcal{O}(I_oI_iMK^2)$ without assuming any special structure of the matrix to be decomposed. For Qin15 [21], the most time-consuming step is to find reliable motion estimation for nearly M patches with an order of M candidate patches. Therefore, its complexity is approximately $\mathcal{O}(I_iN^2M^2K)$. The proposed SPD-MEF approach has a complexity $\mathcal{O}(N^2MK)$, where the N^2 term arises from SPD and the structural consistency check.

The complexities of competing algorithms are summarized in Table IV, from which we can see that SPD-MEF has a higher complexity than Li14 [43] and Pece10 [46], but lower than Lee14 [22]/Oh15 [23], Sen12 [18], Hu13 [20] and Qin15 [21]. It should be noted that the patch size N^2 used in different algorithms may vary, and the number of I_o and I_i may also vary for different algorithms and sequences

(a) Source image sequence by courtesy of Orazio Gallo [37]

(b) Gallo09 (c) Hu12 (d) Hu13

(e) Lee14 (f) Photomatix (g) SPD-MEF

Fig. 8. Comparison of SPD-MEF with Gallo09 [37], Hu12 [19], Hu13 [20], Lee14 [22], and Photomatix [61].

TABLE V

AVERAGE EXECUTION TIME IN SECONDS ON 12 SOURCE SEQUENCES OF SIZE APPROXIMATELY $683 \times 1024 \times 3$

Alg	Sen12 [18]	Hu13 [20]	Lee14 [22]	Qin15 [21]	Oh15 [23]	Pece10 [46]	SPD-MEF
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	1.35 ± 0.07	13.64 ± 0.81

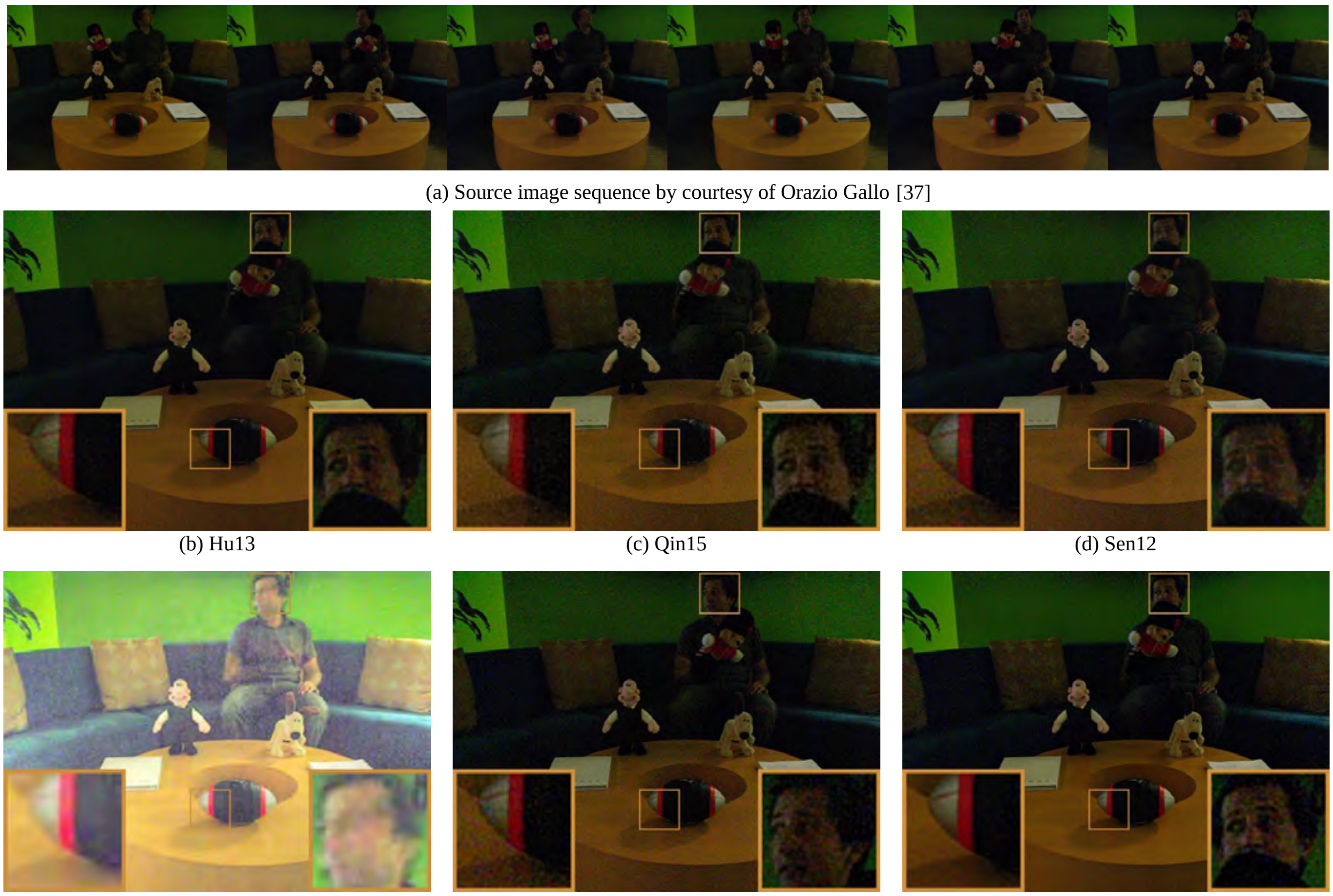
of different contents. To gain a concrete impression, we report the execution time of the above algorithms (except for Li14 [43] whose code is not publicly available) for 12 source sequences of size approximately $683 \times 1024 \times 3$ on a computer with 4G Hz CPU and 32G RAM. From Table V, we have several observations. First, the execution time conforms to the computational complexity analysis. Second, the standard deviations are relatively large for iterative methods, whose iteration numbers depend on the image content. Third, our proposed SPD-MEF algorithm keeps a good balance between fusion performance and computational complexity.

D. Limitations of SPD-MEF

As can be seen from Fig. 7(g), the proposed SPD-MEF algorithm may produce some visible halo artifacts around sharp edges. This is because the mean intensity weights in Eq. (6) may not create smooth enough transitions across exposures near strong edges. Such artifact may be reduced by adding extra constraints that encourage mean intensities of adjacent exposures to be used for fusion. Another solution is to extend SPD-MEF to a multi-scale scheme, which has been used previously to reduce the halo artifacts in HDR imaging and MEF [43]. On the other hand, the selection of the reference image is important for SPD-MEF to deliver satisfactory deghosting results, as in many HDR reconstruction and MEF methods [18], [43], [64]. In certain extreme cases, the moving objects in the reference image are under-/over-exposed, and their structures cannot be properly inferred from either IMF or other exposures. As a result, ghosting artifacts would appear (shown in Fig. 7(g)). The problem is common to most existing deghosting schemes and innovative methods such as image inpainting [65] may come into play.

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel structural patch decomposition (SPD) approach for MEF. Different from most pixel-wise MEF methods, SPD-MEF works on color image patches directly by decomposing them into three conceptually



(a) Source image sequence by courtesy of Orazio Gallo [37]

(b) Hu13 (c) Qin15 (d) Sen12

(e) Oh15 (f) Li14 (g) SPD-MEF

Fig. 9. Comparison of SPD-MEF with Hu13 [20], Qin15 [21], Sen12 [18], Oh15 [23], and Li14 [43]. The result of Oh15 [23] is generated without human intervention for fair comparison.

independent components and by processing each component separately. As a result, SPD-MEF generates little noise in the weighing map and makes better use of color information during fusion. Furthermore, reliable deghosting performance is achieved by using the direction information of the structure vector. Comprehensive experimental results demonstrated that SPD-MEF produces MEF images with sharp details, vivid color appearance and little ghosting artifacts while maintaining a manageable computational cost.

The proposed SPD approach is essentially dynamic range independent. Therefore, it would be interesting to explore its potential use in HDR reconstruction to generate high quality HDR images with little ghosting artifacts. Moreover, the application of SPD is not limited to MEF. As a generic signal processing approach, SPD has been found to be useful in image quality assessment of contrast-changed [66] and stereoscopic images [67]. It is worth considering whether SPD offers any insights that can be transferred to other image processing applications. In addition, although objective quality models for MEF algorithms begin to emerge, the models for objectively comparing MEF algorithms for dynamic scenes are largely lacking. Therefore, it is demanding to switch the focus from developing MEF algorithms for dynamic scenes to developing such objective quality models in order to conduct a fair comparison.

Acknowledgment

The authors would like to thank Dr. Zhengguo Li, Dr. Mingli Song and Dr. Bo Gu for providing the fusion results for comparison, Zhengfang Duanmu and the anonymous reviewers for constructive comments. This work was supported in part by the NSERC Discovery Grant and Steacie Memorial Fellowship programs of Canada, the China NSFC project under contract 61373114, 61661166011 and 61672128, and the Hong Kong RGC General Research Fund (no. PolyU 5313/13E). Kede Ma was also supported by the CSC.

References

[1] K. Ma and Z. Wang, "Multi-exposure image fusion: A patch-wise approach," in *IEEE International Conference on Image Processing*, pp. 1717–1721, 2015.

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

[3] M. D. Grossberg and S. K. Nayar, "Determining the camera response from images: What is knowable?," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1455–1467, 2003.

[4] J.-Y. Lee, Y. Matsushita, B. Shi, I. S. Kweon, and K. Ikeuchi, "Radiometric calibration by rank minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 144–156, 2013.

[5] 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, 2015.

[6] P. J. Burt, *The pyramid as a structure for efficient computation*. Springer, 1984.

[7] 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, 2015.

[8] 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, 2012.

[9] 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, 2009.

[10] Z. Li, J. Zheng, and S. Rahardja, "Detail-enhanced exposure fusion," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4672–4676, 2012.

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

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

[13] S. Li and X. Kang, "Fast multi-exposure image fusion with median filter and recursive filter," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 626–632, 2012.

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

[15] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[16] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures," *Journal of Graphics Tools*, vol. 8, no. 2, pp. 17–30, 2003.

[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[18] 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, p. 203, 2012.

[19] J. Hu, O. Gallo, and K. Pulli, "Exposure stacks of live scenes with hand-held cameras," in *European Conference on Computer Vision*, pp. 499–512, 2012.

[20] 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*, pp. 1163–1170, 2013.

[21] 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, 2015.

[22] 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, 2014.

[23] 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, 2015.

[24] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[25] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *IEEE International Conference on Computer Vision*, pp. 173–182, 1993.

[26] 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, 2014.

[27] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *IEEE International Conference on Computer Vision*, pp. 839–846, 1998.

[28] 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, 2013.

[29] E. S. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Transactions on Graphics*, vol. 30, no. 4, p. 69, 2011.

[30] W. Zhang and W.-K. Cham, "Gradient-directed multiexposure composition," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2318–2323, 2012.

[31] Z. Li, J. Zheng, Z. Zhu, W. Yao, and S. Wu, "Weighted guided image filtering," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 120–129, 2015.

[32] 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, 2012.

[33] R. Shen, I. Cheng, and A. Basu, "QoE-based multi-exposure fusion in hierarchical multivariate gaussian CRF," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2469–2478, 2013.

[34] M. Bertalmio and S. Levine, "Variational approach for the fusion of exposure bracketed pairs," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 712–723, 2013.

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

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

[37] O. Gallo, N. Gelfand, W.-C. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *IEEE International Conference on Computational Photography*, pp. 1–7, 2009.

[38] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[39] A. Eden, M. Uyttendaele, and R. Szeliski, "Seamless image stitching of scenes with large motions and exposure differences," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 2498–2505, 2006.

[40] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, "Non-rigid dense correspondence with applications for image enhancement," *ACM Transactions on Graphics*, vol. 30, no. 4, pp. 1–10, 2011.

[41] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, "The generalized patchmatch correspondence algorithm," in *European Conference on Computer Vision*, pp. 29–43, 2010.

[42] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 319–325, 2003.

[43] 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, 2014.

[44] 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, 2011.

[45] K. Jacobs, C. Loscos, and G. Ward, "Automatic high-dynamic range image generation for dynamic scenes," *IEEE Computer Graphics and Applications*, no. 2, pp. 84–93, 2008.

[46] F. Pece and J. Kautz, "Bitmap movement detection: HDR for dynamic scenes," in *IEEE Conference on Visual Media Production*, pp. 1–8, 2010.

[47] E. A. Khan, A. Akyiiz, and E. Reinhard, "Ghost removal in high dynamic range images," in *IEEE International Conference on Image Processing*, pp. 2005–2008, 2006.

[48] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[49] HDR photography gallery samples [Online]. Available: http://www.easyhdr.com/examples, 2016.

[50] Dani Lischinski HDR webpage [Online]. Available: http://www.cs.huji.ac.il/~danix/hdr/pages/belgium.html, 2016.

[51] HDR projects-software [Online]. Available: http://www.projects-software.com/HDR, 2016.

[52] Martin Čadík HDR webpage [Online]. Available: http://cadik.posvete.cz/tmo, 2016.

[53] HDRsoft gallery [Online]. Available: http://www.hdrsoft.com/gallery, 2016.

[54] Chaman Singh Verma HDR webpage [Online]. Available: http://pages.cs.wisc.edu/~csverma/CS766_09/HDRI/hdr.html, 2016.

[55] MATLAB HDR webpage [Online]. Available: http://www.mathworks.com/help/images/ref/makehdr.html, 2016.

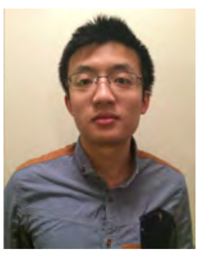
[56] HDR Pangeasoft [Online]. Available: http://pangeasoft.net/pano/bracketeer/, 2016.

[57] Enfuse HDR webpage [Online]. Available: http://www.photographers-toolbox.com/products/lrenfuse.php, 2016.

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

[59] N. D. Bruce, "Expoblend: Information preserving exposure blending based on normalized log-domain entropy," *Computers & Graphics*, vol. 39, pp. 12–23, 2014.

[60] K. Zeng, K. Ma, R. Hassen, and Z. Wang, "Perceptual evaluation of multi-exposure image fusion algorithms," in *6th International Workshop on Quality of Multimedia Experience*, pp. 27–28, 2014.
[61] Photomatrix. commercially-available HDR processing software. [Online]. Available: http://www.hdrsoft.com/, 2015.
[62] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, pp. 369–378, 1997.
[63] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 646–653, 2006.
[64] J. Zheng, Z. Li, Z. Zhu, S. Wu, and S. Rahardja, "Hybrid patching for a sequence of differently exposed images with moving objects," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5190–5201, 2013.
[65] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.
[66] S. Wang, K. Ma, H. Yeganeh, Z. Wang, and W. Lin, "A patch-structure representation method for quality assessment of contrast changed images," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2387–2390, 2015.
[67] J. Wang, S. Wang, K. Ma, and Z. Wang, "Perceptual depth quality in distorted stereoscopic images," *IEEE Transactions on Image Processing*, 2016.



Kede Ma (S'13) received the B.E. degree from University of Science and Technology of China, Hefei, China, in 2012 and M.A.Sc. degree from University of Waterloo, Waterloo, ON, Canada, where he is currently working toward the Ph.D. degree in electrical and computer engineering. His research interests lie in perceptual image processing and computational photography.

Hui Li (S'16) received the B.E. and Master degrees from Shenyang Aerospace University in 2012 and Beihang University in 2015, respectively. He is currently working toward the Ph.D. degree in the Department of Computing, Hong Kong Polytechnic University. His research interests include HDR imaging and multi-exposure image fusion.

Hongwei Yong received the B.Sc. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 2013 and 2016, respectively. His current research interests include low-rank modeling, background subtraction, and video analysis.

Zhou Wang (S'99-M'02-SM'12-F'14) received the Ph.D. degree from The University of Texas at Austin in 2001. He is currently a Professor in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image processing, coding, and quality assessment; computational vision and pattern analysis; multimedia communications; and biomedical signal processing. He has more than 100 publications in these fields with over 30,000 citations (Google Scholar).

Dr. Wang serves as a Senior Area Editor of IEEE Transactions on Image Processing (2015-present), and an Associate Editor of IEEE Transactions on Circuits and Systems for Video Technology (2016-present). Previously, he served as a member of IEEE Multimedia Signal Processing Technical Committee (2013-2015), an Associate Editor of IEEE Transactions on Image Processing (2009-2014), Pattern Recognition (2006-present) and IEEE Signal Processing Letters (2006-2010), and a Guest Editor of IEEE Journal of Selected Topics in Signal Processing (2013-2014 and 2007-2009). He is a Fellow of Canadian Academy of Engineering, and a recipient of 2015 Primetime Engineering Emmy Award, 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, 2013 IEEE Signal Processing Magazine Best Paper Award, 2009 IEEE Signal Processing Society Best Paper Award, 2009 Ontario Early Researcher Award, and ICIP 2008 IBM Best Student Paper Award (as senior author).

Deyu Meng received the B.Sc., M.Sc., and Ph.D. degrees in 2001, 2004, and 2008, respectively, from Xian Jiaotong University, Xian, China. He is currently a Professor with the Institute for Information and System Sciences, School of Mathematics and Statistics, Xian Jiaotong University. From 2012 to 2014, he took his two-year sabbatical leave in Carnegie Mellon University. His current research interests include self-paced learning, noise modeling, and tensor sparsity.

Lei Zhang (M'04-SM'14) received his B.Sc. degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, P.R. China, and M.Sc. and Ph.D. degrees in Control Theory and Engineering from Northwestern Polytechnical University, Xi'an, P.R. China, respectively in 1998 and 2001, respectively. From 2001 to 2002, he was a research associate in the Department of Computing, The Hong Kong Polytechnic University. From January 2003 to January 2006 he worked as a Postdoctoral Fellow in 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. Since July 2015, he has been a Full Professor in the same department. His research interests include Computer Vision, Pattern Recognition, Image and Video Processing, and Biometrics, etc. Prof. Zhang has published more than 200 papers in those areas. As of 2016, his publications have been cited more than 20,000 times in the literature. Prof. Zhang is an Associate Editor of IEEE Trans. on Image Processing, SIAM Journal of Imaging Sciences and Image and Vision Computing, etc. He is a "Web of Science Highly Cited Researcher" selected by Thomson Reuters. More information can be found in his homepage http://www4.comp.polyu.edu.hk/~cslzhang/.