Research Article

Weighted Least Squares Based Detail Enhanced Exposure Fusion

Harbinder Singh,[1] Vinay Kumar,[2] and Sunil Bhooshan[3]

[1] *Department of Electronics and Communication Engineering, Baddi University of Emerging Sciences and Technology, District Solan, Himachal Pradesh 173205, India*

[2] *Grupo de Procesado Multimedia, Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid Leganes, Madrid 28911, Spain*

[3] *Department of Electronics and Communication Engineering, Jaypee University of Information Technology, Waknaghat, District Solan, Himachal Pradesh 173215, India*

Correspondence should be addressed to Harbinder Singh; harbinder.ece@gmail.com

Many recent computational photography techniques play a significant role to avoid limitation of standard digital cameras to handle wide dynamic range of the real-world scenes, containing brightly and poorly illuminated areas. In many of these techniques, it is often desirable to fuse details from images captured at different exposure settings, while avoiding visual artifacts. In this paper we propose a novel technique for exposure fusion in which Weighted Least Squares (WLS) optimization framework is utilized for weight map refinement. Computationally simple texture features (i.e., detail layer extracted with the help of edge preserving filter) and color saturation measure are preferred for quickly generating weight maps to control the contribution from an input set of multiexposure images. Instead of employing intermediate High Dynamic Range (HDR) reconstruction and tone mapping steps, well-exposed fused image is generated for displaying on conventional display devices. A further advantage of the present technique is that it is well suited for multifocus image fusion. Simulation results are compared with a number of existing single resolution and multiresolution techniques to show the benefits of the proposed scheme for variety of cases.

1. Introduction

In recent years several new techniques have been developed that are capable of providing precise representation of complete information of shadows and highlights present in the real-world natural scenes. The direct 8-bit gray and 24-bit RGB representation of visual data, with the standard digital cameras in single exposure settings, often causes loss of information in the real-world scenes because the dynamic range of most scenes is beyond what can be captured by the standard digital cameras. Such representation is referred to as low dynamic range (LDR) image. Digital cameras have the aperture setting, exposure time, and ISO value to regulate the amount of light captured by the sensors. It is therefore important to somehow determine exposure setting for controlling charge capacity of the Charge Coupled Device (CCD). In modern digital cameras, Auto Exposure Bracketing (AEB) allows us to take all the images without touching the camera between exposures, provided the camera is on a tripod and a cable release is used. Handling the camera between exposures can increase the chance of miss-alignment resulting in an image that is not sharp or has ghosting. However, most scenes can be perfectly captured with nine exposures [1], whereas many more are within reach of a camera that allows 5–7 exposures to be bracketed. When the scene's dynamic range exceeds the dynamic range (DR) of camera it is exposure setting that determines which part of the scene will be optimally exposed in the photographed image. The DR of a digital camera is typically defined as the charge capacity divided by the noise [1, 2]. At single exposure setting, either detail in the poorly illuminated area (i.e., shadows) is visible with long exposure or brightly illuminated area (i.e., highlights) with short exposure (see Figure 1). Thus, image captured by the standard digital camera at single exposure setting from a scene containing highlights and shadows is partially over- or underexposed. As a result,
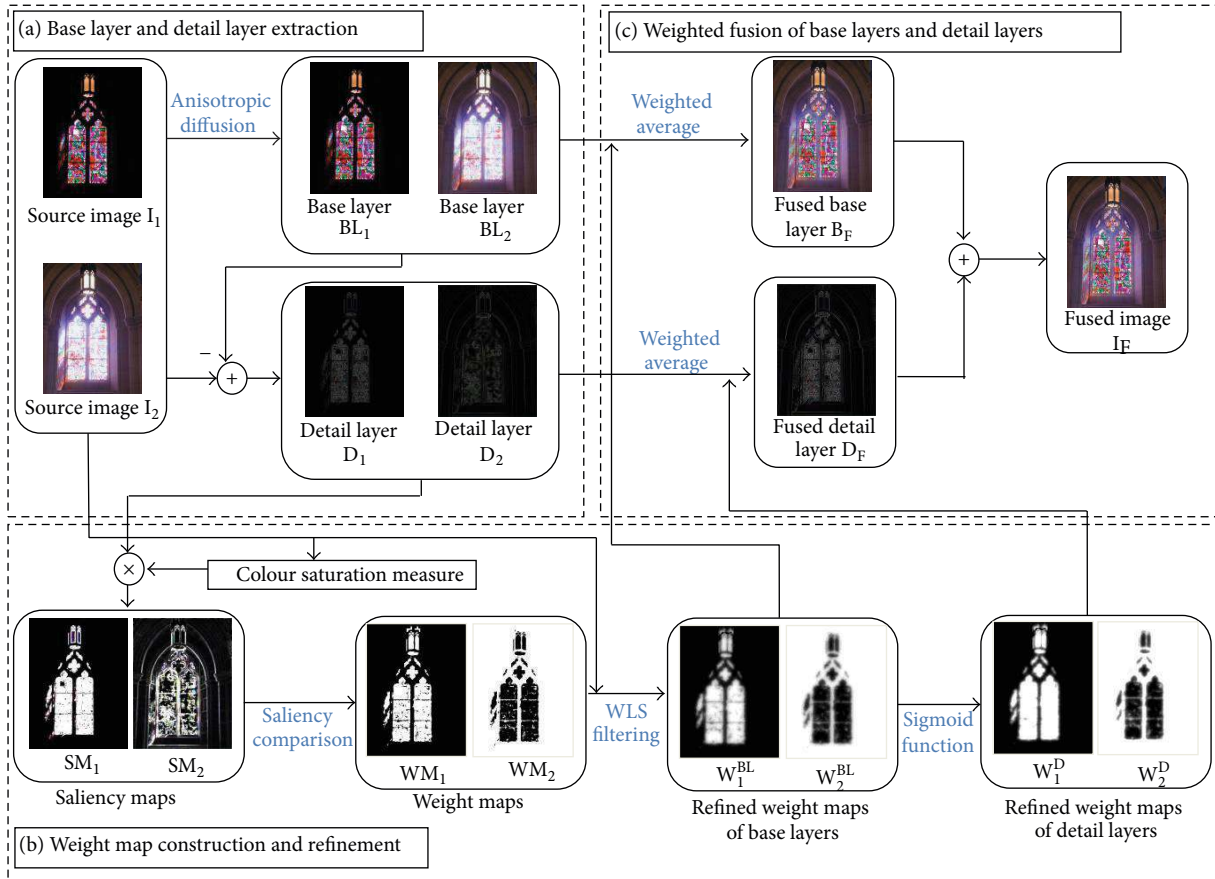
FIGURE 1: Illustration of proposed framework consisting of three principal blocks. (a) Base layer and detail layer extraction. The input images are transformed into two-scale decomposition. (b) Weight map construction and refinement. (c) Weighted fusion of base layers and detail layers. The base layers and detail layers across input image series are fused using simple weighted average approach.

there will always be a need to capture the detail of the entire scene with a sufficient number and value of exposures. The process of collecting complete luminance variations in rapid successions at different exposure settings is known as exposure bracketing.

In principle, there are two major approaches to handle the incapability of existing image capturing devices. The first approach is to develop HDR [3–7] reconstruction from multiexposure images that reconstruct full dynamic range up to 8 orders of magnitude and later tone map these images to adjust their tonal range, to some extent, for depiction on typical display devices. HDR [3–7] imaging is called scene-referred representation which represents the original captured scene values as close as possible. Such representation is sometimes referred to as extrasensory data representation. One of the important applications of HDR capturing techniques for security application is capturing video at entrance of the buildings [1]. Conventional cameras are not able to faithfully capture the interior and exterior of a building simultaneously while HDR camera, which is based on two-phase workflow, would be able to simultaneously record indoor as well as outdoor activities. Other important applications of HDR representation are satellite, scientific, and medical imagery, in which data is analyzed and visualized to record more

than what is visible to the naked eye. On the other hand, because of limited contrast ratio, standard displays (LCD, CRT) and printers are unable to reproduce full dynamic range captured by the HDR devices. In such cases, HDR data needs to be remapped [7] with a lower precision for display on conventional devices. Tone mapping algorithms can be either spatially variant or spatially invariant. In particular, spatially variant methods (also called local operators) [7–10] exploit local adaptation properties of human visual system (HVS), while spatially invariant methods [11–13] exploit global adaptation (also called global operators) of HVS.

Higher bit depths are usually not used because the display devices would not be able to reproduce such images at levels that are practical for human viewing [1]. Although for some real-world scenes low bit depth is sufficient to capture entire detail, there are countless situations that are not accommodated by low bit depth. Although HDR display devices will be developed in the near future, conventional printers may lead to inconsistencies which will be responsible for loss of details in the output. Recently Sunnybrook technologies, BrightSide, and Dolby prototypes of HDR display devices have been proposed [1, 14, 15] that can display HDR data directly. As a result, to avoid these inconsistencies, we must use tone mapping operators [7–13] to prepare HDR imagery

Figure 2: Proposed results for different input multiexposure sequences. (a) Carnival (top: three input exposures, bottom: fusion results), (b) house (top: four input exposures, bottom: fusion results), and (c) bellavita (top: two input exposures, bottom: fusion results). Input image sequences are courtesy of HDRsoft.com, Tom Mertens.

for display on LDR devices. Alternatively, we may directly generate 8-bit low dynamic range (LDR) image that looks like a tone-mapped image [1].

The second approach for the purpose is combining multiexposure images directly into 8-bit single LDR image that does not contain underexposed and overexposed regions [18, 26]. Thus it provides convenient and consistent way for preserving details in both brightly and poorly illuminated areas by skipping the construction of HDR image and the use of tonemapping operators [7–13]. The incorporation of the notion of combining multiple exposures without typical HDR and tone mapping steps is known as "exposure fusion," as shown in Figure 1. The fundamental goal of exposure fusion is often to improve the chance of creating a realistic scene without HDRI representation and tone mapping step. The underlying idea of various exposure fusion approaches [18, 26] is based on the utilization of different local measures for generating weight map to preserve details present in the different exposures. The current manuscript belongs to the second approach. The block diagrammatic representation of the present detail enhanced framework is shown in Figure 1. We have used edge preserving filter based on partial differential equations (PDE) [27] for two-scale decomposition that separates sharp details and fine details across various input images with different exposure levels. The current state-of-the-art method for automatic exposure fusion exploits the capability of edge preserving filter [27, 28] to generate weight function that guides the fusion of different exposures based on two-scale decomposition. We propose WLS filter [20] optimization framework and sigmoid function for weight map refinement of base layers and detail layers, respectively. Farbman et al. [20] has utilized WLS filter to construct a multiscale edge preserving decomposition multiscale tone and detail manipulation. To achieve the optimal contrast in the fused image the current paper develops an appropriate mask based on weak textures and color saturation measure to composite multiexposure images. The method is applicable for the fusion of broad range of textured images. See Figure 2 for an example of our exposure fusion results for typical

scene contain artificial light source (i.e., highlights), shadows, reflections, indoor details, and outdoor details.

Texture features [29] refer to the characterization of regions in an image by their spatial arrangement of color or intensities. Image textures are one way that can be used to help in classification of images [30]. Weak edges or texture information are the ideal indicators to detect over (or under) exposed regions in the image [17]. Raman and Chaudhuri [17] employ a Bilateral Filter (BLT) for compositing multiexposure images, in which weak edges were considered to design weight map. Interestingly, thus an analysis of weak textures seems to be the definition of perceived contrast. We take advantage of such possibility and design the appropriate matting function based on anisotropic diffusion for exposure fusion.

To analyze an image texture, there are primarily two approaches: structural approach and statistical approach. Structural approach uses a set of primitive texture elements in some regular or repeated pattern to characterize spatial relationship. While statistical approach defines an image texture as a quantitative measure of the arrangement of intensities in a region. In general later approach is easier to compute and is more widely used in computer graphics applications, since natural textures are made of patterns of irregular subelements. It has been noticed that simple averaging to fuse details from multiexposure image data yields low contrast in the fused image, especially in brightly and poorly illuminated areas. In the present approach, texture details will decide the contribution of corresponding pixel from different exposures in the fused image. A rich texture details mean a maximum contribution, which tells that image block has higher weight during the fusion process. Such metric is used to quantify the perceived local contrast of an image under different exposure settings and allows discarding underexposed and overexposed pixels. Therefore, to handle underexposed and overexposed regions, we propose a texture feature analysis based on Anisotropic Diffusion (ANI) [27, 28] that has the applicability to design weighting function as shown in Figure 1. Our goal is to exploit the edge preserving property of ANI to produce well-exposed image from

input images captured under different exposure settings. The detailed description of ANI based two-layer decomposition and weight map computation is given in the later sections. Our main contributions in this paper are highlighted follows.

(1) Two-scale decomposition based on anisotropic diffusion is proposed for fast exposure fusion, which does not require optimization of number of scales as required in the traditional multiscale techniques.

(2) A novel weight construction approach is proposed to combine texture features and saturation measure for guiding image fusion process. For weight map construction, we seek to utilize the strength of texture details under the change of exposure setting that takes place between an underexposed and an overexposed image. WLS filtering is proposed for weight refinement. Furthermore, fast sigmoid function based weight map generation for detail layers is proposed that reduce computational complexity of the algorithm.

(3) The important contribution of this paper is the advantages including ease of implementation, quality of compositing, and the provision of detail layer enhancement without introducing artifacts.

The remainder of this paper is structured as follows. Section 2 discusses the current available literature. Section 3 discusses description of separation of large scale variations and smaller scale details (i.e., texture details) based on ANI, consideration of smaller scale details and saturation measure for weight map generation, and the WLS and sigmoid based weight map refinement that produce single well-exposed image using simple weighted average approach. Section 4 discusses the utility of proposed approach for multifocus image fusion and the comparison with the popular single resolution exposure fusion, multiresolution exposure fusion and popular tone mapping operators. Section 5 summarizes the paper with future directions and conclusion.

2. Previous Work

2.1. HDR Imaging. There is a tremendous need to record a much wider gamut than standard 24 bit RGB. The practice of assembling HDR image from multiple exposure images recovers true radiance value present in the real-world scenes [2]. The camera response function recovered from differently exposed images is used to create HDR image whose pixel values are equivalent to the true radiance value of a scene. Radiance maps are stored in a file format that can encode recovered HDR data without losing information. "Floating point tiff" format tends to encode dynamic range up to 79 orders of magnitude and has better precision than the radiance format. Reinhard et al. [1] has provided the description and evaluation of formats available to store true radiance values. However, the success of HDR image capture has shown that it is possible to produce an image that exhibits details in poorly and brightly illuminated areas. Moreover, HDR formats have since found widespread applications in the computer graphics and HDR photography.

The prototypes of HDR display devices provide direct HDR display capabilities by means of a projector or Light Emitting Diode (LED) array that lights the Liquid Crystal Diode (LCD) from behind with a spatially varying light pattern [14, 15]. Unfortunately, conventional display devices (i.e., CRT and flat panel display) have dynamic ranges spanning a few orders of magnitude, much lower than those of the real-world scenes, often less than $100:1$. In order to display HDR images on monitors or print them on paper [31], we must remap the dynamic range of the HDR images to reproduce low dynamic range (LDR) images suitable for human visual system (HVS). In the literature, several tone mapping methods for converting real-world luminances to display luminances have been developed and fulfilling the fast growing demand to be able to display HDR images on low dynamic range (LDR) display devices. Most tone-reproduction algorithms make use of photoreceptor adaptation [32] to achieve visually plausible results. Local operators [7–10] involve the spatial manipulation of local neighboring pixel values based on the observation that HVS system is only sensitive to relative local contrast. Global operators [11–13] do not involve spatial processing. Tonemapping is achieved by applying spatially invariant operator to treat every pixel independently. Both types of techniques have their own advantages and disadvantages in terms of computational cost, easy implementation, halo effects (artifacts), spatial sharpness, and practical application. Reinhard et al. [1] give detailed review of various tone mapping operators.

A simple S-shaped curve (sigmoid function) has been utilized as tone mapping function [33]. The middle portion of such sigmoidal function is nearly linear and thus resembles logarithmic behavior. Moreover, sigmoidal functions have two asymptotes: one for very small values and one for large values. Fattal et al. [9] has introduced gradient based approach to preserve details from HDR image. To simulate the adaptation behavior of human visual system, they have attempted gradient modification at various scales. A reduced, low dynamic range image is then obtained by solving a Poisson equation on the modified gradient field. The algorithm has used local intensity range to reduce the dynamic range in transform domain and preserve local changes of small magnitudes. The method was almost free of artifacts and does not require any manual parameter tweaking.

Recently, dynamic range compression based on two-scale decomposition has been proposed [10]. The base layer was obtained using a nonlinear BLT filter [17] and detail layer was computed by taking difference between the input image and the base layer. Only the contrast of base layer was reduced, thereby preserving fine details.

2.2. Exposure Fusion. In recent years, various fusion algorithms have been developed to combine substantial information from multiple input images into a single composite image. The principal motivation for image fusion is to extend the depth-of-field, extend spatial and temporal coverage, to increase reliability, and extend dynamic range of the fused image and the compact representation of information. Imaging sensor records the time and space varying light intensity information reflected and emitted from object in a

three-dimensional observed physical scene. However, image fusion has a fundamental difficulty in preventing artifacts and preserving local contrast when fusing the characteristics recorded from the incident radiations, such as exposure value, focusing, modality, and environmental conditions. The automated procedure of extracting all the meaningful details from the input images to a final fused image is the main motive of image fusion. To facilitate image fusion, it may be necessary to align input images of the same scene captured at different times, or with different sensors, or with different exposure (EV) settings (called bracketing), or from different viewpoint using local and global registration methods [34, 35]. Normally it is assumed that the input images are captured with the help of tripod mounting. Hence, in general, we expect point-by-point correspondence between different input exposures of a scene. From technical stand point, the fused image reveals all details present in the scene without introducing any artifacts or inconsistencies which would distract the human observer or subsequent image processing stages.Ogden et al. [36] has proposed pyramid solution for image fusion. The pyramid becomes a multiresolution sketch pad to fill in the local spatial information at increasingly fine detail (as an artist does when painting). The Laplacian pyramid contains several spatial frequency bands which depicts certain edge information [37]. Image gradient orientation coherence model based fusion has been proposed for blending flash and ambient images [38, 39]. This model seeks to utilize the properties of image gradients that remain invariant under the change of lighting that takes place between a flash and an ambient image. Region segmentation and spatial frequency have been utilized for multifocus image fusion [40]. A fast multifocus algorithm has recently been developed [23] which utilizes weighted nonnegative matrix factorization and focal point analysis for preserving feature information in the fused image.

Raman and Chaudhuri [17] have utilized edge-preserving filter (i.e., bilateral filter) for the fusion of multiexposure images, in which appropriate matte is generated based on local texture details for automatic compositing process. Goshtasby [41] proposed exposure fusion method based on weights determined by blending function. Information metric was considered to design blending function. Smaller weights were assigned to an image block carrying less information while higher weights were assigned to best-exposed image block. Therefore, an image block was considered best-exposed within an area if it carries more information about the area than any other image blocks. To maximize information content in the fused image, a gradient-ascent algorithm was used to determine optimal block size and width of the blending functions. The size of the block automatically varies with image content as the type of image is changed. Szeliski [42] has used multidimensional histogram as postprocessing operator to achieve optimal contrast enhancement in the fused image; simple averaging was performed to smoothly combine the pixels into a fused image. This method was based on the observation that if the average intensity of the image is maintained during the average operation using histogram equalization, then new image can be created with increased contrast.

Mertens et al. [18] have used multiresolution approach [38] for the fusion of multiexposure image series. The technique was designed to create a well-exposed image without extending the dynamic range and tone mapping of the final image. This approach blends multiple exposures in Laplacian-pyramid code based on quality metrics like saturation and contrast. A part of the technique was stitching of flash and no-flash images, which seems to be suitable for detail enhancement in the fused image. The performance of this multiscale technique is dependent on the number of decomposition levels, that is, the pyramid height. However, the presented approach seems to be computationally expensive. Recently, various fast and effective weighted average based exposure fusion approaches have been proposed. Among these guided filtering based two-scale decomposition fusion approach [43], median filter and recursive filtering based fusion approach [44] and global optimization using generalized random walks for fusion [21] are producing fusion results with better quality. These methods were utilizing different image feature for weight calculation and further refined weight was used to control the contribution of pixels from input exposures. Instead, we use anisotropic diffusion which is effective for two-scale decomposition and weight map generation based on image feature such as weak textures. The major advantage of our technique is that it is based on single-resolution weighted average approach. Generally speaking, due to computational simplicity the present approach can be used in various consumer cameras entering the commercial market. Moreover, we noticed that the present approach can be applied for the multifocus image fusion and has much better results than existing multifocus and multiexposure image fusion methods.

3. WLS Based Exposure Fusion

3.1. Overview. A new type of exposure fusion technique is developed to avoid the limitation of conventional digital camera to handle the luminance variation in the entire scene. The primary focus of this paper is the development of a fast and robust exposure fusion approach based on local texture features computed from edge-preserving filter. Unlike most previous multiexposure fusion methods, we build on ANI, a nonlinear filter introduced by Perona and Malik [27] in 1990 that has the ability to preserve large discontinuities (edges). It derives from magnitude of the gradient of the image intensity and controls diffusion strength in the image to prevent blurring across edges. As such, the algorithm implemented (see Figure 4) includes four steps.

(1) A first step, in our algorithm, is two-scale decomposition based on ANI which is used to separate coarser details (base layer) and finer details (detail layer) across each input exposure.

(2) Weak texture details (i.e. detail layer computed from ANI) and saturation measure are utilized to generate weight mask for controlling the contribution of pixels from base layers separated across all the multiple exposures.

(3) WLS and sigmoid function based weight map refinement is performed for coarser details and finer details computed in the first step, respectively.

(4) Weighted average based blending of coarser details and finer details are performed to form a composite seamless image without blurring or loss of detail near large discontinuities.

3.2. Extraction of Coarser Details and Finer Details. Edge preserving filters have received considerable attention in computational photography over the last decade. BLT [45] and ANI [27] are the most popular edge-preserving operators. Standard BLT uses distances of neighboring pixels in space and range. The space varying weighting function is computed at a space of higher dimensionality than the signal being filtered. As a result, such filters have high computational costs [46]. ANI has led to an excellent tool for smoothing fine details of an image while preserving the coarser details (i.e., edges). It is modeled using partial differential equations (PDEs) and based on nonlinear iterative process. The diffusion equation in two dimensions is defined as follows:

$$\frac{\partial I(i, j, t)}{\partial t} = \text{div} \left[g(\|\nabla I\|) \nabla I \right], \tag{1}$$

where the operator ∇I calculates the image gradient of an input image I, $\|\nabla I\|$ represents the magnitude of the gradient of image intensity, $g(\|\nabla I\|)$ is a spatially varying nonlinear operator that smoothes fine details while avoiding blurring of coarser details, (i, j) specifies spatial position, and t is the iteration parameter.

The diffusion strength in the image is determined by the conduction coefficient which is influenced by the gradient of the image intensity. The principles of conduction coefficient are (i) smoothing the fine textures and (ii) preserving coarser details in the image data. Such type of nonlinear diffusion is achieved by considering image structure. On the other hand, fixed value of conduction coefficient (i.e., $g(\cdot) = 1$) yields isotropic linear diffusion that tends to have constant response for fine textures and strong edges. Therefore to achieve nonlinear diffusion the conduction coefficient is chosen to satisfy $g(i) \rightarrow 0$ when $i \rightarrow \infty$ so that the diffusion process is "stopped" across the region boundaries (i.e., edges) at locations of high gradients.

A diffusion functions $g(\cdot)$ used in our approach can be defined as follows:

$$g(\nabla I) = e^{(-(\|\nabla I\|/K)^2)}, \tag{2}$$

where K is a scale parameter that is determined by empirical constant and the selection of scale parameter may be different for a particular application [27]. In our algorithm the value of $K = 1/7$ was fixed for all cases, which is determined empirically to yield optimally diffused image for fine details extraction.

Anisotropic diffusion [27] for discrete signal is computed as follows:

$$I_s^{t+1} = I_s^t \frac{\gamma}{|\eta_s|} \sum_{p \in \eta_s} g(\nabla I_{s,p}) \nabla I_{s,p}, \tag{3}$$

where I_s^t is a discrete version of input signal, s determine the sample position in the discrete signal, and t determines iterations. The constant γ is a scalar that determines the rate of diffusion, η_s represents the spatial neighborhoods of current sample position s, and $|\eta_s|$ is the number of neighbors.

For discrete image, North (g_N), South (g_S), East (g_E), and West (g_W) spatial locations are considered for the computation of conduction coefficients. In our case, local window of size (3×3) from input image (I) is chosen, which intuitively appears most suitable for the computation of conduction coefficient at low computation cost, but other window sizes are possible as well. So after computing all the possible values of the conduction coefficients for pixel position (s) in the discrete image, the diffused image is obtained as follows:

$$I_s^{t+1} = I_s^t + \frac{\gamma}{|\eta_s|} \left[g_N \cdot \nabla_N I + g_S \cdot \nabla_S I + g_E \cdot \nabla_E I + g_W \cdot \nabla_W I \right]_s^t, \tag{4}$$

where ∇_N, ∇_S, ∇_E, and ∇_W indicate the difference of North, South, East, and West neighbor for pixel position s, respectively.

Let I_n be the nth source image which needs to be operated by an ANI filter. In order to separate coarser and finer details, we first decompose source images into two-scale representations by using anisotropic diffusion. The base layer (i.e., the diffused image defined in (4)) of each source image is obtained as follows:

$$\text{BL}_n = \text{ANI}_{K,t,\gamma}(I_n). \tag{5}$$

Once the base layer is obtained for each nth input image, the detail layer (D_n) can be directly calculated by subtracting the base layer from the source image as follows:

$$D_n = I_n - \text{BL}_n. \tag{6}$$

3.3. Weight Estimation. The motivation behind weight map computation is to yield nonlinear adaptive function for controlling the contribution of pixels from base layers and detail layers computed across all input exposures.

Interestingly, the detail layer (D_n) computed for nth source image in (6) yields analysis of weak textures that seems to be indicator of contrast variation in the image. We adopted such metric to quantify the perceived local contrast of an image under different exposure settings and allow discarding underexposed and overexposed pixels.

Furthermore, in order to accomplish optimal contrast and color details in the fused image, we additionally incorporate the color saturation measure (SAT_n) to our weighting function. In practice for nth RGB source image, SAT_n is computed for nth source image as the standard deviation within the R', G', and B' channel, at each pixel

$$\text{SAT}_n = \sqrt{\frac{1}{3} \left[(R' - \mu)^2 + (G' - \mu)^2 + (B' - \mu)^2 \right]}, \tag{7}$$

where, $\mu = (1/3)(R' + G' + B')$.

As shown in Figure 1, in order to remove the influences of underexposed and overexposed pixels for producing well-exposed image, the two image features, that is, D_n and SAT_n, are combined together by multiplication to estimate combined features (FR_n):

$$FR_n = D_n \times SAT_n. \qquad (8)$$

Then FR_n is convolved with the symmetric Gaussian low pass kernel (U) having 5×5 kernel size (r_U) with standard deviation (σ_U) 5 to construct the saliency maps SM_n:

$$SM_n = |FR_n| * U_{r_U, \sigma_U}. \qquad (9)$$

Next, the saliency maps are compared to determine the weight maps as follows:

$$WM_n^k = \begin{cases} 1, & \text{if } SM_n^k = \max\left(SM_1^k, SM_2^k, \ldots, SM_M^k\right) \\ 0, & \text{otherwise,} \end{cases} \qquad (10)$$

where M is number of source images, SM_n^k is the saliency value of the pixel k in the nth image.

3.4. WLS Based Weight Refinement and Weighted Fusion of Coarser Details and Finer Details.

In this section, we propose WLS optimization framework [20] and sigmoid function [28] based weight map refinement approach to obtain noiseless and smooth weight maps. First, WLS filtering is performed on each weight map WM_n with the corresponding source image serving as the source image for the affinity matrix [20]. The motivation behind weight maps refinement is as follows. The fusion rules (weight map) computed in (10) are hard (the value of weight maps are changing abruptly), noisy, and not aligned with the object boundaries. Weight maps need to be as smooth as possible, since rapid changes in the weight maps will introduce seam and artifacts in fused image.

3.4.1. WLS Optimization Framework.

WLS [20] based edge-preserving operator may be viewed as a compromise between two possible contradictory goals. Given an input image v, we seek a new image w, which, on the one hand, is as close as possible to v, and at the same time, is as smooth as possible everywhere, except across significant gradients in v. To achieve these objectives we seek to minimize the following quadratic functional:

$$\sum_{p'} \left(\left(w_{p'} - v_{p'}\right)^2 + \lambda \left(q_{x,p'}(v) \left(\frac{\partial w}{\partial x}\right)_{p'}^2 \right. \right.$$
$$\left. \left. + q_{y,p'}(v) \left(\frac{\partial w}{\partial y}\right)_{p'}^2 \right) \right), \qquad (11)$$

where the subscript p' denotes the spatial location of a pixel. The goal of the data term $(w_{p'} - v_{p'})^2$ is to minimize the distance between w and v, while the second (regularization) term strives to achieve smoothness by minimizing the partial derivatives of w. The smoothness requirement is enforced in a spatially varying manner via the smoothness weights q_x and

q_y, which depend on v. Finally, λ is responsible for the balance between the two terms, increasing the value of λ results in progressively smoother images w.

Using matrix notation we may rewrite (11) in following quadratic form:

$$(w - v)^T (w - v) + \lambda \left(w^T O_x^T V_x O_x w + w^T O_y^T V_y O_y w \right). \qquad (12)$$

Here V_x and V_y are diagonal matrices containing the smoothness weights $q_x(v)$ and $q_y(v)$, respectively, and the matrices O_x and O_y are discrete differentiation operators.

The vector w that minimizes (12) is uniquely defined as the solution of the linear system

$$(I_d + \lambda L_v) w = v, \qquad (13)$$

where I_d is the identity matrix and $L_v = O_x^T V_x O_x + O_y^T V_y O_y$. Modulo the difference in notation, this is exactly the linear system used in [47], where it was primarily used to drive piecewise smooth adjustment maps from a sparse set of constraints.

In the present approach, O_x and O_y are forward difference operators, and hence O_x^T and O_y^T are backward difference operators, which means that L_v is a five-point spatially inhomogeneous Laplacian matrix. As for the smoothness weights, we define in the same manner as in [47]:

$$q_{x,p'}(v) = \left(\left| \frac{\partial \ell}{\partial x}(p') \right|^\alpha + \varepsilon \right)^{-1},$$
$$q_{y,p'}(v) = \left(\left| \frac{\partial \ell}{\partial y}(p') \right|^\alpha + \varepsilon \right)^{-1}, \qquad (14)$$

where ℓ is the log-luminance channel of the input image v and the exponent α (typically between 1.2 and 2.0) determines the sensitivity to the gradients of v, while ε is a small constant (typically 0.0001) that prevents division by zero in areas where v is constant.

Equation (13) tells us that w is obtained from v by applying a nonlinear operator Z_λ, which depends on v:

$$w = Z_\lambda(v) = (I_d + \lambda L_v)^{-1} v. \qquad (15)$$

In the present approach, $WLS_{\lambda,\alpha,\varepsilon}(v)$ represents the WLS filtering operation. Where λ, α, and ε are the parameters which decide the degree of smoothness, sensitivity to the gradients and small constant of the WLS filter, respectively. In our case WM_n computed in (10) serves as the input image to WLS filter (i.e., $v = WM_n$). More specifically, the coarser version of weight map WM_n will serve as refined weight map for nth base layer W_n^{BL}:

$$W_n^{BL} = WLS_{\lambda,\alpha,\varepsilon}(v). \qquad (16)$$

Once the resulting weight maps for base layer are obtained, sharp and edge-aligned weights are computed based on 1-D sigmoid function for fusing the detail layers. As shown in Figure 1, the spatially smoothed weight maps of base layer are utilized to compute sharp weight mask of detail layer

which preserve texture details in the fused image. Therefore, unlike [43], the proposed solution attempts computationally simple approach to estimate the best possible weight maps for detail layer fusion.

Let W_n^D denote refined weight map for nth detail layer and $\text{SIG}_{a,t',\theta}(W_n^{\text{BL}})$ is the 1-D sigmoid function [28] applied on W_n^{BL}, where $a \in \mathfrak{R}$, $t' \in \mathfrak{R}$, and θ are the weight parameter, independent variable, and the parameter which decide the threshold to further control the degree of sharpness, respectively. Then W_n^D is computed as

$$W_n^D = \text{SIG}_{a,t',\theta}\left(W_n^{\text{BL}}\right). \tag{17}$$

In theory, the 1-D sigmoid is computed as:

$$\text{SIG}_{a,t',\theta}(t) = \frac{1}{1 + e^{(-at'+\theta)}}, \tag{18}$$

where $t' \in \mathfrak{R}$ is the independent variable, $a \in \mathfrak{R}$ is a weight parameter of the sigmoid function, and θ be a fixed threshold to further control the sharpness of sigmoid function.

Once the resulting weight maps W_n^{BL} and W_n^D are obtained, the pixelwise weighted composition of base layers (i.e., fused base layer BL_F), detail layers (i.e., fused detail layer D_F), and the resulting fused image I_F can be directly calculated as follows:

$$\text{BL}_F = \sum_{n=1}^{N} W_n^{\text{BL}} \text{BL}_n, \tag{19}$$

$$D_F = \sum_{n=1}^{N} W_n^D D_n, \tag{20}$$

$$I_F = \text{BL}_F + D_F. \tag{21}$$

4. Experimental Results and Analysis

In order to evaluate the performance and effectiveness of the proposed image fusion approach, we have summarized the comparison of our fusion approach with different exposure fusion, tone mapping, and multifocus image fusion methods. Two objective evaluation metrics (i.e., quality score $Q^{AB/F}$ [48] and visual information fidelity for fusion VIFF [49]) were employed to access the fusion quality and to analyze the effect of free parameters used in the approach. Currently, all experimental results are generated by the MATLAB implementation. Furthermore, to measure distortion in the fused image and strengthen the evaluation capability of $Q^{AB/F}$ and VIFF, we incorporate the Dynamic Range Independent Visible Difference Predictor (DRIVDP) [22]. DRIVDP metric is sensitive to three types of structural changes for distortions measurement (i.e., loss of visible contrast, amplification of invisible contrast, and reversal of visible contrast) between images under a specific viewing condition.

4.1. Comparison with Other Exposure Fusion, Multifocus Image Fusion and Tone Mapping Methods. Figures 2, 3, and 4 depict examples of fused images from the source multi-exposure images. It is noticed that the proposed approach

enhances texture details while preventing halos near strong edges. In order to check effectiveness and robustness of present approach, the algorithm is tested on variety of multiexposure image series. The proposed approach is computationally simple and the results are comparable to several exposure fusion and tone mapping techniques. As shown in Figures 2(a)–2(c) the details from all of the source images are perfectly combined and reveal fine textures while preserving local contrast and natural color. In Figures 3(a)–3(d) we compare our results to the recently proposed approaches. Figure 3(d) depicts the results of optimization framework [16] and Figure 3(e) shows the mate-based fusion results using the edge preserving filter such as BLT [17]. It can be observed that other fusion methods perform well in preserving image details while they fail to reconstruct texture and color details in the brightly illuminated areas. The result of Mertens et al. [18] (see Figure 3(f)) appears blur and loses texture details while in our results (see Figure 3(c)) the fine texture are emphasized that are difficult to be visible in Figure 3(f). This is because of utilization of Gaussian kernel for pyramid generation as it removes Pixel-to-pixel correlations by subtracting a low-pass filtered copy of the image from the image itself to generate Laplacian pyramid and result is a texture and edge details reduction in the fused image. The results produced in Figures 3(d)–3(f) lose visibility in brightly illuminated areas and details are lost in the tree leaves, and the texture on the wall is washed out. Although the results of Raman and Chaudhuri [17] (see Figure 3(e)) exhibit better color details in tree leaves, they appear slightly blurry. In our results (Figure 3(c)) details are preserved in the brightly illuminated areas, yet at the same time fine details are well preserved (tree leaves, wall texture, and lizard).

To further compare our results visually with Mertens et al. [18], iCAM06 [19], WLS [20], and GRW [21], respectively, Figures 4(a), 4(b), 4(c), 4(d), and 4(e) depict experimental results for National Cathedral sequence ($1024 \times 768 \times 3$). Proposed fusion results shown in Figure 4(a) illustrate the ability of enhancing fine texture details. As well as having the ability to produce good color information with natural contrast. This can bring an increased illusion of depth to an image textures. Therefore, enhanced texture details in the fused image let you get everything sharp and yield an accurate exposure that is entirely free of halo artifacts. Although tone mapped results of iCAM06 [19] and WLS [20] have produced comparable results, but they do not preserve contrast from input LDR image series. Figures 4(b) and 4(e) show the results of pyramid approach [18] and GRW optimization framework [21], respectively, which preserve global contrast but losses color information. GRW [21] based exposure fusion is shown in Figure 4(e) which depicts less texture and color details in brightly illuminated regions (i.e., lamp and window glass). Note that Figure 4(a) retains colors, sharp edges, and details while also maintaining an overall reduction in high frequency artifacts near strong edges. The results produced in Figures 4(b)–4(d) were generated by the programs provided by their respective authors. The HDR images for iCAM06 [19] and WLS [20] were generated using HDR reconstruction [2]. The results of GRW [21] shown in Figure 4(e) are taken from its
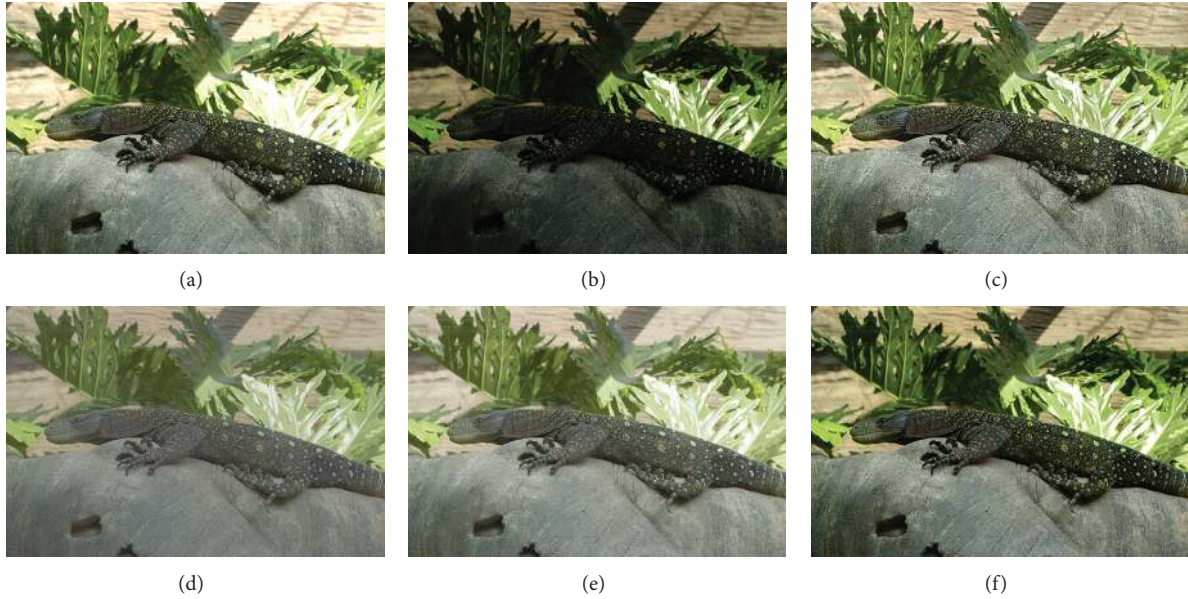
FIGURE 3: Lizard: comparison with popular exposure fusion approaches. (a), (b) Source images, (c) results of our new exposure fusion method. Tree leaves and wall texture appear overexposed and blurry in (d) Kotwal and Chaudhuri [16], (e) Raman and Chaudhuri [17], and (f) Mertens et al. [18]. In the proposed results, it becomes almost possible to preserve texture and strong edge features simultaneously. Moreover, by applying the proposed approach, the fine textures are accurately enhanced. Input image sequence is Eric Reinhard, University of Bristol.

FIGURE 4: National Cathedral: comparison with popular exposure fusion and tone mapping methods. (a) Results of our new exposure fusion method, (b) Mertens et al. [18], (c) iCAM06 [19], (d) WLS [20], and (e) GRW [21]. Note that our method yields enhanced texture and edge features with better color appearance. Input image sequence is courtesy of Max Lyons.

paper. In order to give a relatively fair comparison in our experiments, we have used default sets of parameters for tone mapping [19, 20] and exposure fusion [18] methods.

Figure 5 shows the distortion maps computed from DRIVDP metric proposed by Aydin et al. [22]. This quality assessment metric detects loss of visible contrast (green) and amplification of invisible contrast (blue). The main advantage of this metric is that it yields meaningful results even if the input images have different dynamic ranges. Though we consider here DRIVDP based quality assessment to compare proposed method with one exposure fusion [18] method and two tone mapping methods [19, 20]. We assume that the LDR images are shown in a typical LCD display with

maximum luminance 100 and gamma 2.2. We also assume that for all the LDR images, the viewing distance is 0.5 metres and the number of pixels per visual degree is 30 and peak contrast is 0.0025. Significance of the choice of these parameters can be found in [22]. Figures 5(a)–5(v) show a side-by-side comparison of the loss of visible contrast (green), and amplification of invisible contrast (blue) of proposed results with others methods. To compute visible contrast loss illustrated in Figures 5(d), 5(i), 5(n), and 5(s), respectively, for fused images in Figures 5(c), 5(h), 5(m), and 5(r), the underexposed image (i.e., Figure 5(a)) is used as reference image. Similarly, to compute visible contrast loss illustrated in Figures 5(e), 5(j), 5(o), and 5(t), respectively, for fused
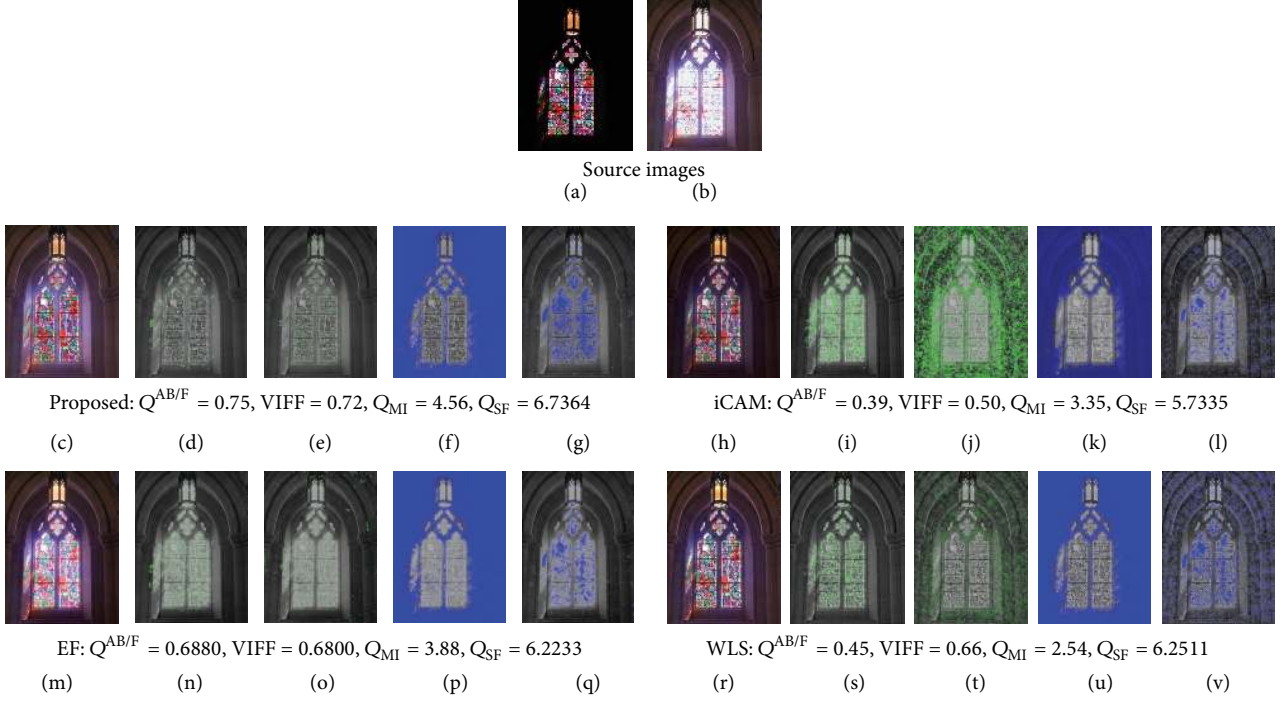
Source images
(a) (b)

Proposed: $Q^{AB/F} = 0.75$, VIFF $= 0.72$, $Q_{MI} = 4.56$, $Q_{SF} = 6.7364$

(c) (d) (e) (f) (g)

iCAM: $Q^{AB/F} = 0.39$, VIFF $= 0.50$, $Q_{MI} = 3.35$, $Q_{SF} = 5.7335$

(h) (i) (j) (k) (l)

EF: $Q^{AB/F} = 0.6880$, VIFF $= 0.6800$, $Q_{MI} = 3.88$, $Q_{SF} = 6.2233$

(m) (n) (o) (p) (q)

WLS: $Q^{AB/F} = 0.45$, VIFF $= 0.66$, $Q_{MI} = 2.54$, $Q_{SF} = 6.2511$

(r) (s) (t) (u) (v)

FIGURE 5: Comparison of proposed results with iCAM06, EF, and WLS on the National Cathedral sequence using DRIVDP [22]. (a), (b) The two source images give good exposures for the paintings on window glass and lamp and the wall, respectively. In a distortion map, green, blue, and gray pixels indicate visible contrast loss, amplification, and no distortion, respectively. Proposed results are more effective in preserving local details and colors than the others. For the paintings on window glass and lamp and the wall, proposed results (see (c)–(g)) depict the least distortion, followed by iCAM06 (see (h)–(l)), Mertens et al. EF (see (m)–(q)), and WLS (see (r)–(v)).

images in Figures 5(c), 5(h), 5(m), and 5(r), the overexposed image (i.e., Figure 5(b)) is used as reference image. We ran invisible amplification metrics on fused images, which are generated using a similar procedure as used for loss of visible contrast metric. The two source images with good exposures, respectively, for the brightly illuminated region (i.e., window) and the poorly illuminated region (i.e., wall) are given in Figures 5(a) and 5(b). The distortion maps for proposed method, iCAM06 [19], Mertens [18], and WLS [20] are given in Figures 5(c)–5(g), Figures 5(h)–5(l), Figures 5(m)–5(q), and Figures 5(r)–5(v), respectively, along with the fused images. In a distortion map, green, blue, and gray pixels indicate contrast loss, amplification, and no distortion, respectively. It can be noticed that the proposed results are more effective in preserving local contrast and color information than the other methods. Please note that visible contrast loss and distortions are the least using the proposed approach. Moreover, to compare the performance of the proposed approach, iCam06, Mertens et al., and WLS, we have employed four fusion quality metrics, that is, $Q^{AB/F}$, VIFF, Mutual Information (Q_{MI}) [50], and Spatial Frequency (Q_{SF}) [51].

$Q^{AB/F}$ [48] evaluates the amount of edge information transferred from input images to the fused image. A Sobel operator is applied to yield the edge strength and orientation information for each pixel. For two input images A and B, and a resulting fused image F (i.e., I_F computed in (21)), the Sobel edge operator is applied to yield the edge strength $e(n', m')$ and orientation $\beta(n', m')$ information for each pixel as

$$e_A\left(n', m'\right) = \sqrt{sb_A^x(m', n')^2 + sb_A^y(m', n')^2},$$

$$\beta_A\left(n', m'\right) = \tan^{-1}\left[\frac{sb_A^x\left(m', n'\right)}{sb_A^y\left(m', n'\right)}\right], \tag{22}$$

where $sb_A^x(m', n')$ and $sb_A^y(m', n')$ are horizontal and vertical Sobel template cantered on pixel $P_A(n', m')$ and convolved with the corresponding pixels of image A. The relative strength and orientation values of an input image A with respect to F are formed as

$$\left(G_{n',m'}^{AF}, A_{n',m'}^{AF}\right)$$

$$= \left(\left(\frac{e_{n',m'}^F}{e_{n',m'}^A}\right)^{\Psi}, 1 - \frac{\left|\beta_A\left(n', m'\right) - \left|\beta_F\left(n', m'\right)\right|\right|}{\frac{\pi}{2}}\right),$$

$$\text{where} \quad \Psi = \begin{cases} 1 & \text{if } e_A\left(n', m'\right) > e_F\left(n', m'\right) \\ -1 & \text{otherwise.} \end{cases}$$
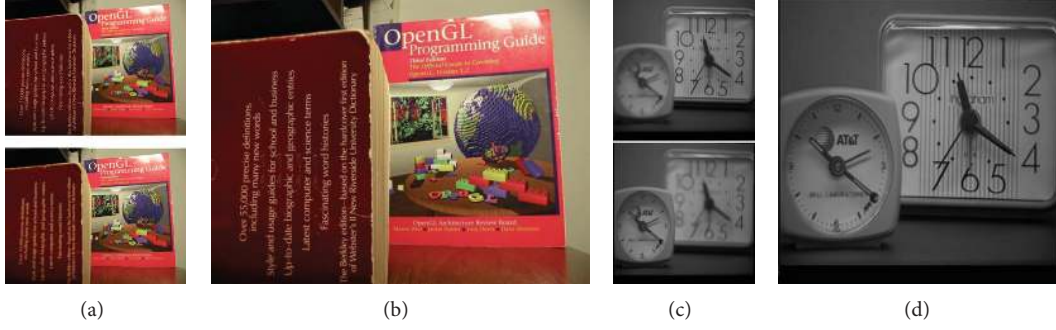
$$\tag{23}$$

FIGURE 6: Multifocus image fusion results: (a), (c) demonstrates the effect of selective focus to capture more details from a particular part of a scene and (b), (d) images generated by the proposed approach. Note how in the fused image all objects appear all-in-focus and enhance the color and texture details present in the foreground and background extracted from the original input images (input sequence is courtesy of Adu and Wang [23]).
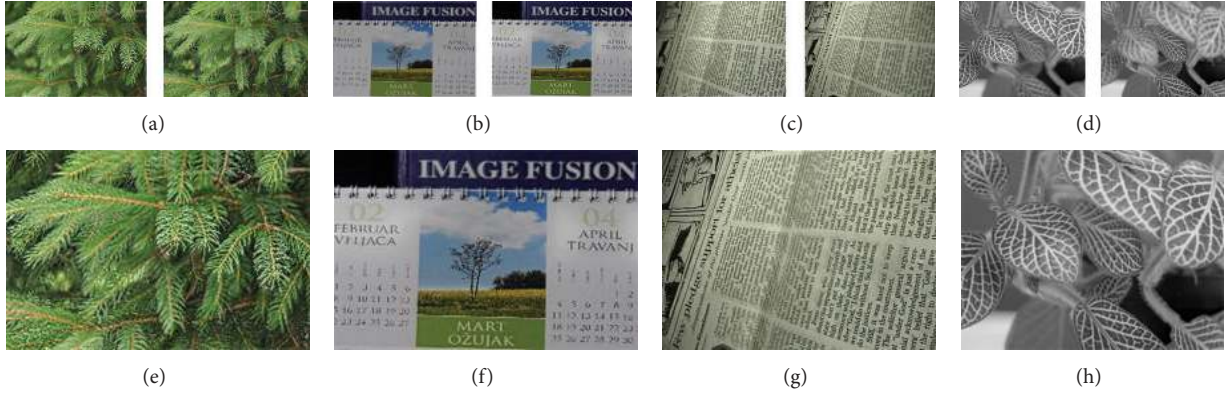

FIGURE 7: Multifocus image fusion results: (a), (b), (c), and (d) demonstrate the effect of selective focus to capture more details from a particular part of a scene and (e), (f), (g), and (h) images generated by the proposed approach. These results demonstrate that present approach has helped to handle typical situations of foreground and background present in the scene (input sequence is courtesy Slavica Savic).

The process of edge information preservation values is defined in [48].

Finally, the $Q^{AB/F}$ is defined as

$$Q^{AB/F} = \frac{\sum_{\forall n',m'} Q^{AF}_{n',m'} w^A_{n',m'} + Q^{AF}_{n',m'} w^B_{n',m'}}{\sum_{\forall n',m'} w^A_{n',m'} + w^B_{n',m'}}, \quad (24)$$

which evaluates the sum of edge information preservation values for both inputs Q^{AF} and Q^{BF} weighted by local importance perceptual factors w^A and w^B. We defined $w^A(n',m') = [e_A(n',m')]^L$ and $w_B(n',m') = [e_B(n',m')]^L$. L is a constant. For the "ideal fusion," $Q^{AB/F} = 1$.

VIFF [49] first decomposes the source and fused images into blocks. Then, VIFF utilizes the models in visual information fidelity (VIF) (i.e., Gaussian Scale Mixture (GSM) model, Distortion model, and human visual system (HVS) model) to capture visual information from the two source-fused pairs. With the help of an effective visual information index, VIFF measures the effective visual information of the fusion in all blocks in each subband. Finally, the assessment

result is calculated by integrating all the information in each subband:

$$\text{VIFF}(I_1, \ldots, I_n, I_F) = \sum_k P'_k \cdot \text{VIFF}_k(I_1, \ldots, I_n, I_F), \quad (25)$$

where P'_k is a weighting coefficient. According to VIF theory, a high VIF yields a high quality test image. Therefore, as VIFF increases, the quality of the fused image improves.

The quality metric Q_{MI} measures how well the original information from source images is preserved in the fused image:

$$Q_{\text{MI}} = 2\left(\frac{\text{MI}(A, F)}{H(A) + H(F)} + \frac{\text{MI}(B, F)}{H(B) + H(F)}\right), \quad (26)$$

where $H(A)$, $H(B)$, and $H(F)$ are the marginal entropy of A, B, and F and $\text{MI}(A, F)$ is the mutual information between the source image A and the fused image F:

$$\text{MI}(A, F) = H(A) + H(F) - H(A, F), \quad (27)$$

where $H(A, F)$ is the joint entropy between A and F, $H(A)$ and $H(F)$ are the marginal entropy of A and F, respectively, and $\text{MI}(B, F)$ is similar to $\text{MI}(A, F)$.
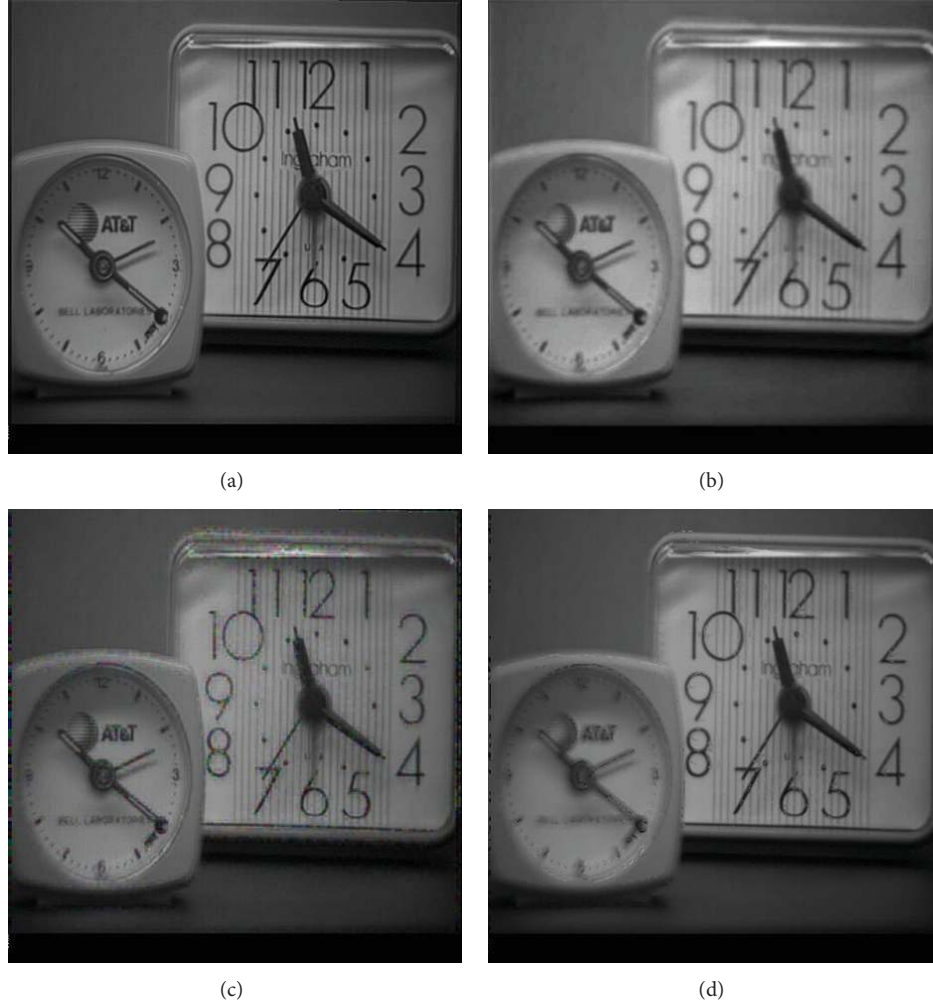
(a)

(b)

(c)

(d)

FIGURE 8: Comparison of Clock image results with recently proposed conventional multifocus image fusion methods. (a) Proposed results, (b) Adu et al. [23], (c) DWT [24], and (d) Tian et al. [25]. It has been found that the proposed approach helps the viewer to observe enhanced texture and edge features simultaneously without depicting visible artifacts. Input image sequence is courtesy of Adu and Wang.

The fourth criterion is Q_{SF}. The spatial frequency, which originated from the human visual system, indicates the overall active level in an image and has led to an effective objective quality index for image fusion [51]. The total spatial frequency of the fused image is computed from row (RF) and column (CF) frequencies of the image block and Q_{SF} is defined as

$$Q_{SF} = \sqrt{RF^2 + CF^2},$$

$$RF = \sqrt{\frac{1}{M'N'} \sum_{m'=1}^{M'} \sum_{n'=1}^{N'} \left(I_F\left(m',n'\right) - I_F\left(m',n'-1\right) \right)^2}, \tag{28}$$

where $I_F(m',n')$ is the gray value of pixel at position (m',n') of image I_F:

$$CF = \sqrt{\frac{1}{M'N'} \sum_{n'=1}^{N'} \sum_{m'=1}^{M'} \left(I_F\left(m',n'\right) - I_F\left(m'-1,n'\right) \right)^2} \tag{29}$$

The quantitative performance analysis using the aforesaid evaluation indices are shown in the caption of Figure 5. The present approach has outperformed the other methods. We can see that the proposed method can preserve more useful information compared with iCam06, Mertens et al., and WLS fusion methods. In particular, evaluation results in Figure 5 have demonstrated that $Q^{AB/F}$, VIFF, Q_{MI}, and Q_{SF} have correspondence with the DRIVDP-based evaluation.

Furthermore, to check the applicability of proposed approach for other image fusion applications, we have presented the experimental results for multifocus image fusion. In Figures 6, 7, and 8, it is demonstrated that proposed method is also suitable for multifocus image fusion to yield rich contrast and texture details. One of the key characteristics of present approach for multifocus image fusion application is illustrated in Figure 6(b): the color details are preserved in the fused image with better visualization of texture details. It can also be noticed in Figure 6(d) that the edges and textures are relatively better than that of
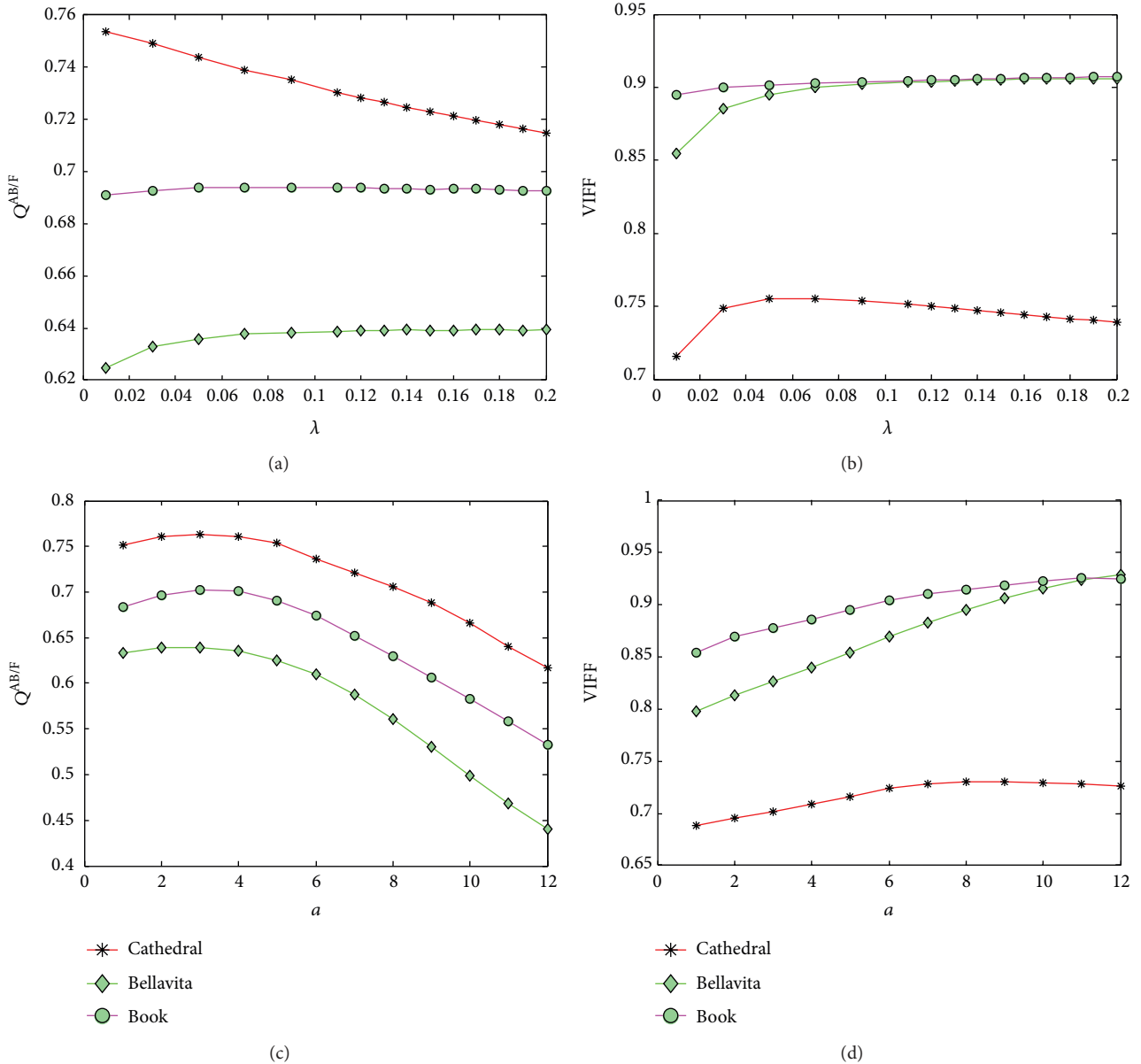
(a)

(b)

(c)

(d)

— ✳ — Cathedral

— ◆ — Bellavita

— ● — Book

— ✳ — Cathedral

— ◆ — Bellavita

— ● — Book

FIGURE 9: Sensitivity analysis of lambda (λ) and free parameter (a)which, respectively, control the smoothness in weight maps computed for base layers and detail layers. (a) Effectiveness of λ on $Q^{AB/F}$, (b) effectiveness of λ on VIFF, (c) effectiveness of a on $Q^{AB/F}$, and (d) effectiveness of a on VIFF.

source images. Fusion results of proposed method on the four standard test scenes (see Figures 7(a)–7(d)) are shown in Figures 7(e)–7(h). Note that, the strong edges and fine texture details are accurately preserved in the fused image without introducing halo artifacts. The halo artifacts may stand out if the detail layer undergoes a substantial boost. Comparisons of Adu [23], DWT [24], Tian et al. [25], and our approach for multifocus image fusion are illustrated in Figures 8(a)–8(d). The result produced in Figure 8(b) is taken from its paper [23]. Results of DWT [24] shown in Figure 8(c) were generated by the MATLAB Wavelet toolbox. For the DWT-based methods, the low-pass subband coefficients and the high-pass subband coefficients are simply merged by

the averaging scheme and the choose-max selection scheme, respectively. The DWT-based fusion algorithm is performed using five-level decomposition and db3 wavelets are used in scale decomposition. The results of Jing et al. shown in Figure 8(d) are generated from the MATLAB code provided by the author. Note that our method (see Figure 8(a)) yields enhanced texture and edge features. We can significantly preserve and enhance fine details separately because our approach excludes fine textures from the base layers.

4.2. Analysis of Free Parameters and Fusion Performance Metrics. Proposed method has eight free parameters, that is, t, K, γ, a, θ, λ, α, ε. We fix $t = 5$, $K = 30$, $\gamma = 1/7$,

(a)   (b)   (c)
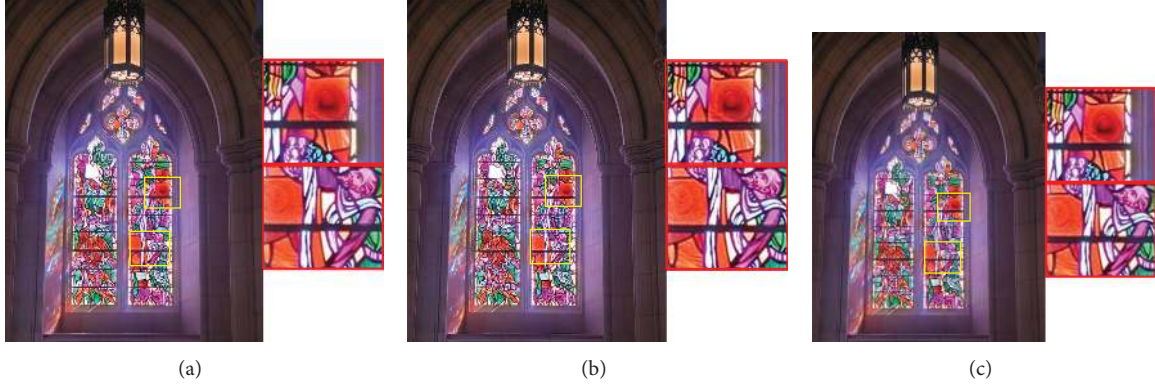
FIGURE 10: Visual inspection of fine detail enhancement in typical lighting situations: The free parameter a in (18) controls detail enhancement and sharpening. We have found that variation of "a" between 3 and 6 is sufficient for fine details extraction and give better results for various typical situations. By choosing higher value of a, the texture details are accurately enhanced. It also shows that selection of much higher value for detail enhancement does not introduce artifacts near object boundaries due to detail enhancement. (a) $a = 1$, (b) $a = 6$, and (c) $a = 12$.

$a = 3$, $\theta = 0.001$, $\lambda = 0.1$, $\alpha = 1.2$, and $\varepsilon = 0.0001$ in all experiments and they are set as default parameters. It is preferred to have a small number of iterations (t) to reduce computational time. The fusion performance is not affected when $t \leq 5$ because present method does not depend much on the exact parameter choice of t. The parameters selection criterion for K, γ, and θ is given in [27, 28], respectively. We have set $K = 30$, $\gamma = 1/7$, and $\theta = 0.001$ as default parameters for all experiments. In the present approach, the fusion performance is dependent on two free parameters, that is, λ and a. To analyze the effect of lambda (λ) and free parameter (a) on $Q^{AB/F}$ [48] and VIFF [49], we have illustrated four plots (see Figures 9(a)–9(d)) for input image sequence of Cathedral ($1024 \times 768 \times 3$), Bellavita ($800 \times 535 \times 3$), and Book ($569 \times 758 \times 3$). The detailed description of $Q^{AB/F}$ and VIFF is given in the previous subsection. To assess the effect of lambda (λ) and free parameter (a) on fusion performance, the $Q^{AB/F}$ and VIFF are experimented.

To analyze the influence of λ and a on $Q^{AB/F}$ and VIFF, other parameters are set to $t = 5$, $K = 30$, $\gamma = 1/7$, $\theta = 0.001$, $\alpha = 2$, and $\varepsilon = 0.0001$. As shown in Figures 9(a) and 9(b), the fusion performance will be worse when the values of λ and a are too large or too small. It should be noticed in Figure 9 that the $Q^{AB/F}$ and VIFF decrease when the λ and a are too large or too small. The visual inspection of effect of a's on "Cathedral" sequence is depicted in Figures 10(a)– 10(c). It can easily be noticed in Figures 10(a)–10(c) that as a increases, the strong edges and textures get enhanced and therefore leads to a detail preserving fusion results. In order to obtain optimal detail enhancement and low computational time, we have concluded that the best results were obtained with $t = 5$, $K = 30$, $\gamma = 1/7$, $a = 3$, $\theta = 0.001$, $\lambda = 0.1$, $\alpha = 1.2$, and $\varepsilon = 0.0001$, which yield reasonably good results and satisfactory subjective performance for all cases.

To further demonstrate the analysis of errors introduced by the free parameter a, four fundamental error performance metrics are adopted, that is, Root Mean Squared Error (RMSE), Normalized Absolute Error (NAE), Laplacian Mean Squared Error (LMSE), and Peak Signal to Noise Ratio (PSNR). The RMSE measure the differences between resulting image and reference image. The error in a pixel $P_{i'}^{*}$ is calculated using Euclidean distance between a pixel in a resulting image with $a > 1$ and the corresponding pixel in the reference image $P_{i'}^{\text{ref}}$ with $a = 1$, that is, $E(P_{i'}^{*}) = \|P_{i'}^{*} - P_{i'}^{\text{ref}}\|$. The total error in a resulting fused image is computed using square root of the Mean Square Error (MSE), that is, $\sqrt{(1/M)\sum_{i'} E(P_{i'}^{*})^2}$.

NAE is a measure of how far is the resulting fused image (when $a > 1$) from the reference fused image (when $a = 1$) with the value of zero being the perfect fit. Large value of NAE indicates poor quality of the image [51]. NAE is computed as follows:

$$\text{NAE} = \frac{\sum_{m'=1}^{M'} \sum_{n'=1}^{N'} \left| I_F\left(m', n'\right) - \widehat{I_F}\left(m', n'\right) \right|}{\sum_{m'=1}^{M'} \sum_{n'=1}^{N'} \left| I_F\left(m', n'\right) \right|}. \tag{30}$$

LMSE is based on the importance of edges detail measurement, which is also the most critical feature for image quality assessment. The large value of LMSE means that image is poor in quality. LMSE is defined as follows:

$$\text{LMSE} = \frac{\sum_{m'=1}^{M'} \sum_{n'=1}^{N'} \left[L\left(I_F\left(m', n'\right)\right) - L\left(\widehat{I_F}\left(m', n'\right)\right) \right]^2}{\sum_{m'=1}^{M'} \sum_{n'=1}^{N'} \left[L\left(I_F\left(m', n'\right)\right) \right]^2}, \tag{31}$$

where $L(I_F(m', n'))$ is laplacian operator:

$$L\left(I\left(m', n'\right)\right) = I_F\left(m' + 1, n'\right) + I_F\left(m' - 1, n'\right)$$
$$+ I_F\left(m', n' + 1\right) + I_F\left(m', n' - 1\right) \tag{32}$$
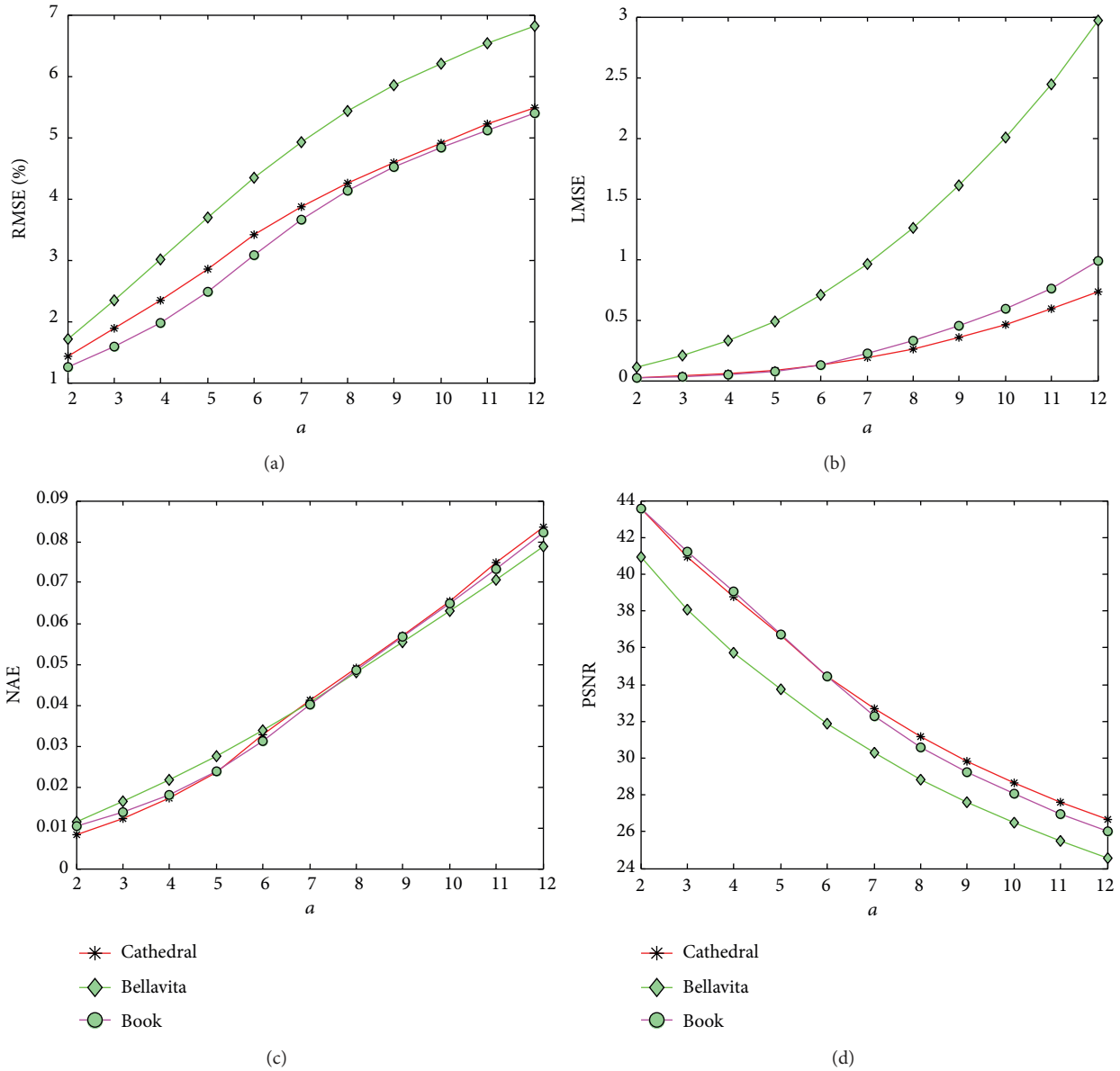$$- 4I_F\left(m', n'\right).$$

(a)

(b)

(c)

(d)

FIGURE 11: Sensitivity analysis of errors introduced by the free parameter a. The errors increases dramatically as a becomes too large but increases slowly when $a \leq 6$. It is observed in (b) and (c) that error increases as a increases but still the deviation range is less. (a) Effectiveness of a on RMSE, (b) effectiveness of a on LMSE, (c) effectiveness of a on NAE, and (d) effectiveness of a on PSNR.

PSNR is the ratio between the maximum possible pixel value of the fused image (MAX_{I_F}) and the MSE, which is computed as follows:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_{I_F}^2}{\text{MSE}} \right). \tag{33}$$

As shown in Figures 11(a)–11(c) the RMSE, LMSE, and NAE increase as parameter a increases. It should be noticed in Figure 11(a) that when $a = 12$, the total error introduced is still less than 7%. It is seen from the computed values of objective measures like NAE and LMSE (see Figures 11(b)-11(c)) for $a = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,$ and 12, for input

image sequence of Cathedral, Bellavita, and Book, the errors increase dramatically as free parameter a in (18) becomes too large but increase slowly when $a \leq 6$. Using graph presented in Figure 11(d), we want to illustrate what can happen if PSNR is used as distortion measure. It has been found that PSNR decreases gradually as free parameter a increases and can be seen that the proposed approach is performing consistently for different value of free parameter (i.e., a) proposed for detail enhancement.

Another interesting interactive tool for manipulating the detail and contrast in the multifocus image fusion has been experimented. Figures 12(a)–12(f) show the results generated for Clock and Leaves image series with the parameters $\lambda =$
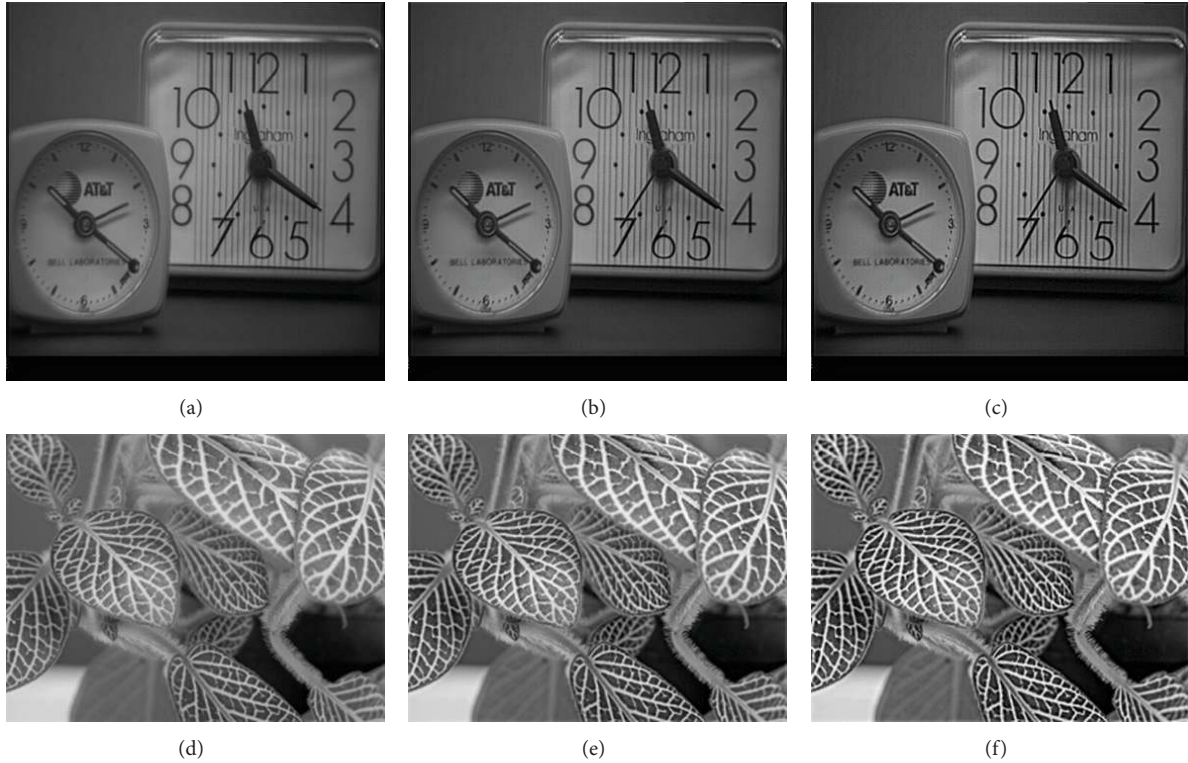
FIGURE 12: Visual inspection of contrast and detail enhancement in multifocus image fusion. The free parameter λ in (15) and a in (18) controls detail enhancement and sharpening, respectively. We are able to achieve aggressive detail enhancement and exaggeration, while avoiding artifacts. Clock image: (a) $\lambda = 0.5$, $a = 7$, (b) $\lambda = 0.5$, $a = 12$, (c) $\lambda = 0.5$, and $a = 17$. Leaves image: (d) $\lambda = 0.5$, $a = 7$, (e) $\lambda = 0.5$, $a = 12$, (f) $\lambda = 0.5$, and $a = 17$.

0.5, $a = 7$ (see Figure 12(a)), $\lambda = 0.5$, $a = 12$ (see Figure 12(b)), $\lambda = 0.5$, $a = 17$ (see Figure 12(c)), $\lambda = 0.5$, $a = 7$ (see Figure 12(d)), $\lambda = 0.5$, $a = 12$ (see Figure 12(e)), and $\lambda = 0.5$, $a = 17$ (see Figure 12(f)), respectively. Here, we demonstrate that we can generate highly detail enhanced fused image from multifocus image series, before objectionable artifacts appear. We found that present approach is very effective for boosting the amount of local contrast and fine details. The effective manipulation range is very wide and will vary in accordance with the texture details present in the input image series: it typically takes a rather extreme manipulation to cause artifacts near strong edges to appear.

5. Conclusion and Future Scope

Our proposed technique constructs a detail enhanced fused image from a set of multiexposure images by using WLS optimization framework. When compared with the existing techniques which use multiresolution and single resolution analysis for exposure fusion the present method perform better in terms of enhancement of texture details in the fused image. Our research was motivated by the edge-preserving property of anisotropic diffusion that has nonlinear response for fine textures and coarser details. The two-layer decomposition based on anisotropic diffusion is used to extract fine textures for detail enhancement. Furthermore, it is interesting to note here that our approach can also be applied for multifocus image fusion problem. More importantly, the information in the resultant fused image can be controlled with the help of proposed free parameters. At last, the future work involves improvement of this method for adaptively choosing the parameters of the WLS filter and checking the utilization for different kinds of image fusion applications.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.
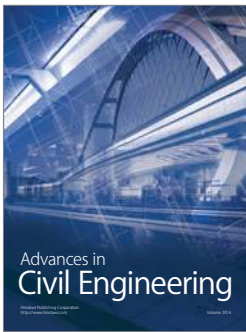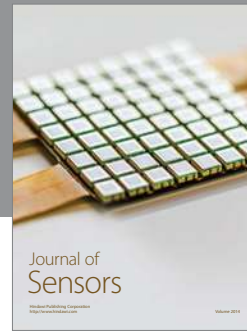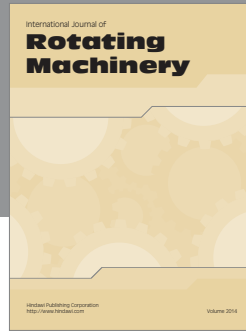
Acknowledgments

References

[1] E. Reinhard, G. Ward, S. Pattanaik, and P. Debvec, *High Dynamic Range Imaging Acquisition, Manipulation, and Display*, Morgan Kaufmann, 2005.

[2] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the Conference on Computer Graphics (SIGGRAPH '97)*, pp. 369–378, August 1997.

[3] S. Mann and R. W. Picard, "Being "undigital" with digital cameras: extending dynamic range by combining differently exposed pictures," in *Proceedings of IST's 48th Annual Conference*, pp. 442–448, May 1995.

[4] K. Jacobs, C. Loscos, and G. Ward, "Automatic high-dynamic range image generation for dynamic scenes," *IEEE Computer Graphics and Applications*, vol. 28, no. 2, pp. 84–93, 2008.

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

[6] A. Tomaszewska and R. Mantiuk, "Image registration for multi-exposure high dynamic range image acquisition," in *Proceedings of the International Conference on Computer Graphics, Visualization and Computer Vision*, Plzen, Czech Republic, 2007.

[7] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 267–276, 2002.

[8] Y. Li, L. Sharan, and E. H. Adelson, "Compressing and Companding high dynamic range images with subband architectures," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 836–844, 2005.

[9] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 249–256, 2002.

[10] F. Durand and J. Dorsey, "Fast bilateral Filtering for the display of high dynamic range images," *ACM Transaction on Graphics*, vol. 21, no. 3, pp. 257–266, 2002.

[11] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 4, pp. 291–306, 1997.

[12] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Computer Graphics Forum*, vol. 22, no. 3, pp. 419–426, 2003.

[13] E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 1, pp. 13–24, 2005.

[14] H. Seetzen, W. Heidrich, W. Stuerzlinger et al., "High dynamic range display system," *ACM Transaction on Graphics*, vol. 23, no. 3, pp. 760–768, 2004.

[15] H. Seetzen, L. A. Whitehead, and G. Ward, "A high dynamic range display using low and high resolution modulator," *In the Society For InFormation Display International Symposium*, vol. 34, no. 1, pp. 1450–1453, 2003.

[16] K. Kotwal and S. Chaudhuri, "An optimization-based approach to fusion of multi-exposure, low dynamic range images," in *Proceedings of the 14th International Conference on Information Fusion*, Chicago, Ill, USA, July 2011.

[17] S. Raman and S. Chaudhuri, "Bilateral filter based compositing for variable exposure photography," in *Proceedings of Eurographics*, Munich, Germany, 2009.

[18] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: a simple and practical alternative to high dynamic range photography," *Computer Graphics Forum*, vol. 28, no. 1, pp. 161–171, 2009.

[19] J. Kuang, G. M. Johnson, and M. D. Fairchild, "iCAM06: a refined image appearance model for HDR image rendering," *Journal of Visual Communication and Image Representation*, vol. 18, no. 5, pp. 406–414, 2007.

[20] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Transactions on Graphics*, vol. 27, no. 3, article 67, 2008.

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

[22] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," *ACM Transactions on Graphics*, vol. 27, no. 3, article 69, 2008.

[23] J. H. Adu and M. Wang, "Multi-focus image fusion based on WNMF and focal point analysis," *Journal of Convergence Information Technology*, vol. 6, no. 7, pp. 109–117, 2011.

[24] P. M. Zeeuw, "Wavelet and image fusion, CWI," Amsterdam, The Netherlands, March 1998, http://www.cwi.nl/~pauldz/.

[25] J. Tian, L. Chen, L. Ma, and W. Yu, "Multi-focus image fusion using a bilateral gradient-based sharpness criterion," *Optics Communications*, vol. 284, no. 1, pp. 80–87, 2011.

[26] J. Shen, Y. Zhao, and Y. He, "Detail-preserving exposure fusion using subband architecture," *The Visual Computer*, vol. 28, no. 5, pp. 463–473, 2012.

[27] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[28] H. Singh, V. Kumar, and S. Bhooshan, "Anisotropic diffusion for details enhancement in multiexposure image fusion," *ISRN Signal Processing*, vol. 2013, Article ID 928971, 18 pages, 2013.

[29] G. Linda Shapiro and C. George, *Stockman, Computer Vision*, Prentice-Hall, Upper Saddle River, NJ, USA, 2001.

[30] K. Laws, *Textured image segmentation [Ph.D. dissertation]*, University of Southern California, 1980.

[31] G. Qiu and J. Duan, "An optimal tone reproduction curve operator for the display of high dynamic range images," in *IEEE International Symposium on Circuits and Systems (ISCAS '05)*, pp. 6276–6279, May 2005.

[32] M. H. Kim, T. Weyrich, and J. Kautz, "Modeling human color perception under extended luminance levels," *ACM Transactions on Graphics*, vol. 28, no. 3, article 27, 2009.

[33] J. Tumblin, J. K. Hodgins, and B. K. Guenter, "Two methods for display of high contrast images," *ACM Transactions on Graphics*, vol. 18, no. 1, pp. 56–94, 1999.

[34] A. Ardeshir Goshtasby, *2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications*, Wiley-Interscience, 2005.

[35] G. Ward, "Fast, robust image registration for. Compositing high dynamic range photographs from hand-held exposures," *Journal of Graphics Tools*, vol. 8, pp. 17–30, 2003.

[36] J. M. Ogden, E. H. Adelson, J. R. Bergen, and P. J. Burt, "Pyramid based computer graphics," *RCA Engineer*, vol. 30, no. 5, pp. 4–15, 1985.

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

[38] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li, "Removing photography artifacts using gradient projection and flash exposure sampling," *ACM Transaction on Graphics*, vol. 24, no. 3, pp. 828–835, 2005.

[39] G. Petschnigg, R. Szeliski, M. Agrawala, M. F. Cohen, H. Hoppe, and K. Toyama, "Digital photoy with flash and no-flash image pairs," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 664–672, 2004.

[40] S. T. Li and B. Yang, "Multifocus image fusion using region segmentation and spatial frequency," *Image and Vision Computing*, vol. 26, no. 7, pp. 971–979, 2008.

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

[42] R. Szeliski, "System and process for improving the uniformity of the exposure and tone of a digital image," U. S. Patent No. 6687400, 2004.

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

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

[45] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the IEEE 6th International Conference on Computer Vision*, pp. 839–846, January 1998.

[46] D. Barash, "A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 844–847, 2002.

[47] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," *ACM Transaction on Graphics*, vol. 25, pp. 646–653, 2006.

[48] C. S. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308–309, 2000.

[49] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Information Fusion*, vol. 14, no. 2, pp. 127–135, 20131.

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

[51] R. Sakuldee and S. Udomhunsakul, "Objective performance of compressed image quality assessments," in *Proceedings of World Academy of Science, Engineering and Technology*, vol. 26, pp. 434–443, 2007.