IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 20, 2020, accepted February 12, 2020, date of publication February 24, 2020, date of current version March 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975896

A Novel Patch-Based Multi-Exposure Image Fusion Using Super-Pixel Segmentation

SHUPENG WANG AND YAO ZHAO
College of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an 710054, China
Corresponding author: Shupeng Wang (wang.shupeng@hotmail.com)

ABSTRACT A novel multi-exposure image fusion method is proposed for solving the problems of color distortion and detail loss through adaptive image patch segmentation. First, we use the super-pixel segmentation approach to divide the input images into the non-overlapping image patches composed of pixels with similar visual properties. Then, the image patches are decomposed into three independent components: signal strength, image structure and intensity. The three components are fused separately based on characteristics of human vision system and exposure level of input image. While, guided filtering is used to remove the blocking artifacts caused by patch-wise processing. In contrast to the existing methods which use fixed-size patches, the proposed method avoids blocking effect and preserves the color attribute of the input images. The experimental results show that the proposed method has advantages both in subjective and objective evaluation over the state-of-the-art multi-exposure fusion methods.

INDEX TERMS Multi-exposure image fusion, super-pixel segmentation, structural patch decomposition, guided filtering.

I. INTRODUCTION

The dynamic range of the natural scene is much larger than that of images captured by ordinary consumptive cameras [1]. The difference between the two dynamic ranges makes it is difficult to retain all the content of natural scene in a single image. In both under-exposed and over-exposed regions of images, a lot of details are lost. There are two solutions to this problem: high dynamic range (HDR) imaging [2] and multi-exposure image fusion (MEF) [3]. HDR imaging usually consists of two main steps: HDR reconstruction and tone mapping [4]. Firstly, multiple low dynamic range (LDR) images with different exposure levels in the same scene are taken, and then the HDR image is reconstructed by inverting the camera response function (CRF). Finally, in order to display on ordinary equipment, HDR image must be converted to LRD image by tone mapping. HDR imaging technology can recover the whole dynamic range of the scene and make all the details visible in a signal image. However, the estimation of CRF itself is a difficult problem [5]. MEF provides a more efficient alternative which can directly generate high quality LDR images without intermediate HDR images [6]. MEF takes a sequence of images with different exposure levels as inputs and synthesizes the fused image that is more informative and perceptually appealing than any of the input images [7].

Most of the existing MEF methods generate a fused image by weighted average of input images with different exposure levels, where one of the key issues is to design a proper weighting scheme. An intuitive idea is that the clearer pixels are assigned the greater weight. The different measures of pixel quality have been proposed based on different assumption to calculate appropriate weights. However, weights obtained based on a single pixel are susceptible to noise and easy to produce visual artifacts in the fused image. In recent years, patch-based MEF methods have attracted more attention [8]. These methods divide multi-exposure images into fixed-size rectangular patches, and perform image fusion process patch-wisely. In contrast to most pixel-wise MEF methods, the patch-based MEF methods can improve color fidelity of the fused image [9]. For the patch-based MEF methods, the division of image patches is an important issue. The MEF methods using fixed-size patches tend to cause the problems such as color distortion and detail loss in the fused image. However, few researchers studied the impact of patch division on the quality of the fused image.

To address this issue, we propose a novel patch-based multi-exposure image fusion method. Different from the existing methods which use fixed-size image patches, we utilize a super-pixel segmentation approach to divide the

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

input images into non-overlapping image patches composed of pixels with similar visual properties. First, the best-exposed image is selected as the reference image from the input multi-exposure images, and a super-pixel segmentation approach [10], [11] is performed on the reference image to obtain the division of image patches. All of the input images use this division to obtain non-overlapping image patches. Then, the image patches are decomposed into three independent components: signal strength, image structure and intensity. The different fusion rules are designed according to the characteristics of these three components. In addition, in order to ensure the spatial consistency of the fused image, guided filtering [12] is used to refine the weight maps, signal strength and intensity component. Three merged components are constructed by using the refined weight maps. Finally the final fused image is reconstructed according to the merged component.

To the best of our knowledge, the paper by Li *et al.* [8] is the most similar work to our method. Their method selects the optimal patch size based on the texture entropy of the input images. However, the shape and size of the image patches are still fixed in a fusion process. In contrast, the patch division approach used in the proposed method can change the shape of the image patches to fit the image content. The proposed fusion method brings two main advantages: 1) The patch division takes into account the characteristics of the image content. As a result, the proposed method avoids the blocking artifacts which are often generated by the patch-based MEF methods. 2) The weight map is calculated patch-wisely based on characteristics of human vision system (HSV) and exposure level of the input images. Therefore, the color attribute of input images is preserved as much as possible. We evaluate the proposed method by comparing with 7 MEF methods on 24 sets of multi-exposure image sequences. The experimental results show that the proposed method produces better fused images.

The rest of the paper is organized as follows. In section 2, the existing MEF methods are briefly reviewed. Section 3 describes the proposed method in detail. The experimental results are discussed and analyzed in section 4. Finally, a conclusion is given in section 5.

II. RELATED WORK

The MEF methods can be classified in two categories: transform domain- and spatial domain-based fusion [8]. The main processes of fusion method based on transformation domain are as follows: First, input images are transformed into the transform domain. Then, the fused coefficients are obtained by applying the fusion rules on the coefficients of input images. Finally, the fused image is reconstructed by inverse transformation. The transformation methods that are often used in MEF include pyramid transform [13], wavelet transform [14], nonsubsampled contourlet transform [15]. However, this kind of methods may produce serious color distortion in the fused image [16]. The pixel-based methods evaluate the quality of each pixel in input images and incorporate the best quality pixels at the same location into the fused image. Song *et al.* [17] proposed an image fusion method by integrating locally adaptive scene detail capture. This method firstly calculates the brightness, contrast and gradient of input images, and then synthesizes the fused image by using a probabilistic model that suppresses reversals in the image luminance gradients. Gu *et al.* [18] proposed a gradient field based image fusion method. This method takes into account that the human visual system is sensitive to contrasts between pixel intensities, not the absolute values. The gradient values of fused images are obtained by maximizing the structure tensor and the fused image is derived from the fused gradient field. Li and Kang [19] proposed a weighted sum based multi-exposure image fusion method. Firstly, this method constructs the weight map using three image features: contrast, brightness and color dissimilarity; then the weight maps were refined by recurrent filtering; finally, the fused image was obtained by weighted sum of the input images. Pixel-based image fusion method is simple and easy to implement. However, these methods only consider single pixel and ignore the relationship between adjacent pixels, which tend to produce the visual artifacts in the fused images [20]. To solve this problem, the patch-based methods design fusion rule by considering image patches instead of single pixel. Goshtasby [21] proposed a fusion method that produces a fused image with maximum information content. This method partitioned the input images into fixed-size image patches and used information entropy to evaluate the amount of information in image patches. The selected images are then blended based on weight map. In the fusion method proposed by Zhang *et al.* [9], a contrast criterion is introduced to measure the quality of exposure and generate weight maps. The fused image is obtained by merging the input images based on weighted average scheme. Ma *et al.* [22] proposed a structural patch decomposition (SPD) approach for MEF. This method decomposes an image patch into three conceptually independent components, and then fuse these three components separately. The fused image is reconstructed using the three fused component. Huang [23] improved the SPD method by designing new fusion scheme based on the quality of image patches. In the above patch-based fusion methods, the size or shape of image patches are fixed to facilitate subsequent processing. However, this results in image patches containing pixels with different color and brightness characteristics. If the same fusion schemes are used to fuse these pixels with different characteristics, the color or detail information of the fused image tend to be lost.

III. MULTI-EXPOSURE FUSION METHOD

In this section, the MEF method based on adaptive patch segmentation is presented in detail. Fig.1 shows the framework of the proposed method. Assume that all input images are already registered. Firstly, the selected reference image is partitioned into non-overlapping image patches by super-pixel segmentation, and the segmentation results are applied to all other input images. Then, each image patch is decom-

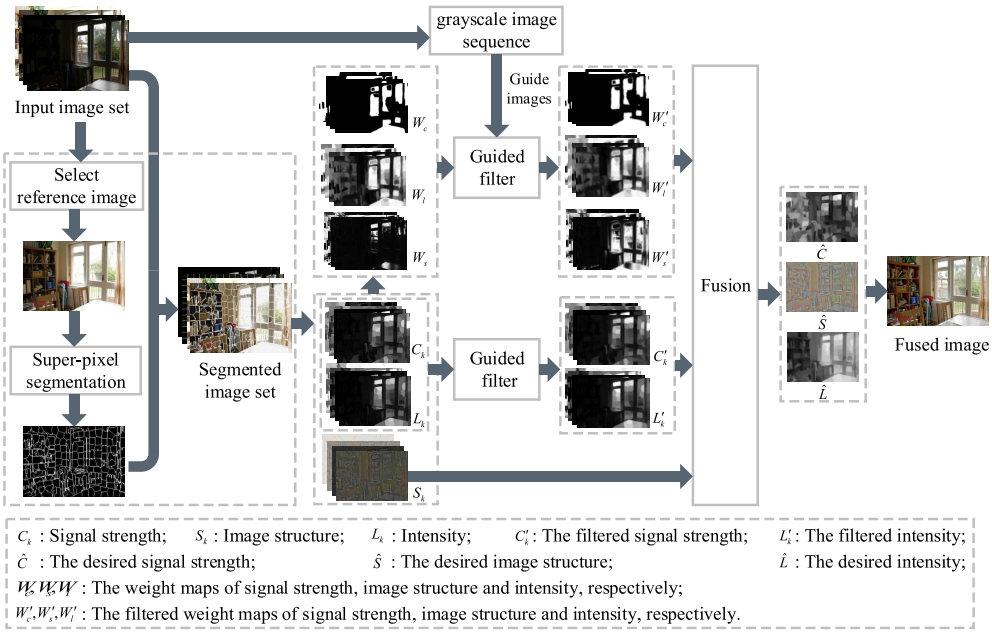


FIGURE 1. Framework of the proposed MEF method.

posed into three independent components by structure patch decomposition: signal strength, image structure and intensity. According to the different characteristics of each component, we design the corresponding fusion rules and use these rules to obtain the fused components. Finally, the fused image is reconstructed from three fused components.

A. ADAPTIVE IMAGE PATCH SEGMENTATION

The super-pixel segmentation approaches can divide the image into irregular image patches composed of pixels with similar visual properties. To achieve the division of the input images, we first select a reference image from the input multi-exposure images. In this paper, the image with the least number of under/over-exposure pixels among all the input images is used as the reference image X_{ref}. Then, the reference image X_{ref} is divided into non-overlapping N image patches by using super-pixel segmentation approach SLIC (Simple Linear Iterative Clustering) [24], which is described as follows:

$$\{P_1, \cdots, P_n, \cdots P_N\} = Slic\left(X_{ref}\right). \tag{1}$$

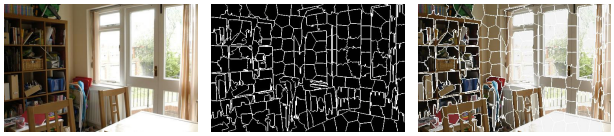
where $Slic(\cdot)$ denotes the super-pixel segmentation operation. P_n represents the spatial location of the n-th image patch. In order to ensure that the segmentation of all input images is consistent, we directly apply the segmentation result of the reference image X_{ref} to the other input images, as shown in Fig. 2.

B. STRUCTURAL PATCH DECOMPOSITION

For all the image patches $(\{X_k(P_n)\}_{k=1,\cdots,K,n=1,\cdots,N})$ obtained by segmentation, we decompose them into three independent components: signal strength, image structure and intensity, where $X_k(P_n)$ denotes the n-th image patch of the k-th input image. For convenience, P_n is omitted in subsequent parts of this paper because the image patches in different spatial location are treated independently. The image patches at the same position of the k-th input source image are directly represented by x_k. Therefore, the patch decomposition can be defined as follows [22]:

$$\begin{aligned} x_k &= \left\|x_k - \mu_{x_k}\right\| \cdot \frac{x_k - \mu_{x_k}}{\left\|x_k - \mu_{x_k}\right\|} + \mu_{x_k} \\ &= c_k \cdot s_k + l_k. \end{aligned} \tag{2}$$

where, $\|\cdot\|$ denotes the l_2 norm of a vector, μ_{x_k} is the mean value of the image patch x_k. c_k, s_k and l_k represent the signal strength, image structure, and intensity component, respectively.



(a) Reference image (b) Mask (c) Segmented image

FIGURE 2. Super-pixel segmentation.

C. WEIGHT MAP CONSTRUCTION

1) SIGNAL STRENGTH

Signal strength is defined as $c_k = \|x_k - \mu_{x_k}\|$, which represents the l_2 norm of the image patch x_k. c_k can be regarded as a measure of the amount of information which x_k contains. The larger the c_k, the more information the image patch x_k contains. In order to preserve the information of the input images in the fused image as much as possible, the fusion

rule of "winner-take-all" is adopted for the signal strength component. The weigh map of signal strength component is defined as follows::

$$w_{c,k} = \begin{cases} 1 & if\ c_k = \max\limits_{1\le k\le K} c_k \\ 0, & otherwise \end{cases} \tag{3}$$

where $w_{c,k}$ denotes the weight map that determines the contribution of the signal strength component of the k-th input image to that of the fused image patch.

2) IMAGE STRUCTURE

Image structure component contains details of texture and structure of image patches. In order to preserve the details of the input images in the fused image as much as possible, we need to evaluate the richness of the perceptible detail contained in the image patches of the different input images and assign a larger weight to the image patch containing more detail. JND refers to the minimum amount of change that can be perceived by HSV [25], [26]. We adopt a JND model to measure signal strength from the perspective of HSV. The perceptible strength of image details can be regarded as saliency weight $J_k\,(i,j)$ which is defined as [23], [27]:

$$\begin{aligned} J_k\,(i,j) &= J_k^l\,(i,j) + J_k^t\,(i,j) \\ &\quad - K_{l,t}\,(i,j) \min\left(J_k^l\,(i,j), J_k^t\,(i,j)\right). \end{aligned} \tag{4}$$

where $J_k^l\,(i,j)$ and $J_k^t\,(i,j)$ represent the luminance weight and texture weight at position (i,j),respectively. $K_{l,t}\,(i,j)$ describes the overlap effect of the weight with values ranging from 0 to 1. The relation between the luminance weight $J_k^l\,(i,j)$ and background luminance is modeled with two parts, i.e., $J_k^l\,(i,j)$ is modeled as root function of average background luminance for low background luminance and as linear function in other case, which is defined as [23], [28]:

$$J_k^l\,(i,j) = \begin{cases} 17\left(1 - \sqrt{\dfrac{\overline{x_k\,(i,j)}}{127}}\right) + 3, & if\ \overline{x_k\,(i,j)} \le 127 \\ \dfrac{3}{128}\left(\overline{x_k\,(i,j)} - 127\right) + 3, & otherwise \end{cases} \tag{5}$$

$$\overline{x_k\,(i,j)} = \frac{1}{32}\sum_{m}^{5}\sum_{n}^{5} x_k\,(i-3+m, j-3+n)\cdot B\,(m,n) \tag{6}$$

$$B\,(m,n) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 0 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 2 & 1 & 1 \end{bmatrix} \tag{7}$$

where $B(m,n)$ is a low-pass filter, $\overline{x_k\,(i,j)}$ denotes the background luminance. Texture weight $J_k^t\,(i,j)$ is usually calculated by local spatial gradients. In this paper, we calculated the gradients in four directions and choose the strongest gradient as the texture weight $J_k^t\,(i,j)$ which is written as [23], [28]:

$$\begin{aligned} J_k^t\,(i,j) &= \max_{h=1,2,3,4}\left\{\left|grad_h\,(i,j)\right|\right\} \\ grad_h\,(i,j) &= x_k\,(i,j) \otimes g_h\,(i,j) \end{aligned} \tag{8}$$

where $\otimes$ is a convolution operator. $g_h\,(i,j)$ is a high-pass filter in the h-th direction($h = 1,2,3,4$), which is defined as follow [23], [28]:

$$g_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 8 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -3 & -8 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$g_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 8 & 3 & 0 & 0 \\ 1 & 3 & 0 & -3 & -1 \\ 0 & 0 & -3 & -8 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}$$

$$g_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 8 & 0 \\ -1 & -3 & 0 & 3 & 1 \\ 0 & -8 & -3 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}$$

$$g_4 = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 8 & 0 & -8 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix} \tag{9}$$

On the other hand, pixels with well-exposure should be given larger weight because they contain more meaningful information for HVS than the over-exposure and under-exposure pixels. In this paper, we use a Gaussian function to construct the exposure weight, which is defined as follows:

$$E_k\,(i,j) = \exp\left(-\frac{\left(x_k^{gray}\,(i,j) - 0.5\right)^2}{2\sigma^2}\right) \tag{10}$$

where $x_k^{gray}\,(i,j)$ denotes the gray value of the pixel $x_k(i,j)$. σ is standard deviation, which describes the degree of spread of x_k^{gray}. σ is an empirical value which is set to 0.2 in all experiments.

In summary, weights of the structure components are constructed by saliency weight $J_k(i,j)$ and exposure weight $E_k(i,j)$

$$w_{s,k}\,(i,j) = J_k(i,j)\cdot E_k(i,j) \tag{11}$$

3) INTENSITY

The weight of intensity component is calculated based on the global mean intensity μ_k of the input image X_k and the local mean intensity l_k of image patch x_k When the average intensity of an image patch is close to the middle value of intensity range, it is considered to have a good exposure and assigned a large weight. Otherwise, it is regarded as

the under-exposure or over-exposure patch and given a small weight. The weight map for intensity component is defined as follow [6, 8, 22]:

$$w_{l,k} = \exp\left(-\frac{(\mu_k - 0.5)^2}{2\sigma_g^2} - \frac{(l_k - 0.5)^2}{2\sigma_l^2}\right) \quad (12)$$

where σ_g and σ_l are standard deviation, which control the spreads along μ_k and l_k dimensions, respectively. This method assumes that all pixels in image have the same desired intensity, i.e. the middle value of intensity range 0.5. In fact, bright areas and dark areas have different intensity. Therefore, we adopt an adaptive expected intensity, i.e.

$$p_g = \alpha \times 0.5 + (1 - \alpha) \times \frac{X_D + X_B}{2} \quad (13)$$

$$p_l(i,j) = \beta \times 0.5 + (1 - \beta) \times \frac{x_D(i,j) + x_B(i,j)}{2} \quad (14)$$

where p_g and p_l are the global expected intensity and the local expected intensity, respectively. X_D and X_B are the mean values of the darkest and brightest input images respectively. x_D and x_B represent the mean values of the darkest and brightest image patches containing pixel (i, j) respectively. α and β controls the tradeoff between prior desired intensity and the average intensity from the input image. In this paper, α and β are set to 0.5. The modified weighting function is defined as follows:

$$w_{l,k}(i,j) = \exp\left(-\frac{(\mu_k - p_g)^2}{2\sigma_g^2} - \frac{(l_k - p_l(i,j))^2}{2\sigma_l^2}\right) \quad (15)$$

D. REFINEMENT

The weight maps of the k-th input image is constructed by combining the weight maps of all image patches of the k-th input image, which is denoted as $W_{c,k}$,$W_{s,k}$,$W_{l,k}$, respectively. Similarly, three component images of the input image are written asC_k,S_k,L_k, respectively. In order to ensure spatial consistency, we refine the signal strength, intensity component and weight maps of input images with guided filter, respectively. It is worth to notice that the structural components are not filtered since the structural components contain a lot of detail of image patches and the filter operator may cause the loss of fine-detail.

Guided filter is an edge-preserving filter which can smooth images without blurring edges. We use the grayscale version of the input images as the guide images. The refined weight maps can be written as:

$$W'_{y,k} = GF\left(X_k^{gray}, W_{y,k}, r, \varepsilon\right), \; y = \{c, s, l\} \quad (16)$$

where $GF(\cdot,\cdot,\cdot,\cdot)$ represents the guided filter operation. $W_{y,k}(y = \{c, s, l\})$ is the weight map of each component of the k-th image. $W'_{y,k}$ is the refined weight map. r is the size of the filter window. ε denotes the regularization parameter which determines the blur degree of the filter. According to the experiments, $r = 3$ and $\varepsilon = 0.15$ are taken in this paper.

Similarly, the refined signal strength C'_k and intensity components L'_k are obtained using guild filter, respectively:

$$C'_k = GF\left(X_k^{gray}, C_k, r, \varepsilon\right) \quad (17)$$

$$L'_k = GF\left(X_k^{gray}, L_k, r, \varepsilon\right) \quad (18)$$

E. FUSION

According to the refined weight maps, refined signal strength, refined intensity components and image structure of the input image, the signal strength $\hat{C}$, image structure $\hat{S}$, and intensity components $\hat{L}$ of the fused image are calculated as follows, respectively.

$$\begin{cases} \hat{C} = \sum_{k=1}^{K} W'_{c,k} C' / \sum_{k=1}^{K} W'_{c,k} \\ \hat{S} = \frac{\bar{S}}{\|\bar{S}\|} \; and \; \bar{S} = \sum_{k=1}^{K} W'_{s,k} S_k / \sum_{k=1}^{K} W'_{s,k} \\ \hat{L} = \sum_{k=1}^{K} W'_{l,k} L' / \sum_{k=1}^{K} W'_{l,k} \end{cases} \quad (19)$$

The fused image $\hat{X}$ is reconstructed from the three components by:

$$\hat{X} = \hat{C} \cdot \hat{S} + \hat{L} \quad (20)$$

IV. EXPERIMENTS

We select 24 sets of multi-exposure source images to verify the performance of the proposed method. The test sets include various scenes such as day and night, indoor and outdoor setting, as listed in Table 1. All experiments are implemented in Matlab2015a on a computer with Intel Core i5, 3GHz CPU, 4GB of RAM, and Microsoft Windows 7 operating system.

A. EXPERIMENTAL RESULTS AND ANALYSIS

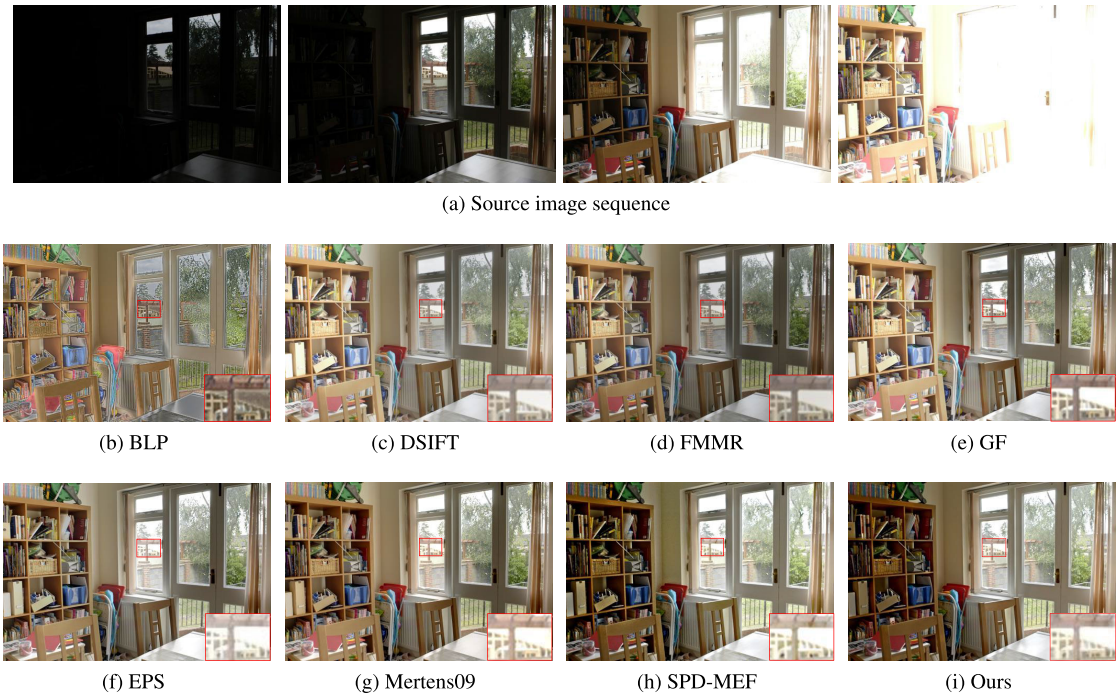
1) SUBJECTIVE ANALYSIS

In this section,we compare proposed method with seven state-of-the-art MEF methods, including BLP [28], DSIFT [29], FMMR [19], GF [30], EPS [31], Mertens09 [32], SPD-MEF [22]. For intuitive analysis of the experimental results, we select four sets of fused images for demonstration among the 24 sets of fused images.

The experimental results of different methods for "House" image sequence are shown in Fig. 3. As we can see, the fused image obtained by BLP method suffers from color distortions and sudden intensity changes. The fused image obtained by FMMR has the brightness inversion, i.e., the bookshelf is brighter than the region outside the window. DSIFT and GF methods produce obvious color distorted as shown in Fig.3(c) and (e). The two chairs of the same color in source images exhibit distinctly different colors in the fused images. The fused image by EPS method loses the color information and local details of the over-exposure area outside the window. In the fused images obtained by Mertens09 and SPD-MEF, the overall appearance is good, but the details of the scene outside the window are blurred. The fused image of the

TABLE 1. Information about input image sequences.

Image set	Size	Image origin	Image set	Size	Image origin
Arno	$339 \times 521 \times 3$	Kede Ma [33]	Mask	$800 \times 1200 \times 3$	HDRsoft [35]
Cafe	$247 \times 371 \times 3$	Fei Kou [6]	Office	$340 \times 512 \times 6$	Matlab [36]
Candle	$364 \times 512 \times 6$	Kede Ma [7]	Pool	$467 \times 700 \times 3$	Pangeasoft [34]
Cave	$384 \times 512 \times 4$	Kede Ma [7]	Preschool	$247 \times 371 \times 3$	Fei Kou [6]
Chinese garden	$340 \times 512 \times 3$	Kede Ma [7]	Room	$467 \times 700 \times 3$	Pangeasoft [34]
Gasmire	$467 \times 700 \times 3$	Pangeasoft [34]	Set	$341 \times 512 \times 3$	Kede Ma [33]
House	$500 \times 752 \times 4$	Mertens [29]	Sports Centre	$247 \times 371 \times 3$	Fei Kou [6]
Kluki	$341 \times 512 \times 3$	Kede Ma [7]	Tower	$512 \times 341 \times 3$	Kede Ma [7]
Land	$1200 \times 800 \times 3$	BLP [26]	Tree	$408 \times 271 \times 3$	Fei Kou [6]
Landscape	$341 \times 512 \times 3$	HDRsoft [35]	Venice	$341 \times 512 \times 3$	HDRsoft [35]
Laurenziana	$512 \times 356 \times 3$	Kede Ma [33]	Window	$384 \times 512 \times 3$	Kede Ma [33]
Lighthouse	$340 \times 512 \times 3$	HDRsoft [35]	Yellow hall	$339 \times 512 \times 3$	Kede Ma [33]



(a) Source image sequence

(b) BLP (c) DSIFT (d) FMMR (e) GF

(f) EPS (g) Mertens09 (h) SPD-MEF (i) Ours

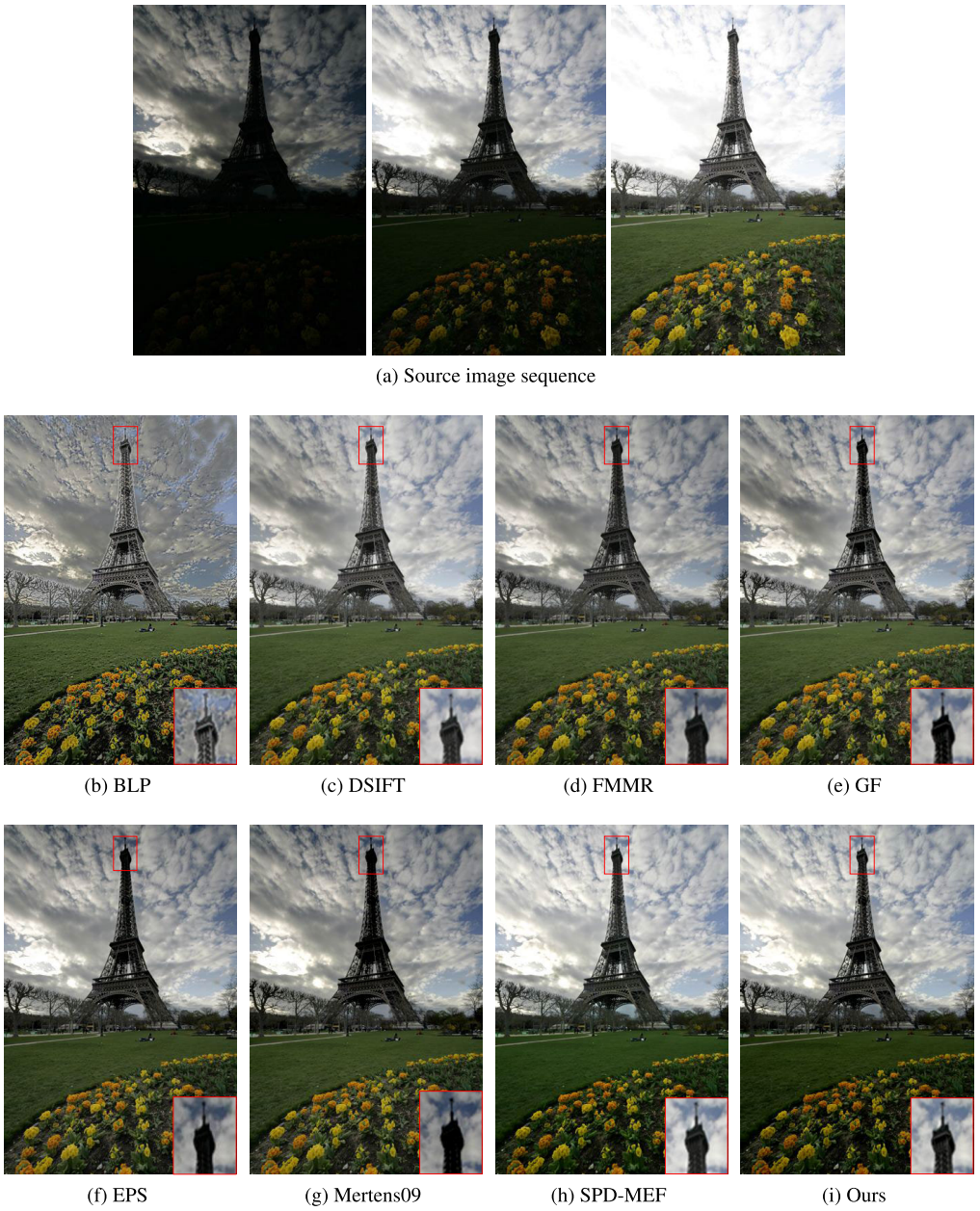
FIGURE 3. Comparison of different methods on the "House" image sequence.

proposed method appears better in detail preservation and brightness distribution than the other methods. Further, the color fidelity of the proposed method is also good.

Fig.4 shows the fused images of "Tower" obtained by the eight different methods. As can be seen from Fig.4 (b), the BLP method produces massive artifacts, especially in the cloud region. In the fused image produced by DSIFT and GF methods, the right side of the tower is obviously brighter than the left, which is inconsistent with the input images. FMMR, EPS and Mertens09 methods not only fail to preserve good contrast in the sky, but also cause color distortion in the lawn. EPS and Mertens09 lose the local detail loss of the tower. The SPD-MEF method performs excellent in color. Compared to the other seven fused results, the proposed method increases the overall contrast while preserving texture details. Besides, the overall appearance of the fused image obtained by the proposed method is appealing.

All fused images of "Chinese Garden" obtained by eight different methods are illustrated in Fig.5. The BLP method produces the image with unnatural colors and artifacts. In the fused images obtained by FMMR, GF and Mertens09, the brightness of sky is dark. In the fused result of EPS, the color is pale. Although the SPD-MEF method performs well in the color saturation and the global contrast, the local detail of the fused image is blurred. The proposed method not only contains rich details, but also achieves excellent performance in global contrast and color saturation compared with the other MEF methods. In addition, the fused image resulted from the proposed method has more natural appearance with respect to the human visual system.

Fig.6 demonstrates the performances of different methods on the "Window" image sequence. As shown in Fig.6 (b), there are obviously black shadows around the lamp. In the fused images obtained by DSIFT and GF methods, the color



(a) Source image sequence

(b) BLP (c) DSIFT (d) FMMR (e) GF

(f) EPS (g) Mertens09 (h) SPD-MEF (i) Ours

FIGURE 4. Comparison of different methods on the "Tower" image sequence.

of the bed is distorted. In the fused images by GF and Mertens09 method, the brightness of wall is too bright. The local details of the magnified area in Fig.6 (f) are blurred. In the fused image obtained by SPD-MEF method, there are obvious black artifacts around the lamp. By contrast, the proposed method better preserves the details and color information, and achieves the best overall visual effect.

2) OBJECTIVE EVALUATION

In order to quantitatively evaluate the performance of the proposed methods, three objective criteria are used. The first criterion is mutual information (MI), defined as the sum of mutual information between each source image and the fused image [37], [38]. The second criterion is correlation coefficient (CC), which measures the degree of linear correlation of the fused image and source images [39]. The third criterion is standard deviation (SD) [39], which measures the contrast in the fused image. For all three criteria, the larger the value is, the better is the image quality.

The MI value reflects the total quantity of information in the fused image which is obtained from the input source images. The comparison results of 8 different MEF methods

(a) Source image sequence

(b) BLP (c) DSIFT (d) FMMR (e) GF

(f) EPS (g) Mertens09 (h) SPD-MEF (i) Ours

FIGURE 5. Comparison of different methods on the "Chinese garden" image sequence.

(a) Source image sequence

(b) BLP (c) DSIFT (d) FMMR (e) GF

(f) EPS (g) Mertens09 (h) SPD-MEF (i) Ours

FIGURE 6. Comparison of different methods on the "Window" image sequence.

on 24 source sequences are listed in Table 2, in which the largest MI value is shown in bold. It is clear that the proposed method performs better than the other methods in most scenes. In other word, the proposed method outperforms the other methods in preserving information from the source images.

TABLE 2. Performance comparison of eight different methods according to MI.

Source sequence	BLP	DSIFT	FMMR	GF	EPS	Mertens09	SPD-MEF	Ours
Arno	4.828	3.881	3.709	3.611	4.885	4.772	7.195	**7.400**
Cafe	2.762	2.522	3.341	2.559	3.786	3.638	4.964	**5.366**
Candle	8.281	6.833	6.448	6.774	9.434	8.723	12.317	**13.557**
Cave	3.619	4.401	4.288	4.238	4.619	4.192	**4.892**	4.335
Chinese garden	3.562	4.122	4.382	3.967	4.798	4.397	**5.239**	5.030
Gasmire	2.116	1.527	2.602	2.168	3.826	2.734	4.777	**4.816**
House	2.080	2.505	2.295	2.443	3.492	3.561	5.445	**5.974**
Kluki	3.244	3.747	3.659	3.458	4.542	4.209	5.710	**6.208**
Land	4.177	5.452	5.136	5.213	4.786	4.777	6.404	**6.728**
Landscape	5.146	4.187	5.351	4.087	5.221	4.625	7.117	**7.448**
Laurenziana	3.967	4.349	4.035	4.475	4.955	4.798	5.679	**5.850**
Lighthouse	3.457	3.016	3.612	3.200	3.380	3.166	6.754	**7.148**
Mask	4.267	4.992	4.745	4.174	4.364	3.938	6.161	**6.378**
Office	5.312	4.940	6.251	5.004	8.617	7.878	10.908	**11.236**
Pool	3.521	3.882	3.706	3.818	4.736	4.583	5.337	**5.452**
Preschool	3.337	3.183	3.633	3.450	3.585	3.888	**5.594**	5.589
Room	4.375	5.09	4.915	5.028	5.274	5.013	5.887	**6.085**
Set	3.673	3.622	4.567	3.793	4.880	4.488	7.032	**7.709**
Sports Centre	3.969	3.753	3.926	3.619	3.877	4.054	5.473	**5.502**
Tower	4.231	4.624	4.717	4.356	4.015	3.939	5.882	**6.019**
Tree	4.152	4.136	4.326	4.137	5.446	5.279	**5.748**	5.561
Venice	3.372	3.503	3.674	3.126	4.430	4.119	5.497	**5.589**
Window	4.682	5.476	4.777	5.160	5.196	5.476	6.417	**6.456**
Yellow hall	3.881	2.723	3.848	3.241	4.374	4.388	7.052	**7.180**

TABLE 3. Performance comparison of eight different methods according to CC.

Source sequence	BLP	DSIFT	FMMR	GF	EPS	Mertens09	SPD-MEF	Ours
Arno	0.726	0.828	0.861	0.593	0.912	0.899	0.937	**0.942**
Cafe	0.398	0.594	0.792	0.596	0.805	0.796	0.869	**0.870**
Candle	0.331	0.566	0.615	0.378	0.618	0.618	0.633	**0.638**
Cave	0.539	0.712	0.768	0.773	0.812	0.812	0.799	**0.850**
Chinese garden	0.813	0.857	0.912	0.865	0.919	0.914	0.953	**0.957**
Gasmire	0.310	0.678	0.816	0.795	0.893	0.845	**0.908**	0.907
House	0.340	0.449	0.431	0.473	0.678	0.683	0.779	**0.785**
Kluki	0.758	0.673	0.752	0.652	0.892	0.873	0.931	**0.934**
Land	0.894	0.888	0.933	0.888	0.925	0.912	0.971	**0.973**
Landscape	0.881	0.782	0.899	0.711	0.947	0.918	0.976	**0.978**
Laurenziana	0.829	0.799	0.872	0.879	0.903	0.910	0.954	**0.957**
Lighthouse	0.608	0.486	0.626	0.444	0.816	0.797	0.922	**0.928**
Mask	0.762	0.851	0.869	0.761	0.893	0.854	0.954	**0.960**
Office	0.730	0.691	0.827	0.743	0.901	0.878	0.922	**0.925**
Pool	0.808	0.846	0.887	0.839	0.911	0.921	0.948	**0.952**
Preschool	0.730	0.762	0.859	0.807	0.848	0.872	0.920	**0.929**
Room	0.869	0.889	0.906	0.898	0.930	0.932	0.936	**0.942**
Set	0.749	0.775	0.864	0.730	0.914	0.889	0.956	**0.962**
Sports Centre	0.746	0.757	0.835	0.755	0.796	0.821	0.857	**0.870**
Tower	0.833	0.900	0.913	0.893	0.919	0.912	0.958	**0.961**
Tree	0.806	0.776	0.838	0.839	0.890	0.882	0.890	**0.901**
Venice	0.720	0.816	0.858	0.790	0.910	0.882	0.942	**0.947**
Window	0.800	0.819	0.873	0.827	0.876	0.819	0.882	**0.903**
Yellow hall	0.898	0.840	0.932	0.888	0.932	0.932	0.976	**0.977**

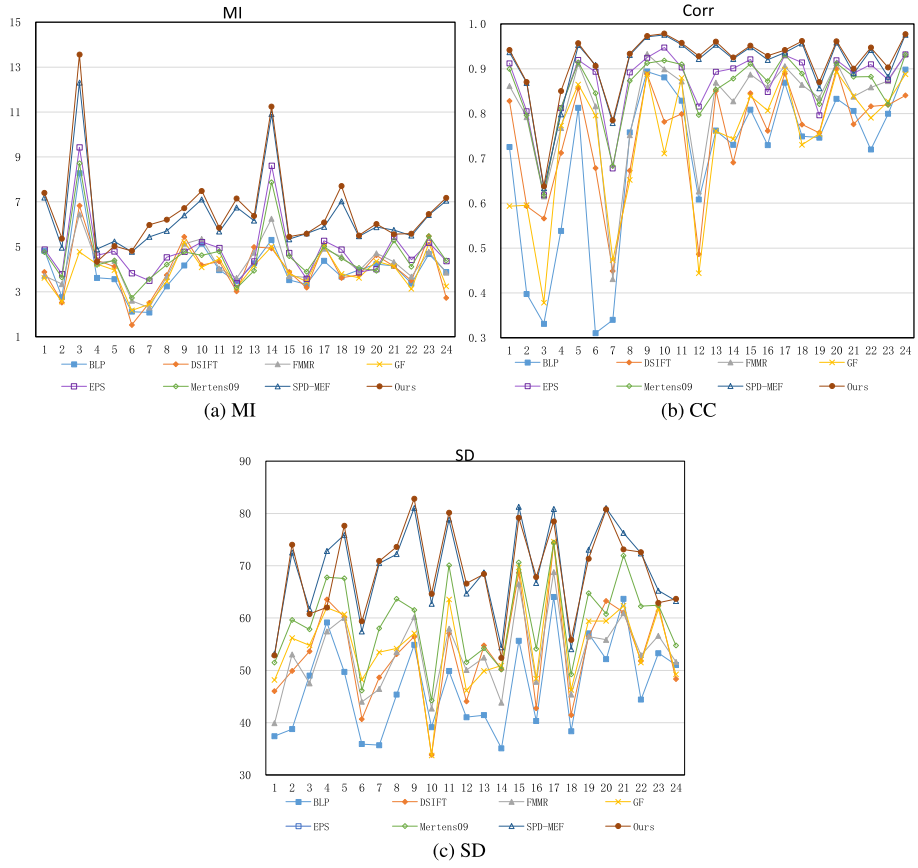
Table 3 lists the performance comparison of the proposed method with 7 other MEF methods using CC metric. The metric CC measures the similarity between the source image and the fused image, ranging from -1 to 1. The larger CC value indicates that the fused image better preserves the information in the source images. Table 3 shows that the proposed method has the best performance. The fused images obtained by BLP have poor performance since they have obvious artifact as shown in Fig. 3-6.

The comparison results of the metric SD are listed in Table 4. The proposed method shows the best performance in 13 sets of image sequences. For the rest of the image sequences, the proposed method ranks the second in most cases. Generally, the fused images obtained by the proposed method have better global contrast than that of the other methods.

In order to make it easy to compare the performance of the different MEF methods using the objective met-

TABLE 4. Performance comparison of eight different methods according to SD.

Source sequence	BLP	DSIFT	FMMR	GF	EPS	Mertens09	SPD-MEF	Ours
Arno	37.420	45.999	39.887	48.156	51.110	51.482	**53.170**	52.832
Cafe	38.762	49.864	53.025	56.179	61.446	59.658	72.525	**74.028**
Candle	49.011	53.630	47.539	54.771	59.519	57.842	**61.714**	60.789
Cave	59.168	63.575	57.439	62.000	72.501	67.779	**72.822**	62.012
Chinese garden	49.721	60.181	60.055	60.681	70.961	67.570	75.879	**77.654**
Gasmire	35.920	40.694	43.996	48.248	50.430	46.165	57.415	**59.384**
House	35.695	48.667	46.436	53.475	58.356	58.051	70.496	**70.919**
Kluki	45.380	53.077	53.501	54.176	63.436	63.652	72.208	**73.603**
Land	54.854	56.445	60.186	57.053	62.068	61.537	81.044	**82.817**
Landscape	39.145	33.841	42.697	33.702	49.237	44.266	62.728	**64.619**
Laurenziana	49.879	57.083	57.978	63.555	70.465	70.090	78.823	**80.151**
Lighthouse	41.058	44.080	50.047	46.170	52.038	51.573	64.659	**66.616**
Mask	41.421	54.750	52.460	49.859	57.728	54.204	**68.746**	68.423
Office	35.088	50.209	43.796	50.905	52.191	50.199	**54.395**	52.354
Pool	55.652	68.680	66.467	69.705	73.448	70.626	**81.282**	79.185
Preschool	40.322	42.757	47.827	48.430	54.786	54.129	66.728	**67.837**
Room	64.024	74.449	68.791	74.625	75.364	74.361	**80.862**	77.538
Set	38.372	41.472	45.357	46.288	48.488	49.204	54.045	**55.806**
Sports Centre	57.056	56.424	56.481	59.423	63.937	64.764	**73.068**	71.319
Tower	52.126	63.262	55.848	59.427	63.355	60.791	**81.019**	80.783
Tree	63.666	61.072	60.926	62.358	72.300	71.952	**76.255**	73.117
Venice	44.413	51.816	52.912	51.508	66.317	62.276	72.367	**72.619**
Window	53.282	62.447	56.620	61.754	62.828	62.447	**65.244**	62.908
Yellow hall	51.037	48.321	51.610	49.220	54.020	54.773	63.269	**63.651**



(a) MI (b) CC

(c) SD

FIGURE 7. Objective performance of eight different methods on three metrics.

rics, the line charts are given in Fig. 7. It is clearly shown in Fig. 7 that the proposed method achieves best performance in most case with respect to all three metrics.

B. IMPACT OF PATCH DIVISION

To illustrate the effect of adaptive patch division, we compare the fused results using different patch division schemes based on the proposed fusion rule. The patches of six different

TABLE 5. Average of objective indicators for different methods.

	Ours-200	Ours-29	Ours-400	Ours-21	Ours-800	Ours-15
MI	6.611	6.202	6.722	6.342	6.799	6.510
CC	0.914	0.912	0.915	0.913	0.915	0.914
SD	68.830	68.544	68.495	68.117	68.326	67.726

(a) Source image sequence

(b) Ours-200 (c) Ours-400 (d) Ours-800

(e) Ours-29 (f) Ours-21 (g) Ours-15

FIGURE 8. Comparison of different patches on the "Candle" image sequence.

sizes are used, where the fixed-sized patches are rectangle of 29×29, 21×21 and 15×15, and adaptive patches are obtained by SLIC algorithm which specify the number of super-pixels as 200, 400, and 800, respectively. The fused images obtained using fixed-sized patches are denoted as Ours-29, Ours-21 and Ours-15, respectively, and the fused images obtained using adaptive patches are Ours-200, Ours-400 and Ours-800. For convenience of comparison, the size of input images is scaled to 512×341 or 314×512. When an image is divided into 200 patches, the average size of patches is approximately equal 29×29, and so on.

Fig.8 shows the fused images of the ''Candle''. The results in Fig.8(b)-(d) are obtained using adaptive patches. The Fig.8(e)-(g) show the fused image produced using fixed-size patches. When using fixed-size patch, the obvious blocking artifacts appears in the fused image as the patch size increase, while the fused images obtained using adaptive patches have less difference. Comparing the results obtained by adaptive patch-based and fixed-size patch-based methods with similar patch size, we can see that adaptive patch-based method have the better performance than fixed-size patch-based method.

Table 5 shows the average metrics of 6 fusion methods on 24 sets of fused images. It can be seen that the smaller the image patch, the larger the MI and CC values, the smaller the SD values. When the patch size is approximately equal, all of the three metrics indicates that adaptive patch-based method outperforms fixed-size patch-based method.

V. CONCLUSION AND FUTURE WORK

In this paper, a multi-exposure image fusion method based on adaptive patch has been proposed. The proposed method uses a super-pixel segmentation approach to divide the input images into the image patches composed of pixels with similar visual properties. Then, the image patches are decomposed into three independent components: signal strength, image structure and intensity. The three components are fused using different fusion rules which are designed based on characteristics of HVS and exposure level of input images. To remove the blocking artifacts caused by patch-wise process, guided filter is performed on signal strength component, intensity component and weight maps. As a result, the proposed method generates little blocking artifacts and preserves well the color attribute of input images. The comparative experiments show that the proposed method outperforms the state-of-the-art multi-exposure fusion methods both in subjective and objective evaluation.

Although the proposed method can produce high-quality fused images, it is not suitable for real-time application. In the

future, both algorithm and implementation will be optimized to improve the efficiency of the fusion method.

ACKNOWLEDGMENT

The author thanks editors and review experts for their time and energy in this article, especially for your valuable advice.

REFERENCES

[1] F. Kou, Z. Wei, W. Chen, X. Wu, C. Wen, and Z. Li, "Intelligent detail enhancement for exposure fusion," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 484–495, Feb. 2018.

[2] Q. Yan, J. Sun, H. Li, Y. Zhu, and Y. Zhang, "High dynamic range imaging by sparse representation," *Neurocomputing*, vol. 269, pp. 160–169, Dec. 2017.

[3] 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.

[4] H.-C. Tsai, H.-J. Lin, and J.-J. Leou, "Multiexposure image fusion using intensity enhancement and detail extraction," *J. Vis. Commun. Image Represent.*, vol. 33, pp. 165–178, Nov. 2015.

[5] 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.

[6] 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.

[7] 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.

[8] Y. Li, Y. Sun, M. Zheng, X. Huang, G. Qi, H. Hu, and Z. Zhu, "A novel multi-exposure image fusion method based on adaptive patch structure," *Entropy*, vol. 20, no. 12, p. 935, Dec. 2018.

[9] W. Zhang, S. Hu, and K. Liu, "Patch-based correlation for deghosting in exposure fusion," *Inf. Sci.*, vols. 415–416, pp. 19–27, Nov. 2017.

[10] R. Malik, "Learning a classification model for segmentation," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Nice, France, Oct. 2003, pp. 10–17.

[11] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art," *Comput. Vis. Image Understand.*, vol. 166, pp. 1–27, Jan. 2018.

[12] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2012.

[13] T. Mertens, J. Kautz, and F. V. Reeth, "Exposure fusion," in *Proc. 15th Pacific Conf. Comput. Graph. Appl.*, Maui, HI, USA, 2007, pp. 382–390.

[14] M. H. Malik, S. Asif, and M. Gilani, "Wavelet based exposure fusion," presented at the World Congr. Eng., London, U.K, Jul. 2008.

[15] G. Yang, M. Li, Y. Rao, and L. Chen, "Detail-enhanced and brightness-adjusted exposure image fusion," *J. Electron. Imag.*, vol. 24, no. 3, May 2015, Art. no. 033017.

[16] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018.

[17] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, "Probabilistic exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 341–357, Jan. 2012.

[18] 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.

[19] S. Li and X. Kang, "Fast multi-exposure image fusion with median filter and recursive filter," *IEEE Trans. Consum. Electron.*, vol. 58, no. 2, pp. 626–632, May 2012.

[20] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, Jan. 2017.

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

[22] 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.

[23] F. Huang, D. Zhou, R. Nie, and C. Yu, "A color multi-exposure image fusion approach using structural patch decomposition," *IEEE Access*, vol. 6, pp. 42877–42885, 2018.

[24] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to State-of-the-Art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[25] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 742–752, Jun. 2005.

[26] H. Hadizadeh, "A saliency-modulated just-noticeable-distortion model with non-linear saliency modulation functions," *Pattern Recognit. Lett.*, vol. 84, pp. 49–55, Dec. 2016.

[27] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1648–1652, Nov. 2010.

[28] 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.

[29] 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.

[30] 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.

[31] Y. Que, Y. Yang, and H. J. Lee, "Exposure measurement and fusion via adaptive multiscale edge-preserving smoothing," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 12, pp. 4663–4674, Dec. 2019.

[32] 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, Mar. 2009, pp. 161–171.

[33] 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.

[34] *Pangeasoft*. Accessed: Nov. 18, 2019. [Online]. Available: http://pangeasoft.net/pano/bracketeer/

[35] *HDRsoft*. Accessed: Nov. 18, 2019. [Online]. Available: http://www.hdrsoft.com/examples2.html

[36] *Mathworks*. Accessed: Nov. 18, 2019. [Online]. Available: http://www.mathworks.com/help/images/ref/makehdr.html

[37] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, pp. 313–315, 2002.

[38] J. Wang, W. Wang, B. Li, G. Xu, R. Zhang, and J. Zhang, "Exposure fusion via sparse representation and shiftable complex directional pyramid transform," *Multimedia Tools Appl.*, vol. 76, no. 14, pp. 15755–15775, Aug. 2016.

[39] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion*, vol. 24, pp. 147–164, Jul. 2015.

SHUPENG WANG received the B.S. and M.S. degrees in communication and information engineering from the Xi'an University of Science and Technology, in 2000 and 2003, respectively, and the Ph.D. degree in pattern recognition and intelligent system from Xidian University, in 2009. He is currently an Associate Professor with the Xi'an University of Science and Technology. His current research interests include image processing and pattern recognition.

YAO ZHAO is currently pursuing the master's degree with the Xi'an University of Science and Technology, China. Her current research interests include multiexposure image fusion and pattern recognition.