Exposure Fusion Using Boosting Laplacian Pyramid

Jianbing Shen, *Senior Member, IEEE,* Ying Zhao, Shuicheng Yan, *Senior Member, IEEE,* and Xuelong Li, *Fellow, IEEE*

***Abstract*—This paper proposes a new exposure fusion approach for producing a high quality image result from multiple exposure images. Based on the local weight and global weight by considering the exposure quality measurement between different exposure images, and the just noticeable distortion-based saliency weight, a novel hybrid exposure weight measurement is developed. This new hybrid weight is guided not only by a single image's exposure level but also by the relative exposure level between different exposure images. The core of the approach is our novel boosting Laplacian pyramid, which is based on the structure of boosting the detail and base signal, respectively, and the boosting process is guided by the proposed exposure weight. Our approach can effectively blend the multiple exposure images for static scenes while preserving both color appearance and texture structure. Our experimental results demonstrate that the proposed approach successfully produces visually pleasing exposure fusion images with better color appearance and more texture details than the existing exposure fusion techniques and tone mapping operators.**

***Index Terms*—Boosting Laplacian pyramid, exposure fusion, global and local exposure weight, gradient vector.**

I. Introduction

THE DYNAMIC range of a natural scene often spans a much larger scope than the capture range of common digital cameras. An exposure image only captures a certain dynamic range of the scene and some regions are invisible due to under-exposure or over-exposure. Variable exposure photography captures multiple images of the same scene with different exposure settings of the camera while maintaining a constant aperture. In order to recover the full dynamic range and make all the details visible in one image, high dynamic range (HDR) imaging techniques [3], [22], [32] are employed to reconstruct one HDR image from an input exposure sequence. These generated HDR images usually have higher fidelity than convectional low dynamic range (LDR) images, which have been widely applied in many computer vision and image processing applications, such as physically-based realistic images rendering [32] and photography enhancement [30]. On the other hand, the current displays are only capable of handling a very limited dynamic range. In order to resolve the contradiction between the HDR characteristic of real world scenes and the conventional LDR display devices, a single HDR image is first reconstructed from multiple exposure images using HDR imaging techniques [3], [29] and then a tone mapping image is generated with existing tone mapping operators [9], [13], [14], [28]. This workflow of two phases is not efficient, while the process of direct exposure fusion combines multiple exposure images into a single LDR image without the aforementioned two phases.

Exposure fusion [18], [20], [23]–[25], [27], [30] is currently a very active research area in the field of computer vision, as it offers the full dynamic range from an input exposure sequence. The task of exposure fusion is slightly different from the traditional HDR imaging technique [3], and it does not need to reconstruct a single HDR image from a set of images under different exposure settings from the same scene [32]. The traditional image fusion techniques are most relevant to our algorithm, but image fusion algorithms focus on preserving the details and can be viewed as an analogy to alpha blending [2], [4], [16]. The purpose of exposure fusion is to acquire the full dynamic range of a scene by blending multiple exposure images into a single high-quality composite image, and preserving the detail and texture information as much as possible.

The general image fusion approaches usually use the multisensor or multispectral images as input [5], [16], while exposure fusion methods use the multiple exposure images as input. Taking a sequence of images with different exposures of a scene as input, our approach produces a detail and texture preserving fusion result using the boosting Laplacian pyramid. The proposed fusion method neither needs to generate any HDR image nor needs to do the tone mapping process. This

Manuscript received September 4, 2012; revised September 4, 2013; accepted November 5, 2013. Date of publication November 22, 2013; date of current version August 14, 2014. This work was supported in part by the Key Program of NSFC-Guangdong Union Foundation under Grant U1035004, in part by the National Basic Research Program of China (973 Program) under Grant 2013CB328805, in part by the National Natural Science Foundation of China under Grant 61272359, Grant 91120302, and Grant 61125106, in part by the Program for New Century Excellent Talents in University (NCET-11-0789), and in part by the Shaanxi Key Innovation Team of Science and Technology (2012KCT-04). This paper was recommended by Associate Editor M. Cetin.

J. Shen and Y. Zhao are with the Beijing Key Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China (e-mail: shenjianbing@bit.edu.cn; bhyzhao@bit.edu.cn).

S. Yan is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore (e-mail: eleyans@nus.edu.sg).

X. Li is with Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, China (e-mail: xuelong_li@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2013.2290435

2168-2267 © 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Fig. 1. Exposure fusion results using the boosting Laplacian pyramid. Our result exhibits more details with enhanced texture information, especially in the cloud background and light glares, and the overall visual appearance contrast. (a) Input sequence. (b) Result by [27]. (c) Our result.

avoids the requirement of knowing the exposure time of capturing different exposure images. Our fusion method produces a more visually pleasing fusion result using the boosting Laplacian pyramid and the exposure quality measurements. As shown in Fig. 1, our result exhibits more details with enhanced contrast and texture information, and achieves a more visually pleasing result than the method in [27], especially in terms of the details of the cloud background, light glares, and road texture.

Compared with previous works on exposure fusion, we can see that our paper has the following research contributions.

1) A new exposure fusion approach is proposed, which is based on the novel boosting Laplacian pyramid and the hybrid exposure weight.
2) A novel hybrid exposure weight measurement is presented, which is based on the local weight, the global weight and the just noticeable distortion (JND)-based saliency weight. Our novel global exposure weight considers the gradient vectors between different exposure images.
3) A new boosting Laplacian pyramid is developed, which is based on the structure of boosting the detail layers and base layers respectively, and the boosting process is guided by the proposed local exposure weight and the JND-based saliency weight.

The paper is organized as follows. In Section II, we summarize the related work on exposure fusion and HDR imaging techniques. In Section III, we present our exposure fusion algorithm by designing a novel hybrid exposure weight. In Section IV, we derive a new boosting Laplacian pyramid by boosting the detail layers and base layers respectively. We give the fusion examples to validate the proposed method and then conclude the paper with a brief discussion in Section V.

II. Related Work

Exposure fusion [6], [11], [15], [23] is originally introduced as a subarea of the more general topic of image fusion approaches, which creates the fusion result with more details of the same scene taken under different exposure settings. However, the general image fusion approaches [16] are referred to the fusion of multisensor and multispectral images, and are usually restricted to merging the gray level images. Exposure fusion is a recently developed technique for producing the details and expanding the dynamic range without the need of creating an intermediate HDR image. For instance, Goshtasby [15] has proposed an exposure fusion method from multiple exposure images of a static scene. Their approach blends the image blocks in an image domain by selecting the uniform image blocks containing the most useful information. In this paper, we focus on the multiple exposure fusion image fusion and present a novel fusion approach using the hybrid exposure weight and the new boosting Laplacian pyramid.

In recent years, a few approaches of multiple exposure fusion have been proposed and are widely used for their simplicity and effectiveness. Mertens *et al.* [23] presented an exposure fusion approach using Gaussian and Laplacian pyramid in a multiresolution fashion, which belongs to the pixel-level fusion approach. Raman and Chaudhuri [25] employed the edge-preserving bilateral filters to generate an exposure fusion result from multiexposure input images. Raman *et al.* [26] further proposed an automatic multiexposure fusion approach without the ghost artifacts by determining the moving objects. Recently, Zhang and Cham [27] presented a tone-mapping-like exposure fusion method under the guidance of gradient-based quality assessment. Later, Song *et al.* [37] synthesized an exposure fusion image using a probabilistic model, which preserves the luminance levels and suppresses reversals in the image luminance gradients. More recently, Hu *et al.* [35] presented a novel registration and fusion approach for the exposure images in the presence of both live scenes and camera motions.

The dynamic ranges of HDR images need to be mapped into the limited dynamic ranges of the dominant display technologies, such as CRTs and LCDs. The works on HDR

imaging and tone mapping techniques [8], [9], [13] are closely related to exposure fusion approaches. The HDR imaging approaches usually create a single HDR image with a good alignment of multiple input exposures and recover the camera response curve [3], [32]. Tone mapping techniques for HDR images have been fascinating for a long time, and many methods have been developed during the last decade. For instance, Reinhard *et al.* [8] presented a tone mapping method to compress the high dynamic range of the real world luminance to the low dynamic range of the photographic print. Recently, Li *et al.* [14] presented a wavelet-based sub-band decomposition for compressing and companding HDR images. In order to remove the halo artifacts caused by the amplitude distortion, they used a symmetrical analysis-synthesis wavelet-based filter bank and local gain control to improve the modified sub-bands. Khan *et al.* [17] proposed a ghosting-free HDR imaging method by a non-parametric model to determine each pixel's weight for the static part of a scene. More recently, Jacobs *et al.* [21] presented an automatic HDR image generation method from LDR images, which removes the ghosting artifacts in the HDR images by a camera-alignment module and a movement detector.

Compared with the traditional tone mapping techniques, the proposed exposure fusion approach does not need to recover the camera response curve and record the exposure time of an input sequence of different exposure images [3]. As described earlier, the conventional two stages include the HDR image creation and tone mapping, and this two-stage process usually needs complex user interactions and tends to miss some color and texture details in the created fusion result. Therefore, it is desirable to produce the fusion result from a multiple exposure sequence input, which is more efficient and effective. The purpose of our exposure fusion technique is to directly produce visually pleasing photographs for the displaying purpose while preserving the details and enhancing the contrast.

III. Exposure fusion algorithm

We now introduce and discuss our exposure fusion method, which is based on a new boosting Laplacian pyramid (BLP). This new pyramid framework is very efficient to fuse multiple exposure images, as it has the advantage of boosting both the structure and texture details while avoiding the phenomenon of color cast. Our method first begins with decomposing an input sequence into multiscale detail and base layers using the proposed BLP. We will give the detailed descriptions on how to construct our boosting Laplacian pyramid in Section IV. Our exposure fusion process is guided by the weight maps for each layer of boosting Laplacian pyramid as follows:

$$R(x, y) = \sum_{i=1}^{N} W_i(x, y) \times F_i(x, y) \tag{1}$$

where $R(x, y)$ denotes the image pixel at coordinate position (x, y) of the fusion result, $W_i(x, y)$ indicates the hybrid exposure weight, and $F_i(x, y)$ is one layer of our boosting Laplacian pyramid ($F_i(x, y) = B_i(x, y) + D_i(x, y)$, $B_i(x, y)$ is the base layer and $D_i(x, y)$ is the detail layer) with the i-th exposure image. N is the number of images in the input sequence. In our implementation, we use different hybrid exposure weight $W_i(x, y)$ for base layers and detail layers in (1), and we will further discuss it in detail in Section IV.

Our goal is to automatically find the useful visibility information from each input exposure image, and then combine these regions together so as to generate a high quality fusion result with more details. We propose three different guidance methods to identify each pixel's contribution to the final fusion components, and we consider both the global and local exposure weight for multiple exposure fusion. Our method is different from the approach in [23], which emphasizes the local cues to determine the weight maps, such as the contrast, saturation and exposure measurement. In contrast, we measure the exposure weight by computing the exposure levels from both the local and global information. The local exposure weight $E(x, y)$ is used to determine the useful pixels needed to be preserved during the fusion process. The global exposure weight is calculated by considering the exposure changes between the current exposure image with another exposure image. We present a global exposure weight method using the gradient vector, since the gradient vector does not change dramatically in the proper exposure image. We then compute the exposure difference with gradient direction from the multiple exposure images, and obtain the global exposure weight $V(x, y)$. In order to preserve the salient regions, we propose the third guidance to compute the saliency weight $J(x, y)$ using the JND model.

In order to obtain a new hybrid exposure quality weight $W_i(x, y)$ by considering the global exposure weight, local exposure weight and saliency weight for the i-th image of the input sequence, we combine the aforementioned three cues together and normalize them as follows:

$$W_i(x, y) = \frac{E_i(x, y) \times V_i(x, y) \times J_i(x, y)}{\sum_{i=1}^{N} E_i(x, y) \times V_i(x, y) \times J_i(x, y)}. \tag{2}$$

A. Local exposure weight

Both under-exposure and over-exposure usually reveal some regions and also make other regions of the image invisible [33]. Therefore, the fusion should be performed on the correct regions so as to increase the image visibility and reduce the visual artifacts. It will be invalid to directly fuse and enhance these regions in under-exposed or over-exposed areas. We use an asymmetric logic function to evaluate the exposure quality [33]. This exposure quality assessment $Q(x, y)$ sets the lightest and darkest regions with zero values, while it assigns other regions with the values between zero and one. Our exposure weight map $E(x, y)$ is the grayscale image of $Q(x, y)$ and represents the exposure quality of the input image $I(x, y)$, which is defined as

$$Q_i(x, y) = 1 - \left| \frac{1}{a} \left\{ log \frac{I_i^{\frac{1}{c}}(x, y)}{1 - I_i^{\frac{1}{c}}(x, y)} - b \right\} \right| \tag{3}$$

$$E_i(x, y) = rgb2gray(Q_i(x, y)) \tag{4}$$

where the parameters a, b, and c are used to control the shape of the asymmetric logic function curve [33]. We set $a = 3.2$,

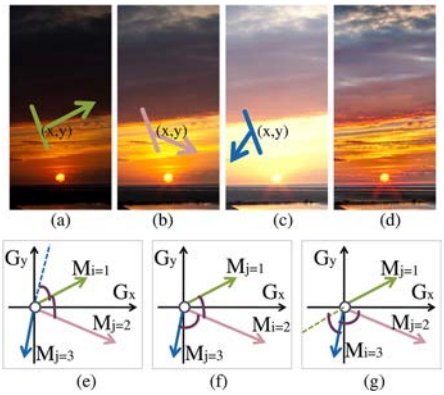

Fig. 2. Illustration of analyzing the gradient vectors with different exposure images. Note that (a), (b) and (c) are the input images. (a) Low level exposure. (b) Middle level exposure. (c) High level exposure. (d) Final fusion result. (e)–(g) Gradient vectors $M_i(x, y)$ and $M_j(x, y)$ at position (x, y) in i-th and j-th images.

$b = -1.3$, and $c = 0.4$ in our implementation. The exposure weight $E(x, y)$ is also used to construct our boosting Laplacian pyramid and we will discuss it in detail in Section IV-A.

B. Global exposure weight

The aforementioned exposure weight is a local weight map, which computes the exposure level only from a single exposure image of the input sequence. However, this local weight map does not utilize the global relationship of measuring the exposure level between different exposure images. Hence we define a global exposure weight $V(x, y)$ to make a better exposure measurement by considering other exposure images from the sequence.

We take advantage of the sine values of the angle between gradient vectors from different exposure images as shown in Fig. 2. We calculate the sine values of gradient vector pairs, which are obtained by selecting any two images from a sequence of N exposure images. Let $< M_i(x, y), M_j(x, y) >$ denotes the angle between gradient vectors $M_i(x, y)$ and $M_j(x, y)$ at a position (x, y) in the i-th and j-th images. The sine value $S_{ij}(x, y)$ is computed and smoothed as follows:

$$S_{ij}(x, y) = \|M_j(x, y)\| \times \sin < M_i(x, y), M_j(x, y) >,$$
$$S_{ij}(x, y) = mean(S_{ij}(x, y)) \quad (5)$$

where $mean(\cdot)$ indicates the mean filter and its window size is normally 9×9 pixels. The mean filter is used to produce the smooth $S_{ij}(x, y)$.

Since the value of $S_{ij}(x, y)$ is in the range of $[0, 1]$, and is very small, we use a factor λ to amplify it. Then the weight map $T_{ij}(x, y)$ is given by

$$T_{ij}(x, y) = 1 - \lambda \times S_{ij}(x, y) \quad (6)$$

where we set $\lambda = 100$ in our implementation.

In fact, $T_{ij}(x, y)$ may be below zero, which will cause the incorrect weight results. So we need the following thresholding function to modify it:

$$T'_{ij}(x, y) = \begin{cases} T_{ij}(x, y), T_{ij}(x, y) > 0; \\ 0, otherwise. \end{cases} \quad (7)$$

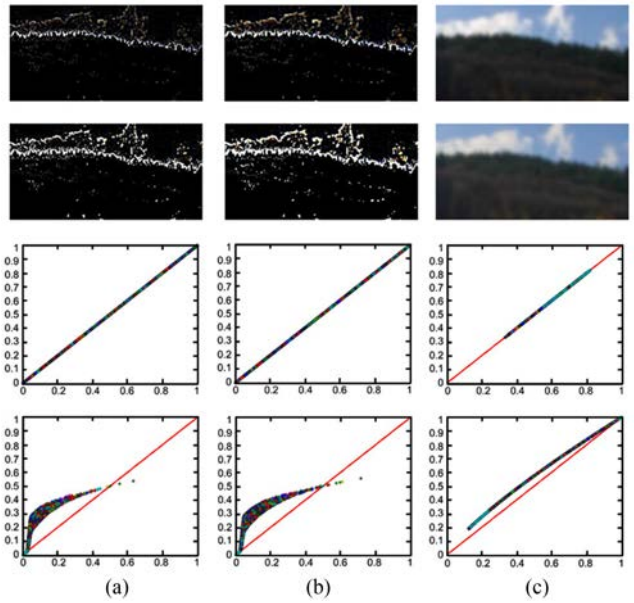

Fig. 3. Magnitude of the RGB color vector in the base layers and the detail layer is boosted along the direction of the RGB color vector. Top row: before boosting. Second row: after boosting. Third row: direction of the RGB color vector. Bottom row: magnitude of the RGB color vector. (a) Level 0. (b) Level 1. (c) Level 2.

Finally, we multiply the weight map of each exposure image to obtain its final global exposure level of the input sequence. Thus, the global exposure weight $V_j(x, y)$ of the j-th exposure image is defined as

$$V_j(x, y) = \prod_{i=1, i \neq j}^{N} T'_{ij}(x, y). \quad (8)$$

C. JND-Based Saliency Weight

JND refers to the maximum distortion that the human visual system (HVS) does not perceive, and defines a perceptual threshold to guide a perceptual image quality measurement task. The JND model helps us to represent the HVS sensitivity of observing an image [12]. It is an important visual saliency cue for image quality measurement. We employ the JND model to define the saliency weight $J(x, y)$, which can pick up the pixels in different exposure images with good color contrast and saturation regions. Furthermore, we can utilize a saliency weight map based function to estimate the level of boosting in our BLP. We use a nonlinear additivity JND model to define the saliency weight as [12], [31]

$$J_i(x, y) = J_i^l(x, y) + J_i^t(x, y) - K_{l,t}(x, y)\min(J_i^l(x, y), J_i^t(x, y)) \quad (9)$$

where $J_i^l(x, y)$ and $J_i^t(x, y)$ are the background luminance masking and the texture masking of the image visibility, respectively. $K_{l,t}(x, y)(0 < K_{l,t}(x, y) < 1)$ accounts for the overlapping effect in masking.

We adopt two different functions to model the relationship between luminance $J_i^l(x, y)$ and the average background luminance [12]. One is a root equation for low background

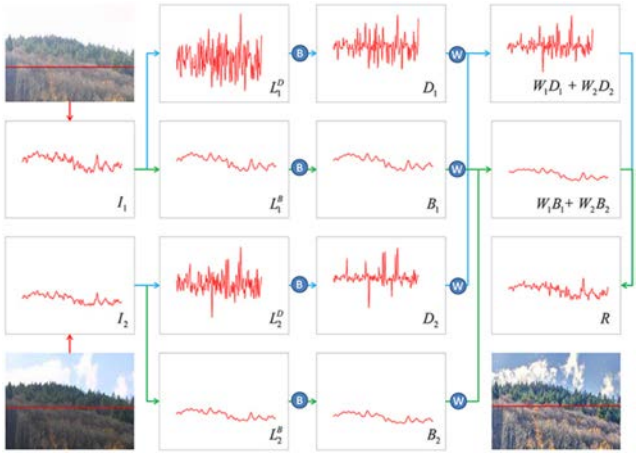

Fig. 4. Boosting and fusion process on the base layer and the detail layers separately. Note that here B in the blue circular icon denotes the boosting operator.

luminance (below 127), and the other is approximated by a linear function for the part over 127, which is defined as

$$J_i^l(x,y)=\begin{cases}17\left(1-\sqrt{\frac{\overline{I(x,y)}}{127}}\right)+3, if\ \overline{I(x,y)}\le 127\\ \frac{3}{128}(\overline{I(x,y)}-127)+3, otherwise.\end{cases}$$

$$\overline{I(x,y)}=\frac{1}{32}\sum_{i=1}^{5}\sum_{j=1}^{5}I(x-3+i,y-3+j)\cdot B(i,j), \tag{10}$$

$$B(i,j)=\begin{bmatrix}1&1&1&1&1\\1&2&2&2&1\\1&2&0&2&1\\1&2&2&2&1\\1&1&1&1&1\end{bmatrix}$$

where $I(x,y)$ is the intensity value and $B(i,j)$ is a weighted low-pass filter.

Texture masking is usually determined by local spatial gradients around the pixel. In order to obtain more accurate JND estimation, edge and nonedge regions should be well distinguished. Based on the fact that the edge structure attracts more attention by HVS [12], we take the edge difference into account to define $J_i^t(x,y)$ as

$$J_i^t(x,y)=\max_{k=1,2,3,4}\{|grad_k(x,y)|\},$$
$$grad_k(x,y)=I(x,y)\otimes g_k(x,y) \tag{11}$$

where $\otimes$ is a convolution operator. $g_k(x,y)$ is the k-th directional high-pass filter, which is defined as follows (k=4):

$$g_1=\begin{bmatrix}0&0&0&0&0\\1&3&8&3&1\\0&0&0&0&0\\-1&-3&-8&-3&-1\\0&0&0&0&0\end{bmatrix},\quad g_2=\begin{bmatrix}0&0&1&0&0\\0&8&3&0&0\\1&3&0&-3&-1\\0&0&-3&8&0\\0&0&-1&0&0\end{bmatrix},$$

$$g_3=\begin{bmatrix}0&0&1&0&0\\0&0&3&8&0\\-1&-3&0&3&1\\0&-8&-3&0&0\\0&0&-1&0&0\end{bmatrix},\quad g_4=\begin{bmatrix}0&1&0&-1&0\\0&3&0&-3&0\\0&8&0&-8&0\\0&3&0&-3&0\\0&1&0&-1&0\end{bmatrix}$$

IV. Boosting Laplacian Pyramid

In order to preserve the color and texture details as much as possible, we propose a novel exposure fusion approach using the aforementioned weight map and the BLP. Our boosting Laplacian pyramid enhances the performance of classical Laplacian pyramids [1], which is based on the standard recursive downsampling and upsampling procedure. The Laplacian pyramid has been employed to decompose images into multiple scales and is widely used for a variety of image processing applications, such as detail enhancement [19] and exposure fusion [23]. Unlike the existing Laplacian pyramid [1], [19] and edge-aware Laplacian pyramid [34], our boosting Laplacian pyramid is designed especially for the task of exposure fusion. We first decompose the input signal into the detail layer and the base layer of each level, and then the detail boosting stage and the base boosting stage are applied separately. Secondly, our boosting process is guided by the aforementioned local exposure weight and JND-based saliency weight with different exposure images. It is very useful to correctly select the salient regions to boost, and the boosting level is controlled by the exposure quality measurement.

In order to avoid the color casting artifacts, we multiply the RGB triplet by a scalar, which keeps the chromaticity unchanged before and after signal boosting and reduces the color distortion. Fig. 3 gives an illustration of our boosting on the magnitude of the RGB color vector along the direction of this color vector. The direction of the RGB color vector lies in the line of 45 degrees diagonal, which means that the direction of this color vector does not change before and after using our boosting Laplacian pyramid (Fig. 3, third row). At the same time, the magnitude of the RGB color vector increases nonlinearly before and after the boosting process (Fig. 3, bottom row). Our boosting method preserves the structure and texture details so as to get a natural and visually pleasing fusion result.

Now we will address the structure of our boosting Laplacian pyramid for exposure fusion. For clarity, we take two exposure images (I_1 and I_2 when $N=2$) as an input sequence for illustration simplicity. Each input exposure image is first decomposed into the detail layers and base layer using the Gaussian pyramid. One base layer (L_{12}^B) and two detail layers (L_{11}^D and L_{12}^D) of the input image I_1 are created when the number of pyramid levels is 3. Then the detail layers are boosted using our detail boosting guidance and (13), and the base layer is boosted using our base boosting guidance and (14). After the above boosting stages, we obtain the enhanced base layer (B_1) and the enhanced detail layers (D_{11} and D_{12}). Using the same decomposition and boosting mechanism, we obtain the boosted base layer (B_2) and detail layers (D_{21} and D_{22}) for the second exposure image. Finally, we obtain the fusion result by combining the weight maps and the boosted base and detail layers as follows:

$$R=\sum_{i=1}^{N=2}W_{i1}D_{i1}+(\sum_{i=1}^{N=2}W_{i2}D_{i2})\uparrow+(\sum_{i=1}^{N=2}W_{i3}B_i)\uparrow \tag{12}$$

where R is the final fusion result, W is the hybrid exposure weight computed by (2), and the symbol $\uparrow$ is the image upsampling operator.

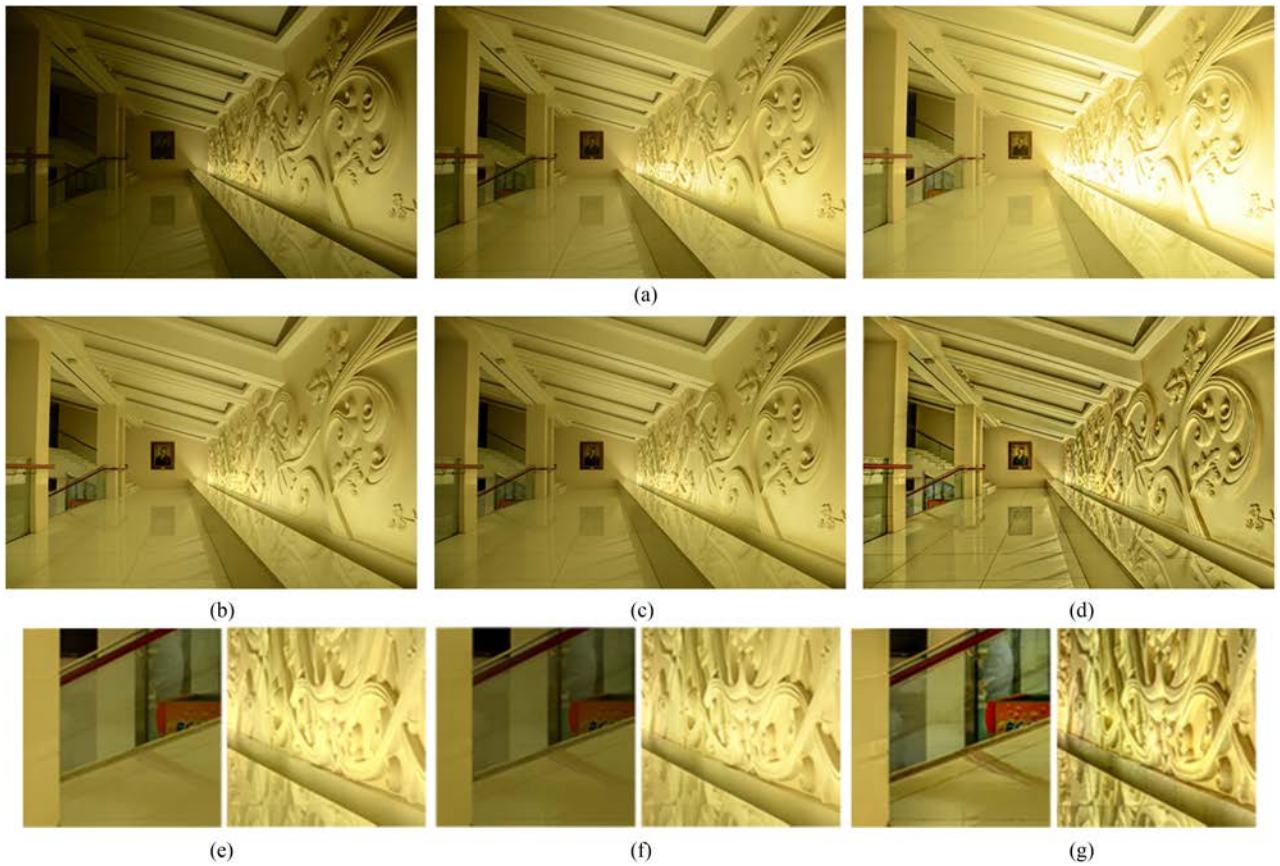

Fig. 5. Performance comparisons on the input sequence with the exposure fusion approach [23] and the gradient directed fusion approach [27]. (a) Input sequence. (b) Result by [23]. (c) Result by [27]. (d) Our result. (e) Close-up of (b). (f) Close-up of (c). (g) Close-up of (d).

The detail layer is obtained by the original signal subtracting the Gaussian pyramid signal, which is based on the standard image Laplacian pyramid [1]. Our BLP exhibits the new advantages of boosting the base layer and detail layer signal according to the boosting guidance, which is computed by our local exposure weight and JND-based saliency weight introduced in Section III.

A. Boosting guidance

As mentioned before, our boosting process is guided by the two image quality measurements, the local exposure weight guidance and the JND-based saliency weight guidance. We first use the exposure weight guidance to select the well exposed regions which need to be boosted using our BLP. This strategy can avoid the visual artifacts. It is beneficial to boost the base layer and detail layers with different extents for each pixel according to the guidance map, since the well-exposure regions and under-exposure or over-exposure regions of the sequence should be enhanced with different amplifying values during the boosting process.

In order to select the useful pixels from the different exposure images, we compute the exposure weight map to guide the boosting level in different pyramid layers. To achieve this goal, we compute the exposure weight map by applying a threshold operation σ to the local exposure weight $E(x, y)$ for the detail layers as follows:

$$G_i^E(x, y) = \begin{cases} 1, E_i(x, y) > \sigma; \\ 0, otherwise. \end{cases} \tag{13}$$

where σ equals 0.01 in our implementation.

For the pixels with $G^E(x, y) = 0$, our boosting process does not operate. Then, we use the saliency weight $J(x, y)$ to further compute the boosting level of each pixel. As mentioned in Section III-C, this saliency weight $J(x, y)$ is a JND-based image quality measurement, and we obtain the updated saliency weight map $G^J(x, y)$ for the base layers by the following equation:

$$G_i^J(x, y) = G_i^E(x, y) \times J_i(x, y). \tag{14}$$

B. Boosting function

As shown in Fig. 4, the input signal is decomposed into the base and detail signal using the Gaussian pyramid. We boost the base and detail layers separately, which is guided by the boosting guidance weight. The detail layer usually contains the structure feature of the original image while the base layer preserves color information. So we apply different boosting functions to enhance these different kinds of layers, and create the final fusion result by combining the boosted detail and base layers using the proposed local exposure weight and saliency

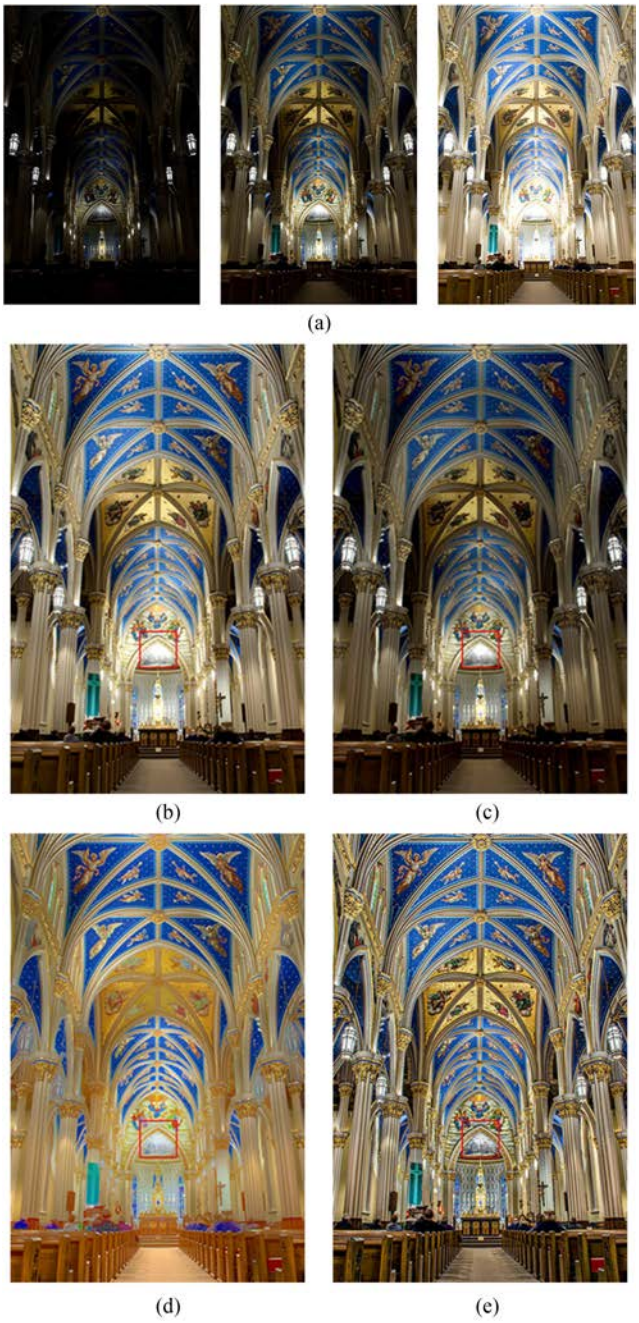

Fig. 6. Comparison of our approach with the exposure methods [23], [27] and the linear windowed tone mapping method [28]. (a) Input sequence. (b) Result by [23]. (c) Result by [27]. (d) Result by [28]. (e) Our result.

weight in Section III. In our implementation, we boost the detail layer $L^D(x, y)$ patch by patch as in [15] and [34].

Detail boosting: Our detail boosting function is inspired by the mechanism of the photoreceptor adaptation model [10], [13]. The classic retinal adaptation mechanism is the multiplicative retinal adaptation model, which multiplies input intensities by a variable between 0 and 1. This adaptation mechanism controls the ability of the human visual system to perceive the spatial pattern vision. The physiological evidence has been presented to measure the intensity-response relationship between different exposure light and varying adaption conditions [10], [13]. Electrophysiological research

Fig. 7. Close-ups of Fig. 6(b)–(e). The regions inside the red boxes in the second and third rows of Fig. 6(b)–(e) are zoomed in. The right bottom close-up is our result. (a) Result by [23]. (b) Result by [27]. (c) Result by [28]. (d) Our result.

has found that the intensity-response relationship of neurones in the vertebrate retina has a sigmoid shape curve. Thus, such intensity-response functions are often modeled by a generalized Naka-Rushton equation [10], [32]

$$D_i(x, y) = \frac{[L_i^D(x, y)]^{\alpha_i}}{[L_i^D(x, y)]^{\alpha_i} + \beta^{\alpha_i}}, \alpha_i = \alpha_0{}^{G_i^J(x,y)} \tag{15}$$

where $D(x, y)$ is the boosting response of the detail layer $L^D(x, y)$, β is the intensity parameter of the half-maximum response value, and α_i controls the response sensitivity according to the saliency weight with standard argument α_0 [32].

Base boosting: The multiple exposure fusion approach is often used to recover the HDR characteristics of a given scene. The pixel values of the same positions can change very drastically for the images which are taken under different exposure conditions. The base layer usually contains the large-scale features and the background color information of the exposure image. We use the geometric average of the importance guidance map to boost the base layer as the following power-like function:

$$B_i(x, y) = [L_i^B(x, y)]^{\gamma_i},$$
$$\gamma_i = \gamma_0 \exp\left(\frac{1}{n}\sum_{(x,y)}^{n} \log\left(G_i^J(x, y) + \epsilon\right)\right) \tag{16}$$

where $B(x, y)$ is the boosting response of the base layer $L^B(x, y)$, and n is the number of pixels in one exposure image. ϵ is a small constant (we set $\epsilon = 1e - 10$), which prevents the logarithm of zero. γ is a parameter that controls the response sensitivity according to the saliency weight $G^J(x, y)$ and the standard argument γ_0 [32].

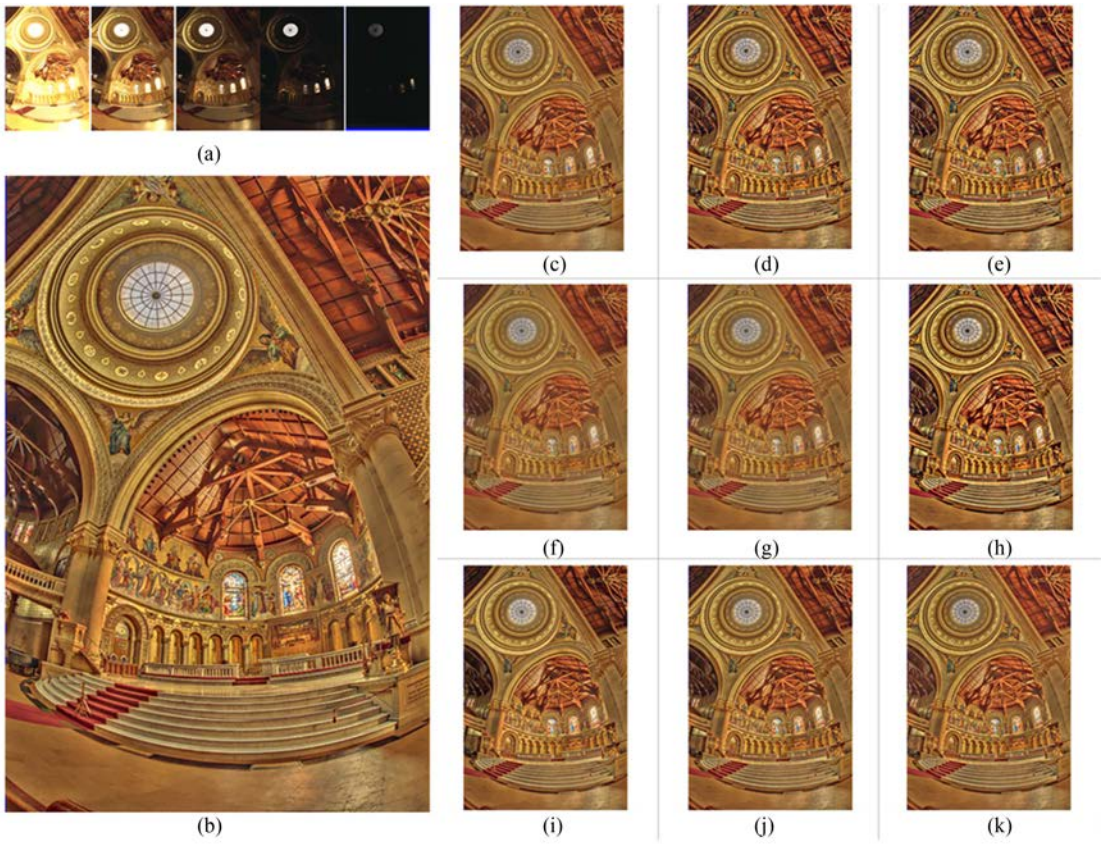

Fig. 11. Sensitivity analysis of the boosting parameters ($\alpha_0, \beta, \gamma_0$) on the Memorial church sequence. The result generated with $\alpha_0 = 0.2$, $\beta = 0.5$ and $\gamma_0 = [1, 1, 1, 1]$ is used as the reference image. (a) Input sequence. (b) Reference image. $\alpha_0 = 0.2; \beta = 0.5; \gamma_0 = [1, 1, 1, 1]$. (c) $\alpha_0 = 1$. (d) $\alpha_0 = 0.001$. (e) $\alpha_0 = 0.0001$. (f) $\beta = 1$. (g) $\beta = 0.8$. (h) $\beta = 0.2$.

V. Experimental Results

We have used several LDR image sequences for comparing with other existing exposure fusion approaches [23], [27] and HDR image tone mapping algorithms [8], [14], [28]. Fig. 5 shows the comparison of our exposure fusion algorithm with other two fusion methods [23], [27] on the city concert hall image sequence. The results of exposure fusion approach [23] in all experiments are generated by direct implementation from its webpage[1] in our comparison. Our result is comparable to both the results by [23] and [27]. However, our fusion result exhibits better contrast and color details than the results by [23] and [27]. In particular, our approach gives more details in the region of the corridor area and the wall with the carved flower pattern [Fig. 5(g)]. Fig. 6 gives a more challenging example, where our result is compared with other image fusion methods [23], [27] and some tone mapping methods [28]. In all our experiments, we use the HDRShop[2] software to generate the HDR images from the corresponding LDR sequence images. The results of linear windowed tone mapping are generated by the MATLAB implementation provided by their website.[3] In order to give a relatively fair fusion quality comparison with other approaches, we run their programs in the default parameter settings in all our experiments. The result by [28] suffers some color artifacts, such as the color distortion in the region of the church ceiling. In contrast, our result gives a more pleasing result with texture details than the results by the approaches in [23], [27], and [28], and close-up comparison is shown in Fig. 7.

In order to better understand the influence of different weight maps during the fusion process, we visualize the local exposure weight $E(x, y)$, global exposure weight $V(x, y)$, and saliency weight $J(x, y)$. As shown in Fig. 8, the color-coded weight maps include the aforementioned local exposure weight [Fig. 8(a)–(c)], global exposure weight [Fig. 8(d)–(f)], and saliency weight [Fig. 8(g)–(i)]. These three weights indicate different aspects of image quality measurements, and the hybrid weight maps denote the overall contribution of the different exposure images to the final fusion result. As mentioned before, the basic principle of exposure fusion is to select the good exposure pixels from the input images, and preserve the detail and texture information as much as possible during the fusion process. Therefore, we define a local weight $E(x, y)$ and a global weight $V(x, y)$ to measure the exposure level of the input images. $E(x, y)$ selects the good exposure pixels while $V(x, y)$ eliminates the bad exposure regions of the current image. Thus, $V(x, y)$ helps $E(x, y)$ to compensate the exposure level measurement. $J(x, y)$ measures the visual saliency according to the background luminance and texture information. Regions with more texture details will get higher weights, which are visualized as the warmer color. By

[1] *http://research.edm.uhasselt.be/~ tmertens/exposure_ fusion*
[2] *http://www.hdrshop.com/*
[3] *http://www.cs.washington.edu/homes/shanqi/*

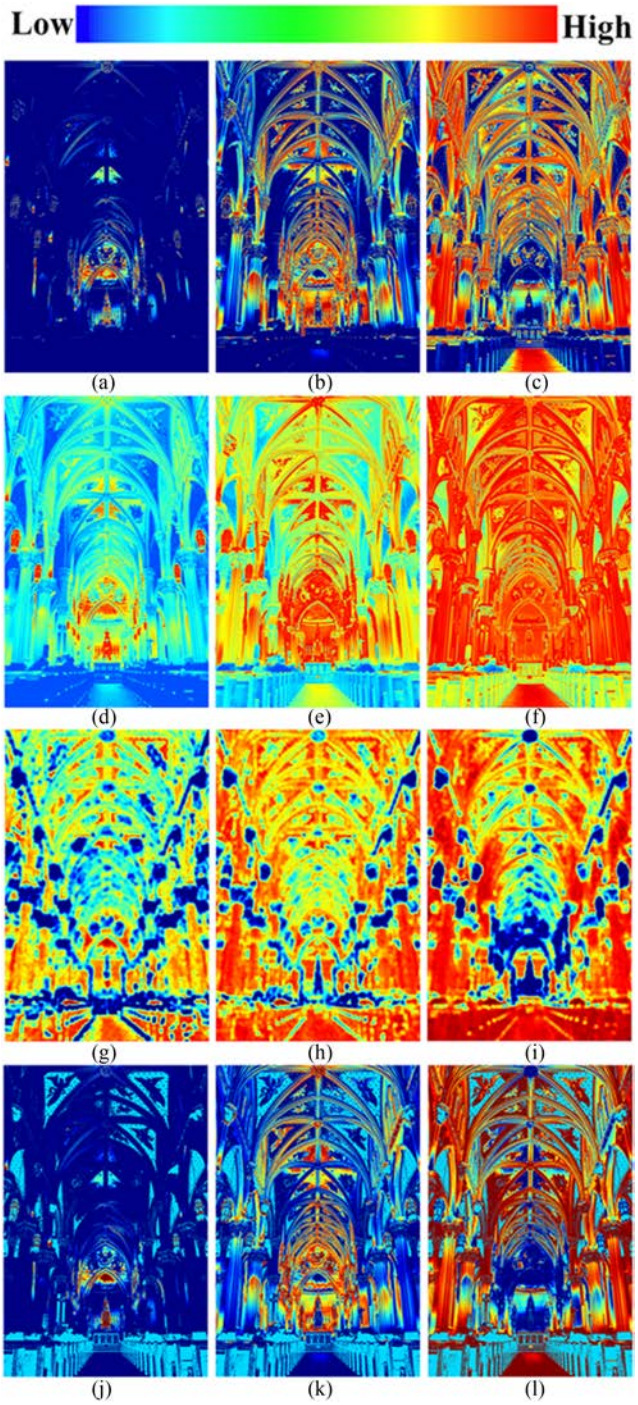

Fig. 8. Visualization of the proposed hybrid weight maps. (a)–(c): Local exposure weight maps E(x,y). (d)–(f): Global exposure weight maps V(x,y). (g)–(i): Saliency weight maps J(x,y). (j)-(l): Hybrid weight map W(x,y). Here the red color refers to large weights and blue means small weights, and the input sequence is shown in Fig. 6(a).

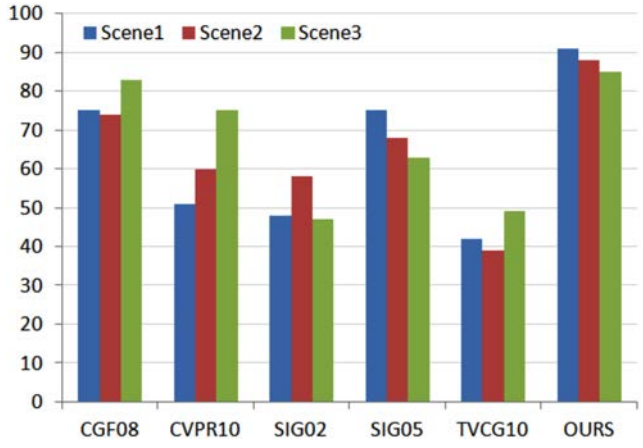

Fig. 9. Results of the user studies of our method and other approaches. Note that we abbreviate exposure fusion approach [23] (CGF08), gradient fusion method [27] (CVPR10), photographic tone reproduction [8] (SIG02), subband compression [14] (SIG05) and linear windowed tone mapping [28] (TVCG10). The input sequences of scenes 1–3 are shown in Figs. 6(a), 11(a), and 1(a), respectively.

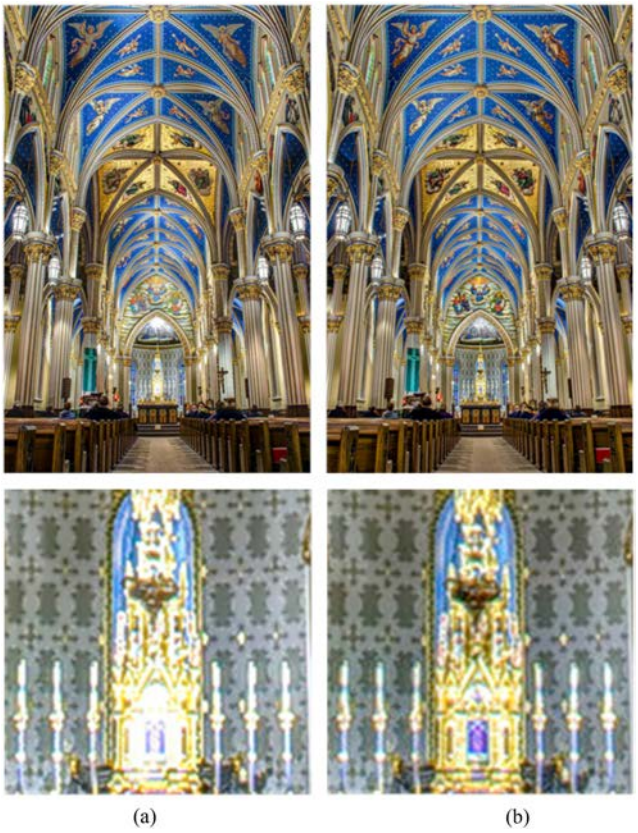

Fig. 10. Comparison results. (a) Our approach without the gradient direction. (b) Our full method including the direction of gradient.

considering three measurements simultaneously, we multiply them all together and then normalize the results to get the final fusion weight maps [Fig. 8(j)–(l)] for each image of the whole sequence.

The performance evaluation of exposure fusion results has so far been qualitative. In order to obtain a quantitative evaluation result, we conduct the user studies to determine which fusion result has the best visual quality. In our user experience study, we show the users 18 exposure fusion result images, which are obtained with three scenes by six different exposure fusion methods and tone mapping approaches. We invite 20 professional photographers to participate in the user experience study, and we ask each user to give one score for each fusion result. As shown in Fig. 9, the total accumulative satisfactory scores by the user study are 91, 88, and 85 for the fusion results by our method, which are higher than the

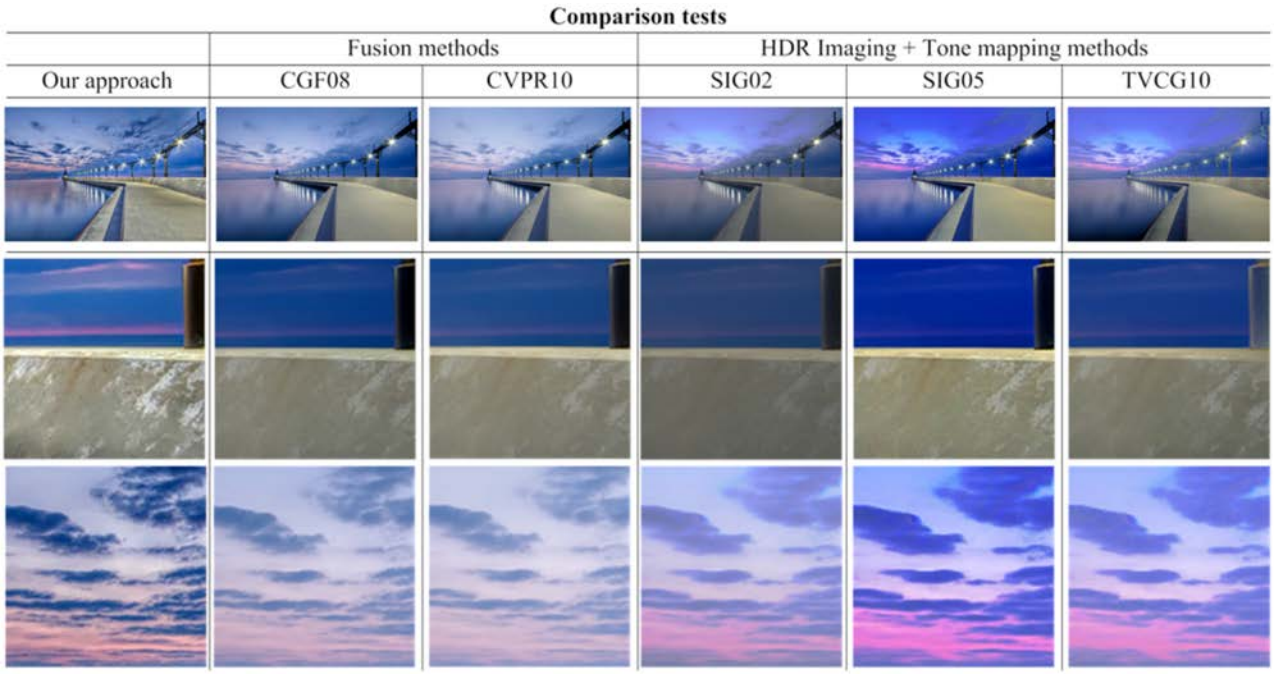

Fig. 12. Comparison with other exposure fusion and tone mapping methods. Top row: the fusion and tone mapping results. Middle and bottom rows: close ups. Note that we abbreviate exposure fusion approach [23] (CGF08), gradient fusion method [27] (CVPR10), photographic tone reproduction [8] (SIG02), sub-band compression [14] (SIG05) and linear windowed tone mapping [28] (TVCG10). The detail of red sunset and riverbank appears in our result (middle row, first column), while this detail is more or less lost in other results (middle row). Moreover, our result preserves the correct color of red sunset, while the results by both the subband compression [14] and linear windowed tone mapping [28] show some color distortion (bottom row).

satisfactory scores with other approaches. Since our approach achieves better visual quality with fine color and texture details, the users prefer to give the higher scores to our fusion results. From this evaluation, we conclude that our approach can acquire the best visual quality in a way well matching the sense of human beings.

In order to justify the effectiveness of considering the gradient direction to be a measure of correct exposure in our method more comprehensively, we perform the comparison experiments by our approach with and without the gradient direction measurement. As shown in Fig. 10, our full algorithm with the gradient direction achieves the better visual quality than the result without it, especially in terms of the fine details and more vivid textures, which are shown in the close ups [Fig. 10(b)]. The overall appearance of our full algorithm and the result without the gradient direction are similar, because our local exposure weight, JND-based saliency weight and BLP together help to achieve the overall visual quality. In contrast, incorporating the gradient direction into the global exposure weight further improves the quality of fusion results with more color and texture details [Fig. 10(b)].

Fig. 11 gives the illustration of the sensitivity analysis with three boosting parameters on the memorial church sequence. We give three groups of results to analyze the sensitivity of each parameter, and each group is generated with only one parameter changed and the other two unchanged. When α_0 gradually decreases, the fusion result contains more details and reveals more information in the under-exposure regions [Fig. 11(c)–(e)]. Our approach achieves the overall visually pleasing results with detail and texture enhancement under different parameter settings. In this paper, we use the default parameters $\alpha_0 = 0.2$, $\beta = 0.5$, and $\gamma_0 = [1, 1, 1, 1, 1]$ to generate all the experimental results. The smaller β is, the more detailed color and texture information in the fusion result there will be [Fig. 11(f)–(h)]. The smaller γ_0 is, the more strongly the base layer will be enhanced [Fig. 11(i)–(k)].

As shown in Fig. 12, we compare and show three representative tone mapping methods, including photographic tone reproduction [8], subband compression [14], and linear windowed tone mapping [28]. We use the HDRShop plugin[4] to generate the results of photographic tone reproduction, and adopt the MATLAB implementations provided by the respective authors to generate the results of sub-band compression[5] and linear windowed tone mapping.[6] In order to give a relatively fair image quality comparison with our exposure fusion method, we use the optimal parameter settings to produce the good quality tone mapping results with their programs in all our comparison experiments. We further compare our approach with two representative exposure fusion approaches, i.e. exposure fusion [23], and gradient fusion [27]. The gradient-directed fusion method is implemented by ourselves, and the final fusion results are produced as good as possible by adjusting the parameter settings in our comparison results. As shown in Fig. 12, our approach achieves the best visual appearance with the color and texture details compared with the previous methods.

[4] *http://www.gregdowning.com/HDRI/tonemap/Reinhard/*
[5] *http://people.csail.mit.edu/yzli/*
[6] *http://www.cs.washington.edu/homes/shanqi/*

VI. Conclusion

This paper has presented a novel exposure fusion approach using BLP to produce a high quality image from multiple exposure images. Our novel BLP algorithm is based on boosting the detail and base signal respectively, and can effectively blend the multiple exposure images for preserving both color appearance and texture structures. A novel hybrid exposure weight is also introduced, which incorporates the local weight, global weight and JND-based saliency weight. Visual inspection and quantitative performance evaluation both demonstrate that the employment of the BLP model has brought a better fusion performance than the traditional fusion approaches. Our proposed exposure fusion approach successfully creates a visually pleasing fusion image with more color and texture details. We therefore believe that our fusion method will suffice to produce the results with fine details for most practical applications.

The comprehensive perceptual study and analysis of exposure fusion algorithms will make an interesting subject for future work. For instance, we can create a benchmark of input exposure images and conduct a user study to compare a representative number of state-of-the-art exposure fusion methods. In terms of the computational efficiency, we would like to extend our exposure fusion algorithm with GPU implementation for real-time applications in future work. There are many more tone mapping approaches as well as public research data of exposure images that we have not mentioned, and we can implement and compare more tone mapping operators in future. However, we believe that the aforementioned experimental results suffice to validate the effectiveness of the proposed approach. We will investigate further extensions of the exposure fusion for detecting the moving objects [22], [35], [36] of the dynamic scenes in the future work.

References

[1] P. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

[2] L. Snidaro, R. Niu, G. L. Foresti, and P. K. Varshney, "Quality-based fusion of multiple video sensors for video surveillance," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 4, pp. 1044–1051, Aug. 2007.

[3] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. SIGGRAPH*, 1997, pp. 369–378.

[4] J. Suo, L. Lin, S. Shan, X. Chen, and W. Gao, "High-resolution face fusion for gender conversion," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 41, no. 2, pp. 226–237, Mar. 2011.

[5] Y. Zhang and Q. Ji, "Efficient sensor selection for active information fusion," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 3, pp. 719–728, Jun. 2010.

[6] S. Mann and R. W. Picard, "On being undigital with digital cameras: Extending dynamic range by combining exposed pictures," in *Proc. 48th Annu. Conf. IS&T*, 1995, pp. 422–428.

[7] M. Uyttendaele, A. Eden, and R. Skeliski, "Eliminating ghosting and exposure artifacts in image mosaics," in *Proc. IEEE CVPR*, 2001, pp. 509–512.

[8] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *Proc. SIGGRAPH*, 2002, pp. 267–276.

[9] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *Proc. SIGGRAPH*, 2002, pp. 257–266.

[10] H. Kolb, "How the retina works," *Amer. Scientist*, vol. 91, no. 1, pp. 28–35, 2003.

[11] R. Raskar, A. Ilie, and J. Yu, "Image fusion for context enhancement and video surrealism," in *Proc. NPAR*, 2004, pp. 85–95.

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

[13] E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Trans. Vis. Comput. Graph.*, vol. 11, no. 1, pp. 13–24, Jan.–Feb. 2005.

[14] Y. Li, L. Sharan, and E. H. Adelson, "Compressing and companding high dynamic range images with subband architectures," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 836–844, 2005.

[15] A. Goshtasby, "Fusion of multiexposure images," *Image Vision Comput.*, vol. 23, no. 6, pp. 611–618, 2005.

[16] M. I. Smith and J. P. Heather, "Review of image fusion technology in 2005," in *Proc. SPIE*, vol. 5782, 2005, pp. 29–45.

[17] E. A. Khan, A. O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Proc. ICIP*, 2006, pp. 2005–2008.

[18] A. Eden, M. Uyttendaele, and R. Szeliski, "Seamless image stitching of scenes with large motions and exposure differences," in *Proc. IEEE CVPR*, 2006, pp. 2498–250.

[19] R. Fattal, M. Agrawala, and S. Rusinkiewicz, "Multiscale shape and detail enhancement from multilight image collections," *ACM Trans. Graph.*, vol. 26, no. 3, article 51, 2007.

[20] S. Raman and S. Chaudhuri, "A matteless, variational approach to automatic scene compositing," in *Proc. IEEE ICCV*, 2007, pp. 1–6.

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

[22] O. Gallo, N. Gelfand, W. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *Proc. IEEE ICCP*, 2009, pp. 1–7.

[23] T. Mertens, J. Kautz, and F. V. Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, 2008.

[24] A. R. Varkonyi-Koczy, A. Rovid, and T. Hashimoto, "Gradient-based synthesized multiple exposure time color HDR image," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 8, pp. 1779–1785, Aug. 2008.

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

[26] S. Raman, V. Kumar, and S. Chaudhuri, "Blind de-ghosting for automatic multiexposure compositing," in *Proc. SIGGRAPH ASIA Posters*, 2009, article 44.

[27] W. Zhang and W. K. Cham, "Gradient-directed composition of multiexposure images," in *Proc. IEEE CVPR*, 2010, pp. 530–536.

[28] Q. Shan, J. Jia, and M. S. Brown, "Globally optimized linear windowed tone mapping," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 4, pp. 663–675, Jul.–Aug. 2010.

[29] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. P. Seidel, and H. P. A. Lensch, "Optimal HDR reconstruction with linear digital cameras," in *Proc. IEEE CVPR*, 2010, pp. 215–222.

[30] T. H. Wang, C. W. Fang, M. C. Sung, and J. J. J. Lien, "Photography enhancement based on the fusion of tone and color mappings in adaptive local region," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3089–3105, Dec. 2010.

[31] 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. 648–1652, Nov. 2010.

[32] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High Dynamic Range Imaging, Second Edition: Acquisition, Display and Image-Based Lighting*. San Mateo, CA, USA: Morgan Kaufmann, 2010.

[33] P. Romaniak, L. Janowski, M. Leszczuk, and Z. Papir, "A no reference metric for the quality assessment of videos affected by exposure distortion," in *Proc. IEEE ICME*, 2011, pp. 1–6.

[34] S. Paris, S. Hasinoff, and J. Kautz, "Local Laplacian filters: Edge-aware image processing with a Laplacian pyramid," *ACM Trans. on Graph.*, vol. 30, no. 4, article no. 68, 2011.

[35] J. Hu, O. Gallo, and K. Pulli, "Exposure stacks for live scenes with hand-held cameras," in *Proc. ECCV*, 2012, pp. 499–512.

[36] W. Zhang and W. K. Cham, "Reference-guided exposure fusion in dynamic scenes," *J. Visual Commun. Image Represent.*, vol. 23, no. 3, pp. 467–475, 2012.

[37] 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.

Jianbing Shen (M'11–SM'12) received the Ph.D. degree in computer science from Zhejiang University, Hangzhou, China, in 2007.

Currently, he is a Full Professor in the School of Computer Science at Beijing Institute of Technology, Beijing, China. He has published more than 30 refereed papers in journals and conference proceedings. His current research interests include computer vision and computer graphics.

Ying Zhao is currently pursuing the master's degree in the School of Computer Science at Beijing Institute of Technology, Beijing, China.

Her current research interests include exposure fusion and high dynamic range image processing.

Shuicheng Yan (M'06–SM'09) is currently an Associate Professor in the Department of Electrical and Computer Engineering at National University of Singapore, Singapore, and the founding lead of the Learning and Vision Research Group. He has authored or co-authored about 200 technical papers over a wide range of research topics. His current research interests include computer vision, multimedia, and machine learning.

He is a recipient of the Best Paper Awards from ACM MM'10, ICME'10, and ICIMCS'09. He also received the PREMIA 2008&2010 Best Student Paper Award, the winner prize of the classification task in PASCAL VOC2010, and the honorable mention prize of the detection task in PASCAL VOC2010. He is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and has been serving as a Guest Editor of special issues for IEEE TRANSACTIONS ON MULTIMEDIA AND COMPUTER VISION AND IMAGE UNDERSTANDING.

Xuelong Li (M'02–SM'07–F'12) is currently a Full Professor with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Shaanxi, China.