Article

Multi-Focus Image Fusion Based on Decision Map and Sparse Representation

Bin Liao [ID]**, Hua Chen * and Wei Mo**

School of Electrical & Electronic Engineering, North China Electric Power University, Beijing 100226, China
* Correspondence: chenhua@ncepu.edu.cn; Tel.: +86-152-3419-8627

check for updates

Abstract: As the focal length of an optical lens in a conventional camera is limited, it is usually arduous to obtain an image in which each object is focused. This problem can be solved by multi-focus image fusion. In this paper, we propose an entirely new multi-focus image fusion method based on decision map and sparse representation (DMSR). First, we obtained a decision map by analyzing low-scale images with sparse representation, measuring the effective clarity level, and using spatial frequency methods to process uncertain areas. Subsequently, the transitional area around the focus boundary was determined by the decision map, and we implemented the transitional area fusion based on sparse representation. The experimental results show that the proposed method is superior to the other five fusion methods, both in terms of visual effect and quantitative evaluation.

Keywords: multi-focus image fusion; low-scale images; decision map; sparse representation; transitional area

1. Introduction

Multi-focus image fusion is a method of combining multiple images with different focal points into a composite image in which all objects are completely focused. The composite image will be more suitable for visual perception, making it easier for humans to further complete image processing tasks. Multi-focus image fusion technology has been widely used in digital photography, computer vision, military reconnaissance, and other fields [1].

With the maturity and improvement of image fusion technology, miscellaneous image fusion methods have emerged in the past few years. As many new fusion algorithms have been proposed recently, we feel inclined to divide the current fusion methods into four categories: multiscale transform (MST) methods, spatial domain methods, sparse representation (SR) methods, and neural network methods. Among the existing transform domain image fusion methods, MST is widely used [2]. A variety of multiscale transforms have been proposed and applied to image fusion. These include the Laplacian pyramid (LP), discrete wavelet transform (DWT) [3,4], dual-tree complex wavelet transform (DTCWT) [5], and discrete cosine harmonic wavelet transform (DCHWT) [6]. The multiscale geometric analysis tools developed in recent years have higher directional sensitivity than wavelets, such as shearlet transform [7], curvelet transform (CVT) [8], nonsubsampled contourlet transform (NSCT) [9], and so on. All of these transform domain fusion methods have a similar "decomposition–fusion–reconstruction" framework. First, the source images are decomposed into a multiscale transform domain to obtain transform coefficients, and the transform coefficients are then fused based on a certain fusion rule. Finally, the fusion coefficients are inversely transformed to reconstruct the fused image.

The inchoate spatial domain fusion methods average source images pixel-by-pixel, which usually causes unsatisfactory results, such as blurry details and low-grade contrast. In order to overcome these drawbacks, some block-based [10,11] and region-based methods [12,13] have been proposed in recent years. The core principle is to select image blocks or regions from source images on the basis of some focus measure metrics [14], such as gradient energy, spatial frequency, image variance, etc. The block-based image fusion methods have a simple process of segmentation and calculation, but the segmentation size directly affects the discrimination of the image block clarity, which easily produces a "block effect". The image fusion methods based on region segmentation are modified. According to the characteristics of image focus information distribution, sophisticated segmentation algorithms are used to split the images to be fused, which can determine the position of the focus region accurately and improve the quality of image fusion. Nevertheless, due to the complexity of the segmentation algorithm used, the efficiency of fusion methods is not high in practical applications. In the last few years, some state-of-the-art algorithms have been presented, such as guided filtering (GF) [15], image matting (IM) [16], and the bilateral filter [17]. To some extent, these methods can not only extract the focused area from the source images more accurately but also maintain consistency with the image source.

The essence of the sparse representation model is a more compact representation of important information in natural signals with a handful of elements. In view of the efficient representation of the sparse representation model, the model is diffusely used in target tracking, face recognition, and image denoising. Numerous multi-focus image fusion methods based on sparse representation have also been proposed [18–24]. For instance, Yang and Li first applied SR to multi-focus image fusion. The method uses the smooth window technique to divide the images to be fused into several image blocks and solves the sparse coefficients using the L1 norm. Finally, these coefficients are merged according to the maximum selection fusion rule to reconstruct the fused image [18]. On this basis, the authors used the synchronous orthogonal matching tracking sparse coding algorithm (SOMP) in the multi-focus image fusion process, which can cause different images that need to be fused to use the same atom for sparse representation to improve the algorithm performance and fusion efficiency [19]. Furthermore, Chen Li et al. combined sparse representation theory with other fusion algorithms for multi-focus image fusion. However, the algorithm has a large amount of sparse decomposition calculation and a high time requirement [20]. Subsequently, Liu et al. attempted to combine MST with the sparse representation model. The proposed fusion structure overcomes the original defects and can be applied to the fusion processing of multiple multi-focus images. However, the algorithm still failed to effectively reduce the computational complexity, and there were still difficulties in practical applications [21]. They also applied the adaptive sparse representation model to image denoising [22]. In [23], the SR model with dictionary learning was used in multi-focus image fusion. A large number of experiments have shown that the fusion methods based on sparse representation are better than the multiscale transformation methods. Nevertheless, as with most existing methods, existing fusion methods based on sparse representation also have potential degradation problems, such as blocking artifacts, artificial edges, and ringing effects.

Finally, the image fusion methods are mainly realized by artificial neural networks. Among them, the pulse-coupled neural network (PCNN) is a cat neural cortex-based biological neural network proposed by Eckhorn et al. [25], which is widely used in various image processing fields, including image fusion. Researchers have developed PCNN-based multi-focus image fusion methods in the multiscale transform domain [26–28] or directly in the spatial domain [29]. The most significant advantage of the PCNN-based fusion method is that its information processing model can simulate the human visual system so that a fusion image conforming to human visual characteristics can be obtained. However, the fusion performance of the PCNN model is often affected by its large number of free parameters, which makes it less stable. In addition, Liu et al. proposed a multi-focus image fusion method based on a deep convolutional neural network (CNN) [30]. It uses a high-quality image patch training deep convolutional neural network and its fuzzy version to encode the mapping, which can

jointly generate activity level measurement and fusion rules and overcome some difficulties faced by certain existing fusion methods.

Based on the analysis and research of existing multi-focus image fusion methods, we propose a new multi-focus image fusion method based on decision map and sparse representation (DMSR), which can not only satisfy the requirements of the visual effect and fusion performance but also make the algorithm robust and adaptive. In our framework, the advantages of fusion methods based on the decision map and sparse representation are combined. Considering that the human visual system does not require much detail in identifying the focused and defocused area of the source images, we generated a sparsity graph using low-scale images of the source images. In the existing multi-focus image fusion methods based on the decision map, each pixel is strictly defined as focused or defocused, which inevitably leads to erroneous judgment in the decision map. In particular, the pixels of the uncertain region are difficult to determine simply as focus or defocus. In order to avoid this defect, we analyzed the sparseness of the corresponding points in the sparsity graph and divided each pixel into three categories—focused, defocused, and uncertain—to generate the initial decision map. Then, the spatial frequency method was used to further divide each point in the uncertain region of the initial decision map into focused or defocused points, and the final decision map was determined. After obtaining the fused image based on the final decision map, the transitional area of the source images was detected according to the final decision map, and the area was processed by the multi-focus image fusion algorithm based on the sparse representation to obtain the transitional area fusion result. Finally, the fused image based on the final decision map and the transitional area fused image were averaged to obtain the final fused image. In order to verify the effectiveness of the proposed method, we performed a large number of experiments using two data sets based on the three target quality indicators. The experimental results show that our method is superior to the other five methods, both in terms of visual effect and quantitative evaluation.

The remainder of this paper is organized as follows. Section 2 describes the specifics of our proposed method. The experimental results, a comparison with the state-of-the-art methods and objective evaluations are demonstrated in Section 3. Finally, Section 4 is the conclusion of this paper.

2. Proposed Fusion Scheme

The newly proposed multi-focus image fusion framework is shown in Figure 1. Obviously, the fusion method consists of two main steps: generating a decision map and performing fusion. In the first step, multi-focus feature analysis of the low-scale images of the two source images is performed to obtain the corresponding clarity score maps. Then, they are normalized to get the initial decision map, and the spatial frequency method is used to obtain the final decision map. Section 2.1 details the creation of the score maps, and the specific process for further obtaining the initial decision map and the final decision map are described in Section 2.2. In the second step, the fused image based on the final decision graph and the transitional area fused image are obtained, respectively, and the two images above are averaged to obtain the final fused image. Among them, the fusion process of the transitional area is based on sparse representation, which is elaborated in Section 2.3.
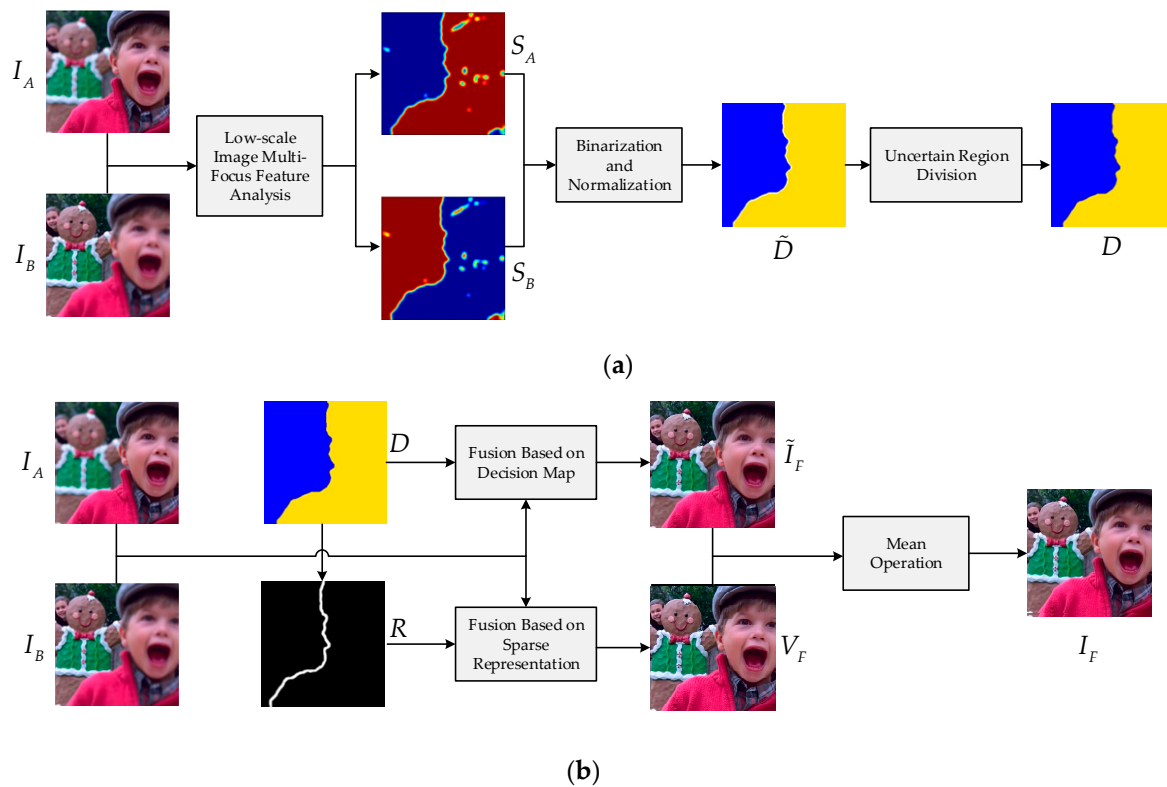
(a)

(b)

Figure 1. Framework of the proposed fusion algorithm. (**a**) The process of generating the decision map; (**b**) fusion process.

2.1. Clarity Score Map

Firstly, wavelet decomposition is performed on the two multi-focus source images by a wavelet basis, and four low-frequency sub-band images of horizontal low-frequency and vertical low-frequency (LL), horizontal low-frequency and vertical high-frequency (LH), horizontal high-frequency and vertical low-frequency (HL), and horizontal high-frequency and vertical high-frequency (HH) are obtained, respectively. Among them, the LL low-frequency sub-band images still maintain the overview and spatial characteristics of the source images and are suitable for the analysis and extraction of the subsequent source image focusing features, so they are selected as the low-scale images of the algorithm, as shown in Figure 2c,d. Next, the sparse representation of low-scale images is carried out, and the corresponding sparsity graphs are generated. Finally, two corresponding clarity score maps are obtained by the image block-based clarity measurement method. The main steps of creating clarity score maps are described as follows:

- The low-scale versions of source images I_A^{LL}, $I_B^{LL} \in R^{H \times W}$ are divided into $\sqrt{n} \times \sqrt{n}$ image patches using the smooth window technique from top left to bottom right, and the sliding step is one. All patches are reshaped into n dimensional column vectors $\{v_A^i\}_{i=1}^N$ and $\{v_B^i\}_{i=1}^N$ ($v_A^i, v_B^i \in R^n, N = (H - \sqrt{n} + 1) * (W - \sqrt{n} + 1)$) via lexicographic ordering.

- Given the global dictionary $\Phi \in R^{n \times K}(n << K)$, each column vector can be represented by the sparse coefficient vector $\{x_A^i\}_{i=1}^N$ and $\{x_B^i\}_{i=1}^N$ ($x_A^i, x_B^i \in R^K$) with the orthogonal matching pursuit (OMP) sparse coding algorithm. The L1 norms of the sparse coefficient vector $\{x_A^i\}_{i=1}^N$ and $\{x_B^i\}_{i=1}^N$ are calculated and reshaped to obtain the sparsity graphs E_A and $E_B \in R^{(H - \sqrt{n} + 1) \times (W - \sqrt{n} + 1)}$.

- Two score maps S_A and $S_B \in R^{H \times W}$ are initialized with all zeroes. For a given pixel (x_i, y_i) in the sparsity graph E, its value measures the activity level of the pixel's down-right corner $\sqrt{n} \times \sqrt{n}$

image patch. For each corresponding pair of $\sqrt{n} \times \sqrt{n}$ image patches, e_A^i and e_B^i in the sparsity graphs, the sum of all the clarity values is calculated as follows:

$$M^i = \sum_{u=0}^{\sqrt{n}-1} \sum_{v=0}^{\sqrt{n}-1} E(x_i + u, y_i + v) \tag{1}$$

where M_A^i and M_B^i denote the sum values, respectively. If $M_A^i \geq M_B^i$, each score value within the $\sqrt{n} \times \sqrt{n}$ corresponding patch is centered at $\left(x_i + \sqrt{n}, y_i + \sqrt{n}\right)$ in the clarity score map S_A add one, and vice versa, as shown in Figure 3. In addition, the total times of the comparison between each corresponding pair of patches are recorded in a weight map W.

- Further, the clarity score maps S_A and S_B are normalized by averaging W at each pixel location. The resulting clarity score maps are shown in Figure 2e,f.
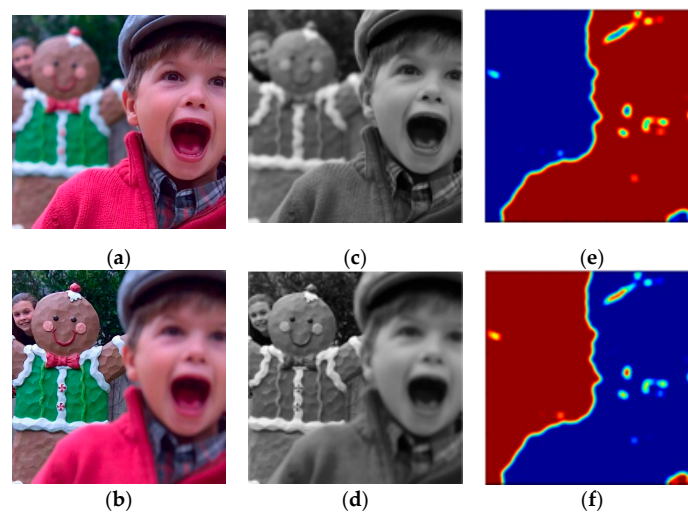

Figure 2. (**a**) Source image I_A; (**b**) source image I_B; (**c**) low-scale image I_A^{LL}; (**d**) low-scale image I_B^{LL}; (**e**) clarity score map S_A; (**f**) clarity score map S_B.
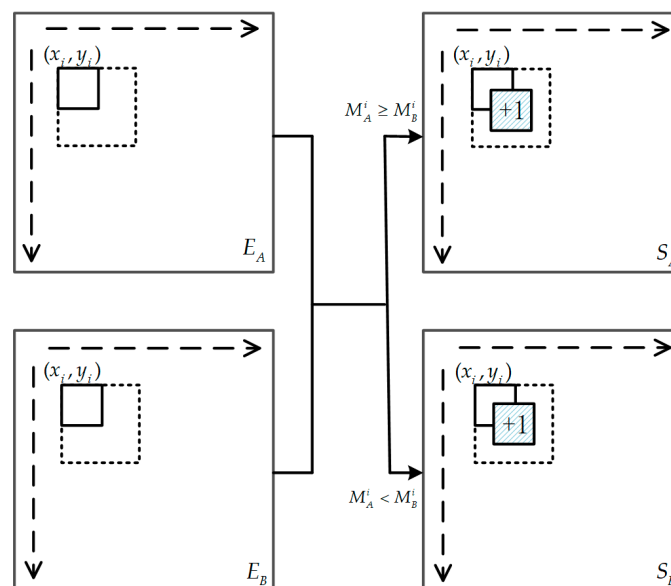

Figure 3. The creation of clarity score maps with the smooth window technique.

2.2. Decision Map

The above clarity score maps are binarized by a given threshold K_1 and denoted as S'_A and S'_B, as shown in Figure 4a,b (the focused pixels are marked as yellow, defocused pixels are marked as blue). It can be observed that there may be some misjudgment areas caused by misclassification in the focused area or the defocused area. Morphological techniques are used to filter out these misclassifications to obtain the standard normalized clarity score maps. The results are shown in Figure 4c,d and denoted as S''_A and S''_B. Thus, we can determine the location of the uncertain area when the focused areas of Figure 4c,d overlap.
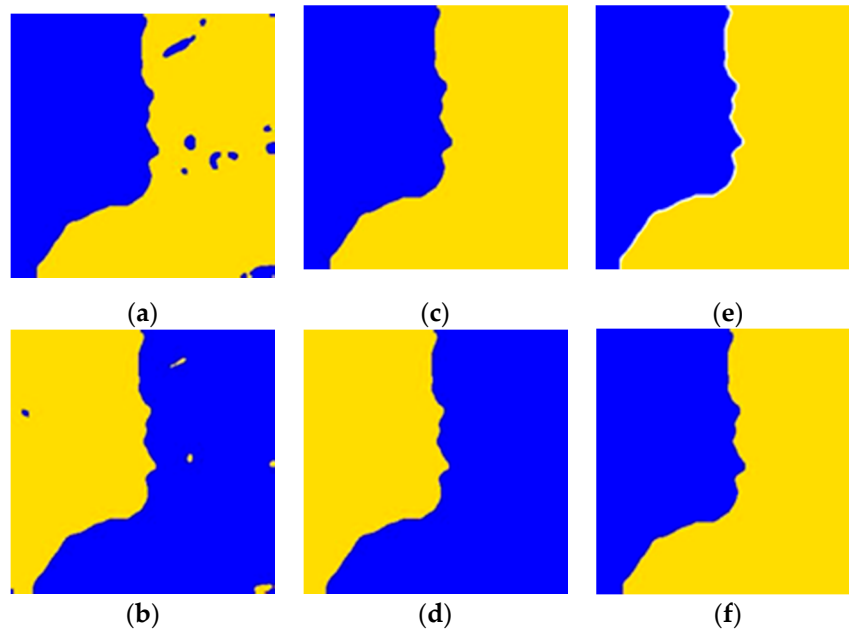

(a)   **(c)**   **(e)**

(b)   **(d)**   **(f)**

Figure 4. (**a**) Binarized clarity score map S'_A; (**b**) binarized clarity score map S'_B; (**c**) standard normalized clarity score map S''_A; (**d**) standard normalized clarity score map S''_B; (**e**) initial decision map $\widetilde{D}$; (**f**) final decision map D.

Finally, the initial decision map is obtained by

$$\widetilde{D}(x,y) = \begin{cases} 1, & \text{if } S''_A(x,y) = 1 \text{ and } S''_B(x,y) = 0 \\ 0, & \text{if } S''_A(x,y) = 0 \text{ and } S''_B(x,y) = 1 \\ \quad 0.5, & \text{otherwise} \end{cases} \tag{2}$$

as shown in Figure 4e, where the white pixels indicate the uncertain area. In order to make the size of the decision map consistent with the source images, the upsampling operation is also carried on to the initial decision map.

The next target is to generate the final decision map. As mentioned above, there is still an uncertain area in the initial decision map $\widetilde{D}$. To obtain the final decision map, further analysis and processing of the uncertain area is needed. We use the spatial frequency method to divide the pixels of the uncertain area in the initial decision map $\widetilde{D}$ into two categories—focused and defocused—to obtain the final decision map containing only the focused area and the defocused area. The spatial frequency method can be described as

$$SF(x,y) = \sqrt{\iint_\Omega (\nabla_x I)^2} + \sqrt{\iint_\Omega (\nabla_y I)^2} \tag{3}$$

where I is the input image, Ω is a 7×7 window centered on the point (x, y), and ∇_x and ∇_y represent the horizontal and vertical differences of the pixel points, respectively. The larger the spatial frequency

value, the higher the clarity value of the point. Thus, points in the uncertain area of the initial decision map $\widetilde{D}$ can be classified according to the following decision rules:

$$\widetilde{D}(x,y) = \begin{cases} 1, & \text{if } SF_A(x,y) > SF_B(x,y) \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

Assuming that the spatial frequency values of the corresponding uncertain pixel points in the two source images are $SF_A(x,y)$ and $SF_B(x,y)$, respectively, and $SF_A(x,y) > SF_B(x,y)$, the pixel point can be determined as the focus point, and vice versa. Based on this, the final decision map D can be obtained, as shown in Figure 4f.

2.3. Fusion

Based on the final decision map D, the fused image $\widetilde{I}_F$ can be simply obtained by

$$\widetilde{I}_F(x,y) = D(x,y)I_A(x,y) + (1 - D(x,y))I_B(x,y). \tag{5}$$

However, in this way, the pixels in the transitional area are actually averaged. This can cause undesirable effects such as the edge-blocking effect and artificial-edge effect. In order to suppress these effects at the same time, it is considered that the pixel classification of the transitional area has the following difficulties: the difference in the clarity of the pixels is small, the gray change is irregular, and the traditional classification methods have difficulty with accurate division. For the transitional area, we choose the fusion method based on sparse representation. The determination of the transitional area and the specific fusion algorithm are as follows:

- The boundary line of the final decision map D is centered, the appropriate radius (3–5 pixels) is set, and the corresponding rectangular area is delineated as the transitional area R.
- Via smooth window technology, each source image is divided into image blocks of size $\sqrt{n} \times \sqrt{n}$; image blocks containing transitional area pixels are converted into column vectors, and all column vectors constitute a vector matrix $V \in R^{n \times p}$ (p is the total number of image patches intersecting the transitional area). At the same time, the matrix Λ is used to record the initial spatial position of each column vector.
- For the j-th patches v_A^j and v_B^j ($1 \le j \le p$), the discrete cosine (DC) components dc_A^j and dc_B^j are extracted first, and then we get v'^j_A and v'^j_B with no DC component.
- The sparse coefficients of each vector in the vector matrix V'_A and V'_B are calculated by the sparse coding algorithm OMP, and the corresponding sparse coefficient matrices C_A and C_B are obtained. Then, each coefficient vector is processed according to the maximum pooling principle:

$$c_F^j(\tau) = c_{\hat{\Gamma}}^j(\tau), \ \hat{\Gamma} = \arg\max_{\Gamma=A,B} \left(\left\| c_\Gamma^j(\tau) \right\| \right) \tag{6}$$

where j is the column index of the sparse coefficient matrix, and τ is the index of the atom in the dictionary Φ.
- The fused vector V'_F without the DC components is obtained by

$$V'_F = \Phi C_F \tag{7}$$

The fused DC component obeys the following rule:

$$dc_F^j = \begin{cases} \frac{dc_A^j + dc_B^j}{2}, & \text{if } 0.85 \le \left| \frac{dc_A^j}{dc_B^j} \right| \le 1.15 \\ \min(dc_A^j, dc_B^j), & \text{otherwise} \end{cases} \tag{8}$$

- The fused vector v_F^j is determined as follows:

$$v_F^j = v'_F^j + dc_F^j \cdot 1 \tag{9}$$

where each column vector v_F^j in V_F is reshaped into a block with size $\sqrt{n} \times \sqrt{n}$ and then overlaid at its recorded position in Λ.

- Finally, the transitional area fused image V_F based on the sparse representation and the fused image $\widetilde{I_F}$ based on the final decision map are averaged to generate the final fused image I_F. As shown in Figure 5, compared with the fused image $\widetilde{I_F}$ based on the final decision map, our final fused image I_F is significantly clearer at the "brim edge" and "sweater texture".
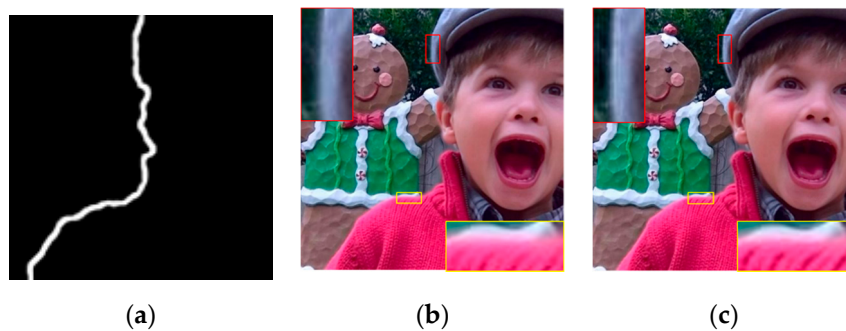

(**a**)　　　　　　　　　　(**b**)　　　　　　　　　　(**c**)

Figure 5. (**a**) The transitional area R; (**b**) the fused image $\widetilde{I_F}$ based on the final decision map; (**c**) the final fused image I_F.

On the fifth step of the algorithm, most of the existing fusion methods calculate the fused DC components using a simple average. However, this easily produces fuzzy effects around some strong edges due to the great change in brightness. The main reason for this is that the energy of the region with high brightness diffuses into the region with low brightness when losing focus. Therefore, we modify the fusion rule for DC components. When the DC components from different source images are close to each other, we choose the average operation; otherwise, the minimal DC component is selected.

3. Experiment and Analyses

This section verifies the effectiveness of the proposed method by experimenting with different types of source images. The fusion results of the proposed method are compared with several existing fusion algorithms, including DCHWT [6], SOMP [19], GF [15], IM [16], and CNN [30].

3.1. Source Images

The experiment was performed on two image datasets. The first one included eight pairs of popular multi-focus source images, as shown in Figure 6 [31]. The other one was composed of 20 pairs of color multi-focus images selected from the Lytro picture gallery, as shown in Figure 7 [32].

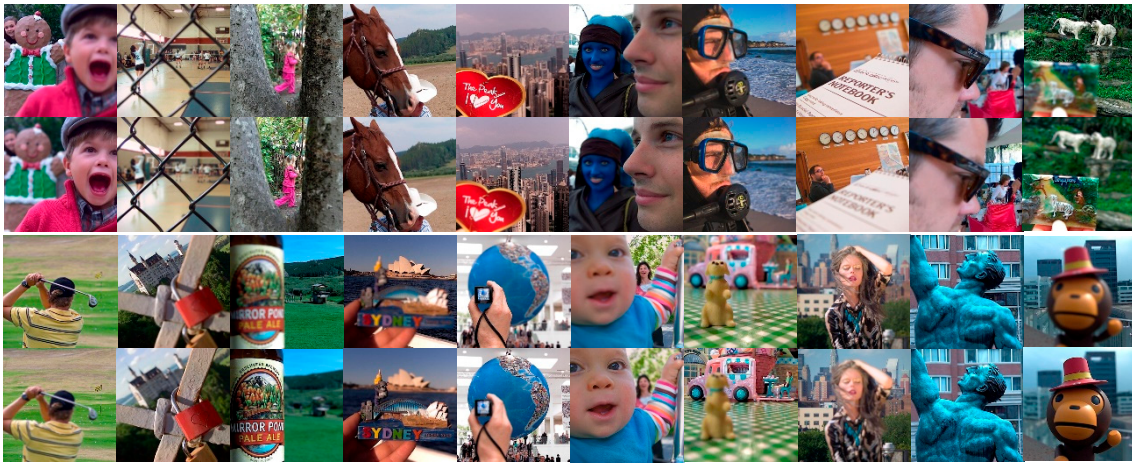**Figure 6.** Eight pairs of popular multi-focus images.

Figure 7. Twenty pairs from the Lytro image dataset.

3.2. Parameter Setting

The 8×8 image patches were used in the computation of sparse coefficients for each pixel location. Besides that, the block size of the sliding window used for clarity level comparison in the clarity score map was also fixed to 8×8. The threshold K_1 for binarizing clarity score maps was set as $K_1 = 0.65$. The overcomplete dictionary Φ used in sparse representation had a size of 64×256, which was trained globally from a large set of natural images. The residue error of the SOMP algorithm was set as $\varepsilon = 5$. The DCHWT method was implemented based on multiscale transform toolboxes downloaded from MATLAB Central [33], and its level of wavelet decomposition was set to 4. The codes for the GF and IM methods can be found on Xu Dongkang's homepage [34], and the codes for the NSCT-PCNN are available on Qu Xiaobo's homepage [35]. The parameters of these methods were set to their recommended values.

3.3. Objective Evaluation Metrics

To evaluate the fusion quality of different fusion methods, three fusion quality metrics were utilized in our experiment. The large value of the fusion quality metric indicates better fusion quality.

1. Normalized mutual information (MI), Q_{MI} [36]: Q_{MI} is used to overcome the deficit of MI [37]. Q_{MI} is defined as

$$Q_{MI} = 2\left[\frac{MI(A,F)}{H(A)+H(F)} + \frac{MI(B,F)}{H(B)+H(F)}\right] \qquad (10)$$

where $H(X)$ is the entropy of image X, and $MI(X,Y)$ is the mutual information between image X and Y. The Q_{MI} measures the amount of information in the fused image inherited from the source images.

2. Petrovic's metric, $Q^{AB/F}$ [38]: $Q^{AB/F}$ evaluates the fusion performance by measuring the amount of gradient information transferred from source images into the fused image. It is calculated by

$$Q^{AB/F} = \frac{\sum_{i,j}\left(Q^{AF}(i,j)W^A(i,j) + Q^{BF}(i,j)W^B(i,j)\right)}{\sum_{i,j}(W^A(i,j) + W^B(i,j))} \qquad (11)$$

where $Q^{AF}(i,j) = Q_g^{AF}(i,j) \cdot Q_o^{AF}(i,j)$. $Q_g^{AF}(i,j)$ and $Q_o^{AF}(i,j)$ are the grad magnitude and orientation at pixel location (i,j), respectively. Q^{BF} is computed similarly to Q^{AF}. $W^A(i,j)$ and $W^B(i,j)$ are the weights of $Q^{AF}(i,j)$ and $Q^{BF}(i,j)$, respectively.

3. The quality index, visual information fidelity for fusion (*VIFF*) [39]: This is a multiresolution image fusion metric based on visual information fidelity. To calculate the *VIFF*, the images are divided into blocks in each sub-band, and visual information in each block is measured using

different models, including the Gaussian scale mixture (GSM) model, the HVS model, and the distortion model. The *VIFF* of each sub-band is then calculated, and an overall quality measure is determined by weighting.

3.4. Experimental Results and Discussions

3.4.1. Evaluation on Popular Multi-Focus Images

In this section, we demonstrate the advantages of the proposed method (DMSR) on popular multi-focus images. An example, the fused images of the "Lab" pair (640 × 480) using different fusion methods is presented in Figure 8c–h. The "Lab" source images are shown in Figure 8a,b. For better comparison, we also present the normalized difference images between the correctly focused source image and the fusion results in Figure 9. It can be observed that the fused images obtained by DCHWT or SOMP methods showed serious artifacts and visible fake edges around the "man". The GF method had ringing artifacts and blurring effects near the "men". The IM method suffered from blurring effects near the "men's hair". The CNN method could achieve better fusion quality, but some small defects could still be found with careful observation, such as imperceptible artificial flaws on the "table" (see the lower middle in Figure 9e). Comparatively, the DMSR produced the best fused image.
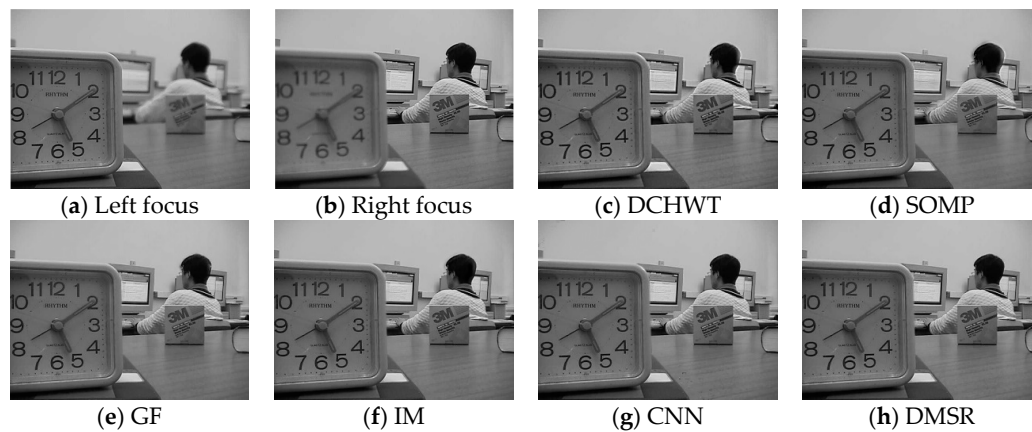

| (**a**) Left focus | (**b**) Right focus | (**c**) DCHWT | (**d**) SOMP |

| (**e**) GF | (**f**) IM | (**g**) CNN | (**h**) DMSR |

Figure 8. The "Lab" source images and fusion results obtained by different fusion methods. DCHWT—discrete cosine harmonic wavelet transform, SOMP—synchronous orthogonal matching tracking sparse coding algorithm, GF—guided filtering, IM—image matting, CNN—convolutional neural network, DMSR—proposed method based on decision map and sparse representation.
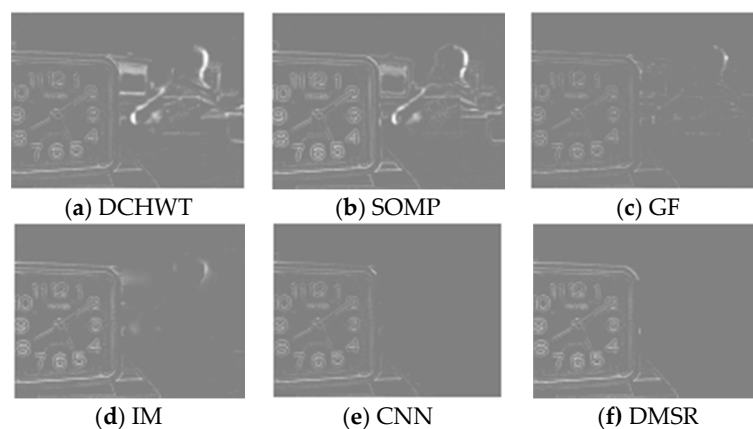

| (**a**) DCHWT | (**b**) SOMP | (**c**) GF |

| (**d**) IM | (**e**) CNN | (**f**) DMSR |

Figure 9. Normalized difference images between each of the fused images and Figure 8b.

Another example, the fusion results of the "Flowerpot" image pair (944 × 736) are shown in Figure 10c–h. The normalized difference images between the correctly focused source image and

the fusion results are shown in Figure 11. Similar to the previous example, the DCHWT and SOMP method produced serious artifacts around the "horologe". The fused image obtained by the GF method suffered from a ringing effect, and the edges of the "horologe" were blurred. The results of the IM method also showed similar artifacts near the "horologe". Although the CNN method performed well overall, it exposed obvious artifacts on the "ground" and the "wall" of the fused image. Comparatively, the DMSR method exhibited the best visual quality.
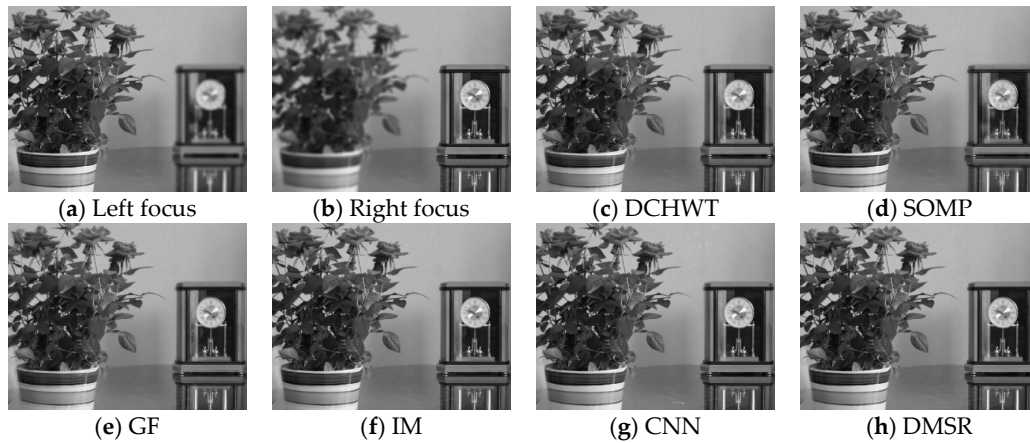

| (**a**) Left focus | (**b**) Right focus | (**c**) DCHWT | (**d**) SOMP |
| (**e**) GF | (**f**) IM | (**g**) CNN | (**h**) DMSR |

Figure 10. The "Flowerpot" source images and fusion results obtained by different fusion methods.

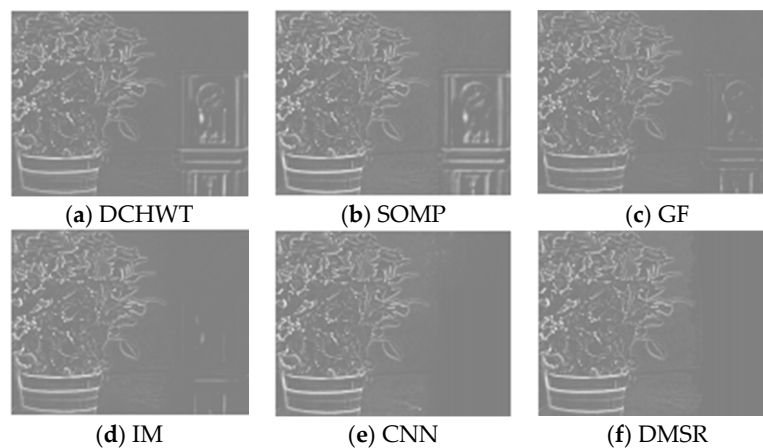| (**a**) DCHWT | (**b**) SOMP | (**c**) GF |
| (**d**) IM | (**e**) CNN | (**f**) DMSR |

Figure 11. Normalized difference images between each of the fused images and Figure 10b.

To evaluate fusion performance more objectively, each pair of popular multi-focus images was fused by six fusion methods. The values of metrics Q_{MI}, $Q^{AB/F}$, and *VIFF* were calculated and are recorded in Table 1, with the best results indicated in bold. It can be seen that the DMSR method outperformed all other methods and won in almost all the quality metrics.

Table 1. Quantitative assessments of different image fusion methods for popular multi-focus images.

Source Images	Metric	Method					
		DCHWT [6]	SOMP [19]	GF [15]	IM [16]	CNN [30]	DMSR
Pepsi	Q_{MI}	1.1167	0.9030	1.2164	**1.3080**	1.2911	1.3033
	$Q^{AB/F}$	0.7315	0.6666	0.7510	0.7550	**0.7591**	0.7583
	VIFF	0.9273	0.9235	0.9498	0.9335	0.9511	**0.9543**
Clock	Q_{MI}	1.0417	0.8485	1.1509	1.1837	1.2058	**1.2351**
	$Q^{AB/F}$	0.7165	0.6510	0.7427	0.7429	0.7464	**0.7478**
	VIFF	0.9279	0.9235	0.9404	0.9322	0.9432	**0.9469**
Lab	Q_{MI}	1.0611	0.9481	1.1928	1.2221	1.2363	**1.2583**
	$Q^{AB/F}$	0.7162	0.6708	0.7552	0.7506	0.7573	**0.7584**

Table 1. *Cont.*

Source Images	Metric	Method					
		DCHWT [6]	SOMP [19]	GF [15]	IM [16]	CNN [30]	DMSR
	VIFF	0.8880	0.8642	0.9190	0.9112	0.9166	**0.9182**
Flowerpot	Q_{MI}	0.8935	0.8142	1.0215	1.0969	**1.1539**	1.1248
	$Q^{AB/F}$	0.6706	0.6484	0.7252	0.7269	0.7326	**0.7345**
	VIFF	0.8309	0.8450	0.8827	0.8891	0.8911	**0.8913**
Bookcase	Q_{MI}	0.9266	0.7737	1.0271	1.0943	1.1104	**1.1353**
	$Q^{AB/F}$	0.6791	0.6374	0.7290	0.7246	0.7342	**0.7357**
	VIFF	0.8600	0.8592	0.8820	0.8781	0.8800	**0.9322**
Leaf	Q_{MI}	0.7825	0.5467	0.8843	0.9883	0.9055	**1.0241**
	$Q^{AB/F}$	0.7176	0.6447	0.7358	**0.7386**	0.7342	0.7353
	VIFF	0.8049	0.8234	0.8071	0.8097	0.8103	**0.8333**
Book	Q_{MI}	1.1069	0.8243	1.1702	1.2093	1.2331	**1.2739**
	$Q^{AB/F}$	0.7061	0.6273	0.7262	0.7228	0.7277	**0.7279**
	VIFF	0.8501	0.8320	0.8497	**0.8584**	0.8496	0.8502
Flower	Q_{MI}	0.9139	0.6207	1.0960	1.1206	1.1263	**1.1404**
	$Q^{AB/F}$	0.6969	0.6239	0.7175	0.7114	**0.7183**	0.7143
	VIFF	0.9192	0.9150	0.9295	0.9243	0.9297	**0.9318**

3.4.2. Evaluation on Lytro Image Dataset

The Lytro image dataset was composed of 20 color multi-focus image pairs of the same size (520 × 520). For visual evaluation, the fused results of the "Lytro17" image pair obtained by different fusion methods are demonstrated in Figure 12. In order to observe the fusion effect of the transitional area more intuitively, some details of the puppy have been intercepted and enlarged. The DCHWT method still exhibited undesirable ringing artifacts around the head, as shown in Figure 12c. The same phenomenon can also be seen in Figure 12d,e,g. As shown in the close-up views of Figure 12f, the IM method suffered from severe blurring effects and false edges. Comparatively, the DMSR methods produced ideal fusion images without perceptible artifacts along the focus boundary.
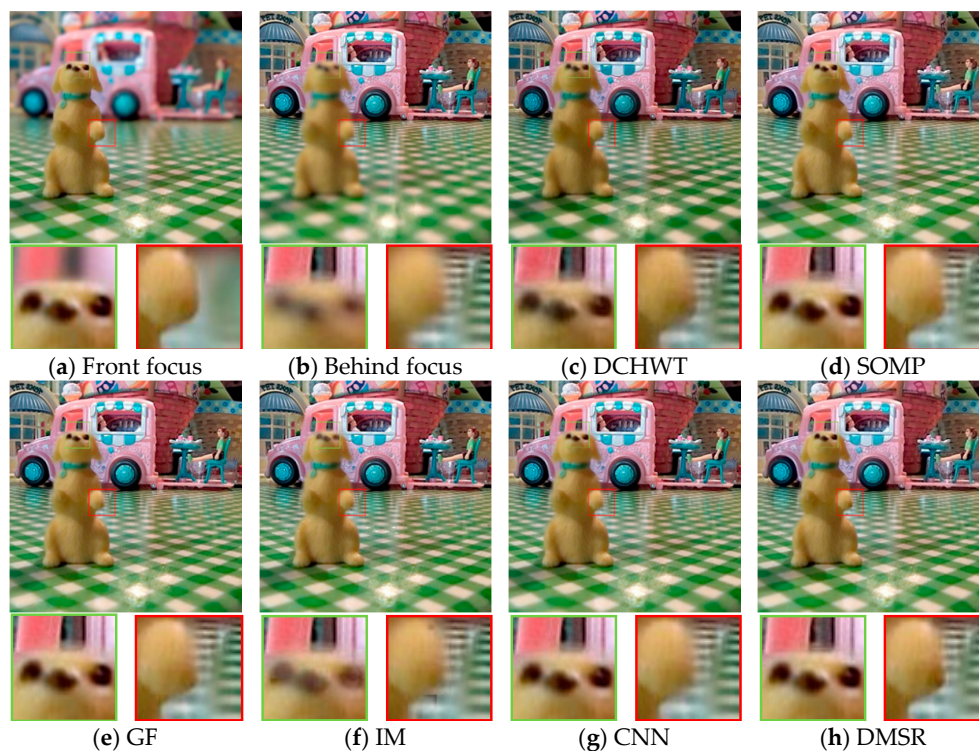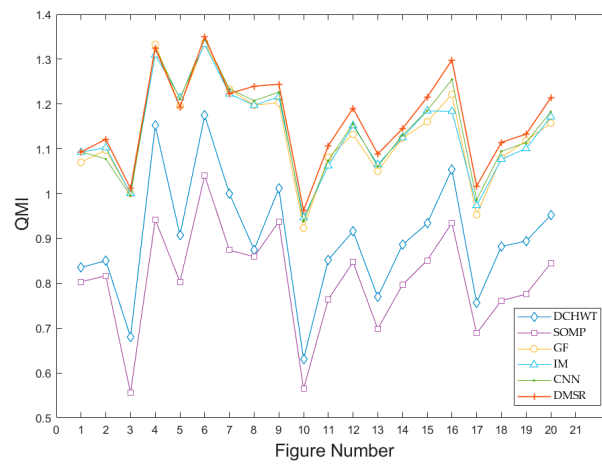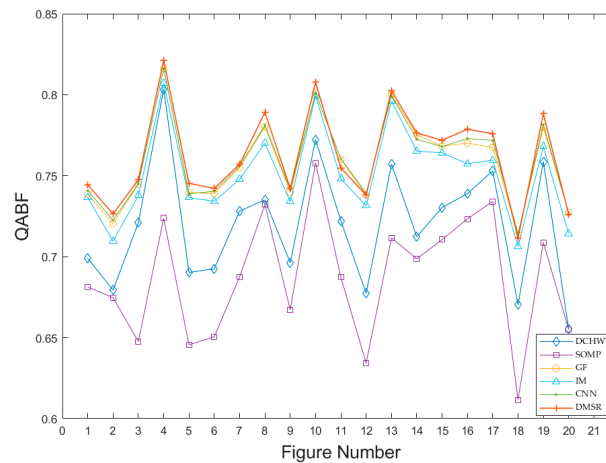

Figure 12. The "Lytro17" image pair and fusion results obtained by different fusion methods.
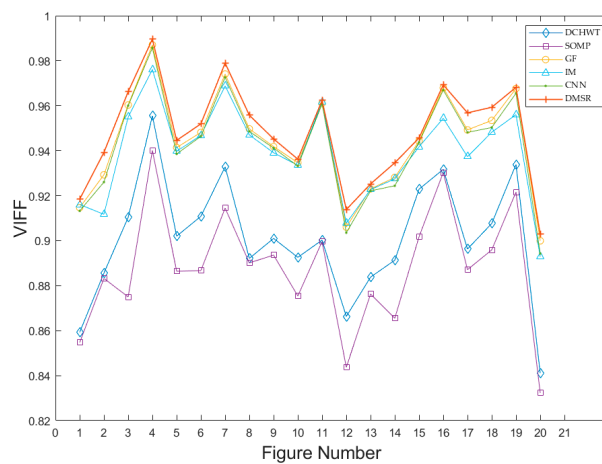
Further, the quantitative assessments of the six methods are shown in Figure 13. The charts show that the proposed method outperformed the others and obtained the best quality metrics.

(**a**)

(**b**)

(**c**)

Figure 13. Quantitative assessment line charts of different image fusion methods for Lytro image dataset: (**a**) Q_{MI}; (**b**) $Q^{AB/F}$; (**c**) visual information fidelity for fusion (*VIFF*).

3.4.3. Evaluation on Three Multi-Focus Images

Our method was also suitable for more than two multi-focus images. The three source images for "Toy" (512×512) are shown in Figure 14a–c, and close-up views are shown at the bottom for better observation. Figure 14d,e show that the fused images obtained by the DCHWT and SOMP methods showed serious blurring effects at the "ball" in the right corner. The GF fusion method produced jagged edges around the "puppet", as shown in Figure 14f. The IM fusion method exhibited slight blurry artifacts in the upper-right corner of the "ball", as shown in Figure 14g. Compared with other methods, the CNN and DMSR performed well. As shown in Figure 14h,i, all focused areas from the source images were merged into the fusion image with imperceptible artifacts. The values of Q_{MI}, $Q^{AB/F}$, and *VIFF* for various fusion methods are presented in Table 2, with the best results indicated in bold.
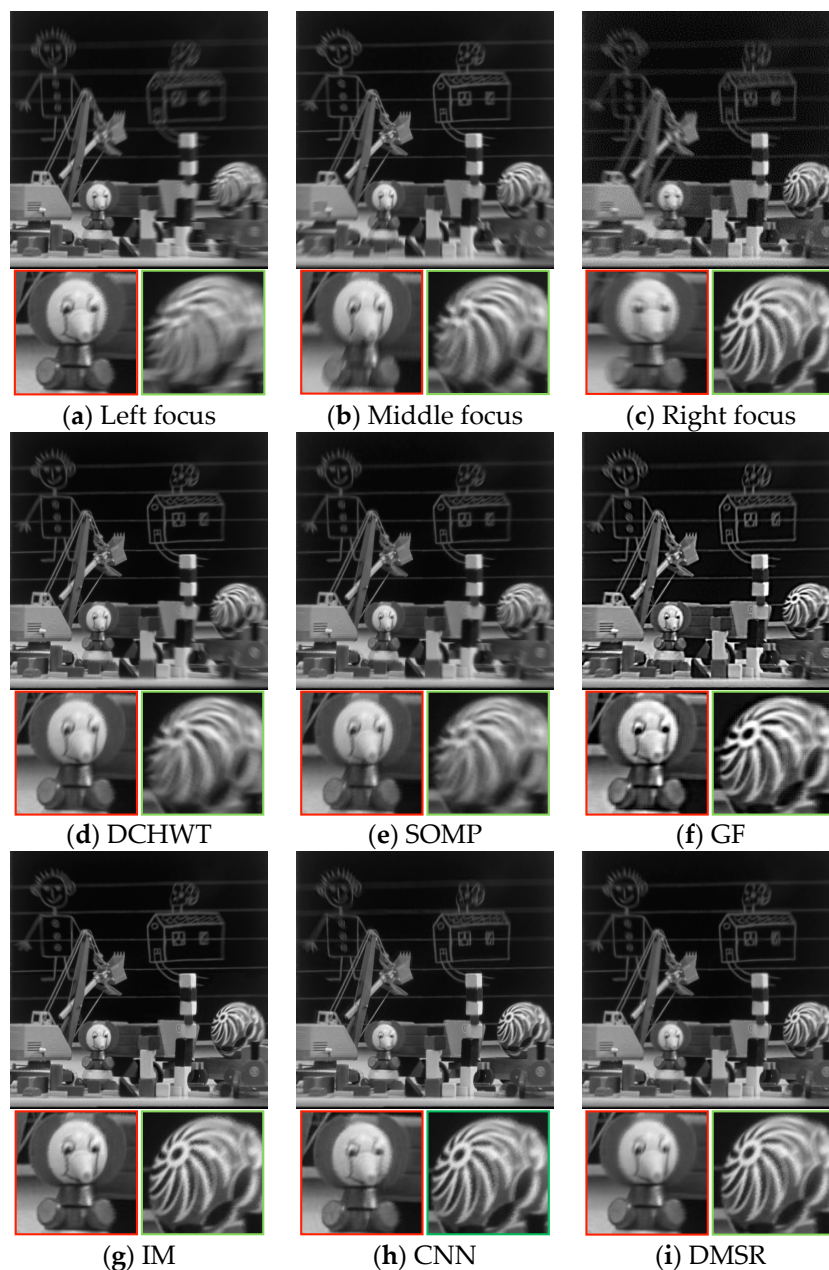

(**a**) Left focus (**b**) Middle focus (**c**) Right focus

(**d**) DCHWT (**e**) SOMP (**f**) GF

(**g**) IM (**h**) CNN (**i**) DMSR

Figure 14. The three multi-focus image and fusion results obtained by different fusion methods.

Table 2. Quantitative assessments of different image fusion methods for three multi-focus images.

Metric	Method					
	DCHWT [6]	SOMP [19]	GF [15]	IM [16]	CNN [30]	DMSR
Q_{MI}	1.1272	0.8180	1.1943	1.1943	1.2561	**1.2588**
$Q^{AB/F}$	0.7452	0.6743	0.7586	0.7409	0.7554	**0.7605**
VIFF	0.9288	0.8720	0.9560	0.9531	0.9600	**0.9631**

4. Conclusions

In this paper, we propose a new multi-focus image fusion method based on decision map and sparse representation. By generating the initial decision map by focusing on feature analysis for low-scale images, not only can the performance be guaranteed but the computational complexity can also be effectively reduced. Aiming at the characteristics of difficult decisions in the transitional area, we used the fusion algorithm based on sparse representation to directly fuse this and effectively reduce the error caused by incorrect judgment while ensuring the quality of fusion. In addition, the fusion method is also generalized to be capable of fusing more than two images. Experimental results show that the fusion method proposed in this paper has better fusion quality than other methods, both in terms of visual perception and objective measurement. In the future, we plan to evaluate whether the method proposed here can be applied to multi-focus image fusion in dynamic scenes.

Author Contributions: B.L. and H.C. conceived and designed the algorithm; B.L. and H.C. performed the experiments; W.M. analyzed the data and contributed reagents/materials/analysis tools; B.L. and H.C. wrote the paper; W.M. provided technical support and revised the paper.

References

1. Wan, T.; Zhu, C.C.; Qin, Z.C. Multifocus image fusion based on robust principal component analysis. *Pattern Recognit.* **2013**, *34*, 1001–1008. [CrossRef]
2. Li, S.T.; Yang, B.; Hu, J.W. Performance comparison of different multi-resolution transforms for image fusion. *Inf. Fusion* **2011**, *12*, 74–84. [CrossRef]
3. Tian, J.; Chen, L. Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure. *Signal. Process.* **2012**, *92*, 2137–2146. [CrossRef]
4. Li, H.; Manjunath, B.S.; Mitra, S.K. Multisensor Image Fusion Using the Wavelet Transform. In Proceedings of the IEEE International Conference on Image Processing, Washington, DC, USA, 23–26 October 1995; pp. 235–245.
5. Lewis, J.J.; O'Callaghan, R.J.; Nikolov, S.G.; Bull, D.R.; Canagarajah, N. Pixel- and region-based image fusion with complex wavelets. *Inf. Fusion* **2007**, *8*, 119–130. [CrossRef]
6. Kumar, B.K.S. Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform. *Signal. Image Video Process.* **2013**, *7*, 1125–1143. [CrossRef]
7. Miao, Q.G.; Shi, C.; Xu, P.F.; Yang, M.; Shi, Y.B. A novel algorithm of image fusion using shearlets. *Opt. Commun.* **2011**, *284*, 1540–1547. [CrossRef]
8. Tessens, L.; Ledda, A.; Pizurica, A.; Philips, W. Extending the depth of field in microscopy through curvelet-based frequency-adaptive image fusion. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, HI, USA, 15–20 April 2007; pp. 861–864.
9. Zhang, Q.; Guo, B.L. Multi-focus image fusion using the nonsubsampled contourlet transform. *Signal. Process.* **2009**, *89*, 1334–1346. [CrossRef]
10. Li, S.; Kwok, J.T.; Wang, Y. Combination of images with diverse focuses using the spatial frequency. *Inf. Fusion* **2001**, *2*, 169–176. [CrossRef]
11. Aslantas, V.; Kurban, R. Fusion of multi-focus images using differential evolution algorithm. *Expert Syst. Appl.* **2010**, *37*, 8861–8870. [CrossRef]

12. Li, M.; Cai, W.; Tan, Z. A region-based multi-sensor image fusion scheme using pulse-coupled neural network. *Pattern Recognit.* **2006**, *27*, 1948–1956. [CrossRef]

13. Li, S.T.; Yang, B. Multifocus image fusion using region segmentation and spatial frequency. *Image Vis. Comput.* **2008**, *26*, 971–979. [CrossRef]

14. Wei, H.; Jing, Z.L. Evaluation of focus measures in multi-focus image fusion. *Pattern Recognit.* **2007**, *28*, 493–500.

15. Li, S.T.; Kang, X.D.; Hu, J.W. Image Fusion with Guided Filtering. *IEEE Trans. Image Process.* **2013**, *22*, 2864–2875. [PubMed]

16. Li, S.; Kang, X.; Hu, J.; Yang, B. Image matting for fusion of multi-focus images in dynamic scenes. *Inf. Fusion* **2013**, *14*, 147–162. [CrossRef]

17. Kumar, B.K.S. Image fusion based on pixel significance using cross bilateral filter. *Signal. Image Video Process.* **2015**, *9*, 1193–1204. [CrossRef]

18. Yang, B.; Li, S. Multifocus image fusion and restoration with sparse representation. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 884–892. [CrossRef]

19. Yang, B.; Li, S. Pixel-level image fusion with simultaneous orthogonal matching pursuit. *Inf. Fusion* **2012**, *13*, 10–19. [CrossRef]

20. Chen, L.; Li, J.B.; Chen, C.L.P. Regional multifocus image fusion using sparse representation. *Opt. Express.* **2013**, *21*, 5182–5197. [CrossRef]

21. Liu, Y.; Liu, S.; Wang, Z.F. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fusion* **2015**, *24*, 147–164. [CrossRef]

22. Liu, Y.; Wang, Z. Simultaneous image fusion and denoising with adaptive sparse representation. *Image Process. IET* **2015**, *9*, 347–357. [CrossRef]

23. Nejati, M.; Samavi, S.; Shirani, S. Multi-focus image fusion using dictionary-based sparse representation. *Inf. Fusion* **2015**, *25*, 72–84. [CrossRef]

24. Li, S.; Yin, H.; Fang, L. Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 3450–3459. [CrossRef] [PubMed]

25. Eckhorn, R.; Reitboeck, H.J.; Arndt, M.; Dicke, P. Feature Linking via Synchronization among Distributed Assemblies: Simulations of Results from Cat Visual Cortex. *Neural Comput.* **2014**, *2*, 293–307. [CrossRef]

26. Qu, X.B.; Yan, J.W.; Xiao, H.Z.; Zhu, Z.Q. Image Fusion Algorithm Based on Spatial Frequency-Motivated Pulse Coupled Neural Networks in Nonsubsampled Contourlet Transform Domain. *Acta Autom. Sin.* **2008**, *34*, 1508–1514. [CrossRef]

27. Broussard, R.P.; Rogers, S.K.; Oxley, M.E.; Tarr, G.L. Physiologically motivated image fusion for object detection using a pulse coupled neural network. *IEEE Trans. Neural Networks* **1999**, *10*, 554–563. [CrossRef] [PubMed]

28. Wang, Z.; Ma, Y.; Gu, J. Multi-focus image fusion using PCNN. *Pattern Recognit.* **2010**, *43*, 2003–2016. [CrossRef]

29. Huang, W.; Jing, Z. Multi-focus image fusion using pulse coupled neural network. *Pattern Recognit.* **2007**, *28*, 1123–1132. [CrossRef]

30. Liu, Y.; Chen, X.; Peng, H.; Wang, Z.F. Multi-focus image fusion with a deep convolutional neural network. *Inf. Fusion* **2017**, *36*, 191–207. [CrossRef]

31. Popular Multi-Focus Images. Available online: http://www.ece.lehigh.edu/SPCRL/IF/image_fusion.htm (accessed on 1 May 2019).

32. Lytro Picture Gallery. Available online: https://mansournejati.ece.iut.ac.ir/content/lytro-multi-focus-dataset (accessed on 1 May 2019).

33. MATLAB Central. Available online: http://cn.mathworks.com/matlabcentral/ (accessed on 5 May 2019).

34. Xu Dongkang's Homepage. Available online: http://xudongkang.weebly.com/ (accessed on 10 May 2019).

35. Qu Xiaobo's Homepage. Available online: http://www.quxiaobo.org/index.html (accessed on 10 May 2019).

36. Aharon, M.; Elad, M.; Bruckstein, A. *rmK*-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal. Process.* **2006**, *54*, 4311–4322. [CrossRef]

37. Elad, M.; Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **2006**, *15*, 3736–3745. [CrossRef]

38. Xydeas, C.S.; Petrović, V. Objective image fusion performance measure. *Electron. Lett.* **2000**, *36*, 308–309. [CrossRef]
39. Han, Y.; Cai, Y.Z.; Cao, Y.; Xu, X.M. A new image fusion performance metric based on visual information fidelity. *Inf. Fusion* **2013**, *14*, 127–135. [CrossRef]