

# Multi-exposure image fusion of dynamic scenes using consistency detection and weighted least squares filtering

Fang Xu,<sup>a,\*</sup> Jinghong Liu,<sup>a</sup> Tenglong Wang,<sup>b</sup> Xuan Wang,<sup>a</sup>  
Hui Sun,<sup>a</sup> and Hang Ren<sup>a</sup>

<sup>a</sup>Chinese Academy of Sciences, Changchun Institute of Optics, Fine Mechanics, and Physics,  
Key Laboratory of Airborne Optical Imaging and Measurement, Changchun, China

<sup>b</sup>Northeast Electric Power Design Institute Company Limited of China Power Engineering  
Consulting Group, Changchun, China

**Abstract.** The high dynamic range image obtained by multi-exposure image fusion (MEF) contains more detailed information, which has broad application prospects and crucial practical significance in many fields. However, due to the introduction of the ghosting artifacts caused by object movement and camera shake, MEF in the dynamic scenes has always been a challenge. To address this problem, we designed a deghosting method for MEF. The over- and under-exposed images are corrected by intensity mapping and consistency detection to obtain the aligned latent images. Then the high- and low-frequency components of the latent images are generated using weighted least squares filtering. The blending weights are calculated based on the image luminance and the exposedness function. These frequency components are integrated into the final deghosted image with more texture details and vivid color. A comprehensive evaluation experiment is carried out, proving that the proposed method has a better visual effect and stable performance than the state-of-the-art deghosting MEF methods. © 2023 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.OE.62.1.013103](https://doi.org/10.1117/1.OE.62.1.013103)]

**Keywords:** multi-exposure image fusion; deghosting; consistency detection; weighted least squares.

Paper 20220770G received Jul. 15, 2022; accepted for publication Jan. 5, 2023; published online Jan. 28, 2023.

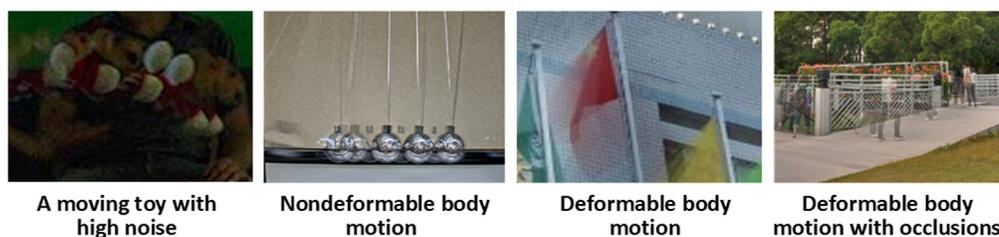
## 1 Introduction

The luminance span in natural scenes is usually large, ranging from starlight at night to dazzling sunlight, with a brightness range of nine orders of magnitude.<sup>1</sup> Restricted by the optical design parameters of the lens, the sensitivity, the full well charge of the detector, and other factors, the dynamic range of the existing imaging equipment is far lower than that of the natural scene. Therefore, it is difficult to record the details of different brightnesses in the background through one single shot. Moreover, many factors easily affect the scene, such as weather, solar altitude angle, clouds, and object characteristics. The bright object in the image is prone to be over-exposed, resulting in a severe loss of information. The dark area may be under-exposed, making the detail unrecognizable. As a result, over- or under-exposed frequently occurs, leading to unsatisfactory imaging results. Multi-exposure image fusion (MEF) technology provides a simple and efficient way to expand the dynamic range of the imaging device. A set of low dynamic range images with different exposures can be integrated into a single high dynamic range (HDR)-like image, retaining the objects' details with different reflectivities in the scene.<sup>2,3</sup>

The most static MEF approaches assume that the source images are perfectly aligned. However, it may be inaccurate in practice, and there are inevitably moving objects with various trajectories. Because of the time difference in the image acquisition, the position and shape of the moving objects are variable in the image sequence with different exposures. If these images are fused directly, some issues, such as blur, ghosts, and translucent areas, may appear in the fused result,<sup>4,5</sup> which will seriously affect the visual quality of the fused image, as shown in Fig. 1.

---

\*Address all correspondence to Fang Xu, [xufang59@126.com](mailto:xufang59@126.com)



**Fig. 1** Different types of fused results with ghosting artifacts.

Therefore, one of the biggest challenges in the current MEF research is to remove the motion blur and ghosts when there is local or global motion in between input exposures.<sup>6</sup> Many MEF studies have been developed to remove the ghosting artifacts and improve the visual quality of the fused image from a different perspective. They lie in two aspects: detecting the ghosting artifacts and eliminating them. Therefore, the current MEF algorithms in the dynamic scenes can be mainly divided into the following categories: moving object removal and moving object retention.

Some MEF approaches remove all moving objects by the static background estimation, where only a tiny part of scene contains moving objects and most parts of the image are static. No reference image is selected in this kind of method. Khan et al.<sup>7</sup> designed a deghosting MEF method and iteratively estimated the probability of each pixel belonging to the background for adjusting the weights. Similar to this iterative process, Pedone and Heikkil<sup>8</sup> increased the likelihood that the pixels belonged to the static background through an energy minimization step. Although some ghosting artifacts can be reduced, their technique cannot cope with camera shake. Zhang and Cham<sup>9</sup> detected whether there was a moving object according to the gradient direction consistency of the input image. However, the robustness may be insufficient when the scene changes frequently. Li et al.<sup>10</sup> approximated the nonuniform pixels based on the intensity mapping (IM) function and bidirectional normalization approach. They further eliminated the ghosts and improved the quality of the fused image using two-round hybrid correction. However, the acceleration of the proposed filter may be an issue. Wang et al.<sup>11</sup> measured the difference between the multi-exposure images based on visual saliency. They introduced bilateral motion detection to improve the moving region determination accuracy and avoid the ghosting artifacts in the fused results. Qu et al.<sup>12</sup> calculated the weight maps using the image's luminance and chromaticity in the YIQ color space. The image difference and superpixel segmentation were utilized to refine the weight maps and decrease the weight of moving objects to eliminate ghosts. Finally, they obtained the fused image with detail enhancement based on the improved Laplace pyramid framework. This method was suitable for the scenes with only a small number of moving objects, and the most were stationary. However, it was time-consuming and could not deal with the scenes with large motion or fast change. Shim<sup>13</sup> adopted the gamma-corrected exposure time ratio between the multi-exposure images to remove the moving object in the fused result. This algorithm was fit for removing some fast-moving objects to construct a ghost-free output. The slow-moving objects may not be eliminated because their background information was missing, and some ghosts were still visible.

The most deghosting MEF methods preserve the moving objects consistent with the reference image in the fused result. They focus on finding the local relationships between the regions affected by the moving objects and reconstructing these pixels. According to the different characteristics of the algorithms, we can divide them into the following categories.

In one class, the dynamic regions of one or more input images are selected as the guide to removing ghosts. Jacobs et al.<sup>14</sup> introduced entropy mapping to detect ghosts in the moving area with low contrast without knowing the camera curve. However, the alignment may fail when the object movements are relatively fast or the texture of the moving region is flat. Granados et al.<sup>15</sup> designed a deghosting approach based on the Markov random field and selected a reference image for guiding the dynamic content. Since the moving region only came from a single image, some translucent areas may be still introduced. Wang and He<sup>16</sup> proposed a deghosting MEF algorithm based on an improved difference strategy. Before detecting ghosts, the exposure of each source image was normalized to the brightness consistent with the reference image.

When the region's brightness was not ideal, they selected two reference images and matched other available exposure.

The second class of methods utilizes a feature matching strategy to remove the ghosts. Zimmer et al.<sup>17</sup> exploited an alignment method based on optical flow and created a super-resolution image. The dense displacement fields with subpixel accuracy were generated. This method relied on a coarse to fine warping scheme to deal with significant displacement. However, since some small objects may disappear on the rough horizontal plane, it is hard to estimate the large displacement of the small objects. Liu et al.<sup>18</sup> introduced an MEF method and used dense scale-invariant feature transform (SIFT) descriptors as the activity level metrics to remove the ghosts in the dynamic scenes, which could be used in both static and dynamic scenes. This method first constructed three weight items: local contrast, exposure quality, and spatial consistency. Then these three parts were combined to estimate a weighted map for each source image and performed weighted fusion. Although their algorithm can create some fused outputs without ghosts, it may not produce satisfactory results when the moving objects appear in multiple input images or the background moves. Based on this research, Hayat and Imran<sup>19</sup> developed an MEF approach based on dense SIFT descriptor and guided filtering. There were two main differences compared with the algorithm in Liu's method.<sup>18</sup> On the one hand, they used histogram equalization and the median filter to calculate the color dissimilarity feature replacing Liu's spatial consistency module. On the other hand, guide filtering was applied to remove the noise and the discontinuity in the initial weights. Although the ghosting artifacts of Hayat's method are fewer than that of Liu's, they are still not entirely removed in some scenes. Zhang et al.<sup>20</sup> proposed two consistencies for matching the reference image with the source images to generate the latent images: mutual consistency based on histogram matching and intraconsistency based on superpixel segmentation. They retained all details and yielded the final deghosted result by the static weighted fusion.

Other algorithms based on image patch matching detect the motion-changed region by the information between the source images. Sen et al.<sup>21</sup> introduced an energy minimization HDR deghosting approach based on image patch matching, which jointly performed the image alignment and reconstruction. It can remove ghosts very well, but the details are slightly worse. Following their work, Hu et al.<sup>22</sup> utilized the histogram and brightness mapping function to design a patch match-based approach. They generated the latent images by minimizing the energy function and solved the ghosting problem in MEF. However, some halos may be produced in some scenarios. Ma et al.<sup>23</sup> performed the pixel consistency mapping based on the directivity of the image patch structure vector. This method outperforms several state-of-the-art algorithms, and there are very slight ghosting artifacts in the dynamic scenes. After that, some research studies<sup>24-26</sup> made some improvements based on this theory, which most can generate high-quality images and overcome the blurring in the fused images for the dynamic cases. In addition, Li's study<sup>25</sup> results in successfully reduced execution time.

In recent years, with the significant success of deep learning in image processing,<sup>27,28</sup> more and more researchers have applied it to MEF, and some have tried to remove the ghosting artifacts in the dynamic scenes. Prabhakar et al.<sup>29</sup> designed a convolution and recursive structure network for the ghost removal in MEF, which can fuse image sequences with different frame numbers. Yan et al.<sup>30</sup> proposed an end-to-end network to remove the ghosts, which integrated the image context and the corresponding gradient information. However, these methods need a large number of training samples, and the training database should have real images in the dynamic scenes. In addition, the model is complex, and the generalization ability is not strong.

In this paper, we present an effective deghosting MEF method in the dynamic scenes for obtaining HDR-like images without preprocessing or motion estimation. It is mainly composed of image alignment and image fusion. First, we utilize IM and consistency detection to obtain the aligned latent images. Then we decompose the latent images based on weighted least squares (WLS) filtering and calculate their corresponding blending weight maps to get the final fused result. The main advantages of this paper are listed as follows.

1. The proposed method can preserve helpful features by selecting an appropriate reference image. The ghosting artifacts caused by the object motion and camera shake are effectively removed and it shows strong robustness to the ghosts in various dynamic scenes.

- Based on WLS filtering and optimized evaluation function derived from both global and local exposedness, the fused image preserves the details of dark and bright regions. It can reflect the complete information of the actual scene and achieve good visual effects.

In addition, a comprehensive evaluation experiment is carried out. We evaluate the performance of the proposed method and the other six dehazing MEF methods on 30 image sequences in the dynamic scenes from subjective and objective aspects. Relevant materials, including test image sequences, fused results, and relevant comparison algorithm codes, have been provided with download links to make it easier to dehazing MEF research in the future.

The remainder of this paper is organized as follows. Section 2 introduces the proposed MEF method, including image alignment and image fusion. We present a comparative study between the proposed method and the other methods in Sec. 3. Section 4 concludes the proposed method.

## 2 Proposed Method

The algorithm framework proposed in this paper is shown in Fig. 2. The proposed method can analyze the relationship between the reference and input images. We align the over- and under-exposed images through IM and consistency detection. The constraint conditions are formulated to correct the local color distortion and obtain the latent images with more accurate details. The latent image's low- and high-frequency components are extracted using WLS filtering. Combined with the brightness of the latent image, the blending weight functions are calculated. Then the final ghost-free image is generated. In the meantime, RGB's three color channels are jointly handled to make the fused image more vivid.

### 2.1 Reference Image Selection

The reference image determines the moving objects that appear in the fused result and further affects its visual quality. Input the multi-exposure image sequence  $I_k$  ( $k = 1, \dots, K$ ),  $K$  is the total number of the input images, the pixel values are first normalized to the  $[0, 1]$  range. To select the appropriate reference image, we rearrange the input images in ascending order according to their brightness and pick the one sorted in the middle as the reference image  $I_{ref1}$ , as shown in Eq. (1). In addition, all source images' overall luminance is calculated, and the image with appropriate exposure is chosen as the reference image  $I_{ref2}$ , which contains relatively few over- and under-exposed pixels. Comparing  $I_{ref1}$  with  $I_{ref2}$ , if their ranking number difference is  $>2$ , it indicates that the overall luminance of the image sequence is relatively high or low, and  $I_{ref1}$  is selected as the reference image. Otherwise,  $I_{ref2}$  is chosen:

$$I_{ref1} = \begin{cases} I_{\frac{K+1}{2}} & K = \text{odd}, \\ I_{\frac{K}{2}} & K = \text{even}. \end{cases} \quad (1)$$

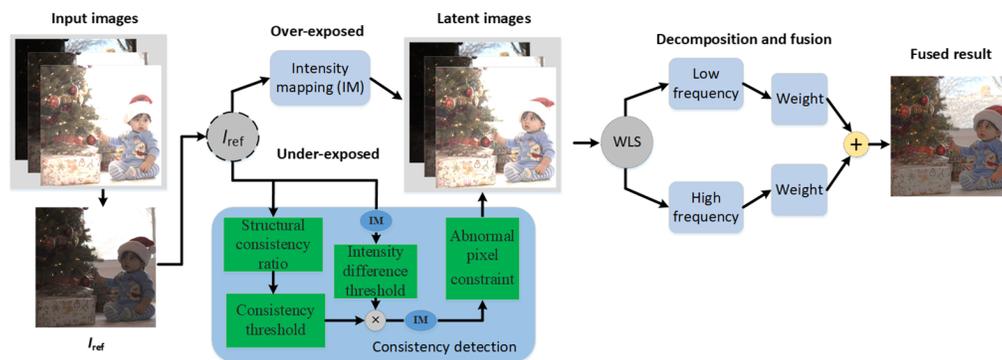


Fig. 2 The algorithm framework proposed in this paper.

After choosing the appropriate reference image, the input image whose luminance is higher than the reference image is taken as the over-exposed image. Otherwise, it is regarded as an under-exposed image.

### 2.2 Intensity Mapping

For the over-exposed image, the IM relationship between these images and the reference image is established to obtain the latent image. It can be obtained by establishing the IM model between the images with different exposures as follows:

$$I_k = \tau(I_{ref}), \tag{2}$$

where  $\tau(I) = g^{-1}(kg(I))$  is the IM function and indicates the radiation relationship between two images.<sup>31</sup>  $I$  is the scene brightness,  $g$  is the camera response function, and  $k$  is the exposure coefficient. The IM relationship can be established using the intensity histogram of the image, as shown in Fig. 3. According to the histogram matching, the luminance of the reference image can be mapped to the same range as the over-exposed image to obtain the latent image of the over-exposed image.

### 2.3 Consistency Detection

The under-exposed image often contains scene information conducive to the final fusion. However, these regions may be saturated in the reference image, resulting in detail loss. We found that if the latent image is calculated directly through the IM relationship, it may lead to information loss or color distortion. Based on the research foundation in Ma,<sup>23</sup> we compute the direction information of the image structure vector to realize motion consistency detection. The moving region consistency detection process is illustrated in Fig. 4. A set of color image patches  $\{g_k\} = \{g_k | 1 \leq k \leq K\}$  are extracted from the same spatial position of multi-exposure image sequence  $I_k$  using the moving window with a fixed step length.  $g_k$  is the vector of the  $CWH$  dimension ( $C$  is three channels,  $W$  and  $H$  are the width and height of the image patch). The signal structure  $S_k$  of the image can be calculated based on these variables. The motion consistency detection could be converted into the direction consistency analysis of the signal structure between the reference image and the source image. The inner product between the signal structure of the reference image  $S_{ref}$  and that of the sequence image  $S_k$  can be calculated as follows:

$$\rho_k = S_{ref}^T S_k = \frac{(g_{ref} - u_{gref})^T (g_k - u_{gk}) + \epsilon}{\|g_{ref} - u_{gref}\| \|g_k - u_{gk}\| + \epsilon}, \tag{3}$$

where  $\|\cdot\|$  denotes the  $l_2$  norm of the vector. The consistency ratio  $\rho_k$  lies in  $[-1, 1]$ , and the larger  $\rho_k$  is, the higher the consistency between  $S_{ref}$  and  $S_k$  is.  $u_{gk}$  is the mean brightness of  $g_k$ ,

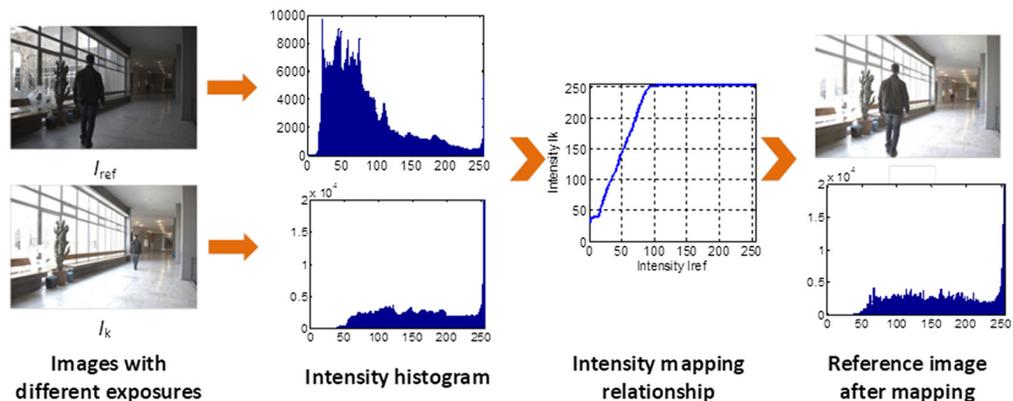
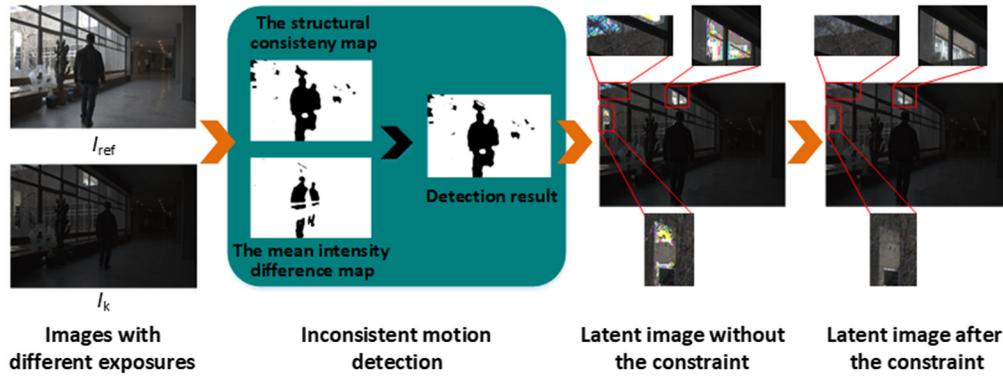


Fig. 3 IM and latent image acquisition.



**Fig. 4** The inconsistent motion detection and the abnormal pixel constraint.

and  $u_{g_{ref}}$  is the mean brightness of  $g_{ref}$ .  $S_k$  is constructed by mean removal and brightness normalization to ensure strong robustness to the exposure and contrast variations. The noise may be included when the structure vector  $S_k$  is scaled to the unit length in the scene with low visibility. Therefore, a parameter  $\epsilon$  close to 0 is introduced, which boosts  $\rho_k$  to be close to 1 regardless of the noise change.

To reject more motion inconsistent pixels, we introduce the structure consistency threshold  $T_\rho$  to binarize  $\rho_k$ . The inconsistent pixels in the image sequence can be observed in the black region of the binary map  $\hat{B}_k$ :

$$\hat{B}_k = \begin{cases} 1, & \rho_k \geq T_\rho, \\ 0, & \rho_k < T_\rho. \end{cases} \quad (4)$$

To detect and correct the inconsistent pixels to the greatest extent, a judgment condition based on mean intensity difference is introduced to reduce the ghosts further. We establish the IM relationship between the under-exposed image  $I_k$  and the reference image  $I_{ref}$ . The  $k$ 'th under-exposed image is mapped to the exposure level of the reference image to obtain the intermediate image. The absolute value of the mean intensity difference of the image patch at the same spatial position between the corresponding intermediate image and the reference image is calculated, and a mean intensity difference threshold  $T_u$  is set:

$$\check{B}_k = \begin{cases} 1, & |u_{gk} - u'_{gk}| < T_u, \\ 0, & |u_{gk} - u'_{gk}| \geq T_u, \end{cases} \quad (5)$$

where  $u'_{gk}$  is the mean brightness of the image patch in the intermediate image, which is obtained by mapping the under-exposed image to the reference image.

The final detection result of motion consistency is

$$B_k = \hat{B}_k \cdot \check{B}_k. \quad (6)$$

The pixel set with  $B_k = 1$  indicates that the motion is consistent with that of the reference image, and the image patch at the same position of the source image is retained. The pixel set with  $B_k = 0$  means that the motion is inconsistent. The inconsistent pixels need to be corrected in combination with the IM function to obtain the latent image  $Q_k$ .

In addition, in some dynamic scenes, it is found that the brightness saturation and color distortion may occur in the local region of the latent image obtained through the above steps, as shown in the three areas marked in the third step in Fig. 4. Therefore, we designed another abnormal pixel constraint to tackle this issue, and the improvement can be observed in the last step in Fig. 4:

$$Q_k = \begin{cases} I_k & Q_{k,1} \geq \tau_1 \text{ or } Q_{k,2} \geq \tau_2 \text{ or } Q_{k,3} \geq \tau_3, \\ Q_k & \text{else,} \end{cases} \quad (7)$$

where  $Q_{k,1}$ ,  $Q_{k,2}$ , and  $Q_{k,3}$  indicate the RGB's color channels of the latent image.  $\tau_1$ ,  $\tau_2$ , and  $\tau_3$  are the color thresholds on three channels. The latent image mapped from the reference image to the under-exposed image can be obtained through the above steps.

### 2.4 Image Decomposition Using WLS Filtering

After the input images are aligned, the luminance image  $\tilde{Q}_k$  of the latent image is obtained through the RGB color channel weighted sum. Then WLS filtering is used to accurately estimate the low-frequency information of  $\tilde{Q}_k$ , and the high-frequency component can be calculated further. Their blending weights are, respectively, calculated for the final fusion.

WLS filtering is an effective edge-preserving filter that can smooth the image while maintaining the image edge.<sup>32</sup> Compared with other filters, WLS filtering can make the best compromise between image blurring and sharpening. Therefore, we select WLS filtering as a low-pass filter to estimate the low-frequency information of the latent images with different exposures. Given an input image  $\tilde{Q}_k$ , we calculate its corresponding filtered low-frequency image  $\tilde{L}_k$ , which is smooth and close to the image  $\tilde{Q}_k$  as much as possible. The image  $\tilde{L}_k$  can be calculated by the following formula:

$$\left( \underset{\tilde{L}_k}{\operatorname{argmin}} \left( (\tilde{L}_k - \tilde{Q}_k)^2 + \lambda \left( w_x \left( \frac{\partial \tilde{L}_k}{\partial x} \right)^2 + w_y \left( \frac{\partial \tilde{L}_k}{\partial y} \right)^2 \right) \right) \right), \tag{8}$$

$$w_x = \left( \left| \frac{\partial V_k}{\partial x} \right|^\alpha + \epsilon \right)^{-1}, \tag{9}$$

$$w_y = \left( \left| \frac{\partial V_k}{\partial y} \right|^\alpha + \epsilon \right)^{-1}, \tag{10}$$

where the first item  $(\tilde{L}_k - \tilde{Q}_k)^2$  ensures the similarity between the input and output images. The smaller the distance between them is, the higher the similarity between the filtered and the latent images is. In the second item, the degree of smoothness is achieved by calculating the partial derivative, where  $w_x$  and  $w_y$  are the smoothing factors.  $\lambda$  is the regularization factor that maintains the balance between the two subformulas. The higher the  $\lambda$  value is, the smoother the output image is.  $V_k$  is the logarithmic form of the input image  $\tilde{Q}_k$ , namely  $V_k = \log(\tilde{Q}_k)$ . The index  $\alpha$  indicates the sensitivity to the gradient of the input image, and  $\epsilon$  is a positive constant close to 0.

Once the low-frequency component  $\tilde{L}_k$  of the latent image is obtained, according to each channel of the latent image, the high-frequency component  $\tilde{H}_k$  can be extracted:

$$\tilde{H}_k = Q_k - \tilde{L}_k. \tag{11}$$

### 2.5 Blending Weight Calculation and Fusion

The degree of exposure plays a significant role in preserving and revealing the details in MEF. Based on a general exposedness function, we modify it slightly for low- and high-frequency components to precisely reflect the global and local structures. The low-frequency component is a comprehensive measure of the intensity of the whole image. We construct the blending weights of this component to evaluate the exposure appropriateness based on the local and global luminance.<sup>33</sup> For the former, the local mean intensity at each pixel position is utilized as the exposure feature to assess the local exposure quality. However, to maintain the structural consistency between the low-frequency component and its weight map, the low-frequency component itself is applied as the local exposure feature. For the latter, the global exposure quality assessment of the  $k$ 'th image  $\tilde{G}_k$  is calculated according to the luminance sum of the whole image and the image size. The blending weights are obtained by combining these two parts:

$$W_k^L = \exp\left(-\left(\frac{(\tilde{L}_k - 0.5)^2}{2\sigma_L^2} + \frac{(\tilde{G}_k - 0.5)^2}{2\sigma_G^2}\right)\right), \quad (12)$$

where  $\sigma_L$  and  $\sigma_G$  are the Gaussian spread parameters. The first item preserves the local structure with good exposure, and the second item encourages the consistency of spatial luminance of the whole image.

For the high-frequency component of the image, the mean luminance of the small local neighborhood at each pixel position is computed to evaluate its exposure quality. The weight of the high-frequency component of the  $k$ 'th latent image is calculated as follows:

$$W_k^H = \exp\left(-\frac{(\tilde{h}_k - 0.5)^2}{2\sigma_H^2}\right), \quad (13)$$

where  $\tilde{h}_k$  is generated by convolving  $\tilde{Q}_k$  with mean filtering.

After constructing the weights of all components of the latent images, the low-frequency component  $\tilde{L}_k$  and the high-frequency component  $\tilde{H}_k$  are combined by weighted summation. To reserve more color characteristics, the image color information is implicitly retained when calculating the mean luminance. The RGB color channels are jointly processed during fusion to obtain the final fused image with vivid color and ghost-free:

$$F = \sum_{k=1}^K W_k^L \tilde{L}_k + \sum_{k=1}^K W_k^H \tilde{H}_k. \quad (14)$$

### 3 Experiment Results and Analysis

#### 3.1 Experiment Setup

To verify the performance of the proposed deghosting MEF method, we test it in various representative dynamic scenes. The multi-exposure images used in the experiment cover indoor, outdoor, and different times, including moving object with different amplitudes and camera shakes. The test source images used in this paper are mainly from below.

The proposed algorithm is compared with six typical deghosting MEF algorithms to evaluate their effectiveness in the dynamic scenes. We select the method to be compared according to the following principles: the algorithm has been developed in recent years, and its influence is high in this field, i.e., the number of citations is relatively great. The source code is publicly available and can be freely downloaded. We select six representative deghosting MEF methods in our comparative study. Table 1 lists the details of the chosen method, including the origin and code download links. For all of them, the default parameter settings are the same as those in the original literature.

The parameters in the proposed method need to be set, including structural consistency threshold  $T_\rho$ ; mean brightness difference threshold  $T_u$ ; color threshold  $\tau_1$ ,  $\tau_2$ , and  $\tau_3$ ; WLS

**Table 1** The methods for the comparison.

Method	Year	Link of source code
Liu <sup>18</sup>	2015	<a href="https://github.com/yuliu316316/DSIFT-EF">https://github.com/yuliu316316/DSIFT-EF</a>
Hayat <sup>19</sup>	2019	<a href="https://github.com/ImranNust/Source-Code">https://github.com/ImranNust/Source-Code</a>
Ma <sup>23</sup>	2017	Ref. 34
Qi <sup>24</sup>	2020	<a href="https://github.com/zhiqinzhu123/HDR-Multi-exposure-image-Fusion-Source-Code">https://github.com/zhiqinzhu123/HDR-Multi-exposure-image-Fusion-Source-Code</a>
LH20 <sup>25</sup>	2020	<a href="https://github.com/xiaohuiben/fmmef-TIP-2020">https://github.com/xiaohuiben/fmmef-TIP-2020</a>
LH21 <sup>26</sup>	2021	<a href="https://github.com/xiaohuiben/MESPD_TCSVT-2021">https://github.com/xiaohuiben/MESPD_TCSVT-2021</a>

filtering parameters  $\lambda$  and  $\alpha$ ; and Gaussian spread parameters  $\sigma_L$ ,  $\sigma_G$ , and  $\sigma_H$ .  $T_\rho$  and  $T_u$  should retain consistent motion and remove inconsistent movement as much as possible. Through the analysis of the latent image generated in the various dynamic scenes,  $T_\rho = 0.8$  and  $T_u = 0.2$  are selected, which can ensure the minor color distortion and residual shadow. The color thresholds  $\tau_1$ ,  $\tau_2$ , and  $\tau_3$  can be set to 0.98, 0.98, and 0.9 on the fused visual effect empirically. The selection of  $\lambda$  and  $\alpha$  will be described in Sec. 3.4 in detail. For Gaussian spread parameters for calculating the blending weights, the values of  $\sigma_L$ ,  $\sigma_G$ , and  $\sigma_H$  are inherited from the research results in Ma<sup>23</sup> and Nejati.<sup>33</sup> They are equal to 0.5, 0.2, and 0.12, respectively.

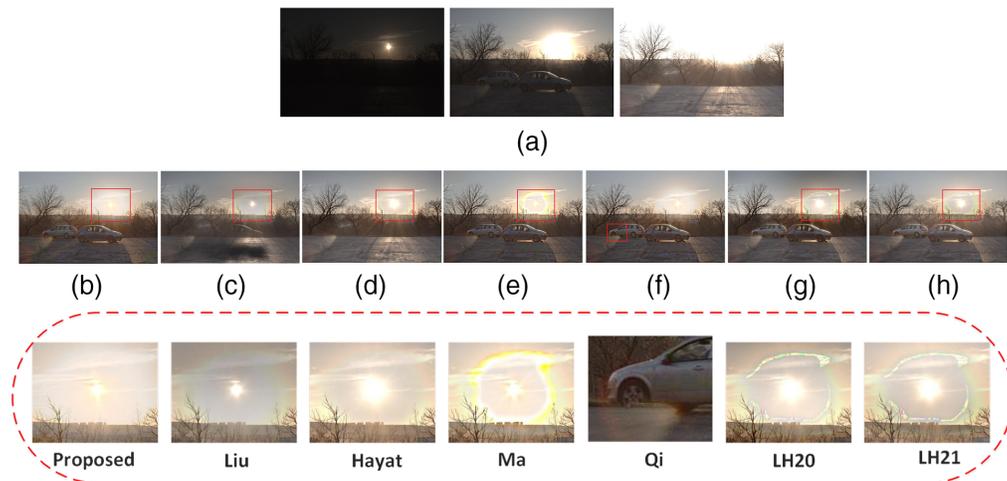
### 3.2 Subjective Qualitative Evaluation

High-quality fused images should have an excellent visual effect and without ghosting artifacts or residual shadows. In this part, combined with the image dataset given in Table 2, a test dataset with different exposures in the dynamic scenes is established, including 30 groups of image sequences. We have disclosed the source image sequences and 210 fused results from the proposed method and six comparison methods, and interested readers can freely download them (in Ref. 38). Three sets of fused results are chosen for detailed discussion and analysis as shown in Figs. 5–7.

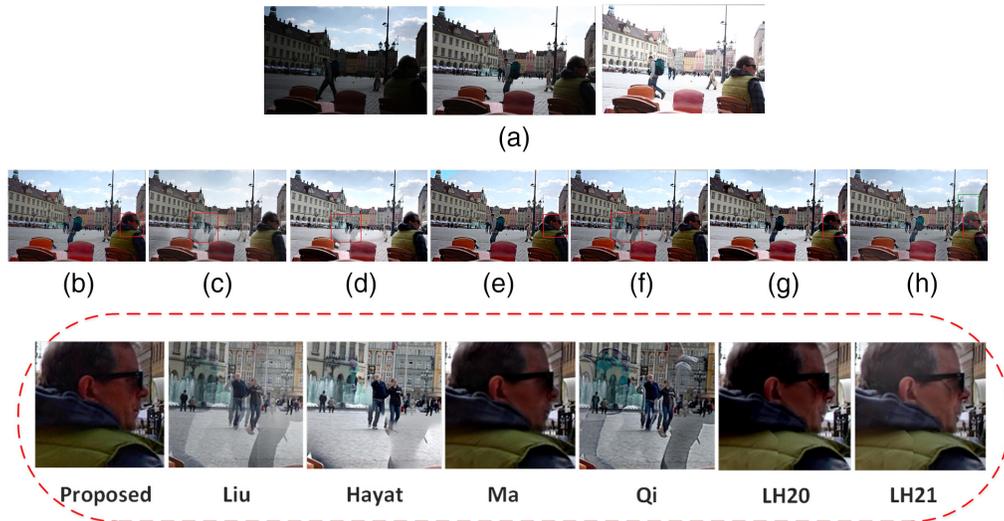
Figure 5 illustrates the final fused images are different when different reference images are determined. For example, there is no vehicle in the reference image of Hayat,<sup>19</sup> and the reference images selected by other methods all have moving cars. The fused image given by Liu<sup>18</sup> has a poor effect on the ghost removal, and there are some translucent and dark regions. Other methods are relatively effective in ghost removal. Observing the local enlarged view in Hayat,<sup>19</sup> the details of the central location of the Sun are not recovered well, and the central point of the Sun and the nearby cloud are unclear. Ma,<sup>23</sup> LH20,<sup>25</sup> and LH21<sup>26</sup> have color distortion near the Sun, and

**Table 2** Image dataset in the dynamic scenes.

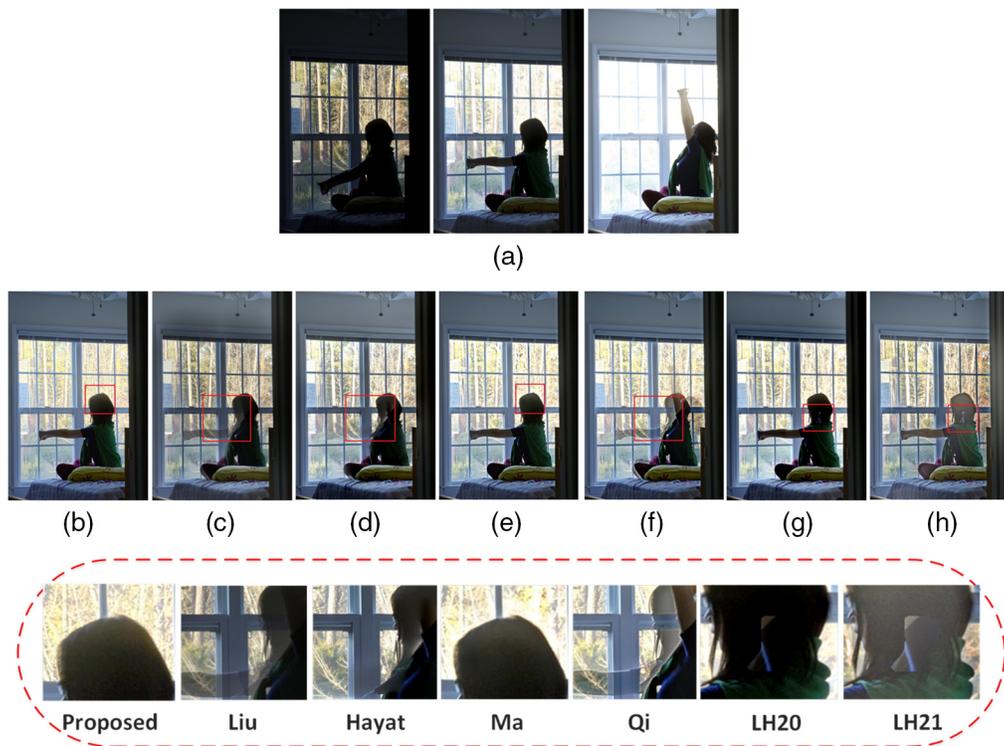
Dataset	Year	Image sequences	Total number	Link of source code
Dataset <sup>35</sup>	2016	17	153	Ref. 36
IQASet <sup>23</sup>	2020	20	84	<a href="https://github.com/h4nwei/MEF-SSIMd">https://github.com/h4nwei/MEF-SSIMd</a>
HDRRNN <sup>29</sup>	2021	84	588	Ref. 37



**Fig. 5** The qualitative comparison of different methods on the “cars” image sequence and local enlarged view: (a) source image sequence, (b) proposed, (c) Liu, (d) Hayat, (e) Ma, (f) Qi, (g) LH20, and (h) LH21.



**Fig. 6** The qualitative comparison of different methods on the “Wroclav” image sequence and local enlarged view: (a) source image sequence, (b) proposed, (c) Liu, (d) Hayat, (e) Ma, (f) Qi, (g) LH20, and (h) LH21.



**Fig. 7** The qualitative comparison of different methods on the “lady” image sequence and local enlarged view: (a) source image sequence, (b) proposed, (c) Liu, (d) Hayat, (e) Ma, (f) Qi, (g) LH20, and (h) LH21.

the color of the upper part of the sky in LH20<sup>25</sup> is dark. In addition, a small part of the front wheel is missing in their fused result. Although no color distortion occurs near the Sun in Qi,<sup>24</sup> there is a faint halo around the cars, and the part of the front wheel is lost. The comprehensive comparison results show that the performance of the proposed method is the best, which can preserve more details. Moreover, the halo effect is suppressed.



**Fig. 8** The fused result on the “flag” image sequence: (a) source image sequence and (b) the fused result.

As can be seen in Fig. 6, the proposed method, Ma,<sup>23</sup> LH20,<sup>25</sup> and LH21<sup>26</sup> are effective in ghost removal. The ghosts are not entirely removed and still exist in Liu,<sup>18</sup> Hayat,<sup>19</sup> and Qi.<sup>24</sup> Moreover, the luminance of the ground in Hayat<sup>19</sup> is still high, and the details are not well recovered. Although the ghost removal effect is good in LH20,<sup>25</sup> it is a little poor in the detail recovery of local regions, such as the luminance of the table and the ground is still high. There is partial distortion in the color recovery of the sky area in Ma.<sup>23</sup> On the whole, the details of the proposed method and LH21<sup>26</sup> are well preserved in the bright and dark areas. However, the color of some areas of the sky in LH21<sup>26</sup> is not unevenly recovered. The color of the region below the cloud is darker, and the hierarchy is slightly worse, as shown in the green box in Fig. 6(h). In the local enlarged view, it can be found that the facial contour and mouth area of the man in Ma<sup>23</sup> are slightly poor in detail. There are some blocking problems on the face and clothes of the man in LH20<sup>25</sup> and LH21.<sup>26</sup> According to the details of the whole image and the local enlarged view, the visual quality of the proposed method is the best.

Figure 7 illustrates the fused results from the lady who swings her arms in the indoor scene. We can see that except for the proposed method and Ma,<sup>23</sup> other methods have ghosts to varying degrees. Among them, the ghosts in Liu,<sup>18</sup> Hayat,<sup>19</sup> and Qi<sup>24</sup> are serious. Although the details are recovered well in LH21,<sup>26</sup> there are discontinuous blocks in the head and chest area of the woman. The same problem exists in LH20.<sup>25</sup> Although there are no apparent ghosts in Ma,<sup>23</sup> there is a noticeable halo on the top of her head and dislocation in the chest area. Therefore, the performance of the proposed method is optimal on the whole from the image sequence and the local enlarged view.

In addition, we give two examples of deformable body motion to demonstrate the effectiveness of the proposed method, as shown in Figs. 8 and 9. The fused results show that the ghost artifacts can be effectively removed, and the texture and color details can be recovered well.

### 3.3 Objective Quantitative Comparison

There are few objective quantitative indicators for MEF in the dynamic scenes, and the most researchers only adopted subjective evaluation.<sup>11,12,14,20,39-41</sup> Some studies use the HDR-VDP-2 model to predict the quality of the fused image,<sup>42,43</sup> and Qmos (mean opinion score) in this model can evaluate the distortion between the fused image and the reference image. If the score is larger, the image quality is higher. Otherwise, the image quality is worse. The Qmos scores of 210 fused images are calculated and listed in Table 3, where the bold font indicates the highest



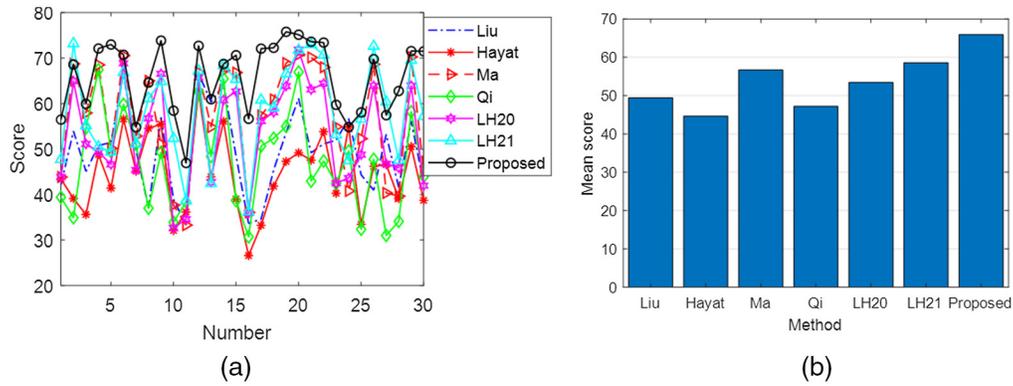
**Fig. 9** The fused result on the “candle” image sequence: (a) source image sequence and (b) the fused result.

**Table 3** Quantitative comparisons of different deghosting methods on Qmos score.

Sequence	Liu	Hayat	Ma	Qi	LH20	LH21	Proposed
Arch	42.1539	43.4479	43.8774	39.4118	43.8846	47.7516	<b>56.4972</b>
Brunswick	53.9745	39.1341	68.7604	34.9105	64.9440	<b>73.2189</b>	<i>68.6200</i>
Building	45.0889	35.6322	57.9390	54.6528	51.1381	55.1796	<b>59.9514</b>
Campus	50.7968	48.8127	68.5215	67.4566	49.3895	50.5085	<b>72.1113</b>
Cars	51.4744	41.4154	50.1511	49.4603	46.6479	49.3922	<b>73.0031</b>
Forest	66.6202	56.6106	70.5719	59.8429	68.8709	66.8502	<b>70.7423</b>
Lady	49.7846	45.1670	52.8730	51.0440	45.4192	51.1117	<b>54.8218</b>
Llandudno	37.1335	54.5771	<b>65.0071</b>	37.0104	56.7933	61.1436	<i>64.6298</i>
MarketMires2	56.9971	55.4768	51.1316	49.3215	66.5522	64.8403	<b>73.8640</b>
Men	38.9525	32.1070	37.6922	33.9290	32.6753	52.3410	<b>58.4556</b>
Office	33.3883	36.1874	33.2980	38.4133	34.6107	38.6722	<b>46.9505</b>
Playground	66.6578	62.6573	66.9124	62.6548	65.7661	67.2283	<b>72.7180</b>
ProfJeonEigth	60.3252	43.9002	54.8428	48.3970	42.7288	42.4942	<b>60.9715</b>
River	64.1139	56.1138	67.2512	65.4758	60.7846	68.4059	<b>68.7286</b>
Square	48.4465	39.0180	66.8230	38.6824	62.6895	65.3453	<b>70.6387</b>
Street	33.6765	26.5961	36.0854	30.7094	35.9272	36.2446	<b>56.6738</b>
Suoe	34.6099	33.2567	57.4005	50.6865	56.1674	60.8040	<b>72.0869</b>
Suon	45.3769	41.8485	60.9157	52.3933	58.0983	59.1014	<b>72.2730</b>
Suos	53.6372	47.2952	68.9225	54.9949	63.8709	66.5662	<b>75.7899</b>
Suow	61.1277	49.2031	70.7916	66.9492	71.8296	71.7604	<b>75.1509</b>
t1	49.2015	47.5557	70.1511	42.9625	63.0951	72.9974	<b>73.6079</b>
t10	51.1690	53.8533	68.0086	47.3814	64.3980	70.6989	<b>73.4272</b>
Tate3	51.9602	40.3044	54.9352	42.4331	42.5136	53.1564	<b>59.7365</b>
Tour	<b>56.7095</b>	54.8627	40.8084	49.5950	43.6533	47.4784	<i>54.8626</i>
Wroclav	44.3532	33.4461	52.2975	32.3816	48.7993	56.5951	<b>58.0914</b>
YW	40.9657	46.2136	68.6462	47.7800	63.9147	<b>72.5969</b>	<i>69.7789</i>
z179	53.2283	46.6198	40.3606	31.0169	46.7720	<b>60.3851</b>	<i>57.4731</i>
zSantas	41.6042	39.1722	39.6670	34.1192	45.8995	47.5281	<b>62.7781</b>
zstack_ghost	56.1666	50.5479	70.2718	58.2366	63.9338	69.5758	<b>71.5528</b>
zStreetDay	42.3889	38.7459	45.1246	44.0013	41.9762	57.2124	<b>71.5070</b>

score and the italic one means the second-highest score. It can be seen that the overall score of the proposed method is the best on all test image sequences.

To more intuitively view the performance of each method on the test image sequences, the Qmos scores in Table 3 are drawn in Fig. 10(a). From the distribution of each curve, we can see the performance of these methods on 30 groups of test image sequences in the dynamic scenes.



**Fig. 10** Quantitative comparisons of 30 image sequences in the dynamic scenes: (a) Qmos score and (b) mean score.

The mean scores of each method are drawn in Fig. 10(b). It can be seen that the proposed method can obtain an ideal performance over the other six methods.

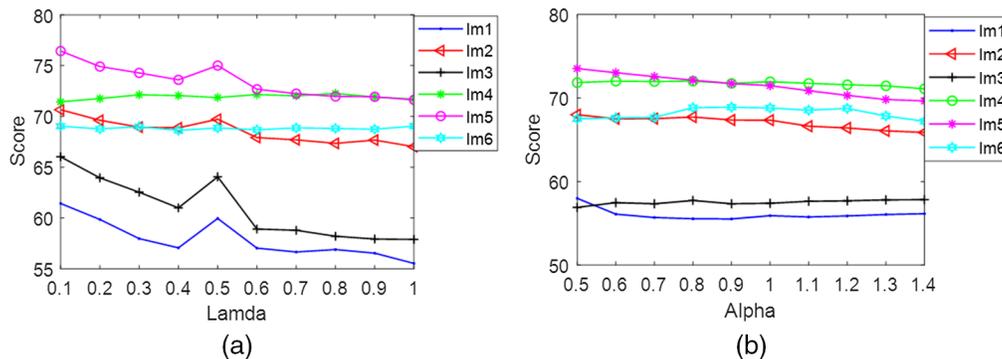
### 3.4 Selection of $\lambda$ and $\alpha$

In WLS filtering,  $\lambda$  controls the smoothness of the output image, and  $\alpha$  decides the sensitivity of gradient change. We randomly select six image sequences for the test to determine the optimal parameters. According to the default value  $\lambda = 1$  and  $\alpha = 1.2$  used in Ref. 32, we first fix  $\alpha = 1$  and change  $\lambda$  from 0.1 to 1.0. Then we fix  $\lambda = 1$  and change  $\alpha$  from 0.5 to 1.4. Qmos scores are calculated under different parameters, respectively, and illustrated in Fig. 11.

We found that the change of  $\lambda$  and  $\alpha$  has no significant impact on Qmos score. Although the Qmos score will increase a little with the decrease of these two parameters, the smaller parameters may lead to the deterioration of texture details in some fused images. Therefore, we set  $\lambda = 0.5$  and  $\alpha = 1$  to make a good balance.

### 3.5 Computational Efficiency Comparison

In addition to the evaluation mentioned above, we also test the computational efficiency of different methods. All of them are implemented in MATLAB 2018a with a 2.71-GHz CPU (Intel® Core™ i5-7200U) and 8 GB RAM. Table 4 lists the running time of all methods on the source images with a size of  $1024 \times 682$  and three exposures in the dynamic scene. The proposed algorithm is implemented without any optimization and ranks in the middle among all comparison methods. The proposed method is expected to have a higher speed after optimization and C++ implementation.



**Fig. 11** Qmos scores on different parameters: (a) fixing  $\alpha$  and changing  $\lambda$  and (b) fixing  $\lambda$  and changing  $\alpha$ .

**Table 4** The computational efficiency of different methods.

Method	Time (s)	Method	Time (s)
Liu	8.0484	LH20	4.2080
Hayat	7.0016	LH21	5.0264
Ma	14.7634	Proposed	10.3538
Qi	13.2259		

## 4 Conclusion

This paper designs a deghosting MEF method for HDR imaging expansion in the dynamic scenes. First, we analyze the relationship between the source images and the selected reference image. IM and structural consistency detection are carried out for over- and under-exposed images. The ghosts can be reliably removed, and the latent images at different exposures are obtained. Second, WLS filtering is used to estimate the latent images' low-frequency components, and then the corresponding high-frequency components are extracted. Third, blending weights are calculated to fuse the low- and high-frequency components to generate the final fused image with more details. The comprehensive comparative experiment shows that the proposed method can perform better than state-of-the-art MEF algorithms in the dynamic scenes. It can not only deal with objects with different amplitudes or even large motions but also slightly camera shake. By comparing the calculation time with other methods, we need to optimize further and accelerate the algorithm model to improve its practicability.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62175233). The authors declare no conflicts of interest.

## Data and Materials Availability

Data underlying the results presented in this paper are available in Ref. 38.

## References

1. L. Huang et al., "Multi-exposure image fusion based on feature evaluation with adaptive factor," *IET Image Process.* **15**(13), 3211–3220 (2021).
2. T. Nie et al., "Multi-exposure fusion of gray images under low illumination based on low-rank decomposition," *Remote Sens.* **13**(2), 204 (2021).
3. J. Zhang and S. Hu, "Graphic-processing-unit-accelerated real-time exposure fusion method using pixel-level optimal exposure criterion," *Opt. Eng.* **51**(7), 071404 (2012).
4. O. T. Tursun et al., "The state of the art in HDR deghosting: a survey and evaluation," *Comput. Graph. Forum* **34**(2), 683–707 (2015).
5. F. Xu et al., "Multi-exposure image fusion techniques: a comprehensive review," *Remote Sens.* **14**(3), 771 (2022)
6. Z. Xu et al., "Multi-exposure image fusion for dynamic scenes with ghosting removal," *J. Electron. Imaging* **31**(2), 023036 (2022).
7. E. A. Khan, A. O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *IEEE Int. Conf. Image Process.*, pp. 2005–2008 (2006).
8. M. Pedone and J. Heikkil, "Constrain propagation for ghost removal in high dynamic range images," in *VISAPP*, pp. 36–41 (2008).
9. W. Zhang and W. K. Cham, "Gradient-directed composition of multi-exposure images," in *IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 530–536 (2010).

10. Z. Li et al., "Selectively detail-enhanced fusion of differently exposed images with moving objects," *IEEE Trans. Image Process.* **23**(10), 4372–4382 (2014).
11. Z. Wang, Q. Liu, and T. Ikenaga, "Robust ghost-free high-dynamic-range imaging by visual saliency based bilateral motion detection and stack extension based exposure fusion," *IEICE Trans. Fundam. Electron. Commun. Computer Sci.* **E100**(11), 2266–2274 (2017).
12. Z. Qu et al., "Algorithm of multi-exposure image fusion with detail enhancement and ghosting removal," *J. Electron. Imaging* **28**(1), 013022 (2019).
13. S. O. Shim, "Estimation of gamma-corrected exposure time ratio in multi-exposure images for removal of moving objects," *Appl. Opt.* **59**(13), 4076–4080 (2020).
14. K. Jacobs, C. Loscos, and G. Ward, "Automatic high-dynamic range image generation for dynamic scenes," *IEEE Comput. Graphics Appl.* **28**(2), 84–93 (2008).
15. M. Granados et al., "Automatic noise modeling for ghost-free HDR reconstruction," *ACM Trans. Graphics* **32**(6), 201 (2013).
16. C. Wang and C. He, "A novel deghosting method for exposure fusion," *Multimedia Tools Appl.* **77**(24), 31911–31928 (2018).
17. H. Zimmer, A. Bruhn, and J. Weickert, "Freehand HDR imaging of moving scenes with simultaneous resolution enhancement," *Comput. Graphics* **30**(2), 405–414 (2011).
18. Y. Liu and Z. Wang, "Dense SIFT for ghost-free multi-exposure fusion," *J. Vis. Commun. Image Represent.* **31**, 208–224 (2015).
19. N. Hayat and M. Imran, "Ghost-free multi exposure image fusion technique using dense sift descriptor and guided filter," *J. Vis. Commun. Image Represent.* **62**, 295–308 (2019).
20. W. Zhang et al., "Motion-free exposure fusion based on inter-consistency and intra-consistency," *Inf. Sci.* **376**, 190–201 (2017).
21. P. Sen et al., "Robust patch-based HDR reconstruction of dynamic scenes," *ACM Trans. Graphics* **31**(6), 203 (2012).
22. J. Hu et al., "Exposure stacks of live scenes with hand-held cameras," in *Proc. Eur. Conf. Comput. Vision (ECCV)*, pp. 499–512 (2012).
23. K. Ma et al., "Robust multi-exposure image fusion: a structural patch decomposition approach," *IEEE Trans. Image Process.* **26**(5), 2519–2532 (2017).
24. G. Qi et al., "A precise multi-exposure image fusion method based on low-level features," *Sensors* **20**(6), 1597 (2020).
25. H. Li et al., "Fast multi-scale structural patch decomposition for multi-exposure image fusion," *IEEE Trans. Image Process.* **29**, 5805–5816 (2020).
26. H. Li et al., "Detail-preserving multi-exposure fusion with edge-preserving structural patch decomposition," *IEEE Trans. Circuits Syst. Video Technol.* **31**(11), 1–12 (2021).
27. Y. Zhang et al., "Neural network-based image reconstruction in swept-source optical coherence tomography using undersampled spectral data," *Light Sci. Appl.* **10**(9), 1658–1671 (2021).
28. X. Li et al., "Unsupervised content-preserving transformation for optical microscopy," *Light Sci. Appl.* **10**(3), 390–400 (2021).
29. K. R. Prabhakar, S. Agrawal, and R. V. Babu, "Self-gated memory recurrent network for efficient scalable HDR deghosting," *IEEE Trans. Comput. Imaging* **7**, 1228–1239 (2021).
30. Q. Yan et al., "High dynamic range imaging via gradient-aware context aggregation network," *Pattern Recognit.* **122**, 108342 (2022).
31. M. D. Grossberg and S. K. Nayar, "Determining the camera response from images: what is knowable?" *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(11), 1455–1467 (2003).
32. Z. Farbman et al., "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graphics* **27**(3), 1–10 (2008).
33. M. Nejati et al., "Fast exposure fusion using exposedness function," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, pp. 2234–2238 (2017).
34. K. Ma, "Robust multi-exposure image fusion: a structural patch decomposition approach," <https://ece.uwaterloo.ca/~k29ma/> (2017).
35. O. Tursun et al., "An objective deghosting quality metric for HDR images," *Comput. Graph. Forum* **35**(2), 139–152 (2016).
36. O. Tursun, "An objective deghosting quality metric for HDR images," <https://user.ceng.metu.edu.tr/~akyuz/files/eg2016/index.html> (2016).

37. K. R. Prabhakar, "Self-gated memory recurrent network for efficient scalable HDR deghosting," <https://val.cds.iisc.ac.in/HDR/HDRRNN/index.html> (2021).
38. F. Xu, "MEFData of dynamic scenes," Github, 2022, <https://github.com/xfupup/MEFData-of-dynamic-scenes>.
39. W. Zhang and W. K. Cham, "Reference-guided exposure fusion in dynamic scenes," *J. Vis. Commun. Image Represent.* **23**(3), 467–475 (2012).
40. C. Lee, Y. Li, and V. Monga, "Ghost-free high dynamic range imaging via rank minimization," *IEEE Signal Process. Lett.* **21**(9), 1045–1049 (2014).
41. W. Zhang, S. Hu, and K. Liu, "Patch-based correlation for deghosting in exposure fusion," *Inf. Sci.* **415**, 19–27 (2017).
42. R. Mantiuk et al., "HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graphics* **30**(4), 1–14 (2011).
43. M. Chang et al., "Robust ghost-free multi-exposure fusion for dynamic scenes," *J. Electron. Imaging* **27**(3), 033023 (2018).

**Fang Xu** received her PhD from the University of Chinese Academy of Sciences in 2018. She is a research assistant at the Chinese Academy of Sciences of Changchun Institute of Optics, Fine Mechanics, and Physics. Her research interests include image processing and computer vision.

**Jinghong Liu** is a researcher and PhD tutor at the Chinese Academy of Sciences of Changchun Institute of Optics, Fine Mechanics, and Physics. Her current research interests include image processing and airborne localization.

**Tenglong Wang** is an engineer at the Northeast Electric Power Design Institute Company Limited of China Power Engineering Consulting Group. His research interests include electrical power system planning and image processing.

**Xuan Wang** received his PhD from the University of Chinese Academy of Sciences in 2017. He is working as an associate researcher at Chinese Academy of Sciences, Changchun Institute of Optics, Fine Mechanics, and Physics. His research interests include aerial target positioning, target tracking, and image processing.

**Hui Sun** received his PhD from the University of Chinese Academy of Sciences in 2022. He is working as an associate researcher at Chinese Academy of Sciences, Changchun Institute of Optics, Fine Mechanics, and Physics. His research interests include image processing and airborne target geo-location.

**Hang Ren** received his PhD from the University of Chinese Academy of Sciences in 2010. He is working as an associate researcher at Chinese Academy of Sciences, Changchun Institute of Optics, Fine Mechanics, and Physics. His research area is image processing and design of CCD drive circuits.