Digital Signal Processing 106 (2020) 102821

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Multi-focus image fusion using learning based matting with sum of the Gaussian-based modified Laplacian

Yadong Xu [a], Beibei Sun [a,*], Xiaoan Yan [b], Jianzhong Hu [a], Minglong Chen [c]

[a] School of Mechanical Engineering, Southeast University, Nanjing, 211189, China
[b] School of Mechatronics Engineering, Nanjing Forestry University, Nanjing, 210037, China
[c] The First Affiliated Hospital of Nanjing Medical University, Nanjing, 210029, China

ARTICLE INFO

Article history:
Available online 28 July 2020

Keywords:
Multi-focus image fusion
SGML
Sliding window strategy
Image matting technique

ABSTRACT

In most traditional multi-focus image fusion algorithms, the focus measures used to detect the focused areas of multiple images are unstable and sensitive to noise. Although many modified methods implement more sophisticated strategies to cope with this problem, the complexity of these methods turns out to be a problem for mobile devices. In order to cope with these problems, a novel multi-focus image fusion method based on image matting technique is put forward in this paper. This method is simple yet effective. Firstly, a new focus measure named the sum of the Gaussian-based modified Laplacian (SGML) is proposed to estimate the focus map of source images. Then, the initial segmentation map can be obtained via a novel sliding window strategy. Further, with the rough segmentation map as an input, learning based image matting technique is performed to extract the exact boundaries between focused and defocused regions. Finally, by combining the information of the focus areas, an all-in-focus image can be obtained. The numerous experiments have revealed that the proposed approach yielded better results in comparison with some competing techniques both in subjective and objective evaluation.
© 2020 Elsevier Inc. All rights reserved.

1. Introduction

Owing to the limited depth of field in the camera's optical lens, the optical imaging system cannot clearly capture all objects in the focus range at the same time. Image fusion technique is an available measure to produce a panoramic image by merging the multi-focus images from the same scene [8,37]. Therefore, it has been widely used in the field of computer vision and target recognition.

In the past few decades, many image fusion algorithms have been proposed, and most of them focus on two aspects: spatial domain fusion and transform domain fusion [22]. One aspect is to perform image fusion in the spatial domain, that is, direct pixel-based fusion is performed on the source images. This method belongs to the spatial domain fusion. The other aspect requires the image to be transformed from the spatial domain to another space, which is so called transform domain fusion.

Transform domain fusion is consistent with human visual characteristics and can provide human with visually sensitive and relevant information. The multi-scale transform domain method first applies multi-scale decomposition scheme to decompose the source images into high-frequency coefficients and low-frequency coefficients. Then the coefficients are fused according to different fusion rules. Finally, the inverse transform is performed to get the enhanced fusion result. The choice of transformation domain and the design of fusion rules are the two most critical factors in the fusion process. Typical transform domain methods include pyramid transform [30–32], wavelets transform [24–26], non-subsampled shearlet transform [27–29], sparse representation [38,41] and so on.

Spatial domain fusion can be classified into pixel-based categories [33,34] and block-based categories [23,35,36]. Pixel-based methods are simple and straightforward, which preserve image details well and are easy to implement. However, the pixel-based fusion algorithms only consider focus information from a local neighborhood, which may cause contrast distortion or block artifacts. Block-based fusion methods can avoid some disadvantages of pixel-based fusion algorithms. However, the selection of block size is a problem. Because when the block size is large, the precision of the focused regions cannot be guaranteed. While when the block size is small, the focused areas may not be detected. Therefore, the choice of block size may limit the performance of fusion [4,10].

Learning-based methods have become a new research hotspot, which take the image fusion process as a classification task. In

* Corresponding author.
E-mail address: bbsun@seu.edu.cn (B. Sun).

https://doi.org/10.1016/j.dsp.2020.102821
1051-2004/© 2020 Elsevier Inc. All rights reserved.

many works [11,12], convolutional neural networks are used as the classifiers to extract the focused and defocused regions. Although these methods are effective, excessive training time may also be a problem.

To improve the subjective and objective evaluation of the fusion result, we present a new matting-based multi-focus image fusion approach. First of all, a new focus measure named the sum of Gaussian-based modified Laplacian (SGML) is presented to extract the focused information of source images. Then a novel sliding window strategy is introduced to generate an initial decision map, which approximately divides the source images into three regions: the focused areas, the unfocused areas and the uncertain areas. With the acquisition of the initial decision map, learning based matting technique is implemented to obtain the specific focused areas. Finally, by combining the information of the focus areas, an all-in-focus image can be constructed. Compared with traditional multi-focus fusion methods, our method has the following advantages:

- The proposed focus measure is more powerful in detecting the focused area of the source images in comparison with plenty of frequently-used gradient focus measures.
- A novel sliding window strategy is introduced, which can roughly divide the source image into three regions.
- Accurate boundary between focused and defocused areas can be extracted with learning based matting.
- The method is simple yet effective and is expected to be used in mobile devices in the future.

The rest of this paper is organized as follows. In section 2, related works will be briefly reviewed, and the motivation of our work will be presented. The proposed multi-focus image fusion algorithm will be described in detail in Section 3. Section 4 will give the experiment results and discussions. Finally, we will conclude the work in section 5.

2. Related works and motivation

In this section, we will first review the recently published works of image fusion, and the motivation of our work will be presented later.

2.1. Related work

In general, image fusion refers to the process of merging multiple images with complementary information into a single image with optimal information [43]. Multi-focus image fusion technique can integrate images with different focus information from the same scene into an enhanced image, improve the subjective visual quality of the input images, and thus can support the extraction of valuable information for some computer vision tasks [46].

Recently, many related methods have been proposed, and these methods have achieved pretty good results. Li et al. [6] proposed to utilize morphological filtering to get the rough segmentation results, and then used robust image matting algorithm to extract the certain focused areas of source images. The fused result can be constructed with the extracted information afterward. However, morphological filtering may sometimes fail to determine the focused area of the source image accurately. Ma et al. [44] introduced the boosted random walks-based algorithm to achieve the task of multi-focus image fusion. Ma et al. [45] proposed a joint sparse representation based method, which combined sparse representation and optimum theory for multi-focus image fusion. This method was experimentally verified to be effective, but the time consuming caused by iterative processing is a problem. Yang et al. [39,40] applied the sparse representation theory to various image fusion applications and achieved good fusion results. Kim et al. [42] presented a clustering-based dictionary learning method for multi-modal image fusion, which can be used in different image fusion scenarios.

Liu et al. [7] used the dense scale-invariant feature transform (SIFT) as image local features for image fusion. The method was verified to be competitive in comparison with other fusion methods. However, the high computational complexity is a problem. In [4], Bai et al. proposed an effective quad-tree decomposition strategy combined with the sum of the weighted modified Laplacian to produce one entirely all-in-focus image. Experiments showed that the proposed approach can achieve pretty good results. But the method is complicated compared with related approaches. Zhang et al. [3] fulfilled the task of extracting the focus information of source image by treating it as a boundary finding task. He presented a new morphological focus-measure, with which he obtained the accurate focused region of source images. The algorithm worked well when the boundary regions of images are simple. However, the final fusion image may be degraded when the boundary regions of the source images are very complex. Our work is partially inspired by [6]. Compared with [6], the focus measure proposed in this paper is more robust in detecting the focused areas of source images, and our method is easier to implement.

2.2. Motivation

In order to discover the focused areas accurately and improve the subjective and objective evaluation of the fusion result, we present a new fusion method using learning based matting with the sum of the Gaussian-based modified Laplacian. The main innovations of our approach can be concluded as follows: (1) Propose a new focus-measure called the sum of the Gaussian-based modified Laplacian (SGML), which is experimentally verified to be better than the sum of modified Laplacian (SML) in its ability to detect focused areas in this paper. (2) Introduce a novel sliding window scheme to get the initial segmentation map. (3) Learning based matting technique is performed to obtain accurate focus areas with the initial decision map. (4) The proposed method is simple and easy to implement, and is expected to be applied to mobile devices.

3. Proposed method

In this section, we will present the specific steps of our method. To this end, we first introduce our proposed focus-measure to extract focus information of input images. Then, we employ a simple sliding window strategy to obtain the initial segmentation map. Finally, learning based matting is performed to get the final fused image. The flowchart of our algorithm can be seen in Fig. 1.

3.1. A novel focus measure: the sum of the Gaussian-based modified Laplacian

The focus measure plays a pivotal role in differentiating the focused area from the unfocused region in most fusion algorithms. The choice of focus measure will affect the visual effect of fusion result [3]. Based on this, many focus area extraction methods have been put forward. For example, Kovacs et al. [47] introduced an automatic estimation method to extract focus regions. This method is effective and does not require any explicit knowledge about the input image. In general, the commonly used gradient focus measures include variance, the energy of image gradient (EOG), energy of Laplacian of the image (EOL), sum of the modified Laplacian (SML) and spatial frequency (SF) [1,2]. In some areas with small gradients, the difference between focused and unfocused blocks

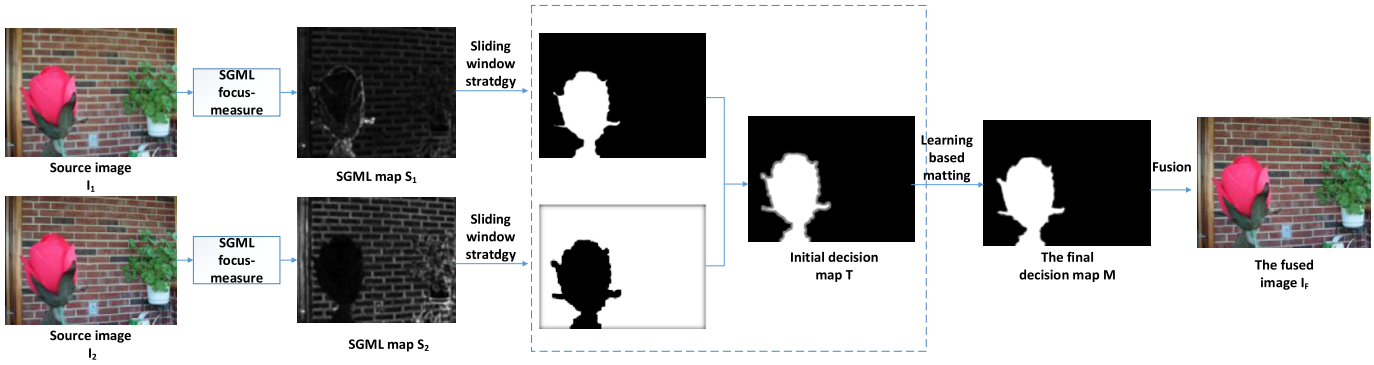


Fig. 1. The flowchart of the proposed method.

will also be very small [4]. So, it is difficult to differentiate the corresponding focused blocks from the unfocused blocks precisely with these commonly used focus measures.

It has been experimentally verified that SML outperformed the other focus measures in [2]. We wonder whether we can get a more effective focus measure if the SML is weighted at each location. Because the nearby gradients will enhance the strength of gradients at each location in this way. The Gaussian kernel is chosen as the weight matrix. The reason we chose it is that it is simple and symmetrical. In addition, it conforms to such a rule that as the distance from the pixel increases, the influence on the pixels around the pixel decreases. In this paper, we name the modified version of SML as the sum of the Gaussian-based modified Laplacian (SGML).

3.1.1. Sum of the modified Laplacian

Laplacian is a second-order derivative operator, which is often used to extract edge information. However, the second derivatives in the x and y directions may possess opposite signs and prone to cancel each other. In order to overcome this shortcoming, Nayar et al. [5] proposed the modified Laplacian (ML):

$$ML(x,y) = \left|2f(x,y) - f(x-s,y) - f(x+s,y)\right| + \left|2f(x,y) - f(x,y-s) - f(x,y+s)\right|, \tag{1}$$

where s denotes changeable step.

Then, the focus measure SML at the point (x, y) is computed as the sum of modified Laplacian values, in a small window around (x, y), that are greater than a threshold value. As is shown in Eq. (2):

$$SML(x,y) = \sum_{i=x-r}^{x+r} \sum_{j=y-r}^{y+r} ML(i,j) \quad \text{for } ML(i,j) > T, \tag{2}$$

where T denotes the threshold value and it is often set to 0; r determines the window size for calculating the focus measure.

3.1.2. Sum of the Gaussian-based modified Laplacian

As mentioned earlier, in order to increase the robustness of ML, we propose a modified version, which is named the Gaussian-based modified Laplacian (GML).

To begin with, the ML gradients of the source images can be acquired in terms of Eq. (1). Then, the Gaussian-based modified Laplacian can be obtained by convolving modified Laplacian and Gaussian kernel at each pixel location (x, y). The mathematical expression is as follows:

$$GML(x,y) = ML(x,y) * G_\sigma(x,y),$$
$$G_\sigma(x,y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right), \tag{3}$$

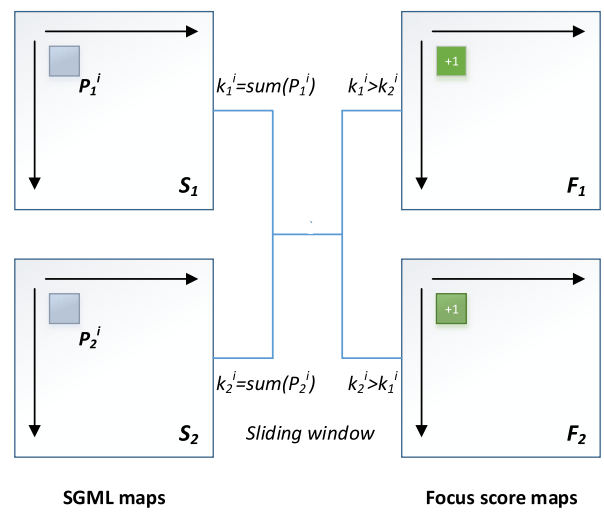


Fig. 2. The process of constructing score maps based on sliding window strategy.

where $*$ denotes convolution operator; σ represents the standard deviation of Gaussian kernel, and it is set as 2 in this paper.

Then, we define the sum of the Gaussian-based modified Laplacian (SGML) as in Eq. (4):

$$SGML(x,y) = \sum_{i=x-r}^{x+r} \sum_{j=y-r}^{y+r} GML(i,j). \tag{4}$$

Where r determines the window size and it is set to 7 in this paper.

3.2. Strategy of initial segmentation

As we have obtained the focus maps S_1 and S_2, the next stage is to compare S_1 and S_2 using an appropriate strategy to obtain an initial decision map. The purpose of this task is to classify the source image area based on the focus information, which is similar to region segmentation of the source image. We introduce a simple sliding window strategy that divides the source image into three regions, namely the focused area, the unfocused area, and the uncertain area. The specific steps are as follows:

(1) Divide the focus maps S_1 and S_2 into blocks of size $n \times n$ from top left to bottom right, as depicted in Fig. 2. Suppose the original size of the input image is $H \times W$ and the step length is one pixel, then there will be $Q = (H-n+1) \times (W-n+1)$ patches denoted as $\{P_1^i\}_{i=1}^{Q}$ and $\{P_2^i\}_{i=1}^{Q}$. As in [7], we set n as 8 in this paper.

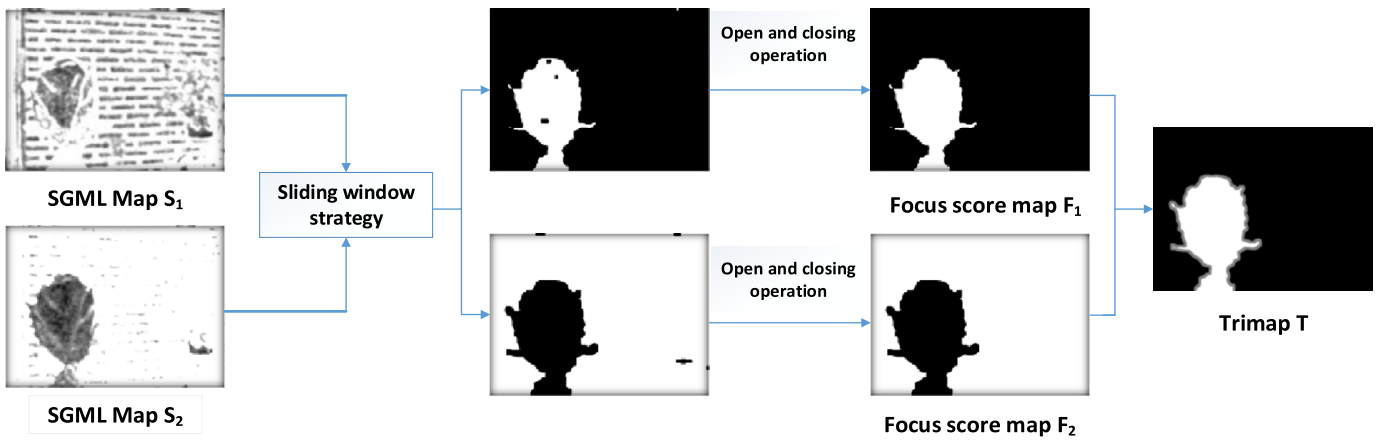


Fig. 3. The construction process of initial segmentation map.

(2) Create two "zero-valued" maps F_1 and F_2 to store focus scores, which are the same size as S_1 and S_2. As shown in Fig. 2, compute the sum of all coefficients in the corresponding pair of blocks P_1^i and P_2^i. Let k_1^i and k_2^i denote the obtained total values, respectively. If $k_1^i > k_2^i$, the coefficient within the corresponding block in S_1 plus one and vice versa. After iterating the above process for all corresponding blocks in $\{P_1^i\}_{i=1}^Q$ and $\{P_2^i\}_{i=1}^Q$, we can obtain the two score maps F_1 and F_2.

(3) Divide corresponding source images into three categories to get the initial segmentation maps. The specific classification rules of a pixel at the position (x, y) in the relevant input images I_1 and I_2 are as follows:

$$I_1(x,y) = \begin{cases} 1, & \text{if } F_2(x,y) = 0 \\ 0, & \text{if } F_1(x,y) = 0 \\ 0.5, & \text{otherwise,} \end{cases} \tag{5}$$

$$I_2(x,y) = \begin{cases} 1, & \text{if } F_1(x,y) = 0 \\ 0, & \text{if } F_2(x,y) = 0 \\ 0.5, & \text{otherwise.} \end{cases} \tag{6}$$

(4) Opening and closing operations of morphology filter [9] is used to remove dispersive points surrounded by the focused area. Then the initial segmentation map can be obtained by

$$T(x,y) = \begin{cases} 1, & \text{if } I_1(x,y) = 1 \\ 0, & \text{if } I_2(x,y) = 1 \\ 0.5, & \text{otherwise.} \end{cases} \tag{7}$$

The construction process of initial segmentation map can be seen in Fig. 3.

3.3. Image matting technique for fusion

After obtaining the focus information of source images, the next stage is to extract the specific focused region. Image matting is an effective technique to precisely distinguish the foreground from the background [13]. Based on image matting technique, the observed image $I(x, y)$ can be regarded as the combination of background $B(x, y)$ and foreground $F(x, y)$:

$$I(x,y) = \alpha(x,y)F(x,y) + \big(1 - \alpha(x,y)\big)B(x,y), \tag{8}$$

where $\alpha(x, y)$ is the opacity of the front named alpha matte and its value ranges from 0 to 1. $\alpha(x, y) = 0$ implies that $I(x, y)$ is in background, and $\alpha(x, y) = 1$ indicates that $I(x, y)$ is in foreground. When $0 < \alpha(x, y) < 1$, it suggests that these pixels are mixed with foreground and background. In this paper, learning based matting method is implemented to get the exact decision map of source images. Zheng et al. [9] assumed that the alpha matte value can be predicted by a linear combination of the alpha values of its corresponding neighboring pixels. The alpha values of all pixels can be estimated simultaneously through minimizing the quadratic function:

$$\underset{\alpha}{arg\,min} \left\|\alpha - F^T\alpha\right\|^2 + c\left\|\alpha_l - \alpha_l^*\right\|, \tag{9}$$

where α_l represents the vector of the alpha variables of the labeled pixels; α_l^* denotes the vectors of the already known alpha values of the labeled pixels; F represents the foreground of the input image, which can be estimated by the alpha coefficients during the sampling process; The accurate solution to this function can refer to [9]. After the implementation of image matting, the exact boundary between focused regions and defocused regions can be obtained. To prevent artifacts, the guided image filter is used to ensure spatial consistency of decision map, where the initial segmentation map T serves as the guidance image:

$$M(x,y) = GF_{r,\varepsilon}\big(\alpha(x,y), T(x,y)\big), \tag{10}$$

where $GF_{r,\varepsilon}(\cdot)$ denotes the guided filter operator, and r and ε are the parameters, which are set to 5 and 0.3, respectively.

The final fused image I_F can be obtained by the following pixel-wise weighted-averaging rule:

$$I_F = M(x,y)I_1(x,y) + \big(1 - M(x,y)\big)I_2(x,y). \tag{11}$$

As is shown in Fig. 4, Fig. 4(a) is the decision map obtained by learning based matting; Fig. 4(b) and Fig. 4(c) show the focused area of corresponding source image, respectively; Fig. 4(d) gives the fused result with our proposed method.

4. Experiment and discussions

4.1. Experiment settings

In our experiment, we used 30 pairs of public multi-focus images to verify the effectiveness of the algorithm. Fig. 6 shows a portion of the test images, including 5 grayscale images and 10 color images.

The proposed approach is compared with two competing transform domain fusion approaches, including multi-focus image fusion in DCT domain (DCT) [14] and multi-focus image fusion using

Fig. 4. The decision map obtained by learning based matting (a), the focused regions of corresponding source images (b and c) and the all-in-focus image obtained by our proposed approach (d).

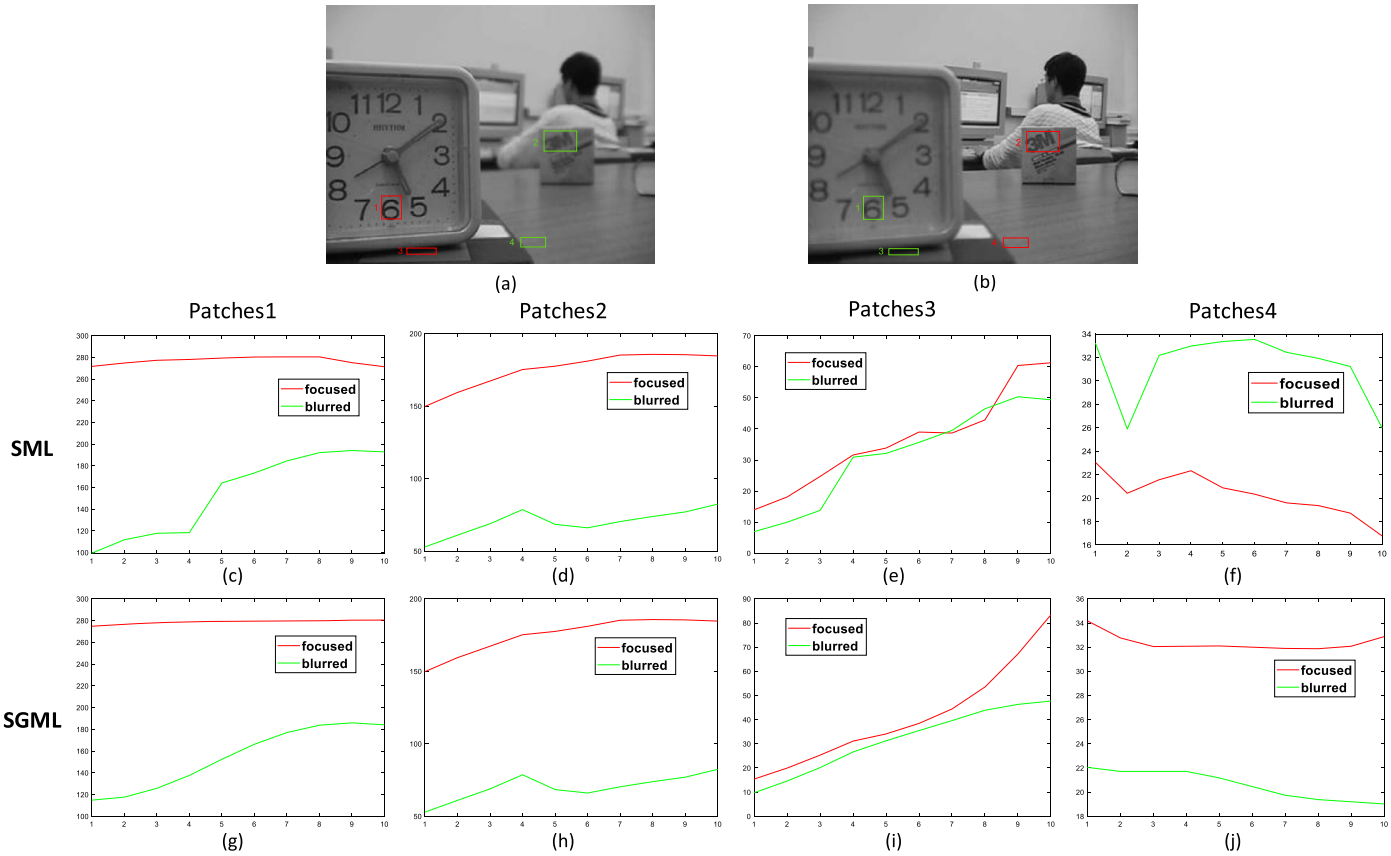


Fig. 5. Comparison between SML and SGML. (a and b) are the two source images, where the red blocks represent focused areas, and the green patches denote unfocused areas; the middle row (c, d, e, f) and bottom row (g, h, i, j) show the result of SML focus measure and SGML focus measure, respectively. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

PCNN (PCNN) [15], and five competing fusion methods, including the dense SIFT-based approach (DSIFT) [7], the boundary finding based approach (BF) [3], Image matting based approach (IM) [6], Quadtree-based approach (QB) [4] and a recently proposed multi-scale fusion approach (MWGF) [16].

We also quantified the performance of multiple fusion methods through five evaluation indicators, including feature mutual information-based metric Q_F [17], edge-based metric Q_E [18], phase congruency-based metric Q_P [19], structural similarity-based metric Q_S [20] and human perception based metric Q_H [21]. Q_F extracts the feature information of source images, such as gradient or edge. Q_E assesses the amount of edge information transferred from source images to the fused image. Q_P measures the amount of edge and corner information transferred from the source images to the fused image. Q_S reflects the similarity of different images. Q_H compares the contrast features of the source images with those of the fused image. For all these evaluation indicators, the larger the score value, the better the fusion effect.

4.2. Comparison between SML and our proposed focus measure

Fig. 5 gives a brief example of the comparison between SML and SGML. SML and SGML are utilized for each of the labeled areas (Patches1, Patches2, Patches3, Patches4) at the top of Fig. 5 to detect the corresponding focused areas. We sampled a row of pixels in the labeled regions to illustrate. The x-axis denotes the sampled pixels marked by the red rectangle, and the y-axis denotes the corresponding focus-measure value of the pixel. It can be observed that SML and SGML can both identify the focused patch correctly for Patches1 and Patches2. However, SML makes the wrong decision for Patches3 and Patches4 in that Patches3 and Patches4 have relatively low contrast compared with Patches1 and Patches2. So it can be concluded that SGML is better than SML in detecting focused areas of images.

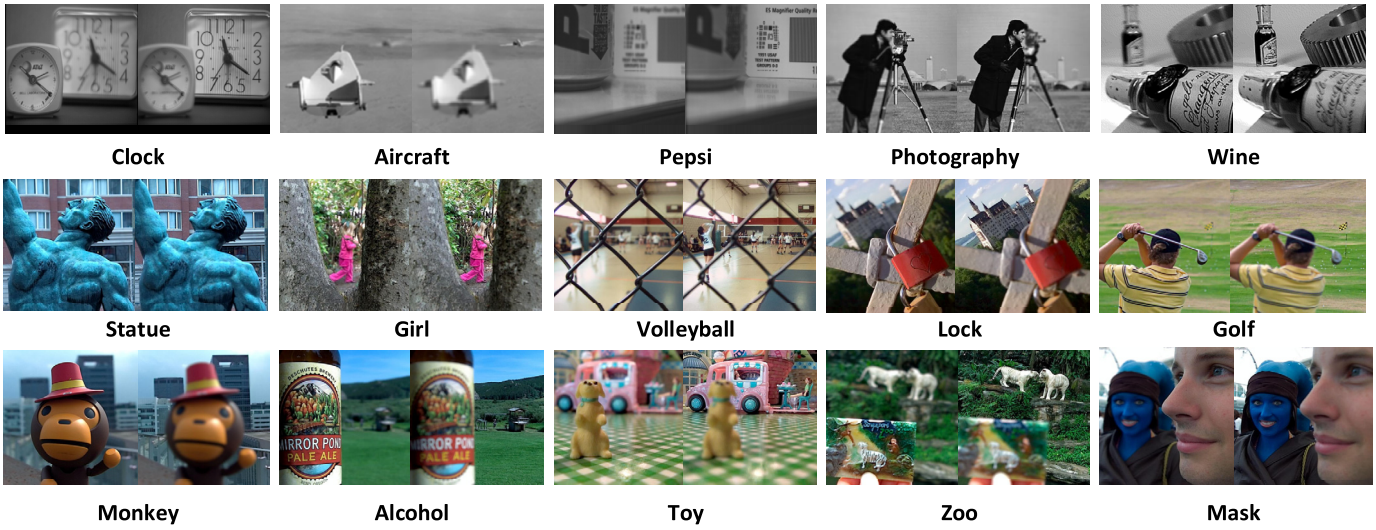


Fig. 6. A portion of the images used in our experiment.

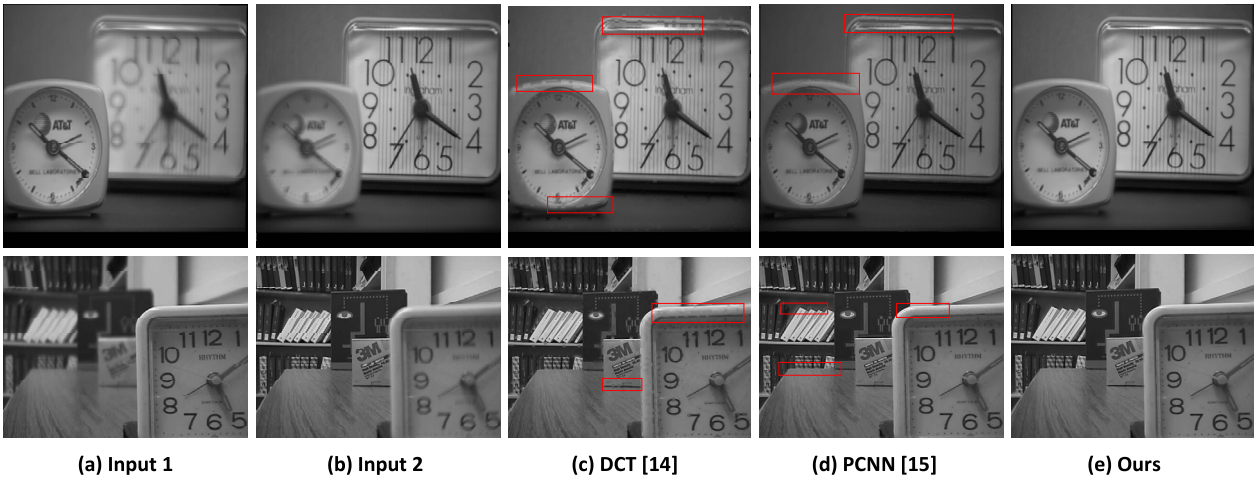


Fig. 7. The comparison between DCT based method, PCNN based method and the proposed method. (a) and (b) denote the source images, and (c), (d) and (e) are the results obtained from DCT-based method, PCNN-based method and our method, respectively.

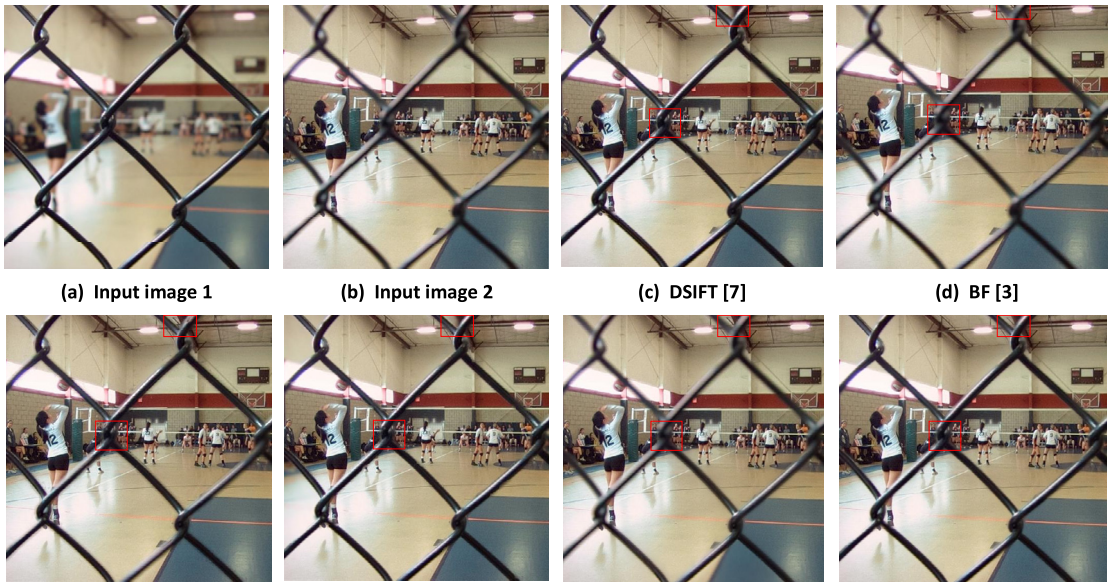
4.3. Objective and subjective comparison

Fig. 7 gives a comparison result between our algorithm and two transform domain algorithms. (a) and (b) are the input multi-focus images. The source image has different focus positions, resulting in different levels of blur. (c)–(e) are the fused results by DCT based fusion algorithm (DCT), PCNN based fusion algorithm (PCNN), and our algorithm, respectively. As can be seen in Fig. 7(c), there are apparent distortions near the clock in the images, and in Fig. 7(d), the pixels near the clock and table are blurred obviously. While in Fig. 7(e), the detail of the pixels around the clock and desk is preserved well, and there is no blur in the fused images.

The comparison example between our approach and five spatial domain approaches DSIFT, BF, IM, QB and MWGF are shown in Fig. 8. The example is pretty complicated, for that the focused region and defocused region in Fig. 8(a) and (b) cross each other. The BF, QB, and MWGF cannot detect all of focused parts in source images and only some of the dispersive focused areas from Fig. 8(a) are fused, producing different levels of blur in the fused images (see Fig. 8(d), (f) and (g)). Although DSIFT and IM can produce more visually pleasing fusion results, some distortions still exist in some small regions (as can be seen area marked by the red rectangle in Fig. 8(c) and Fig. 8(e)). It is not hard to observe that our method can detect all of focused regions in source images precisely and no blur exists in our fused image (see Fig. 8(h)).

Fig. 9 shows the comparison between DSIFT, BF, IM and the proposed method. The first two columns of Fig. 9 depict the source images. The corresponding focus maps and fused results obtained with DSIFT, BF, IM and our method are displayed in last four columns. Although DSIFT, BF and IM can achieve pretty good fusion results in most cases, they sometimes fail to detect the edges of focused regions accurately, which may result in distortions and blur around the boundaries (see the fifth row in Fig. 9(c), the first and third rows in Fig. 9(d) and the Seventh row in Fig. 9(e)). It can be seen that our method can overcome these shortcomings and extract the boundaries between focused and defocused regions accurately (see Fig. 9(f)).

Fig. 10 gives the comparison example between QB, MWGF and our proposed method. Fig. 10(a) and (b) are the source images. Fig. 10(c), (d), (e) are the fusion results by QB, MWGF, and our methods, respectively. It can be seen that QB fails to identify the focused region around the dog's eye marked by the red rectangle in Fig. 10(c), causing the distortion of surrounding pixels. In Fig. 10(d), the focused region around the dog's paw marked by the red rectangle is blurred. Our method successfully combines the small focused regions of source images and produces better visual effect.



(a) Input image 1 (b) Input image 2 (c) DSIFT [7] (d) BF [3]

(e) IM [6] (f) QB [4] (g) MWGF [16] (h) Ours

Fig. 8. The comparison of fusion result obtained from different methods. (a)-(b) are the source images. (c)-(h) are the results obtained by DSIFT, BF, IM, QB, MWGF and our method respectively.

Table 1
Average evaluation metrics of different methods on 30 pairs of multi-focus images.

Methods metrics	DCT	PCNN	DSIFT	BF	IM	QB	MWGF	Ours
Q_F	0.6343	0.5039	0.6570	0.6213	0.6422	0.5723	0.4974	**0.6663**
Q_E	0.6043	0.6018	0.6712	0.5369	0.6700	0.6411	0.5362	**0.6931**
Q_P	06411	0.2457	**0.8085**	0.8034	0.7737	0.4632	0.5102	0.7994
Q_S	0.9011	0.2524	0.9332	0.8300	0.9323	0.6390	0.7323	**0.9384**
Q_H	0.6500	0.5362	0.7292	**0.7738**	0.7381	0.7147	0.6129	0.7415

Table 2
Average time consumed by different approaches (second).

Fusion approaches	DCT	PCNN	DSIFT	BF	IM	QB	MWGF	Ours
Average time	0.87	2.6	11.2	1.65	3.01	1.35	3.8	**3.1**

Notes: The cost time of our method is highlighted in bold.

As is shown in Fig. 11(a) and (b). There is a slight movement of the student's head area between the two source images. Fig. 11(c)–(j) are the fused results obtained by different fusion methods. Fig. 11(c) and (d) are the merged results by two transform domain methods: DCT based method and PCNN based method. It is evident that the pixels around the man's head marked by the red rectangle are blurred, and the whole picture is seriously distorted (see Fig. 11(c) and (d)). Fig. 11(e)–(i) are the fused results produced by DSIFT, BF, IM, QB and MWGF, respectively. Although the five spatial-domain methods can distinguish between focused and defocused areas accurately, some blur still exists in discrete areas (as the corner of the block marked by the red rectangle in Fig. 11(e)–(i)). In contrast, the proposed method can well extract the focused regions from the source images and can present pleasing visual effect around the boundary regions (see Fig. 11(j)).

Finally, we will assess the performance of multiple fusion approaches in quantitative with the five previously mentioned quality metrics Q_F, Q_E, Q_P, Q_S and Q_H. The average indicator scores by multiple fusion algorithms on 30 pairs of multi-focus images are listed in Table 1. It can be seen that the proposed approach outperforms the other approaches in terms of Q_F, Q_E and Q_S. For Q_P and Q_H, our method performs well too in comparison with the other methods. Overall, our method performs best on average in terms of these five quality indicators.

4.4. Comparison of the computational cost

The comparison of computational efficiency is also included in this paper. All the running codes have been running in the same computer with MATLAB R2018b on a PC with Intel(R) Core(TM) i5-7300HQ CPU @2.50 GHz processor. The average consumption time of different methods is shown in Table 2. It is obvious that the DCT-based fusion costs the least time, which means the DCT-based method of image fusion is more efficient and time-saving. While DSIFT requires much more time than the other seven methods, for that dense SIFT descriptor is complex compared with other methods. The consuming time of the proposed method is almost the same as IM in that both of the two methods use the matting-based technique for image fusion. The matting technique can effectively determine the boundary between the focused and defocused areas, but it is comparatively time-consuming. Our future work will focus on optimizing this technique and improving the efficiency of the algorithm. In general, our approach is competent for mobile devices. In addition, with the rapid development of hardware equipment, real-time applications based on this method will no longer be a problem.

Fig. 9. The results of proposed approach and three spatial-domain fusion approaches: DSIFT, BF and IM. (a)-(b) are the original images. (c)-(f) are the focus maps and fused results acquired by DSIFT, BF, IM and the proposed approach, respectively.

4.5. More fusion examples

The proposed method can also be extended to process more than two input images. Take three images as an example, we first apply the proposed method to fuse the first two images, and then merge the third image with the fused image to produce the final fusion result. Fig. 12 gives an example of three source images processed by our method. The three input images are all partially

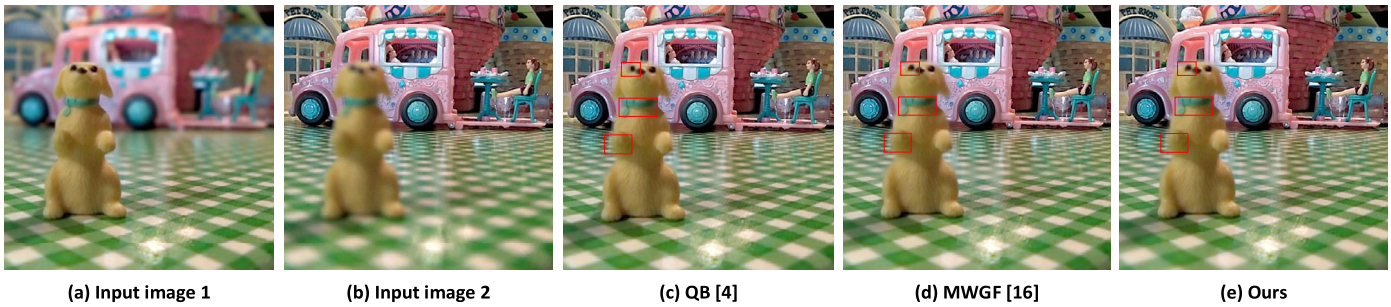


Fig. 10. The results of our approach and two spatial-domain approaches: QB and MWGF. (a)-(b) are the input images. (c)-(e) are the fusion results by QB, MWGF and our approach, respectively.

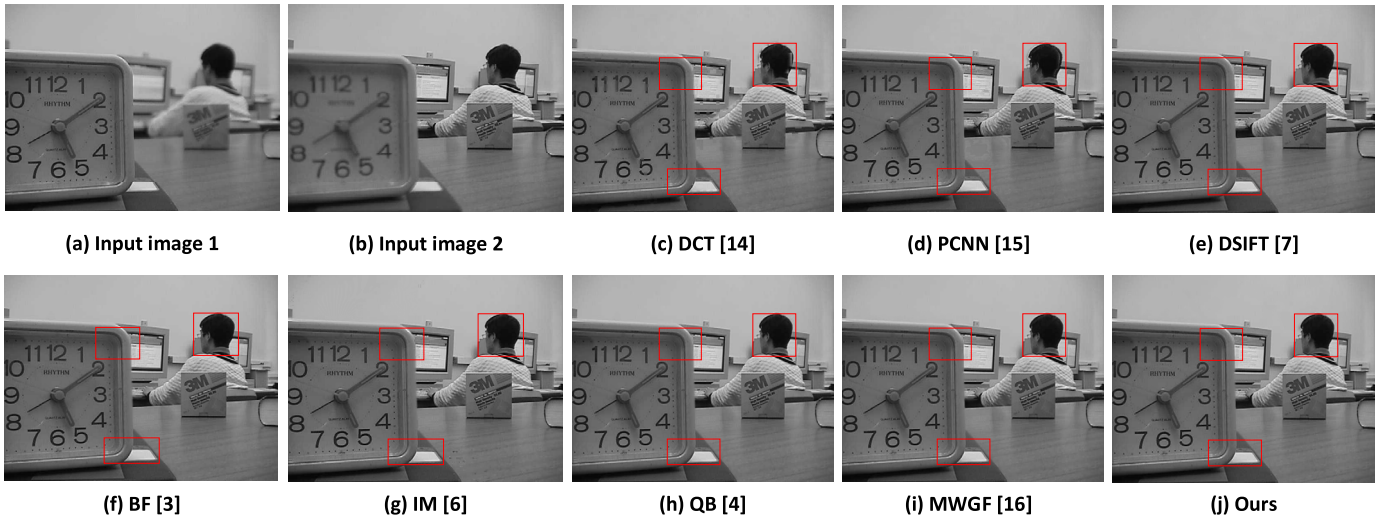


Fig. 11. The results of our algorithm and two transform-domain algorithms: DCT based algorithm, PCNN based algorithm and five spatial-domain fusion algorithms: DSIFT, BF, IM, QB, and MWGF, respectively.

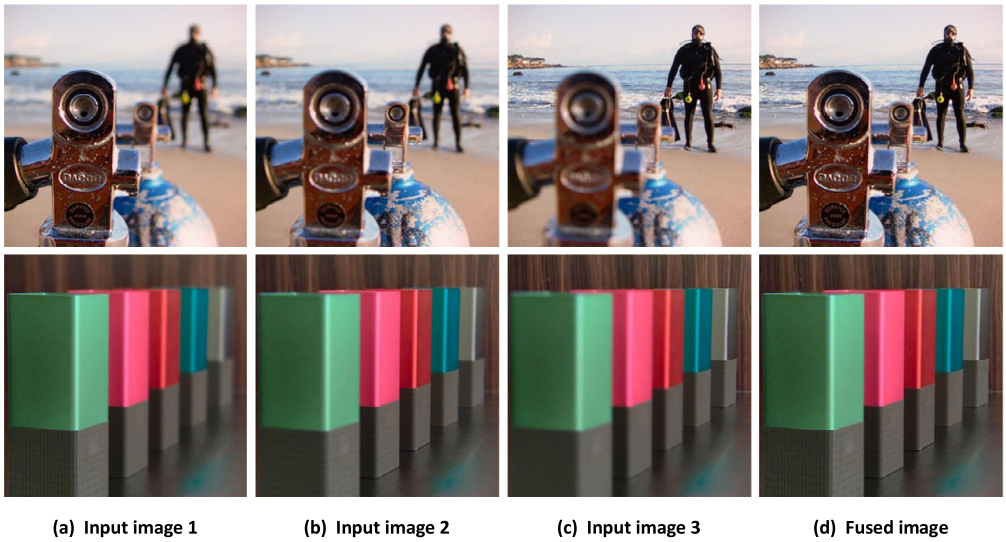


Fig. 12. Fusion results of proposed method for three multi-focus input images.

focused. It is evident that the fused images contain all the focus information of the source images. Besides, no blur exists in the fused image. Similarly, this method can also be extended to the fusion of multiple multi-focus images.

4.6. Noisy image fusion

In order to verify the robustness of the proposed algorithm, we also performed experiment on noisy images. We add Gaussian white noise with variance of 0.01 to the original images, as shown in Fig. 13(a) and Fig. 13(b). As is shown in Fig. 13(c) and (d), the obtained fusion results produce blurs and ghosts, and the noise intensity of the fused image is amplified. In comparison, pixel based fusion methods achieve better results. There exist noticeable pixel distortions in Fig. 13(g), (h) and (i). The method proposed in this paper can distinguish the boundary between the focused area and the defocused area, and can integrate the partially focused source image into a fully focused image. Fig. 14 gives another example.

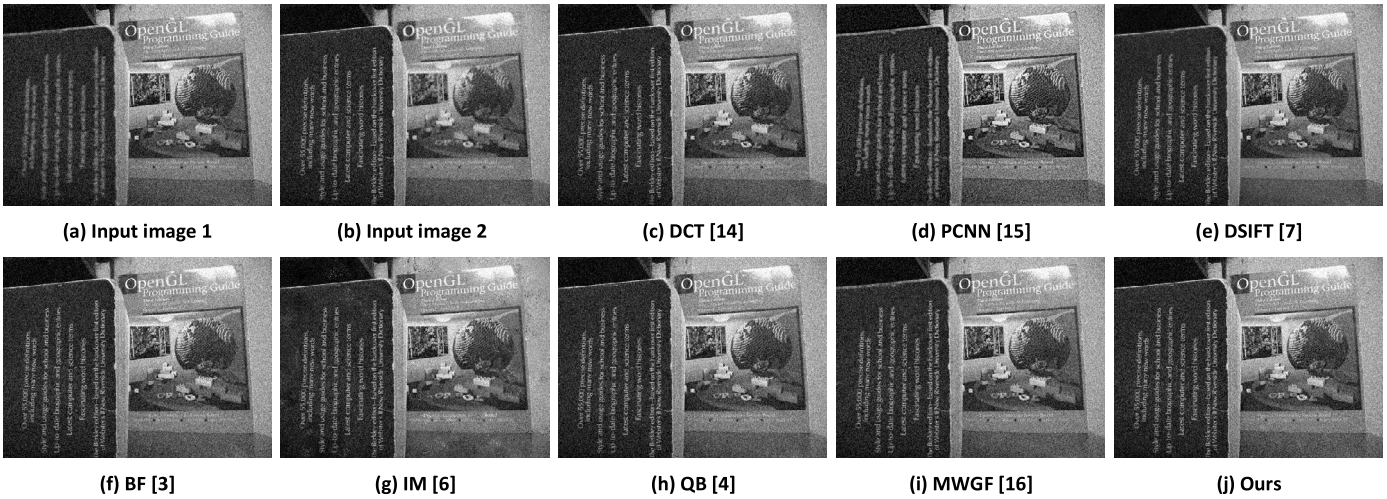


Fig. 13. Fusion results of different methods on noisy image set.

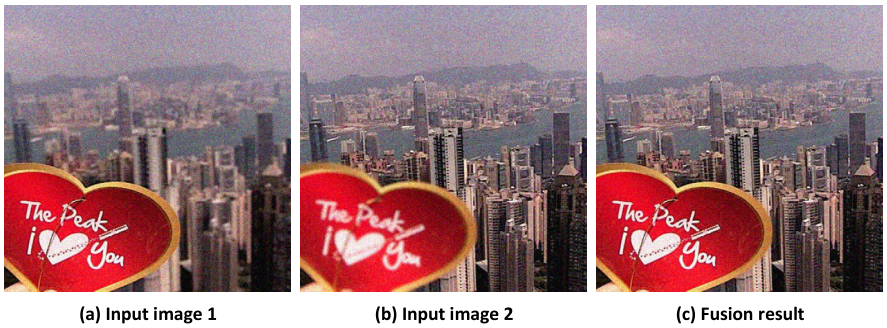


Fig. 14. Fusion result of proposed method on noisy image set.

It is evident that the fused result produced by proposed method is satisfactory. For example, the details of the focused loving heart and Skyscrapers in the source images are well preserved.

5. Conclusion

The process of image fusion can be summarized as extracting the information of focused areas from source images and then combining the information to form an all-in-focus image. Boundary detection between focused and defocused areas is a crucial step for image fusion. In this paper, we firstly present a new gradient focus measure named sum of the Gaussian-based modified Laplacian (SGML), which measures activity level between focused and unfocused areas of source images. Then, combined with the new focus-measure, a novel sliding window strategy is introduced to obtain a rough segmentation result. Further, with the initial segmentation result as an input, learning based image matting technique is performed to get the specific focused region of the source image. Finally, the all-in-focus image can be obtained by combining the focused areas. With the experimental analysis, we can draw the conclusion that the proposed method can extract the accurate boundary between focused regions and defocused regions. It can also be extended to process multiple multi-focus images and noise images. Furthermore, compared with the state-of-the-art competing techniques, the proposed approach yields comparative and even better performance both in objective and objective evaluation. Our future work will focus on improving the efficiency of the algorithm, and making it more suitable for hardware and mobile devices.

CRediT authorship contribution statement

Yadong Xu: Methodology, Validation, Writing - original draft. **Beibei Sun:** Writing - review & editing. **Xiaoan Yan:** Software. **Jianzhong Hu:** Software. **Minglong Chen:** Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

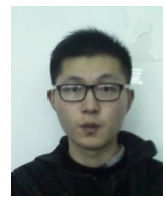
Acknowledgment

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFB2006404), in part by the Fundamental Research Funds for the Central Universities (Grant No. 2242019K3DN05), in part by the cooperative research project of Southeast University and Nanjing Medical University (Grant No. 3202009351). The author would like to appreciate the editors and the anonymous reviewers for their insightful comments.

References

[1] J. Ma, Z. Zhou, B. Wang, et al., Multi-focus image fusion based on multi-scale focus measures and generalized random walk, in: 2017 36th Chinese Control Conference, CCC, IEEE, 2017, pp. 5464–5468.
[2] W. Huang, Z. Jing, Evaluation of focus measures in multi-focus image fusion, Pattern Recognit. Lett. 28 (4) (2007) 493–500.
[3] Y. Zhang, X. Bai, T. Wang, Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure, Inf. Fusion 35 (2017) 81–101.

[4] X. Bai, Y. Zhang, F. Zhou, et al., Quadtree-based multi-focus image fusion using a weighted focus-measure, Inf. Fusion 22 (2015) 105–118.
[5] S.K. Nayar, Y. Nakagawa, Shape from focus, IEEE Trans. Pattern Anal. Mach. Intell. 16 (8) (1994) 824–831.
[6] S. Li, X. Kang, J. Hu, et al., Image matting for fusion of multi-focus images in dynamic scenes, Inf. Fusion 14 (2) (2013) 147–162.
[7] Y. Liu, S. Liu, Z. Wang, Multi-focus image fusion with dense SIFT, Inf. Fusion 23 (2015) 139–155.
[8] X. Yan, H. Qin, J. Li, et al., Multi-focus image fusion using a guided-filter-based difference image, Appl. Opt. 55 (9) (2016) 2230–2239.
[9] Y. Zheng, C. Kambhamettu, Learning based digital matting, in: 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 889–896.
[10] I. De, B. Chanda, Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure, Inf. Fusion 14 (2) (2013) 136–146.
[11] H.T. Mustafa, J. Yang, M. Zareapoor, Multi-scale convolutional neural network for multi-focus image fusion, Image Vis. Comput. 85 (2019) 26–35.
[12] Y. Liu, X. Chen, H. Peng, et al., Multi-focus image fusion with a deep convolutional neural network, Inf. Fusion 36 (2017) 191–207.
[13] J. Wang, M.F. Cohen, Image and video matting: a survey, Found. Trends Comput. Graph. Vis. 3 (2) (2008) 97–175.
[14] M.B.A. Haghighat, A. Aghagolzadeh, H. Seyedarabi, Multi-focus image fusion for visual sensor networks in DCT domain, Comput. Electr. Eng. 37 (5) (2011) 789–797.
[15] Z. Wang, Y. Ma, J. Gu, Multi-focus image fusion using PCNN, Pattern Recognit. 43 (6) (2010) 2003–2016.
[16] Z. Zhou, S. Li, B. Wang, Multi-scale weighted gradient-based fusion for multi-focus images, Inf. Fusion 20 (2014) 60–72.
[17] M. Haghighat, M.A. Razian, Fast-FMI: non-reference image fusion metric, in: 2014 IEEE 8th International Conference on Application of Information and Communication Technologies, AICT, IEEE, 2014, pp. 1–3.
[18] C.S. Xydeas, V. Petrovic, Objective image fusion performance measure, Electron. Lett. 36 (4) (2000) 308–309.
[19] J. Zhao, R. Laganiere, Z. Liu, Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement, Int. J. Innov. Comput. Inf. Control 3 (6) (2007) 1433–1447.
[20] C. Yang, J.Q. Zhang, X.R. Wang, et al., A novel similarity based quality metric for image fusion, Inf. Fusion 9 (2) (2008) 156–160.
[21] Y. Chen, R.S. Blum, A new automated quality assessment algorithm for image fusion, Image Vis. Comput. 27 (10) (2009) 1421–1432.
[22] B. Zhang, X. Lu, H. Pei, et al., Multi-focus image fusion based on sparse decomposition and background detection, Digit. Signal Process. 58 (2016) 50–63.
[23] X. Ma, S. Hu, S. Liu, et al., Multi-focus image fusion based on joint sparse representation and optimum theory, Signal Process. Image Commun. 78 (2019) 125–134.
[24] J. Tian, L. Chen, Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure, Signal Process. 92 (9) (2012) 2137–2146.
[25] Y. Liu, J. Jin, Q. Wang, et al., Region level based multi-focus image fusion using quaternion wavelet and normalized cut, Signal Process. 97 (2014) 9–30.
[26] H. Nasir, V. Stanković, S. Marshall, Singular value decomposition based fusion for super-resolution image reconstruction, Signal Process. Image Commun. 27 (2) (2012) 180–191.
[27] S.D. Ramlal, J. Sachdeva, C.K. Ahuja, et al., Multimodal medical image fusion using non-subsampled shearlet transform and pulse coupled neural network incorporated with morphological gradient, Signal Image Video Process. 12 (8) (2018) 1479–1487.
[28] B. Cheng, L. Jin, G. Li, General fusion method for infrared and visual images via latent low-rank representation and local non-subsampled shearlet transform, Infrared Phys. Technol. 92 (2018) 68–77.
[29] P. Ganasala, V. Kumar, Feature-motivated simplified adaptive PCNN-based medical image fusion algorithm in NSST domain, J. Digit. Imaging 29 (1) (2016) 73–85.
[30] W. Wang, F. Chang, A multi-focus image fusion method based on Laplacian pyramid, Am. J. Clin. Pathol. 6 (12) (2011) 2559–2566.
[31] J. Sun, Q. Han, L. Kou, et al., Multi-focus image fusion algorithm based on Laplacian pyramids, JOSA A 35 (3) (2018) 480–490.
[32] S. Singh, R. Rajput, Multiple image fusion using Laplacian pyramid, Int. J. Comput. Appl. Eng. Sci. 3 (12) (2014) 9442–9446.
[33] V. Aslantas, A.N. Toprak, A pixel based multi-focus image fusion method, Opt. Commun. 332 (2014) 350–358.
[34] N. Rehman, S. Ehsan, S. Abdullah, et al., Multi-scale pixel-based image fusion using multivariate empirical mode decomposition, Sensors 15 (5) (2015) 10923–10947.
[35] X. Qiu, M. Li, L. Zhang, et al., Guided filter-based multi-focus image fusion through focus region detection, Signal Process. Image Commun. 72 (2019) 35–46.
[36] H. Li, L. Li, J. Zhang, Multi-focus image fusion based on sparse feature matrix decomposition and morphological filtering, Opt. Commun. 342 (2015) 1–11.
[37] M.A. Rahman, S. Liu, C.Y. Wong, et al., Multi-focal image fusion using degree of focus and fuzzy logic, Digit. Signal Process. 60 (2016) 1–19.
[38] X. Li, H. Shen, L. Zhang, et al., Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning, IEEE Trans. Geosci. Remote Sens. 52 (11) (2014) 7086–7098.
[39] B. Yang, S. Li, Multi focus image fusion and restoration with sparse representation, IEEE Trans. Instrum. Meas. 59 (4) (2010) 884–892.
[40] B. Yang, S. Li, Pixel-level image fusion with simultaneous orthogonal matching pursuit, Inf. Fusion 13 (1) (2012) 10–19.
[41] S. Liu, G. Wu, H. Liu, et al., Image restoration approach using a joint sparse representation in 3D-transform domain, Digit. Signal Process. 60 (2016) 307–323.
[42] M. Kim, D.K. Han, H. Ko, Joint patch clustering-based dictionary learning for multimodal image fusion, Inf. Fusion 27 (2016) 198–214.
[43] In general, image fusion refers to the process of merging multiple images with complementary information into a single image with optimal information.
[44] J. Ma, Z. Zhou, B. Wang, et al., Multi-focus image fusion using boosted random walks-based algorithm with two-scale focus maps, Neurocomputing 335 (2019) 9–20.
[45] X. Ma, S. Hu, S. Liu, et al., Multi-focus image fusion based on joint sparse representation and optimum theory, Signal Process. Image Commun. 78 (2019) 125–134.
[46] O.J. Kwon, S. Choi, D. Jang, et al., All-in-focus imaging using average filter-based relative focus measure, Digit. Signal Process. 60 (2017) 200–210.
[47] L. Kovacs, T. Sziranyi, Focus area extraction by blind deconvolution for defining regions of interest, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 1080–1085.



Yadong Xu received the B.S. degree in mechanical manufacturing and automation from Nanjing Forestry University, Nanjing, China, in 2016. He is currently pursuing the Ph.D. degree in mechanical engineering from Southeast University, Nanjing, China. His current research interests include digital image processing, intelligent fault diagnosis and pattern recognition.

Beibei Sun received the Ph.D. degree in mechanical engineering from Southeast University, Nanjing, China in 2001. She is currently a full Professor and a Doctoral Supervisor with the School of Mechanical Engineering, Southeast University. Her main research interests include digital image processing, machine learning, dynamics of complex electromechanical systems, and dynamic optimization design of construction machinery.

Xiaoan Yan was born in 1989. He received the M.S. degree in mechanical engineering from North China Electric Power University, Baoding, China, in 2015 and the Ph.D. degree in mechanical engineering from Southeast University, Nanjing, China, in 2019. He is currently a lecturer at School of Mechatronics Engineering, Nanjing Forestry University, Nanjing, China. His research interests include signal processing, intelligent fault diagnosis, pattern recognition, health status assessment and residual life prediction. E-mail: yanxiaoan89@sina.com

Jianzhong Hu received the B.S. and the M.S. degree in mechanical engineering from Huazhong University of Science and Technology, Wuhan, China, and the Ph.D. degree in mechanical engineering from Southeast University, Nanjing, China in 1993, 1999 and 2003, respectively. He is currently an associate professor with the School of Mechanical Engineering, Southeast University. His research interests include signal processing, intelligent fault diagnosis, pattern recognition, health status assessment and residual life prediction.

Minglong Chen, chief physician, professor, and doctoral supervisor. He is currently the deputy director of the Department of Cardiology at the First Affiliated Hospital of Nanjing Medical University (Jiangsu Provincial People's Hospital), a distinguished professor of Nanjing Medical University, and the dean of the Affiliated Hospital of Xuzhou Medical University.