Multi-exposure fusion for high dynamic range scene[①]

Shen Xiaohe (申小禾)[②]* ** ***, Liu Jinghong*

(* Changchun Institute of Optics Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, P. R. China)
(** University of Chinese Academy of Sciences, Beijing 100049, P. R. China)
(*** Beijing Institute of Aerospace Control Devices, Beijing 100039, P. R. China)

Abstract

Due to the existing limited dynamic range a camera cannot reveal all the details in a high-dynamic range scene. In order to solve this problem, this paper presents a multi-exposure fusion method for getting high quality images in high dynamic range scene. First, a set of multi-exposure images is obtained by multiple exposures in a same scene and their brightness condition is analyzed. Then, multi-exposure images under the same scene are decomposed using dual-tree complex wavelet transform (DT-CWT), and their low and high frequency components are obtained. Weight maps according to the brightness condition are assigned to the low components for fusion. Maximizing the region Sum Modified-Laplacian (SML) is adopted for high-frequency components fusing. Finally, the fused image is acquired by subjecting the low and high frequency coefficients to inverse DT-CWT. Experimental results show that the proposed approach generates high quality results with uniform distributed brightness and rich details. The proposed method is efficient and robust in varies scenes.

Key words: multi-exposure fusion, high dynamic range scene, dual-tree complex wavelet transform (DT-CWT), brightness analysis

0 Introduction

The dynamic range of digital cameras is usually at 3 orders of magnitude, while the dynamic range of the nature world can be up to 9 orders of the magnitude. When the dynamic range of the nature scene exceeds that of a camera, all details in the scene are difficult to show in one image. By adjusting shutter speed and aperture[1], some regions of the image are always overexposed or underexposed. Exposure fusion is a solution to this problem, which fuses well-exposed regions of input images into a new image. This method can generate satisfactory results.

At present, many image fusion methods have been proposed. Exposure fusion methods include spatial domain method[2-5] and multi-resolution analysis method[6]. Spatial domain method usually evaluates the exposure condition by analyzing the features of blocks and pixels, such as image entropy, gradient, saturation, and brightness. Li, et al.[2] estimated the weight maps by measuring local contrast, brightness and color dissimilarity and then using recursive filter for refining the weight maps. Paul, et al.[5] proposed a multi-exposure

fusion method in gradient domain. At present, spatial domain method is wildly used in multi-exposure fusion. Multi-resolution analysis method decomposes the image to different matrices at different scales. Thus it has demonstrated good performance in detail processing. Mertens, et al.[6] decomposed input images using Laplacian pyramids and chose the weight for every image through saturation, contrast, and well exposedness. Gaussian pyramids are constructed for the weight maps. The Laplacian and Gaussian pyramids are blended at each level and the final result is calculated.

The existing multi-exposure fusion methods don't research on the relationship between the overall intensity of an image and the exposure condition of every pixel in this image. In this paper, brightness condition of a multi-exposure sequence is analyzed. The dual-tree complex wavelet transform (DT-CWT)[7,8] is employed for decomposing images into low and high frequency components. Low frequency components reflect the main energy and intensity; therefore, the weight maps for them are based on the brightness analysis. High frequency components reflect image details and Sum Modified-Laplacian (SML) reflects image details and definition[9]. Thus, maximum region SML is adopted

for high frequency fusion. Experimental results demonstrate that the proposed method has a good performance with high definition and suitable brightness.

This paper includes 4 sections. In Section 1, the characters of brightness in a multi-exposure sequence is analyzed. In Section 2, the proposed exposure fusion method is explained in detail. Experimental results and comparison are discussed in Section 3. Finally, conclusion is provided in Section 4.

1 Brightness analysis

Every image of a multi-exposure sequence contains pixels with appropriate exposure. Otherwise, the image is not suitable for fusion. In an image with low brightness, pixels that have high intensities are well exposed; likewise, in an image with high brightness, the pixels that have low intensities are well exposed. Fig. 1 demonstrates such a phenomenon.

The purpose of multi-exposure fusion is fusing the appropriately exposed regions of all images into one. The overexposed or underexposed regions should be
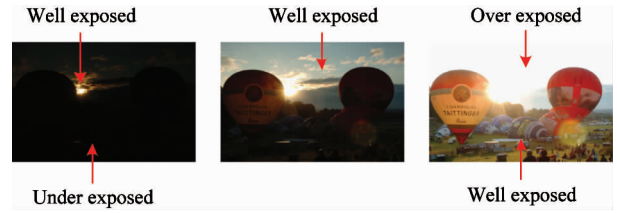

Fig. 1 A multi-exposure sequence

considered with a less weight than that of the well-exposed regions. The well-exposed pixels should have greater weight than other pixels. In this paper, functions that measure the relation between intensity and weight are established which are shown in Fig. 2.

Two thresholds p_l and p_h are set for estimating the brightness. In one image, if more than half of the pixels have values lower than p_l, it is considered that it is an image with low brightness. Similarly, if more than half of the pixels have values greater than p_h, it is regarded as an image with high brightness. Otherwise, it is assumed that it is an image with medium brightness. In this paper, p_l and p_h are set to 64 and 196, respectively.
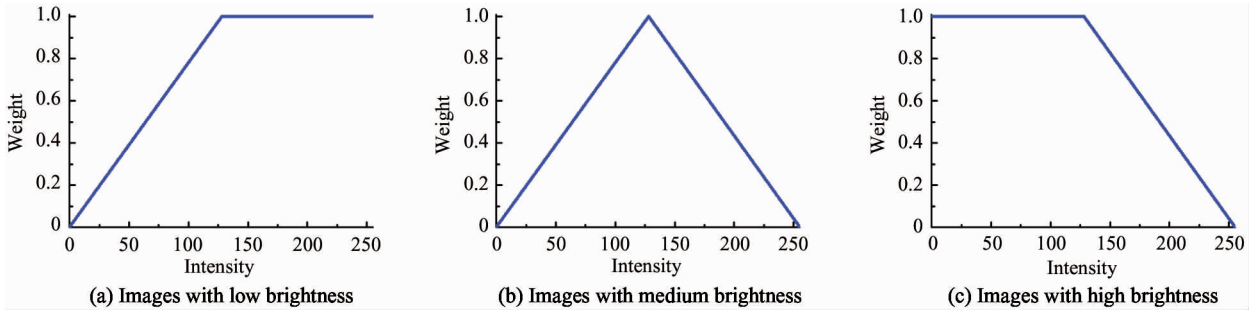

Fig. 2 Weight functions

2 Multi-exposure fusion method

It is assumed that a multi-exposure sequence has registered and the camera does not move. Input multi-exposure sequence is decomposed using the DT-CWT. The R, G, and B channels are decomposed respectively. The DT-CWT of a two-dimensional image $I(i, j)$ which is formulated using scaling function and the complex wavelet function is shown as

$$I(i, j) = \sum_{r \in R^2} a_{L,r} \varphi_{L,r}(i, j)$$
$$+ \sum_{\theta' \in \theta} \sum_{l \leqslant L} \sum_{r \in R^2} d_{l,r}^{\theta'} \psi_{l,r}^{\theta'}(i, j) \quad (1)$$

where L indicates the level of decomposition. R^2 indicates two-dimensional space. $\theta = \{ \pm 15°, \pm 45°,$ and $\pm 75° \}$. $a_{l,r}$ and $d_{l,r}^{\theta'}$ indicate the approximate part and

detail part of the signal respectively. $\varphi(i, j)$ indicates scaling function, while $\psi(i, j)$ indicates complex wavelet function, and (i, j) indicates coordinate of every pixel in an image.

Fig. 3 gives the algorithm flowchart. First, the input sequence is decomposed into high and low frequency coefficients by DT-CWT. Then the illumination of each input image is estimated to generate an illumination map, and a weight function is chosen accordingly. The box demonstrates the low frequency coefficients fusion which shall be introduced in Section 2.1. High frequency coefficients fusion shown in the other box is explained in Section 2.2. The final fusion result is obtained from the fused high frequency coefficients and the fused low frequency coefficients using the inverse DT-CWT.
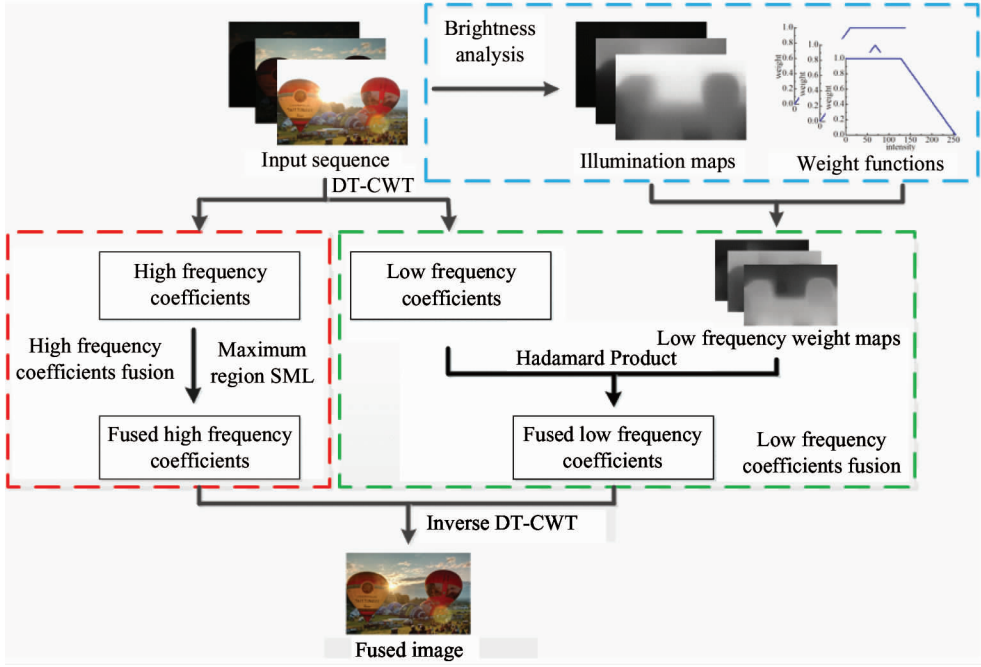
Fig. 3   Algorithm flowchart

2.1   Low frequency coefficients fusion

The fusion rules for low frequency coefficients should focus on the overall intensity rather than each single pixel. Therefore, in this study, the illumination of every image is estimated first and then the weights are distributed accordingly. The envelope filter[10] is adopted for illumination estimation because it eliminates the weak fluctuations of sudden changes, meanwhile retains their main trends. Eq. (2) expresses the envolpe filter, where parameter δ is assigned according to Eq. (3).

$$Y_v^+ = \max\{\delta(\nabla S_v) \cdot Y_{v-1} + (1 - \delta(\nabla S_v)) \cdot S_v, S_v\}$$
$$Y_v = \max\{\delta(-\nabla S_{v+1}) \cdot Y_{v-1}^+ + (1 - \delta(-\nabla S_{v+1}))$$
$$\cdot S_{v+1}, S_v\} \tag{2}$$

$$\delta(\nabla S) = \begin{cases} \delta_0 & \nabla S \geqslant -1/T \\ -\dfrac{\delta_0}{\nabla S \cdot T} & \text{if} \quad \nabla S < -1/T \end{cases} \tag{3}$$

where $S_{v=1,2,\cdots,V}$ indicates the input signal, $Y_{v=1,2,\cdots,V}$ represents the output signal, $Y_{v=1,2,\cdots V}^+$ means the intermediate signal, and $\nabla S_{v=1,2,\cdots,V}$ notes the gradient of signal. The number of signals is noted as V, and $-1/T$ is the threshold.

Based on the filtering results, the estimated illumination maps $\{E_k\}$ ($k = 1, 2, \cdots, K$, K is the number of input images) are obtained. For example, Fig. 4 shows the estimated illumination maps of Fig. 1.

Fig. 4   Estimated illumination maps of Fig. 1

The estimated illumination maps and the low frequency coefficients are not in the same scale. Therefore, the former are scaled down to the same scale as that of the latter and the rescaled estimated illumination maps $\{E_k^A\}$ are obtained. Calculating the weight functions of every estimated illumination map, the weight maps $\{W_k\}$ are acquired. The final weight maps $\{W_k^F\}$ for low frequency, shown in Eq. (4), are ob-tained by normalizing $\{W_k\}$.

$$W_k^F(i, j) = \frac{W_k(i, j)}{\sum_{k=1}^{K} W_k(i, j)}, \qquad (k = 1, 2, \cdots, K) \tag{4}$$

where (i, j) indicates the matrix coordinate of weight maps. Fig. 5 displays the final weight maps for low frequency of Fig. 1.

Fig. 5 Low frequency weight maps of Fig. 1

Low frequency coefficients A_k of every image and their weight maps $\{W_k^F\}$ are multiplied by the Hadamard product ($*$) and then summed up. Thus, the fused low frequency coefficients A^F are obtained, as shown in

$$A^F = \sum_{k=1}^{K} W_k^F * A_k \qquad (5)$$

2.2   High frequency coefficients fusion

High frequency coefficients reflect image details like edge and texture. The higher the high frequency coefficient is, the more obvious the illumination changes are and the richer the edges and textures are. Sum Modified-Laplacian reflects image details and definition correctly. Compared with other measurements of definition, SML has better ability to distinguish definition. SML also performs excellently for image fusion. In order to get results with better visual effect and extensive details, "Maximum region SML" is adopted for high frequency coefficients fusion.

Decomposed by DT-CWT with L decomposion level, an image generates $6 \cdot L$ high frequency coefficient matrices $\{D_{(l,\theta)}\}$ ($l = 1,2,\cdots,L$, $\theta = \pm15°$, $\pm45°$, $\pm75°$). Therefore, an image sequence with the number of K produces $K \cdot 6 \cdot L$ such matrices $\{D_{(l,\theta)}^k\}$. For each $\{D_{(l,\theta)}^k\}$, its Sum Modified-Laplacian $ML_{(l,\theta)}^k(i,j)$ is formulated in

$$ML_{(l,\theta)}^k(i,j) = |L_x * D_{(l,\theta)}^k(i,j)| + |L_y * D_{(l,\theta)}^k(i,j)| \qquad (6)$$

where (i, j) indicates the coordinate. $*$ means Hadamard product. L_x and L_y indicate second-order difference operator at x-direction and y-direction, which are shown in

$$L_x = \begin{bmatrix} -1 & 2 & -1 \end{bmatrix} \qquad (7)$$

$$L_y = \begin{bmatrix} -1 & 2 & -1 \end{bmatrix}^{\mathrm{T}} \qquad (8)$$

Region SML of $\{D_{(l,\theta)}^k\}$ is formulated in

$$SML_{(l,\theta)}^k(i,j) = \sum_{n=-(N-1)/2}^{(N-1)/2} \sum_{m=-(M-1)/2}^{(M-1)/2} [ML_{(l,\theta)}^k(i+m, j+n)]^2 \qquad (9)$$

$M \times N$ is a neighborhood of (i, j). $M \times N$ is usually set to 3×3 or 5×5. In this study 3×3 is selected. For k high frequency coefficients matric $\{D_{(l,\theta)}^k\}$

in θ direction with l decomposion level, the high frequency coefficient which has the largest region SML, are chosen as the fused high frequency coefficient. This rule is formulated in

$$D_{(l,\theta)}^F(i,j) = D_{(l,\theta)}^k(i,j) \quad \text{if}$$
$$SML_{(l,\theta)}^k = \max\{SML_{(l,\theta)}^k, \ k = 1,2,\cdots,K\} \qquad (10)$$

3   Experiment and discussion

Fig. 6(a) and (b) show the multi-exposure sequence "Desk" with the size of 780×520. Fig. 5(c) ~ (g) show the fusion results using methods proposed by Refs[2-6]. The result is shown in Fig. 6(h). In Fig. 6(d) and (e), there exists a black shadow at the right upper corner. This phenomenon exists in Fig. 6(c), but not serious. At the right upper corner in the source images is wall space, and it should not be black. Fig. 6(d) and (e) perform inaccurately in terms of local brightness. Fig. 6(g) is darker than other results. The result shows that the intensity is appropriate and suitable for human vision.

In order to evaluate the proposed algorithm in this paper objectively, spatial frequency (SF), figure definition (FD), average gradient (AG), and edge information transferred from source images to the fused image ($Q^{ab/f}$) [11] are adopted for evaluating the performance of different exposure fusion methods.

Spatial frequency (SF) is the time of repeated changes in the image function per unit length. SF reflects the change characteristics in the spatial domain and the sharpness of the fused image. Figure definition (FD) reflects the clarity of images. This paper adopts rate of gray value for measure figure definition. Average gradient (AG) is the average gray gradient of an image. AG measures the features of small details, namely, contrast and texture change of the fused image. SF, FD and AG are image quality evaluation metrics and they measure the fused image quality. For SF, FD and AG, a larger value indicates a better result. $Q^{ab/f}$ calculates how much edge information is transferred from the input images to the fused image. $Q^{ab/f}$ is used to measure the performance of fusion algorithm used to from the perspective of similarity. It generates

a quality value that ranges from 0 to 1, and a higher value indicates better quality.

The objective evaluation metrics of Fig. 6 is shown in Table 1. Based on the 4 evaluation criteria, it can be seen that the proposed method performs the best compared with other 5 methods.
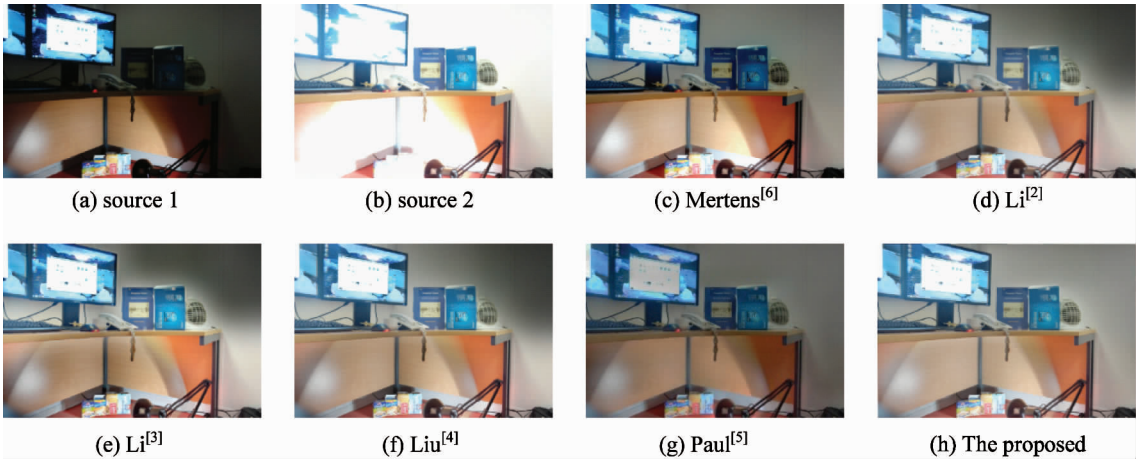

(a) source 1	(b) source 2	(c) Mertens[6]	(d) Li[2]
(e) Li[3]	(f) Liu[4]	(g) Paul[5]	(h) The proposed

Fig. 6　"Desk" sequence and its exposure fusion results using 6 different methods

Table 1　Objective evaluation metrics of "Desk" sequence

Indices	Mertens[6]	Li[2]	Li[3]	Liu[4]	Paul[5]	The proposed
SF	20.308	17.963	22.162	18.509	12.966	22.723
FD	5.785	5.131	6.259	5.277	4.109	6.752
AG	4.387	3.963	4.837	4.079	3.108	5.042
$Q^{ab/f}$	0.879	0.863	0.888	0.869	0.640	0.902

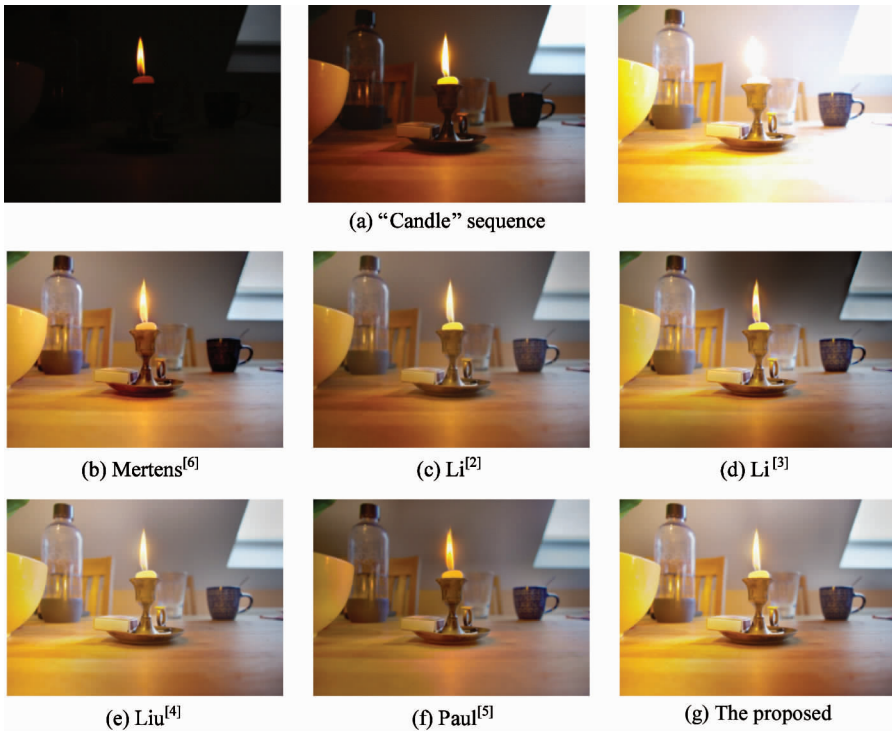| (a) "Candle" sequence		
(b) Mertens[6]	(c) Li[2]	(d) Li[3]
(e) Liu[4]	(f) Paul[5]	(g) The proposed

Fig. 7　"Candle" sequence and its exposure fusion results using 6 different methods

Fig. 7(a) shows the multi-exposure sequence "Candle", and Fig. 7(b) ~ (g) display the results using the 6 methods. The image size is 512×364. In Fig. 7(b) ~ (e), the brightness of the wall around the window is non-uniform and inappropriate, especially in Fig. 7(d). In Fig. 7(f), the global brightness is darker too and the proposed result has a good visibility.

The objective evaluation metrics of Fig. 7 is shown in Table 2. As can be seen, the results have the best SF, FD and AG and $Q^{ab/f}$.

Table 2   Objective evaluation metrics of "Candle" sequence

Indices	Mertens[6]	Li[2]	Li[3]	Liu[4]	Paul[5]	The proposed
SF	8.440	6.613	8.137	6.778	7.064	8.810
FD	2.906	2.323	3.032	2.651	2.727	3.317
AG	2.456	2.004	2.471	2.296	2.163	2.726
$Q^{ab/f}$	0.828	0.802	0.844	0.816	0.805	0.861

The proposed algorithm is widely used for various multi-exposure sequences and can obtain satisfying results. More fusion results are shown in Fig. 8. The proposed multi-exposure fusion method has a good robustness in various scenes.
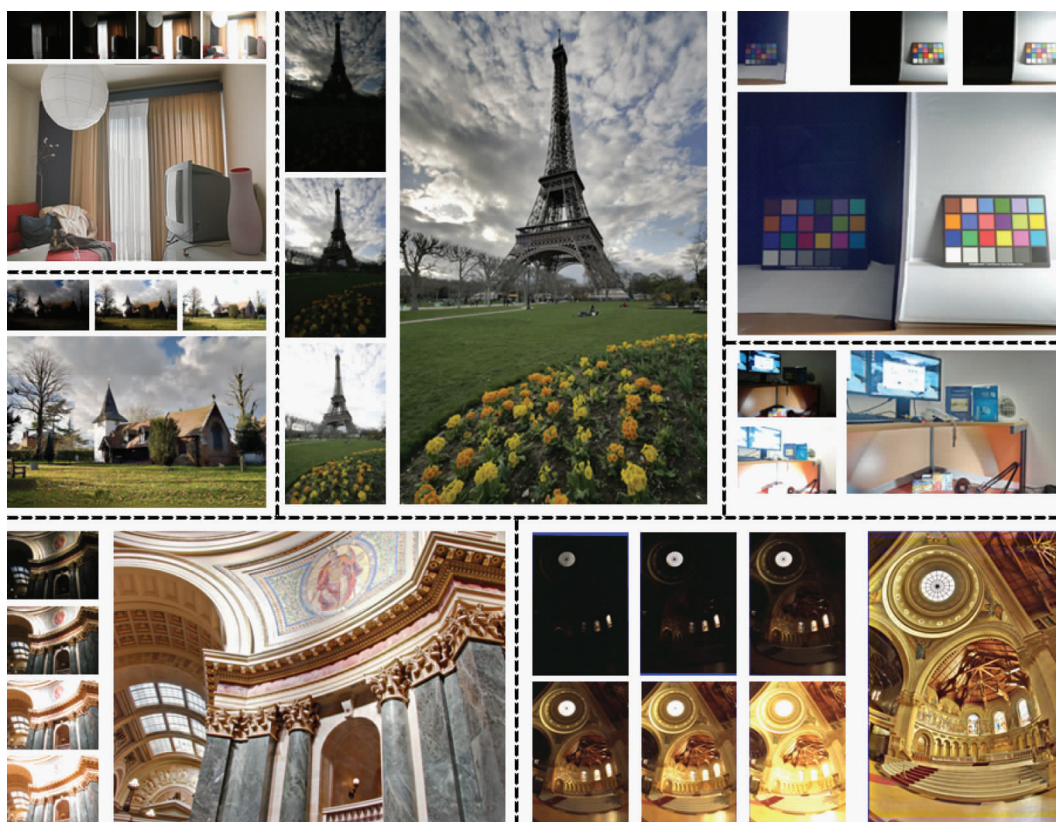

Fig. 8   Results of different image sequences

4   Conclusion

A multi-exposure fusion algorithm is proposed for high dynamic range scene. The multi-exposure images are decomposed into low and high frequency component by DT-CWT. 3 different weight functions depending on brightness analysis of the images are then proposed for low frequency coefficients fusion, which generate fusion results with suitable brightness. For high frequency coefficients fusion, maximizing region Sum Modified-Laplacian is adopted. Thus, the image details are preserved effectively. Experiments demonstrate that the proposed algorithm generates high-quality images in various scenes. The obtained images have suitable overall brightness and preserve details and edges effectively.

The proposed algorithm is efficient and robust in varies scenes.

References

[1] Shen X H, Liu J H, Chu G S. Auto exposure algorithm for aerial camera based on histogram statistics method. *Journal of Electronics & Information Technology*, 2016, 38(3): 541-548

[2] Li S T, Kang X D. Fast multi-exposure image fusion with median filter and recursive filter. *IEEE Transactions on Consumer Electronics*, 2012, 58(2): 626-632

[3] Li S T, Kang X D, Hu J. Image fusion with guided filtering. *IEEE Transactions on Image Process*, 2013, 22(7): 2864-2875

[4] Liu Y, Wang Z F. Dense SIFT for ghost-free multi-exposure fusion. *Journal of Visual Communication and Image Representation*, 2015, 31(C): 208-224

[5] Paul S, Sevcenco S I, Agathoklis P. Multi-exposure and Multi-Focus Image Fusion in Gradient Domain. *Journal of Circuits*, *Systems and Computers*, 2016, 16(6): 123-131

[6] Mertens T, Kautz J, Reeth F V. Exposure fusion: a simple and practical alternative to high dynamic range photography. *Computer Graphics Forum*, 2009, 28(1): 161-171

[7] Kingsbury N. The Dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement. In: Proceeding of the 9th European Signal Processing Conference, Island of Rhodes, Greece, 1998. 1-4

[8] Selesnick I W, Baraniuk R G, Kingsbury N. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 2005, 22(6): 123-151

[9] Zhao H, Li Q, Feng H. Multi-focus color image fusion in the HSI space using the sum-modified-laplacian and a coarse edge map. *Image & Vision Computing*, 2008, 26(9):1285-1295

[10] Shaked D, Keshet R. Robust recursive envelope operators for fast Retinex, HPL-2002-74 (R. 1). Haifa: Hewlett Packard Research Laboratories, 2004

[11] Petrovic V, Xydeas C S. Objective image fusion performance characterisation. In: Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China, 2005. 17-21

Shen Xiaohe, born in 1990. She received the Ph. D degree from the University of the Chinese Academy of Sciences in 2017. She received the B. S. degree in applied physics from Jilin University in 2012. She works in Beijing Institute of Aerospace Control Devices now. Her current research interests include automatic exposure control and exposure fusion.