Fast Multi-Focus Image Fusion Using Inner Channel Prior

TINGNA KUANG, FEIHONG YU, AND XUXIANG NI
College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China
Corresponding author: Feihong Yu (feihong@zju.edu.cn)

ABSTRACT In this paper, we propose a simple but effective sharpness prior—inner channel prior—to detect the focused area of multi-focus images. The inner channel prior is a kind of location feature of natural image sharpness. It is based on a key observation—the high-frequency information of the saturation channel is located inside the object. The inner channel prior indicates the focus degree inside objects, which can obtain better color fidelity and sharper object. Using the proposed prior combined with the multi-scale image fusion, we can directly obtain a sharp image with an extended depth of field focusing on all the objects. Results on a variety of defocused image sequences demonstrate the availability of the proposed prior. In addition, the proposed prior is insensitive to overexposure and can better maintain the color of multi-focus images.

INDEX TERMS Image fusion, multi-focus images, focus measurement, depth of field, multi-scale transform.

I. INTRODUCTION

It is challenging to obtain an image in which all the captured objects are focused, for the finite depth of field (DOF) of photographic lens. Only at a particular distance from the camera can the object be focused and have acceptable sharpness, whereas the sharpness decreases gradually as the object moves away from the sharp focus plane. To obtain an all-in-focus image, we need to fuse the images captured from the same scene at separate object distances.

Image fusion is highly desired in computational photography and computer vision applications, such as remote sensing [1], medical diagnosis [2], [3], micro-image fusion [4], biochemical analysis [5], image dehazing [6], [7] and microscopy [8]. The fusion of multi-focus images is a major application of image fusion. The main idea of multi-focus image fusion is to extract the important and essential information from the input images, then fusing these information into one image which theoretically consists of all the information of input images [9], as shown in Fig. 1.

However, gathering focused objects is a challenging problem, since there is no golden criteria to measure the sharpness of an object. Therefore, many focus measurements [10] have been proposed to distinguish sharp image blocks from defocused image blocks, including variance, energy of image gradient (EOG), Tenenbaum's algorithm (Tenengrad) and

(a) Source Image A   (b) Source Image B   (c) Fused Image

FIGURE 1. An example of multi-focus image fusion (520 × 520), the fused image was processed by our proposed method.

sum-modified-Laplacian (SML) [11]. Experiments [10] show that SML can provide a better fusion performance than other focus measures. Meanwhile, how to fuse a sharp image as many details as possible is also a tough problem. Image fusion has made significant progress [12] in recent years. Sun *et al.* [13] utilize the region-level method, Laplacian pyramids, to fuse multi-focus images, obtaining a sharp image with more original information than the method at pixel level. However, this method cannot obtain a clear image in the case of fusing images with fewer details. Chen *et al.* [14] utilizes the image matting algorithm, making full use of the spatial information to refine the weight maps with state-of-the-art performance, even in some cases of obvious misregistration. However, this method could have some wrong judgements inside objects by using edge information based image matting algorithm.
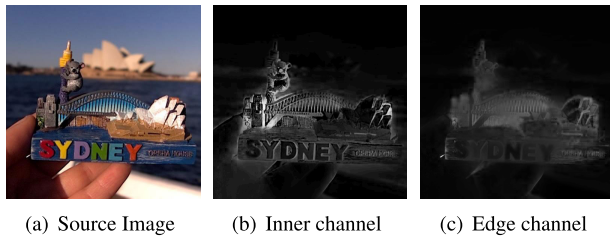

The associate editor coordinating the review of this manuscript and approving it for publication was Gangyi Jiang.

(a) Source Image   (b) Inner channel   (c) Edge channel

FIGURE 2. An example of inner channel and edge channel.

In this paper, we propose a novel prior—inner channel prior—for focus detection. The inner channel prior is based on the statistics of sharp images. We find, the color saturation of object varies inside the object, thus the high spatial frequency information of the saturation channel mainly locates inside the object. Generally, the high spatial frequency information of the image primarily appears at the edge of the object [15], which is called the edge channel, as shown in Fig.2. Therefore, the high frequency information of saturation can directly produce an sparse focus map of the sharp object. Then the guided filter [16] is applied to propagate the sparse map to the entire image. In addition, the high spatial frequency information of value channel, which mainly appears at the edge of the object, is used to deal with the blurry edge of multi-focus images.

The main contributions of this paper are described as follows:

1) We propose the inner channel prior to measure the focus degree inside objects, avoiding the misjudgement caused by the edge information based measurement.

2) Inspired by the sum-modified-Laplacian (SML), we select the guided filter instead of sum operator to propagate the sparse sharpness map to the full image. Guided filter-modified-Laplacian operator would increase the focus detection accuracy.

3) We utilize the edge channel to refine the focus detection at the edge of objects and some tiny objects containing a small number of pixels.

4) Combining the image pyramid fusion method, we can generate a high-quality fused image. Experiments demonstrate that our approach is physically valid and efficient.

5) Using the thread programming, we can simultaneously process the R,G,B channel, which would reduce the time cost of the algorithm.

II. RELATED WORK

Multi-focus image fusion aims to acquire an all-in-focus image, which can be used in diverse tasks of image processing [17]. From the literature, multi-focus image fusion (MFIF) methods can be divided into 3 categories: spatial domain, transform domain and neural network [18].

The spatial domain algorithms directly cope with pixels and pixel-intensity based operations [19], which can be divided into pixel-based and region-based [20]. The dense SIFT method [21] is a typical method of spatial domain

method. Spatial domain algorithm can be time-efficient but sensitive to noise. Besides, it tends to have fixed limitations, such as image blurring and spatial distortion [22], which would lead to the artifacts in the fused image. However, these problems can be tackled in transform domain.

The transform domain algorithms convert the original images into the feature domain, in which these images can be fused with higher quality [23]. The representative transform domain algorithms consist of the nonsubsampled contourlet transform (NSCT) technique [24], the sparse representation technique [25], [26], wavelet transform [27] and gradient domain technique [28]. The gradient domain method transforms the source images into the gradient domain, then the image fusion is implemented by gradient process. After processing the gradients, image reconstruction is executed to obtain the sharp image. However, transform domain algorithms are time-consuming and computational expensive.

The deep learning algorithms have become increasingly popular techniques for MFIF in the past few years. Neural network is a feature model of deep learning which can learn to extract the characteristic descriptors for an image in separate levels of abstraction [29]. The deep learning methods can obtain an all-round good performance regarding all the fusion metrics but cannot present the obvious strength in every single metrics compared with the traditional methods [20]. Despite the great advancement achieved in recent years, there are still several major issues in the area of multi-focus image fusion [12], [20], such as defocus effects, color distortion and process efficiency. In this paper, we plan to cope with these challenging problems.

III. INNER CHANNEL PRIOR

The inner channel prior is based on the following observation on sharp images: An all-in-focus natural image has a vivid color with large saturation range, while a defocused image appears more muted and grey with smaller saturation range. Meanwhile, the saturation of object color varies inside the object, thus the high spatial frequency information of the saturation channel mainly locates inside the object (Fig. 3(b)), whereas the high spatial frequency information of the value channel primarily appears at the edge of the object (Fig. 3(c)).

A. DEFINITION

To formally describe this observation, we first introduce the HSV color model. Compared with RGB using primary color, HSV is closer to the way human perceive color. HSV model has three ingredients: hue, saturation and value. This color space describes colors in terms of their saturation and their brightness value. According to [30], the ingredient value is the prime part that is used really in image processing. However, we mainly show the novel characteristic of the saturation component in this paper.

HSV model can be derived from RGB model by the following mathematical relationship:
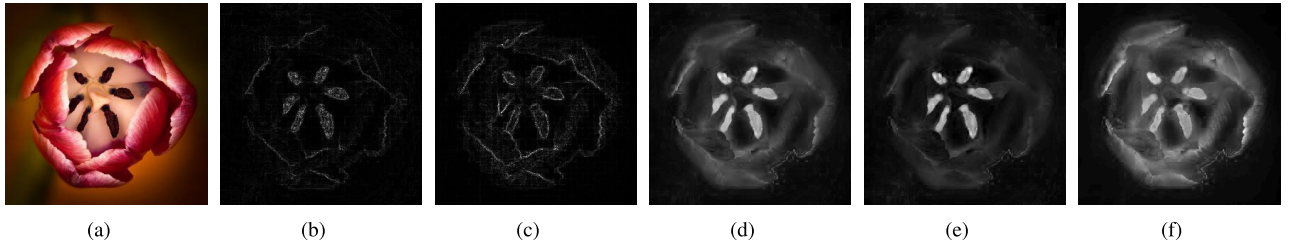
$$max = \max(R, G, B), min = \min(R, G, B)$$

FIGURE 3. (a) An arbitrary image (285 × 285). (b) The high frequency information of the saturation ingredient. (c) The high frequency information of the value ingredient. (d) inner channel + edge channel. (e) inner channel. (f) edge channel.

$$s = \begin{cases} 0, & max = 0 \\ \dfrac{max - min}{max} = 1 - \dfrac{min}{max}, & \text{otherwise} \end{cases}$$
$$v = max \qquad (1)$$

where R, G, B are the value of RGB channel, respectively. *max* and *min* are the mathematical operators, return the maximum and minimum of the variables, respectively. *max* and *min* are the maximum and minimum of R, G, B, respectively. s and v are the saturation component and value component of the image, respectively.

Since this prior is used for focus detection, we choose the focus evaluation to acquire the high frequency information of the saturation. According to [10], we choose the ML (Modified Laplacian) operator to perform the computation. The sparse focus map ML^{in} is given by,

$$ML^{in}(x, y) = | 2I^{sat}(x, y) - I^{sat}(x - 1, y)$$
$$- I^{sat}(x + 1, y) | + | 2I^{sat}(x, y)$$
$$- I^{sat}(x, y - 1) - I^{sat}(x, y + 1) | \qquad (2)$$

where $I^{sat}(x, y)$ is value of the saturation component of the input image I at (x, y). (x, y) is the pixel index. $ML^{in}(x, y)$ is the sparse map of inner channel.

Then we apply the Guided filter [16] to propagate the focus estimates from a sparse high-frequency map to the entire image (Fig. 3(e)) for acquiring inner channel I^{in} , using the guided image I, which is the original input image. The inner channel is defined as

$$I_i^{in} = \text{guidedFilter} \left(ML_i^{in} \right)$$
$$= \frac{1}{|\omega|} \sum_{k, i \in \omega_k} (a_k I_i + b_k), \quad \forall i \in \omega_k \qquad (3)$$

where i is the pixel index, I_i^{in} is the value of inner channel at pixel i, ω_k is a window centered at the pixel k. $|\omega|$ is the size of the window ω_k, (a_k, b_k) are some linear coefficients assumed to be constant in ω_k, defines as $a_k = \frac{\frac{1}{|\omega|} \sum i \in \omega_k ML_i^{in} I_i - \mu_k \overline{ML_k^{in}}}{\sigma_k^2 + \epsilon}$, $b_k = \frac{1}{|\omega|} (\sum_{i \in \omega_k} ML_i^{in} - a_k \sum_{i \in \omega_k} I_i)$, $\mu_k = \frac{1}{|\omega|} \sum_{i \in \omega_k} I_i$, $\sigma^2 = E(I^2) - E(I)^2$. For the detailed derivation of Eq.(3), readers can refer to [31].

B. FILTER PARAMETER
There are two parameters in guided filter influencing the resultant image: regulation ϵ and radius r. ϵ is a regularization

parameter penalizing large a_k. The edge of the output image would be better preserved with smaller ϵ, and in this paper $\epsilon = 10^{-4}$, according to [31]. r is the radius of the window ω_k. The output image would be smoother with larger radius r. This can be seen in Fig. 4.

However, the radius can't be too large, since large radius would smooth the edge of the fused image and degrade the resultant image qualification, as shown in Fig. 5. So in this paper, we set radius to 15.

Using the definition of the inner channel, our observation indicates that if I is a sharp image with clear objects, the signal of I' inner channel mainly located inside the object. The signal of the inner channel locating inside the object mainly due to two factors: a) it is pretty rare to see pure colors in nature, which means there are mixed colors with saturation varied. b) most objects are three-dimensional. The saturation perceived by the sensor can be slightly changed because of the angle of the incident ray.

To verify how excellent the inner channel is, we collect a sharp image database consisting of 150 images from heliconsoft.com and Flocker.com. Fig. 6 shows several sharp images and the corresponding inner channels and edge channels.

As shown in the Fig. 6, the inner channel can indicate the focus degree of objects better than the edge channel, which could be clearly seen in the picture of ''mite'' (the 3rd picture), ''wing'' (the 4th picture) and ''bee'' (the 5th picture). However, there are many edges and tiny objects in the natural images, such as the 6th and 7th image, so we utilize the edge channel to refine the focus map of natural images.

IV. MULTI-FOCUS IMAGE FUSION
Multi focus fuison produces the desired image by picking up merely the best parts in the multi-focus image series. This procedure is conducted by a set of focus measures, which are then united into a full weight map. It is convenient to refer the input image series as a focus stack. The focused image is then produced by collapsing the focus stack under the guidance of full weight map.

A. FOCUS DETECTION
All the images in the focus stack comprise defocused regions because of the DOF of the cameras. Such regions should be abandoned whereas the regions including sharp objects and details should be preserved. We plan to fulfill this by carrying
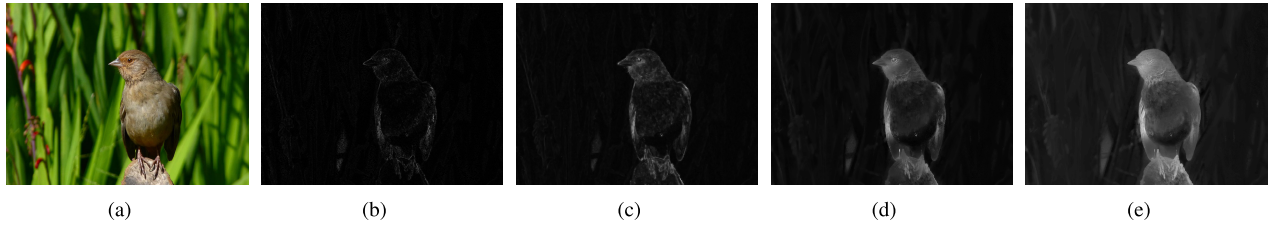
FIGURE 4. (a) a 800 × 600 sharp image and the sparse focus map (b) and the inner channel using (c) *r* = 3, (d) *r* = 15, (e) *r* = 30.
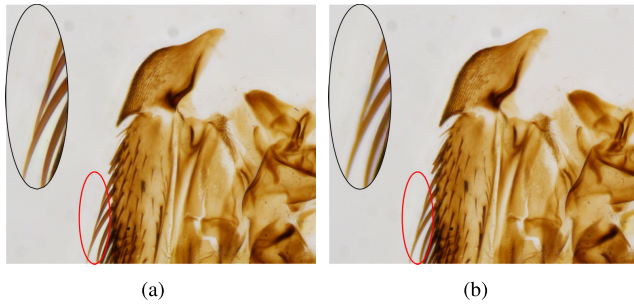

FIGURE 5. The resultant image of 2448 × 2048 image sequences using (a) *r* = 15, (b) *r* = 30.

out following metrics. The larger pixel value of the metrics means the pixel is better focused.

1) INNER CHANNEL

We apply a *ML* operator combined with guided filter to the saturation component of each image, yielding an indicator I^{in}. It tends to assign a high weight to significant elements inside objects.

2) EDGE CHANNEL

There are burring effect locating at the edge of the object, which needs to be addressed to generate a sharp fused image. Thus we introduce the edge channel I^{ed}, which is computed by a *ML* operater combined with guided filter. This parameter specify the interesting elements at the edge of the object (Fig. 3(f)). Meanwhile, this measure can be used to compensate for the limitation of inner channel, such as grey objects (saturation equals to zero) and some tiny objects consisting of very few pixels.

The definition of edge channel I^{ed} is similar to that of inner channel I^{in}.

$$ML^{ed}(x, y) = |\, 2I^{val}(x, y) - I^{val}(x - 1, y)$$
$$-I^{val}(x + 1, y)\,| + |\, 2I^{val}(x, y)$$
$$-I^{val}(x, y - 1) - I^{val}(x, y + 1)\,| \quad (4)$$

where $I^{val}(x, y)$ is value of the value component of the input image I at (x, y). (x, y) is the pixel index. $ML^{ed}(x, y)$ is the sparse map of edge channel.

$$I_i^{ed} = \text{guidedFilter}\left(ML_i^{ed}\right)$$
$$= \frac{1}{|\omega|} \sum_{k, i \in \omega_k} (a_k I_i + b_k), \quad \forall i \in \omega_k \quad (5)$$

where i is the pixel index, I_i^{ed} is the value of edge channel at pixel i, ω_k is a window centered at the pixel k. $|\omega|$ is the size of the window ω_k, (a_k, b_k) are some linear coefficients assumed to be constant in ω_k.

For each pixel, we combine the information processed by the two separate measures to form a full focus map using summation. We select pixels over a linear combination, as we want to apply all qualities defined by these measures. Similar to weighted items of a linear combination, we can alter the significance of each measure using following power mathematical relatioship:

$$\widehat{W_{i,j,k}} = \left(I_{i,j,k}^{in}\right)^{\alpha} + \left(I_{i,j,k}^{ed}\right)^{\beta} \quad (6)$$

where α and β are the corresponding weighting exponents of its focus measure. The subscript i, j, k denotes pixel (i, j) in the k-th image. The $\widehat{W_{i,j,k}}$ (Fig. 3(d)) is the full focus map.

B. WEIGHT PARAMETER

$I_{i,j,k}^{in}$ aims to maintain color of the object and obtain sharp object, and $I_{i,j,k}^{ed}$ aims to deal with blurry effects. The larger α and β indicates the better corresponding performance. However, larger α and β means higher computational complexity and lower processing efficiency. The performance of image fuison with separate β are shown in Fig. 7.

We cannot obtain an image with clear edge in the case of $\beta = 0$, whereas an clear image can be acquired while $\beta = 1$ and $\beta = 2$, as shown in Fig. 7.

α should be large in some cases, such as an image has different objects at same location but the fused image solely need the object at the top. As shown in Fig. 8.

$\alpha = 1$ and $\beta = 1$ can obtain the image with sharp object and edge. Larger α and β would be computational expensive since it utilize multiplication to do the computation. So α and β all equal to one in this paper. The guided filter enables the neighbouring region share the value of sparse map, which will enhance the weight of the defocused region (Fig. 9(a)), so we need to emphasise the sharp objects and details (Fig. 9(b)) by the following function

$$W_{i,i,k} = [\left(I_{i,j,k}^{in}\right)^{\alpha} + \left(I_{i,j,k}^{ed}\right)^{\beta}] * [\left(ML_{i,j,k}^{in}\right)^{\alpha} + \left(ML_{i,j,k}^{ed}\right)^{\beta}] \quad (7)$$

where $W_{i,j,k}$ is the full weight map. α and β are the corresponding weighting exponents of its focus measure. The subscript i, j, k denotes pixel (i, j) in the k-th image. $I_{i,j,k}^{in}$ and $I_{i,j,k}^{ed}$ are the value of inner channl and edge channel
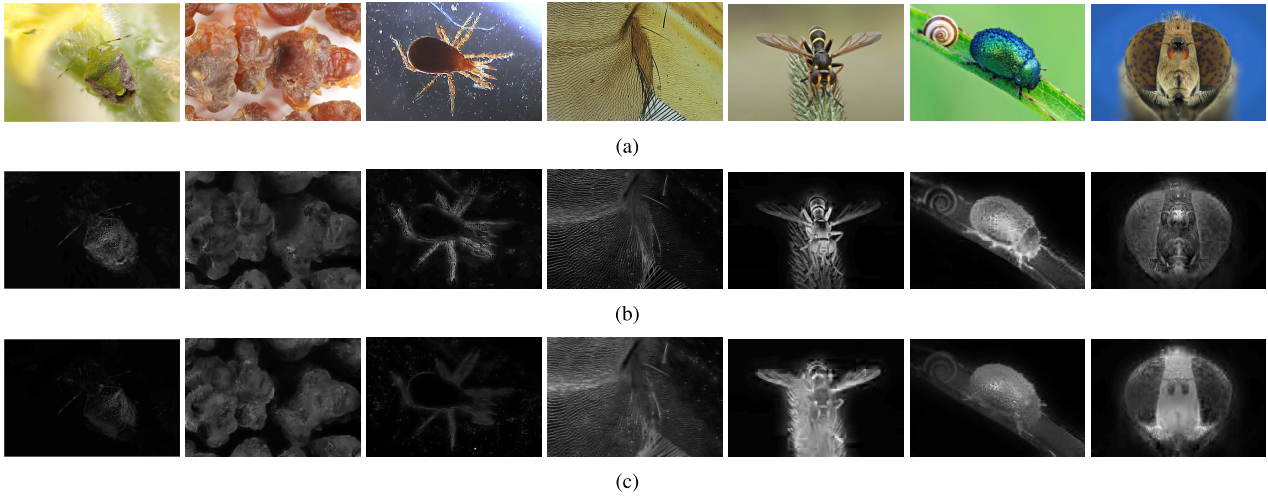
FIGURE 6. (a) Example images in our sharp image database. (b) The corresponding inner channels. (c) The corresponding edge channels.
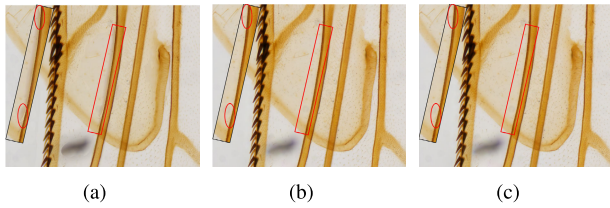

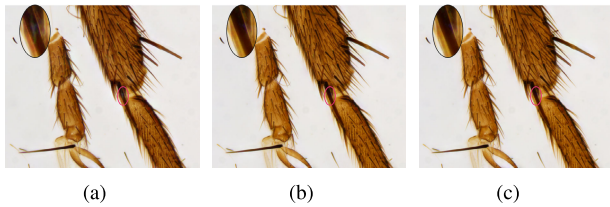**FIGURE 7.** The fused image using (a) $\alpha = 1$, $\beta = 0$. (b) $\alpha = 1$, $\beta = 1$. (c) $\alpha = 1$, $\beta = 2$.

FIGURE 8. The fused image using (a) $\alpha = 0.5$, $\beta = 1$. (b) $\alpha = 1$, $\beta = 1$. (c) $\alpha = 2$, $\beta = 1$.
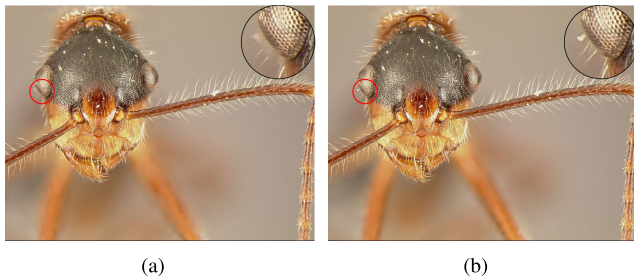

FIGURE 9. (a) The fused image guided by full focus map $\widehat{W_{i,j,k}}$. (b) The fused image guided by full weight map $W_{i,j,k}$.

respectively. $ML_{i,j,k}^{in}$ and $ML_{i,j,k}^{ed}$ are the sparse map of inner channel and edge channle respectively.

The full weight map $W_{i,j,k}$ is used to guide the image fusion.

C. IMAGE FUSION

We use a technique inspired by Mertens *et al.* [32], which seamlessly fuses N images guided by N normalized weight maps, and works at multiple scales using a pyramid image decomposition. We adapt this technique to our case.

Define the l-th level in a Laplacian pyramid decomposition of k-th image as L_k^l, and G_k^l is the same definition in a Gaussian pyramid decomposition.

We build the Laplacian pyramid of every RGB image for fusion and the Gaussian pyramid of the saturation and value components for generating the full weight map. Then, we generate the full focus map $\widehat{W_k}$ using the original-image level of the Gaussian pyramid, as it boasts the full information. We separately compute the sparse focus map at each scale of Gaussian pyramid, forming a sparse map pyramid Gs_k^l.

$$Gs_k^l = \left(ML_k^{in,l}\right)^\alpha + \left(ML_k^{ed,l}\right)^\beta \qquad (8)$$

where Gs_k^l is the l-th level of the k-th image of the sparse map pyramid. α and β are the corresponding weighting exponents of its focus measure. $ML_k^{in,l}$ and $ML_k^{ed,l}$ are the corresponding sparse map of inner channel and edge channel at l-th level of the k-th image.

To generate the Focus pyramid F_k^l, we build the Gaussian pyramid of the $\widehat{W_k}$, forming a full focus map pyramid Gw_k^l. The Focus pyramid is obtained by multiplying Gs_k^l by Gw_k^l respectively.

$$F_k^l = Gs_k^l * Gw_k^l \qquad (9)$$

where F_k^l, Gs_k^l and Gw_k^l is the corresponding focus map, sparse map and weight map at l-th level of k-th image.

The mask pyramid is obtained by applying following function to the Focus pyramid series of image sequences:

$$M_{i,j}^l = \underset{k}{\mathrm{argmax}} \left(F_{i,j,k}^l\right) \qquad (10)$$

where $M_{i,j}^l$ is the value of l-th level at (i,j) pixel. *argmax* is the mathematical operator which return the index of the variables. F is the focus map pyramid. The subscript i,j,k denotes pixel (i,j) in the k-th image.

I_1, I_2 : Input source images
$\widehat{Wp_1}, \widehat{Wp_2}$: Gaussian pyramids of full focus map
F_1, F_2 : Focus pyramids of image
M : Mask pyramid

LP : build Laplacian pyramid
GP : build Gaussian pyramid
FP : build Focus pyramid
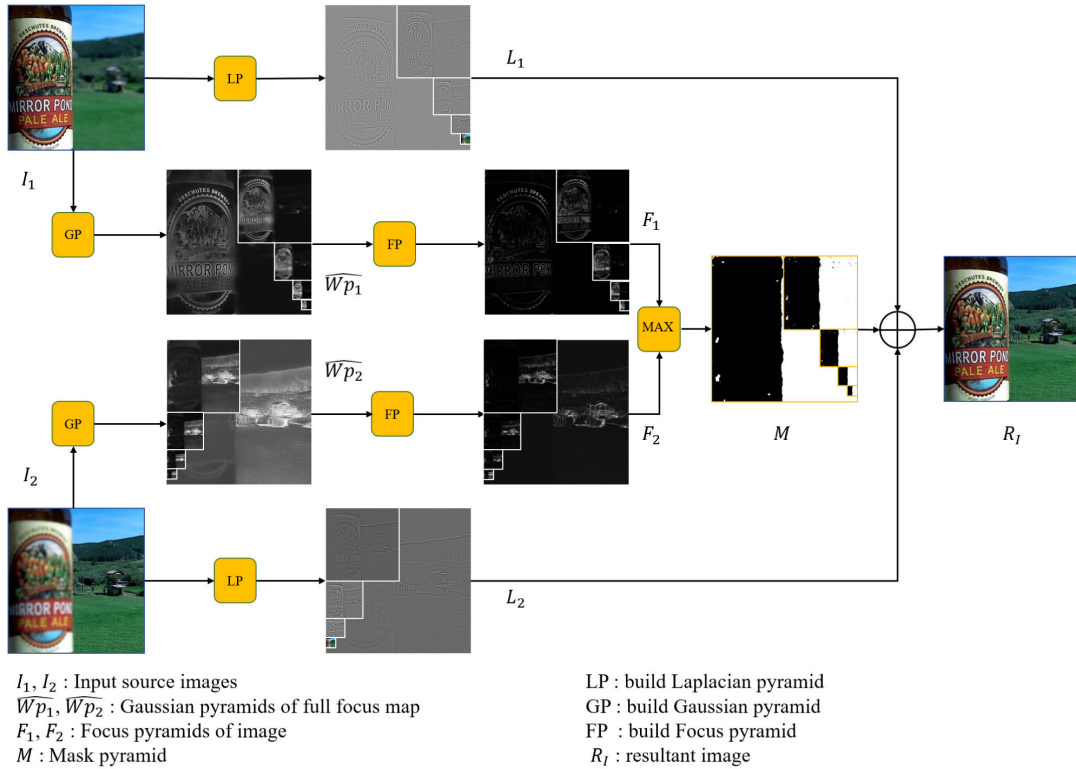R_I : resultant image

FIGURE 10. The full procedure of our proposed method.

Then the all-in-focus image pyramid is obtained under the guidance of the mask pyramid M. The intensity value at pixel (i, j) of the l-th level in sharp image pyramid is the intensity value at pixel (i, j) of the same level of k-th multi-focus image, where k is the intensity value at pixel (i, j) of the l-th level in mask pyramid. The full procedure is shown in Fig. 10. The R,G,B channel of image would be seperately processed, using thread programming in visual studio 2017.

V. EXPERIMENTS

We compare the proposed method with 7 other multi-focus fusion techniques: BRW [33], MGFF [34], NSCT-SR [35], DSIFT [35], GFDF [36], MFF-Net [37] and CNN [29]. The 'Lytro' [38] dataset is introduced in our experiments, which is normally used for MFIF.

A. QUANTITATIVE EVALUATION

Quantitative analysis has been performed using four widely used image metrics: structural similarity based metric Q_s [39], average gradient (AG) [40], mean square deviation (MSD) [37] and linear index of fuzziness (LIF) [41]. The picked 4 metrics are concisely introduced as follows.

1) STRUCTURAL SIMILARITY BASED METRIC Q_s

The metric Q_s simultaneously takes the original images and edge images into consideration.

$$Q = \frac{1}{|W|} \sum_{w \in W} \left[\begin{array}{c} \lambda(w)Q_0(A, F \mid w) \\ +(1 - \lambda(w))Q_0(B, F \mid w) \end{array} \right]$$

$$Q_W = \sum_{w \in W} c(w) \left[\begin{array}{c} \lambda(w)Q_0(A, F \mid w) \\ +(1 - \lambda(w))Q_0(B, F \mid w) \end{array} \right]$$

$$Q_E = Q_W(A, B, F) \cdot Q_W(A', B', F')^{\alpha} \quad (11)$$

where $Q_0(A, F \mid w)$ refers to the UIQI [42] map obtained by a sliding window w operation using A and F. The saliency weight $\lambda(w)$ is computed based upon the local variances of the source images. $c(w)$ is the normalized saliency of the family of all windows. A', B' and F' correspondingly refers to the edge images of A, B and F. α is a weight parameter controling the weight of the edge term. A larger Q_s indicates a better image fusion performance.

2) AVERAGE GRADIENT (AG)

The metric AG takes the gradient information to measure the quality of fused images:

$$AG = \frac{1}{(M-1)(N-1)}$$
$$\times \sum_{m=1}^{M-1} \sum_{n=1}^{N-1} \frac{1}{4} \sqrt{\left(\frac{\partial I(m, n)}{\partial m} \right)^2 + \left(\frac{\partial I(m, n)}{\partial n} \right)^2}$$
$$(12)$$

where $\frac{\partial I(m,n)}{\partial m}$ and $\frac{\partial I(m,n)}{\partial n}$ correspondingly denote the gradients of the image in horizontal and vertical directions. M and N are the corresponding total pixel number of the row and column of the image. A larger AG refers to more and sharper boundaries in the fused image.

TABLE 1. The Quantitative Comparison of Different MFIF Techniques on the 'Lytro' Dataset.

Metrics	CNN	DSIFT	GFDF	BRW	MGFF	MFF-Net	NSCT-SR	Proposed
Q_s	0.2119	0.2111	0.2102	0.2167	0.2158	0.2164	0.2098	**0.2172**
AG	9.6709	**9.7886**	9.7727	0.7792	0.7751	9.7754	9.6917	*9.7808*
MSD	0.5967	0.5987	0.5976	06008	05986	0.5997	0.5961	**0.6010**
LIF	0.3875	0.3870	0.3872	0.3868	0.3874	0.3870	0.3874	**0.3860**
Time(s)	139.33$_{(512\times512)}$	37.18$_{(512\times512)}$	1.33$_{(512\times512)}$	0.57$_{(512\times512)}$	1.69$_{(512\times512)}$	33.02$_{(512\times512)}$	51.09$_{(512\times512)}$	**0.777**$_{(2448\times2048)}$

3) MEAN SQUARE DEVIATION (*MSD*)

The metric *MSD* indicaes the image detail richness by computing the difference between intensity and mean intensity of fused image, a larger *MSD* means a clearer resultant image.

$$MSD = \frac{1}{(M-1)(N-1)}$$
$$\times \sqrt{\sum_{m=1}^{M-1}\sum_{n=1}^{N-1}(I(m,n) - \bar{I})^2} \quad (13)$$

where M and N are the corresponding total pixel number of the row and column of the image. m, n are the pixel indexes. $\bar{I}$ is the mean of the image.

4) LINEAR INDEX OF FUZZINESS (*LIF*)

The metric *LIF* evaluate the enhancement of the resultant image. A smaller *LIF* refers to better enhancement of the fused image.

$$LIF = \frac{2}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\min\{p_{mn}, (1 - p_{mn})\}$$
$$p_{mn} = \sin\left[\frac{\pi}{2}\left(1 - \frac{I(m,n)}{I_{max}}\right)\right] \quad (14)$$

where $I(m,n)$ is the intensity of pixel (m,n) and I_{max} is the maximum intensity. m, n are the pixel indexes. M and N are the corresponding total pixel number of the row and column of the image.

The quantitative comparisons over 20-pair test images are shown in Table 1, the larger *AG*, *MSD*, Q_s and smaller *LIF* denote the better performance. The best results are highlighted in bold. The proposed method remarkably outforms the other methods in all quality metrics except the AG metric, because the DSIFT method introduces some edges near the focused/defocused boundary, which are nonexistent in source images, as shown in Fig. 11(d).

5) TIME COMPLEXITY

All the chosen experiments, except MFF-Net, are performed in MATLAB R2020a on a PC with Intel(R) Core(TM) i7-8550U CPU @ 2.00GHz and 16-GB RAM. MFF-Net is performed in Python 3.8, and the proposed method is performed in visual studio 2017 with OpenCV 3.4.3. The parameters of different fusion methods are set to their default value.

The time complexity comparisons over test images are shown in Table 1. BRW method is the fastest algorithm among selected mehthods, which costs 0.57s to fuse two

520 × 520 images. The proposed method cost 0.77s to fuse two 2448 × 2048 images. We use the BRW to fuse two 1944 × 1296 images, and it cost 4.61s to obtain the fused image.

We utilize the thread programming to simultaneously process the R, G, B channel, which would save the two-thirds time. Fast guided filter [16] is utilized to maximize the time efficiency of the algorithm. Moreover, we use the summation to combine the inner channel with edge channel, which would be more efficient compared with multiplication.

B. QUALITATIVE ANALYSIS

1) LYTRO DATASET

Fig. 11 shows the visual comparison on the 'Lytro' dataset. We select three-pair images to show the strength of the proposed method in three different situation: focused/defocused boundary with small defocused area and large defocused area and the acceptably sharp region.

The top row ('Lytro-09') shows the strength of the proposed method to deal with the focused/defocued boundary with small defocused area. Source image A (Fig. 11(a)) shows the focused/defocued boundary and source image B (Fig. 11(b)) shows the sharp grassland near the boundary. BRW (Fig. 11(f)) and the proposed method (Fig. 11(j)) acquire an acceptably sharper boundary and sharper grassland, which is closer to the source image, particularily compared with DSIFT (Fig. 11(d)), CNN (Fig. 11(c)), MGFF (Fig. 11(g)) and NSCT-SR (Fig. 11(i)).

The middle row ('Lytro-19') presents the power of proposed method to cope with the acceptably sharp region obtain the acceptably sharp region. Source image A presents the small defocused region. The CNN (Fig. 11(c)), GFDF (Fig. 11(e)), BRW(Fig. 11(f)) and NSCT-SR (Fig. 11(i)) fail in the case of fusing two large defocused area image. However, the proposed method (Fig. 11(j)) acquires a clearer area similar to that in the source image A (Fig. 11(a)).

The bottom row ('Lytro-13') exhibit the ability of the proposed technique to address the focused/defocued boundary with large defocused area. There are sharp region consisting of grass in source images, as shown in the enlarged region. BRW(Fig. 11(f)) and the proposed method (Fig. 11(j)) obtain the sharp region and maintain the sharp boundary between foreground and background. MGFF (Fig. 11(g)) and NSCTSR(Fig. 11(i)) obtain the sharp region and blurry edge.

2) COLLECTED DATASET

We respectively collected four datasets that include 22 defocused images of stone, 6 defocused images of flower,
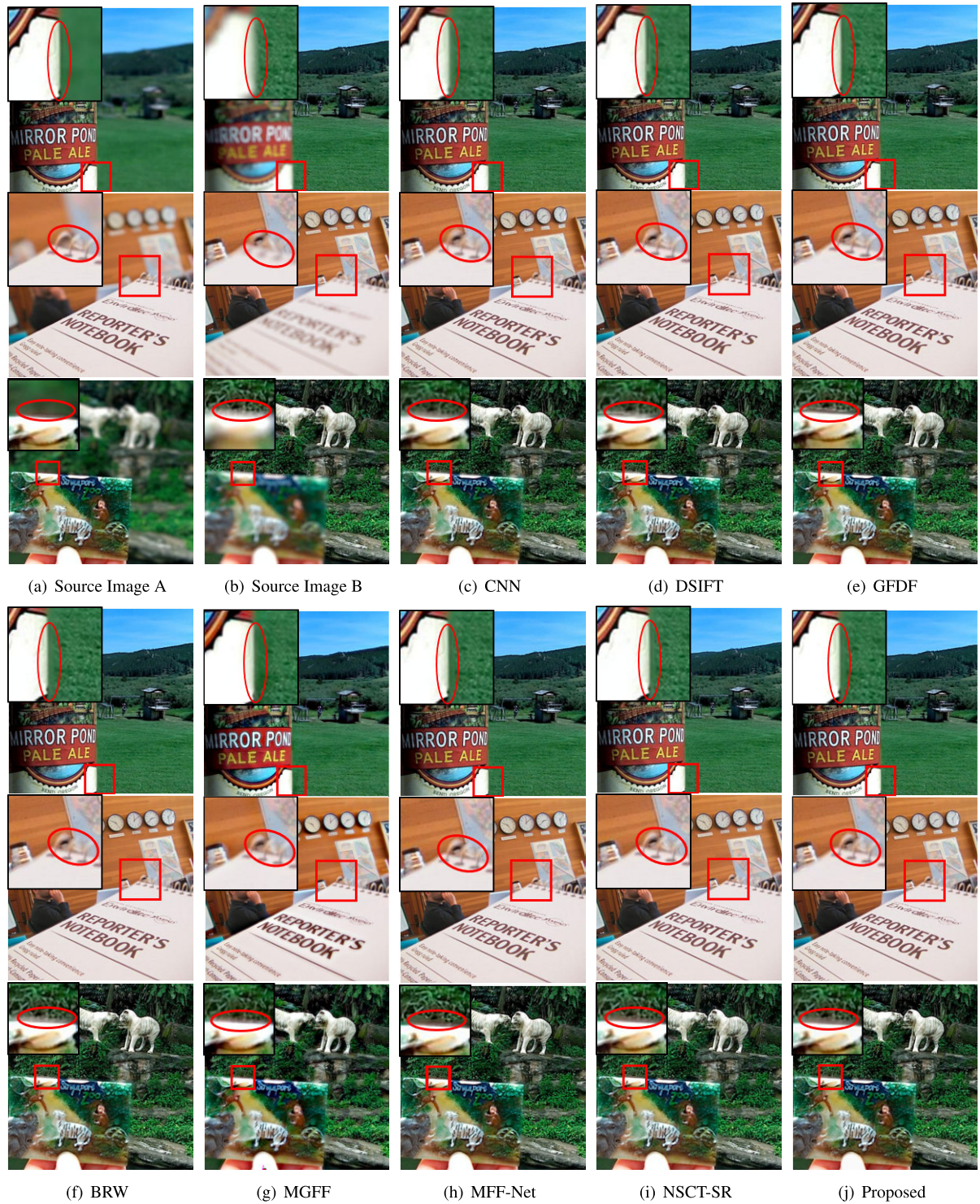
FIGURE 11. The fusion result of different methods on the 'Lytro' dataset. Compared with other techniques, both the edge of the object and background are clearer in proposed method, as shown in the enlarged region.

TABLE 2. The Quantitative Comparison of Different MFIF Techniques on the collected Dataset.

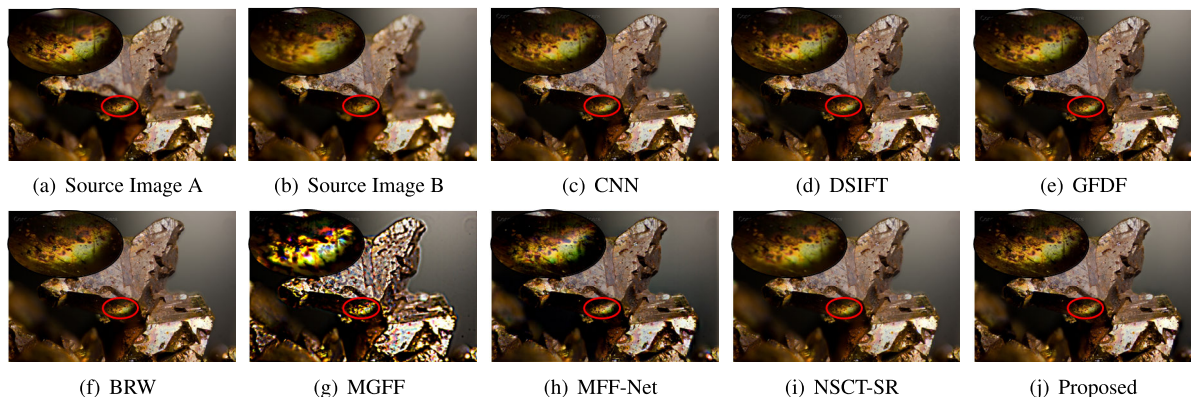| Metrics	CNN	DSIFT	GFDF	BRW	MGFF	MFF-Net	NSCT-SR	Proposed
Q_s	0.1933	0.1927	0.1932	0.1937	0.1922	0.1935	0.1934	**0.1939**
AG	1.1415	1.2155	1.1791	1.2585	**1.3555**	1.1465	1.1380	1.2767
MSD	0.7258	0.7120	0.7178	0.7381	0.7227	0.7362	0.7293	**0.7448**
LIF	0.6120	0.6115	0.6162	0.6088	0.6134	0.6075	0.6092	**0.5926**

FIGURE 12. The fusion result of different method on addressing color distortion. Compared with other techniques, the proposed method can obtain better color fidelity, as shown in the enlarged region.
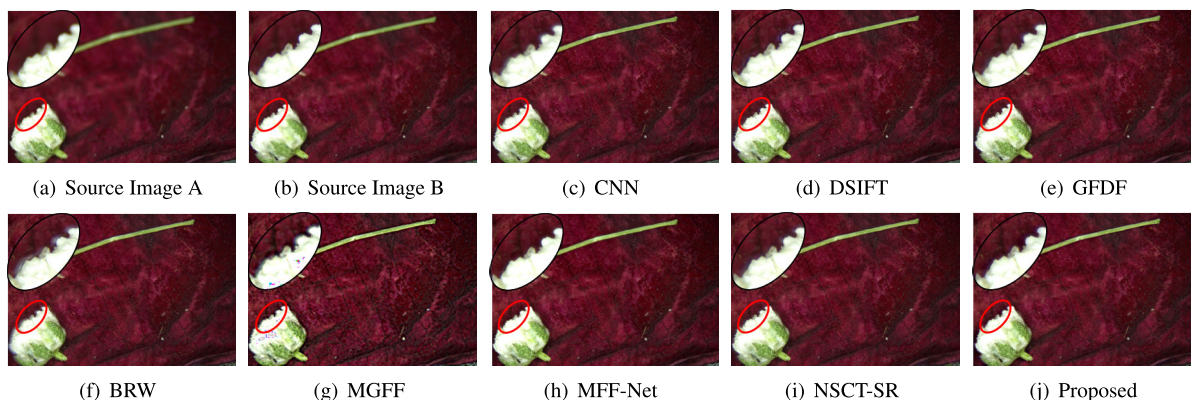

FIGURE 13. The fusion result of different method on addressing blurry effects. Compared with other techniques, both the edge of the object and background are clearer in proposed method, as shown in the enlarged region.
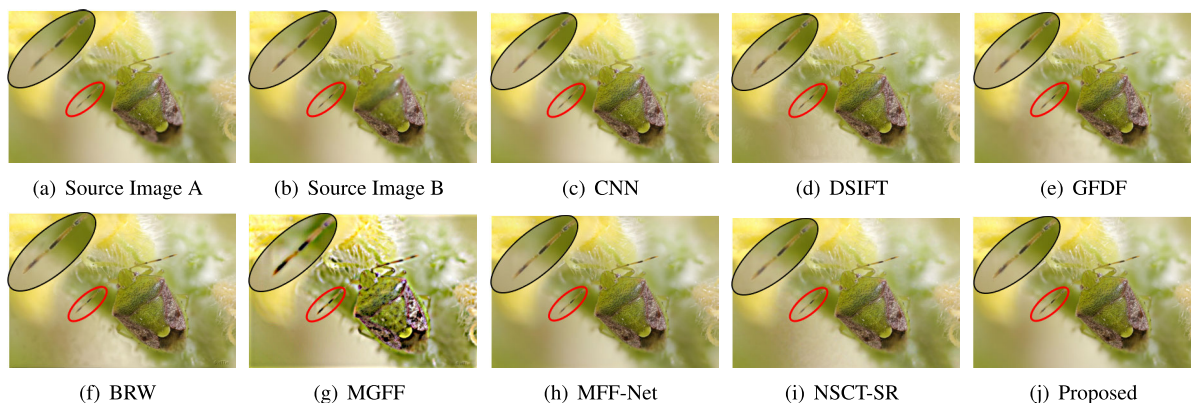

FIGURE 14. The fusion result of different method on fusing tiny objecct, Compared with other techniques, the proposed method acquires sharper object and better color fidelity, as shown in the enlarged region.

8 defocused images of insect and 9 defocused images of mental mesh to show the strength of our method, as shown in Fig. 12-Fig. 15. The source images have the clear region of interest.

a: COLOR DISTORTION

Fig. 12 demonstrates merit of the proposed method to cope with the color distortion. Source image A (Fig. 12(a)) shows the sharp colorful area. BRW (Fig. 12(f)), GFDF (Fig. 12(e)) and the proposed method (Fig. 12(j)) obtain the sharp region

with right color. CNN (Fig. 12(c)), MGFF (Fig. 12(g)) and NSCT-SR (Fig. 12(i)) fail to acquire the clear region on the lower left area. MFF-Net (Fig. 12(h)) introduces the color distortion in the green area whereas DSIFT (Fig. 12(d)) introduces the color distortion in the brown area.

b: BLURRY EFFECTS

Fig. 13 mainly shows the ability of the proposed technique to deal with blurry effects. Source image B (Fig. 13(b)) shows the sharp area. The results of CNN (Fig. 13(c)), GFDF
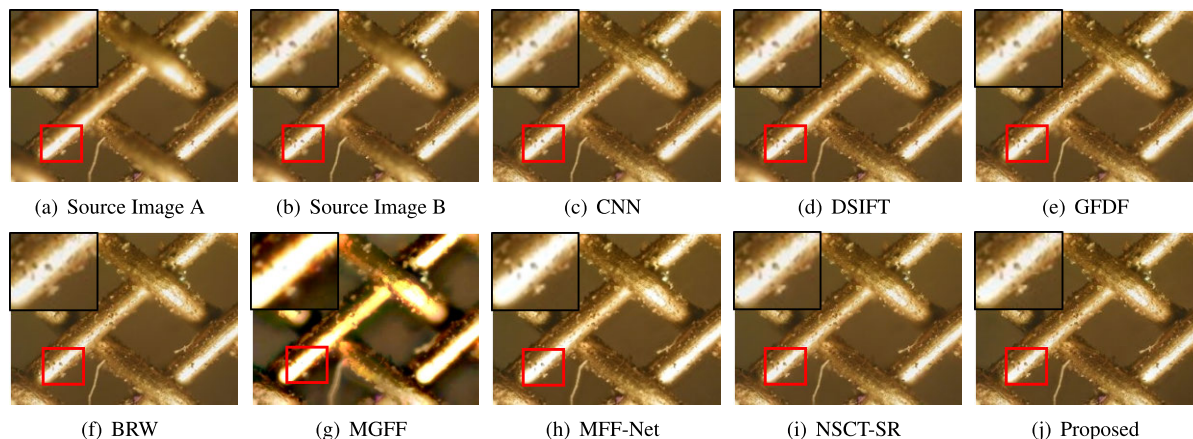
FIGURE 15. The fusion result of different method on addressing overexposure. Compared with other techniques, both the edge of the object and background are clearer in proposed method, as shown in the enlarged region.

(Fig. 13(e)), MGFF (Fig. 13(g)) and NSCT-SR (Fig. 13(i)) all have white shade around the white flower. BRW (Fig. 13(f)) and DISFT (Fig. 13(d)) introduce another tiny edge near the object. MFF-Net (Fig. 13(h)) and proposed method (Fig. 13(j)) can acquire the sharp boundary similar to the source image.

c: TINY OBJECT

Fig. 14 presents the ability of our technique coping with tiny object. Source image B (Fig. 14(b)) shows the sharp object. BRW (Fig. 14(f)), GFDF (Fig. 14(e)) and the proposed method (Fig. 14(j)) obtain the sharper object compared with others. MGFF (Fig. 14(g)) fails to acquire the clear image in this case.

d: OVEREXPOSURE

Fig. 15 shows the strength of the proposed method dealing with overexposure. Source image A (Fig. 15(a)) shows the overexposure region and source image B (Fig. 15(b)) shows the clear area. The overexposure would affect the focus metric, which would fuse more defocused region into the resultant image, as shown in CNN (Fig. 15(c)), GFDF (Fig. 15(e)) and NSCT-SR (Fig. 15(i)). MGFF (Fig. 15(g)) and MFF-Net (Fig. 15(h)) fail to recognize the boundary of the overexposure region. Only DSIFT (Fig. 15(d)) and the proposed method (Fig. 15(j)) acquire the clear region similar to source image.

The quantitative evaluation of the new dataset are show in Table 2. The best results are highlighted in bold. The proposed method remarkably outforms the other methods in all quality metrics except AG metric, because the MGFF method fails to obtain sharp image and introduces a lot of noise, as shown in Fig. 12- Fig. 15.

VI. CONCLUSION

In this paper, we have proposed a very simple but effective prior, called inner channel prior, for multi-focus image fusion. The inner channel prior aims to detect the sharp region inside objects, whereas the edge channel aims to obtain the

sharp region at the edge of objects. Combined with edge channel, inner channel can acquire the sharp area of the whole image. Since the saturation is introduced in the inner channel, this proposed method can avoid the color distortion occuring in the fusion process. Meanwhile, this method can be time-effective by the multi-threaded optimization. Some experiments have shown the high efficiency and effectiveness of the prior compared with existing typical methods.

Since the inner channel prior is based on the HSV model, it would be invalid when the image sequences are gray-scale images. Gray-scale images don't contain the saturation component, so the inner channel prior can't be used to measure the focus degree inside objects. Moreover, as the inner channel is obtained by using guided filter, its accuracy may decrease when the halo artifacts are introduced by the guided filter.

As for another common challenging problem in image fusion, misregistration, is not addressed in this paper. We leave this problem for future research.

REFERENCES

[1] S. G. Simone, A. Farina, F. C. Morabito, S. B. Serpico, and L. Bruzzone, "Image fusion techniques for remote sensing applications," *Inf. Fusion*, vol. 3, no. 1, pp. 3–15, 2002.

[2] Z. Wang and Y. Ma, "Medical image fusion using *m*-PCNN," *Inf. Fusion*, vol. 9, no. 2, pp. 176–185, 2008.

[3] C. Panigrahy, A. Seal, and N. K. Mahato, "MRI and SPECT image fusion using a weighted parameter adaptive dual channel PCNN," *IEEE Signal Process. Lett.*, vol. 27, pp. 690–694, 2020.

[4] Z.-G. Jiang, D.-B. Han, J. Chen, and X.-K. Zhou, "A wavelet based algorithm for multi-focus micro-image fusion," in *Proc. 3rd Int. Conf. Image Graph. (ICIG)*, 2004, pp. 176–179.

[5] R. Van de Plas, J. Yang, J. Spraggins, and R. M. Caprioli, "Image fusion of mass spectrometry and microscopy: A multimodality paradigm for molecular tissue mapping," *Nature Methods*, vol. 12, no. 4, pp. 366–372, 2015.

[6] M. Zheng, G. Qi, Z. Zhu, Y. Li, H. Wei, and Y. Liu, "Image dehazing by an artificial image fusion method based on adaptive structure decomposition," *IEEE Sensors J.*, vol. 20, no. 14, pp. 8062–8072, Jul. 2020.

[7] Z. Zhu, H. Wei, G. Hu, Y. Li, G. Qi, and N. Mazur, "A novel fast single image dehazing algorithm based on artificial multiexposure image fusion," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–23, 2021.

[8] Y. Song, M. Li, Q. Li, and L. Sun, "A new wavelet based multi-focus image fusion scheme and its application on optical microscopy," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Dec. 2006, pp. 401–405.

[9] M. Amin-Naji, A. Aghagolzadeh, and M. Ezoji, "Ensemble of CNN for multi-focus image fusion," *Inf. Fusion*, vol. 51, pp. 201–214, Nov. 2019.

[10] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 493–500, 2007.

[11] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, Aug. 1994.

[12] S. Bhat and D. Koundal, "Multi-focus image fusion techniques: A survey," *Artif. Intell. Rev.*, no. 6, pp. 1–53, 2021.

[13] J. Sun, Q. Han, L. Kou, L. Zhang, K. Zhang, and Z. Jin, "Multi-focus image fusion algorithm based on Laplacian pyramids," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 35, no. 3, pp. 480–490, 2018.

[14] Y. Chen, J. Guan, and W.-K. Cham, "Robust multi-focus image fusion using edge model and multi-matting," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1526–1541, Mar. 2018.

[15] A. Sengupta, A. Seal, C. Panigrahy, O. Krejcar, and A. Yazidi, "Edge information based image fusion metrics using fractional order differentiation and sigmoidal functions," *IEEE Access*, vol. 8, pp. 88385–88398, 2020.

[16] K. He and J. Sun, "Fast guided filter," 2015, *arXiv:1505.00996*. [Online]. Available: http://arxiv.org/abs/1505.00996

[17] Y. Yang, W. Zheng, and S. Huang, "Effective multifocus image fusion based on HVS and BP neural network," *Sci. World J.*, vol. 2014, pp. 1–10, Jan. 2014.

[18] P. Kaur and M. Kaur, "A comparative study of various digital image fusion techniques: A review," *Int. J. Comput. Appl.*, vol. 114, no. 4, pp. 26–31, Mar. 2015.

[19] M. Nejati, S. Samavi, N. Karimi, S. R. Soroushmehr, S. Shirani, I. Roosta, and K. Najarian, "Surface area-based focus criterion for multi-focus image fusion," *Inf. Fusion*, vol. 36, pp. 284–295, Jul. 2017.

[20] Y. Liu, L. Wang, J. Cheng, C. Li, and X. Chen, "Multi-focus image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 64, pp. 71–91, Dec. 2020.

[21] Y. Liu, S. Liu, and Z. Wang, "Multi-focus image fusion with dense SIFT," *Inf. Fusion*, vol. 23, pp. 139–155, May 2015.

[22] R. Garg, P. Gupta, and H. Kaur, "Survey on multi-focus image fusion algorithms," in *Proc. Recent Adv. Eng. Comput. Sci. (RAECS)*, Mar. 2014, pp. 1–5.

[23] D. K. Sahu and M. P. Parsai, "Different image fusion techniques—A critical review," *Int. J. Mod. Eng. Res.*, vol. 2, no. 5, pp. 4298–4301, 2012.

[24] H. Wang, "Multi-focus image fusion algorithm based on focus detection in spatial and NSCT domain," *PLoS ONE*, vol. 13, no. 9, Sep. 2018, Art. no. e0204225.

[25] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 884–892, Apr. 2010.

[26] Z. Zhu, H. Yin, Y. Chai, Y. Li, and G. Qi, "A novel multi-modality image fusion method based on image decomposition and sparse representation," *Inf. Sci.*, vol. 432, pp. 516–529, Mar. 2018.

[27] C. Panigrahy, A. Seal, N. K. Mahato, O. Krejcar, and E. Herrera-Viedma, "Multi-focus image fusion using fractal dimension," *Appl. Opt.*, vol. 59, no. 19, pp. 5642–5655, 2020.

[28] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," in *Proc. ACM SIGGRAPH Papers*, 2004, pp. 294–302.

[29] Y. Liu, X. Chen, H. Peng, and Z. F. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.

[30] W. Bao and X. Zhu, "A novel remote sensing image fusion approach research based on HSV space and bi-orthogonal wavelet packet transform," *J. Indian Soc. Remote Sens.*, vol. 43, no. 3, pp. 467–473, Sep. 2015.

[31] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[32] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion," in *Proc. 15th Pacific Conf. Comput. Graph. Appl. (PG)*, 2007, pp. 382–390.

[33] J. Ma, Z. Zhou, B. Wang, L. Miao, and H. Zong, "Multi-focus image fusion using boosted random walks-based algorithm with two-scale focus maps," *Neurocomputing*, vol. 335, pp. 9–20, Mar. 2019.

[34] D. P. Bavirisetti, G. Xiao, J. Zhao, R. Dhuli, and G. Liu, "Multi-scale guided image and video fusion: A fast and efficient approach," *Circuits, Syst., Signal Process.*, vol. 38, no. 12, pp. 5576–5605, Dec. 2019.

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

[36] X. Qiu, M. Li, L. Zhang, and X. Yuan, "Guided filter-based multi-focus image fusion through focus region detection," *Signal Process., Image Commun.*, vol. 72, pp. 35–46, Mar. 2019.

[37] H. Ma, Q. Liao, J. Zhang, S. Liu, and J.-H. Xue, "An α-matte boundary defocus model-based cascaded network for multi-focus image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 8668–8679, 2020.

[38] M. Nejati, S. Samavi, and S. Shirani, "Multi-focus image fusion using dictionary-based sparse representation," *Inf. Fusion*, vol. 25, pp. 72–84, Sep. 2015.

[39] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. Int. Conf. Image Process.*, vol. 3, 2003, p. 173.

[40] G. Cui, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition," *Opt. Commun.*, vol. 341, pp. 199–209, Apr. 2015.

[41] X. Bai, F. Zhou, and B. Xue, "Noise-suppressed image enhancement using multiscale top-hat selection transform through region extraction," *Appl. Opt.*, vol. 51, no. 3, pp. 338–347, 2012.

[42] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

TINGNA KUANG received the B.S. degree from Tianjin University, Tianjin, China. She is currently pursuing the M.S. degree in optical engineering with Zhejiang University. Her research interests include image processing and computer vision.

FEIHONG YU received the Ph.D. degree in optical engineering from Zhejiang University, in January 1993.

He joined the Center of Display Research, The Hong Kong University of Science and Technology, in 1995. In 1997, he joined the Institute of Material Research and Engineering, National University of Singapore. Since July 1999, he has been a Full Professor with the Optical Engineering Department, Zhejiang University. His current research interests include physical optics, optical CAD technology, optical instrumentation, digital image processing, the design and implementation of the hardware, and software of optical image sensors.

XUXIANG NI received the Ph.D. degree in optical engineering from Zhejiang University, in 1998. He is currently an Associate Professor with the Department of Optical Engineering, Zhejiang University. His research interests include optical instrumentation and signal processing.

• • •