ABSTRACT Based on the analysis of the multi-scale-based and sparse representation-based multi-focus image fusion method, we believe that a multi-scale fusion method should consider the directivity of high/low frequency sub-bands and low computational complexity, ensuring the effectiveness and efficiency. Therefore, we propose a novel multi-focus image fusion algorithm based on multi-scale and multi-directional dictionaries. In the proposed method, source images are decomposed by multi-scale transform (MST) to obtain high/low frequency sub-bands. For the high-frequency part, a direction contrast-based fusion rule is presented. For the low-frequency part, the image patches are divided into strong/weak information patches. The strong information patches are merged with a fusion approach based on directional dictionaries, while the weak information patches are fused using weighted average. Finally, the fused image is obtained by performing the inverse MST. The experimental results show that the proposed approach extracts more important effective information from source images, and the computational efficiency is greatly improved.

INDEX TERMS Multi-focus image fusion, multi-scale transform, directionality, fusion rule.

I. INTRODUCTION
It is well known that optical sensors can only focus on targets with the same depth-of-field (DOF), and targets that are defocus are fuzzy. These images are called single-focus images. Since the need of images with all clear targets is getting higher and higher in the follow-up process of machine vision and in the fields of military and public safety, the development of multi-focus image fusion technology, to obtain an all-in-focus image from single-focus images, becomes more and more important.

Numerous studies have attempted to explain multi-focus image fusion from different point of view [1]–[3]. In recent years, there has been an increasing amount of literature on multi-scale decomposition methods for their good multi-resolution and multi-directional characteristic of high frequency [4], [5]. As early as 1990s, [6], [7] used discrete wavelet transform (DWT) to process multi-focus images, better extracting detailed information from source images such as edge contours. However, the traditional wavelet transform can only capture high frequency information in three directions (horizontal, vertical and diagonal). Therefore, dual-tree complex wavelet transform (DTCWT) [8], curvelet (CVT) [9], non-subsampled contourlet transform (NSCT) [10], non-subsampled shearlet transform (NSST) [11], fast finite shearlet transform (FFST) [12] and other MSTs were proposed successively. These MSTs have multiple directions in the high frequency part and directional selectivity, therefore good results are achieved.

We find that the MST-based fusion methods usually have the following characteristics: (1) most of fusion strategies are based on 'max-absolute' rule for high-frequency part; (2) weighted average rule is used for low-frequency part [13], [14]. We believe that the high frequency fusion rule doesn’t take into account surrounding parameter information; the low frequency fusion rule hasn’t self-adaptability and doesn’t make use of direction information. This kind of methods are abbreviated as MST(L,M,R,A) in the paper, where MST is the multi-scale transformation used, L is decomposition level, M means ‘max’, R is the negative power of the threshold value and A means ‘average’.

Along with the multi-scale methods, several attempts have been made to adaptive sparse representation (SR) based on dictionary learning [15]–[17]. For example, in 2010, Yang and Li applied SR to multi-focus image fusion, in which the sparse representation coefficients are fused [18]. In 2014, Liu and Wang proposed ASR model, in which instead of a single redundant dictionary, a set of more compact
sub-dictionaries are used. Each sub-dictionary is adaptively learned from numerous high-quality image patches that are selected in a given direction [19]. In 2018, a fusion scheme based on image cartoon-texture decomposition and sparse representation is proposed. A spatial-based morphological structure preservation method is proposed for cartoon components. A sparse representation method is proposed for texture components and dictionaries with strong representation ability are trained [20].

Improved approaches based on the combination of MST and SR have been attempted. In 2013, Wang et al. proposed a fusion method combining NSCT and SR [21]. In 2015, Liu et al. adopted complementary advantages of MST and SR, and proposed a general image fusion framework(NSST is not included) [22]. In 2017, Qiu et al. proposed a fusion method based on NSST and SR [23]. This kind of methods is abbreviated in this paper as MST(L,Ms,R,D), where Ms means the absolute values of coefficients as activity level measurement rule for high-frequency part, D means dictionary coefficient max rule for low-frequency part.

Compared with MST(L,M,R,A), MST(L,Ms,R,D) methods has the following characteristics: (1) the high frequency part has multi-scale, directionality and fast algorithm. (2) The fusion rule for low frequency part is adaptive, but lacks directivity.

Based on the above analysis, we propose a novel multi-focus image fusion approach based on multi-scale and multi-directional dictionaries. The proposed method makes better use of the complementation of the MST-based and SR-based fusion approaches, and makes up the shortcomings of the MST(L,Ms,R,D) methods by full using of direction information of the high/low frequency part.

To be more specific, there are three main contributions in this new method.

1) A novel fusion rule for high frequency sub-band. This rule is based on direction contrast and uses direction weight of each direction of each scale. Since the high-frequency sub-bands that are obtained by the MST decomposition have multi-directional characteristics, taking full use of the direction information in high frequency fusion rule will greatly improve the fusion results.

2) In the low frequency part, image patches containing useful information can be divided into strong information patch and weak information patch. Many patches with small variance are called weak information patches and the fusion rule for them can simply be weighted average. This approach for non-sub-sampled MSTs greatly decreases the number of image patches to be processed, reducing computational cost greatly.

3) A new fusion method for strong information patches is proposed in the low-frequency part. These patches with direction information are adaptively sparsely represented by directional dictionaries and the “L1-max” fusion rule is performed on the sparse representation coefficients. In this method based on the directional dictionaries, the strong information patches are divided into several finer direction groups, and the fusion rule makes the best of the direction information of the low-frequency image patches.

The rest of this paper is organized as follows. Section 2 combines high and low frequency fusion rule to get a complete multi-focus image fusion model. Section 3 discusses the high frequency sub-band fusion strategy based on direction contrast, including the determination of the MST decomposition level. In Section 4, the learning method for directional dictionaries is introduced. The concepts of strong/weak information patches of low-frequency sub-band and the fusion rules for these two kinds of patches are proposed. The fusion performance of the proposed method is verified by seven sets of experiments in Section 5. Section 6 summarizes the main content of this paper.

### II. MULTI-FOCUS IMAGE FUSION METHOD BASED ON MST DIRECTION

Suppose there are $M$ source images $\{I_1, I_2, \cdots, I_M\}$, and the size of the source images is $X, Y$. The proposed method is summarized in Algorithm 1.

To the best of our knowledge, this is the first attempt to combine the multi-direction in the high frequency part and the directional dictionaries in the low frequency part for multi-focus image fusion.
III. HIGH FREQUENCY SUB-BAND FUSION RULE
The key problem of high frequency fusion is to determine the number of MST decomposition level and full utilization of the directivity of high frequency sub-bands.

A. MST DECOMPOSITION LEVEL
The determination of the decomposition level is a key to the MST-based fusion methods [24]. If the decomposition level is too small such as 1 or 2, there will be a lot of spatial details that are not separated from the source images; if the decomposition level is too large, more than 4, the high frequency sub-band fusion will be more sensitive to noise and mis-registration [22]. On the basis of [22], the decomposition level is set to 3 or 4 in the paper, making the method more robust to mis-registration.

B. HIGH FREQUENCY SUB-BAND FUSION RULE BASED ON DIRECTION CONTRAST
After MST decomposition, the energy of the high-frequency part of the source images is concentrated on a few coefficients [25]. These sparse high-frequency coefficients represent detail information such as edges and textures at different scales and different directions. It should be pointed out that due to the vanishing moment property of the MSTs, at the same position the coefficient energy of the clear region is greater than that of the fuzzy region [4]. Therefore, ’max absolute’ fusion rule is commonly used for the high-frequency fusion. This method is simple, but the correlation between pixels in image is not taken into account, which can’t well extract the detailed information from the source images, and reduces the contrast of the fusion image.

Considering the correlation between image pixels, a fusion rule based on traditional contrast was proposed in [26]. Traditional image contrast is defined as

\[ HC = (HB)^{-1}(HP - HB) = (HB)^{-1}(HH). \]  

(1)

where HP is the local gray level of the image, HB and HP−HB are respectively equal to the low frequency and high frequency component after image transformation. Traditional contrast takes into account the correlation between pixels of image points, ignoring the correlation between the coefficients in the region and the directivity of the high frequency sub-bands.

In order to comprehensively consider the correlation between regional pixels and make better use of contrast with different scales and different directions, the direction contrast is defined by the regional window \( B^{p,q}(0, 1) \) matrix and the direction weight \( \omega \). The direction weight accords to the directional characteristics of the high frequency sub-bands, and every direction at every scale of the high frequency sub-bands have its own weight.

Let \( \omega^{p,q} \) is the direction angle of the \( q \)-th high frequency sub-band at the \( p \)-th scale. The elements of regional window matrix \( B^{p,q} \) with size \( N \times N \) is defined as 0 except the elements located between \( s1 = \tan(\alpha)r(n) + 1 \) and \( s2 = \tan(\alpha)r(n) - 1 \), where the value is set to 1. In the formula above, \( r(n) \) is from 1 to \( N \), \( r(n) \) is from \(-N/2\) to \( N/2 \).

Then the direction weight of the q-th directional sub-band at the p-th scale can be defined as

\[ \omega^{p,q} = B^{p,q} \odot E, \]  

(2)

where \( E = UV, U = (1, 2, \cdots, N/2+1), (N/2, \cdots, 2, 1)^T \), \( V = U^T, \{\lfloor \cdot \rfloor \} \) denotes floor operation).

Recall that \( \{H_1^{p,q}, H_2^{p,q}, \cdots, H_M^{p,q}\} \) are the high frequency sub-bands. Using the convolution of them with respect to \( \omega^{p,q} \), we have

\[ cH_{lm}^{p,q} = \omega^{p,q} * \left| H_{lm}^{p,q} \right|. \]  

(3)

It can enhance directional features of \( H_{lm}^{p,q} \) and reduce noise. Now the direction contrast can be defined:

\[ C_{lm}(x, y) = \sum_{p=1}^{P} \sum_{q=1}^{Q} \left| cH_{lm}^{p,q}(x, y) \right| \sum_{i,j \in N(x,y)} L_{lm}(i,j), \]  

(4)

where \( N(i, j) \) is the neighborhood window with pixel point \( (x, y) \) as the center and size is \( N \times N \). The direction contrast associates the low frequency sub-band with the high frequency sub-bands, and is obtained by adding up contrast information of all the high frequency sub-bands. So the contrast information of the source images \( I_m \) is clearly reflected by the matrix \( C_{lm} \).

The high frequency sub-bands fusion rule is established based on direction contrast matrix \( \{C_{lm}\}_{m=1}^{M} \). The high frequency sub-bands \( \{H_1^{p,q}, H_2^{p,q}, \cdots, H_M^{p,q}\} \) are fused by

\[ H_{ld}^{p,q}(x, y) = H_{ld}^{p,q}(x, y), d = \arg \max_m \left\{ \left| C_{lm}(x, y) \right| \right\}. \]  

(5)

This high-frequency fusion method more comprehensively considers the correlation between regional pixels and the contrast of different scales and different directions, which can better extract the effective information from the high frequency sub-bands of the source images.

IV. LOW FREQUENCY SUB-BAND FUSION RULE
The two key problems of low-frequency fusion are the classification of strong/weak information patches, and the fusion method of strong information patches based on directional dictionaries.

A. STRONG/WEAK INFORMATION PATCHES
In the MST-based multi-focus fusion methods, the decomposition level should be set to 3 or 4, therefore a large amount of basic information that reflects the overall characteristics of the image still remain in some low frequency patches. Figure.1 and Figure.2 show the percentage of the low-frequency image patches, whose variance all less than 10−3 after six MSTs (DWT, DTCWT, CVT, NSCT, NSST, FFST) respectively.
We can see that for the MSTs with down-sampling, such as DWT, DTCWT, CVT, etc., most of the low frequency image patches contain useful information. Due to the down-sampling the size of low frequency part is much smaller than that of source images, which makes the information in the low frequency sub-band highly concentrated. Furthermore, for non-subsampled MSTs, such as NSCT, NSST, FFST, etc., the size of the low-frequency sub-band is the same as that of the original images, and the useful information is scattered. Therefore, the low-frequency sub-bands of this kind of MSTs contain many patches with small variance.

The percentage of patches with variance greater than $10^{-3}$ and have at least one of the four direction information is showed in Figure.3 and Figure.4. It is observed that the percentage of patches which have direction information goes from zero to 90%, depending on the different MST used.

This analysis suggests that, to make full use of the direction information while reducing the computational cost, the image patches of the low frequency part should be divided into weak/strong information patch depending on the variance of each image patches. Given a threshold, if the variance of image patch less than threshold, this image patch is weak information patch (smooth image patch); otherwise, it is strong information patch (image patch containing lots of useful information). Since the smooth weak information patches contain less detailed information, the simple weighted average rule can be directly performed. For the strong information patches, the sparse representation fusion rule based on directional dictionaries is used to better utilize direction information of strong information patches. This divided and conquer method can greatly improve the fusion results and computational efficiency.

**B. DIRECTIONAL DICTIONARIES LEARNING**

In order to make better use of the direction information of strong information patches, a fusion method based on directional dictionaries is proposed.

As we all know, the high frequency of each scale after the image decomposed by MST contains different directional sub-bands, indicating the direction information of the high frequency part of the image. For example, DWT contains three directions that are horizontal, diagonal ($45^\circ, 135^\circ$) and vertical. Similarly, FFST also contains four different parts. Figure.5 is the frequency domain segmentation of FFST, and the frequency domain plane is divided into different regions, horizontal cones and vertical cones. The number of direction angles of high frequency sub-bands in each scale is greater than or equal to 4. In consequence, the number of main directions of the low-frequency sub-band should be less than or equal to 4. In order to consider the directivity of the low-frequency sub-band more comprehensively, we learn the directional dictionaries with four direction angles in the low frequency part. Thus, the low-frequency image patches with direction can be better represented in its corresponding directional dictionary, making the dictionary learning process more efficient.

To be specific, we randomly sample 60,000 image patches from thirty natural images that don’t contain the images used in the experiment. The smooth image patches with small variance are removed to obtain 53,896 image patches with structural information. According to the principle of supervised classification and consistent gradient information, four groups, each group has one direction, are obtained from these patches. The corresponding directional dictionary is respectively learned from one of the four groups, and the number of atoms of each sub-dictionary is set to 256. In order to represent image patches with no clear single direction, such
Let directional dictionary \([27], [28]\). The results of the directional image patches. The reference patch, the directional sub-dictionary is selected, largest variance is taken as the reference patch. According to of them is higher than the threshold, they are regarded as patches at the same position, if the maximum variance \(M\) is calculated over \(l\). As patches are overlapped, the value of each pixel in \(L_F\) is averaged over its accumulation times.

\[
M = \max_{l} \left\{ \frac{1}{2} \sum_{m=1}^{M} \left\| c_m^l \right\|_2 \right\}
\]

where \(\bar{c}_m^k\) is the mean value over all the elements of \(c_m^k\). \(1\) denotes an all-one valued \(H \times 1\) vector. Pick out \(c_f^k\) which has largest variance among \(\{c_1^k, c_2^k, \ldots, c_M^k\}\). Then the adaptively selected dictionary \(\{D_i\}_{i=1}^4\) is selected based on its gradient information [19].

3) For each column vector \(c_m\), calculate the sparse coefficient vector \(\alpha_m^k\) by

\[
\min_{\alpha_m} \left\{ \alpha_m^k \right\}_0 \text{ s.t. } \left\| c_m - D_f^k \alpha_m^k \right\|_2 < \varepsilon,
\]

where \(\varepsilon > 0\) is error tolerance. Merge the \(M\) obtained sparse vectors \(\{\alpha_1^k, \alpha_2^k, \ldots, \alpha_M^k\}\) by

\[
f = \arg \max_m \left\{ \left\| \alpha_m^k \right\|_1, m = 1, 2, \ldots, M \right\}, \quad c_f^k = c_{f}^k,
\]

\[
\alpha_f^k = \alpha_{f}^k, \alpha_f^k, \alpha_f^k, \text{ the fused sparse vector, } c_f^k,\text{ is the merged mean value of } \{\bar{c}_m^k\}_{m=1}^M.
\]

4) The fused result \(c_f^k\) of \(\{c_1^k, c_2^k, \ldots, c_M^k\}\) is calculated by

\[
c_f^k = D_f^k \alpha_f^k + \bar{c}_m^k \cdot 1.
\]

For each \(c_f^k\), reshape it into \(H \times H\) patch \(l_f^k\).

Let \(L_F\) denote low frequency fused coefficient, and then plug \(l_f^k\) into its original position \(k\) in \(L_F\). As patches are overlapped, the value of each pixel in \(L_F\) is averaged over its accumulation times.
V. EXPERIMENTS

A. EXPERIMENTAL PREPARATION

1) SOURCE IMAGES

As shown in Figure 8, seven pairs of source images are employed to verify the effectiveness and efficiency of the proposed fusion algorithm [29].

2) EXPERIMENTAL SETTING

In experiments, six MSTs and, which are DTW, DTCWT, CVT, NSCT, NSST, FFST are used to exhibit the superiority of the proposed fusion method, and the comparison of different fusion methods based on the same MST is performed in each group of experiments. These comparisons are MST(L,M,R,A), MST(L,M,R,D), MST(L,M,R,dD), MST(L,dM,R,D), MST(L,Ms,R,D) proposed in [22], [23], MST(L,dM,R,dD) proposed in this paper and ASR proposed in [19]. Where L is decomposition level, M means 'max', A means 'average', D means dictionary coefficient max rule for low-frequency part, R is the negative power of the threshold value, dD means the low frequency fusion rule based on directional dictionaries, dM means high frequency fusion rule based on direction contrast, Ms means the absolute values of coefficients as activity level measurement rule for high-frequency part, ASR means a SR-based method.

The image filters for all the DTCWT-methods are selected as 'LeGall 5-3' and 'Qshift-06'; we use the 'pyrex' as the pyramid filter and the 'vk' as the directional filter for all the NSCT-methods, and the four decomposition levels are selected as 3, 3, 4 and 4, respectively; for all the NSST-methods, the size of shearing filter matrixes at four decomposition levels are respectively set to 30, 30, 36 and 36, the error tolerance $\varepsilon = 0.1$ [22]. The patch size $H = 8$ and the neighborhood window size $N = 9$. All experiments are performed on MATLAB R2017a. The computer processor is Intel(R) Xeon(R) Silver 4110CPU.

3) SELECTION OF THE NEGATIVE POWER R

This section verify the validity with the addition of variance thresholds. Compare MST(L,dM,R,dD) with MST(L,dM,0,dD), where MST(L,dM,R,dD) is fusion method proposed in this paper, R is the negative power of the variance limit, the variance thresholds $\delta = 10^{-R}$, $R = 2, 3, 4$; in MST(L,dM,0,dD) fusion method, each low-frequency image patch is represented on the directional dictionary. In the experiment, the 'Disk' images with different fuzzy region are used to evaluate the fusion performance of different variance thresholds. In terms of visual effects, the fused images are almost no difference and the images are relatively clear in Figure 9.

Except for the visual effects, we also compare the comprehensive evaluation indexes of the fused results. The evaluation indexes of the methods based on NSCT are listed in Table 1. The bold fonts of Table 1 mean better fusion performance. It can be seen that at the same MST and decomposition level, most of evaluation index values obtained by NSCT(L,dM,3,dD) are optimal.
threshold $10^{-3}$, the strong and weak information patches can be divided accurately. The similar experiments are carried out for other MSTs and other source images, and it is proved that $10^{-3}$ is the best. To avoid lengthy experiments, only a set of images and a table are presented.

In general, MST(L,dM,3,dD) has the best fusion performance. The low-frequency sub-band patches with variance less than $10^{-3}$ should be merged by the weighted average rule and the low-frequency sub-band patches with variance more than $10^{-3}$ should be merged by fusion rule based on directional dictionaries, which can enhance the fusion result.

4) EVALUATION METRICS

In general, image fusion results can be subjectively and objectively evaluated. On the one hand, the subjective approach is based on psycho-visual. Usually the fusion results are less visually different, therefore it is hard to properly evaluate the results. So two subjective evaluation metrics peak signal to noise-ratio(PSNR) [30] and root mean square error (RMSE) [31] that calculate the difference between the original image and the fusion image can be used as auxiliary. On the other hand, six commonly used objective evaluation metrics are selected to quantitatively measure the experimental fusion results. These are mutual information metric (MI) [32], human perception based fusion metric ($Q_G$) [34], gradient based fusion metric ($Q_F$) [34], phase congruency based fusion metric ($Q_P$) [35], structural similarity-based fusion metric ($Q_Y$) [36] and standard deviation (SD) [37]. As we known, MI is used to measure the amount of information obtained from the source images; $Q_G$ reflects the major features of the human visual system (visual information fidelity); $Q_F$ evaluates the range of gradient information extracted from the source images, and measures the clarity degree of the fused image; $Q_P$ reflects the information of image salient features such as edges and corners; $Q_Y$ measures the level of structural information of the source images preserved in the fusion image; SD is used to measure the richness of the texture information of the fused image. Except RMSE, larger values for other metrics mean better fusion results. The closer RMSE is to zero, the more successful the method is.

| TABLE 1. Quantitative assessments of Figure.10 (a)-(h). |
|---|---|---|---|---|---|---|---|
| Metric | a | b | c | d | e | f | g |
| MI | 0.8730 | 0.8797 | 0.8796 | 0.8796 | 0.8744 | 0.8744 | 0.8727 | 0.8727 |
| $Q_{CB}$ | 0.7926 | 0.7713 | 0.7569 | 0.7494 | 0.7419 | 0.7400 | 0.7415 | 0.7715 |
| $Q_F$ | 0.6105 | 0.6148 | 0.6200 | 0.6200 | 0.6227 | 0.6229 | 0.6220 | 0.6220 |
| $Q_Y$ | 0.8610 | 0.8609 | 0.8605 | 0.8605 | 0.8613 | 0.8614 | 0.8613 | 0.8613 |
| $Q_Y$ | 0.9299 | 0.9330 | 0.9316 | 0.9316 | 0.9299 | 0.9299 | 0.9296 | 0.9296 |
| PSNR | 83.5685 | 83.5641 | 83.5736 | 83.5736 | 83.6779 | 83.8049 | 83.7622 | 83.7622 |
| RMSE | 10.5207 | 10.5249 | 10.5268 | 10.5268 | 10.6461 | 10.6470 | 10.6424 | 10.6424 |

| TABLE 2. Fusion time of MST(L,dM,0,dD) and MST(L,dM,3,dD) on ‘Disk’. |
|---|---|---|---|---|
| MST | (3,dM,0,dD) | (3,dM,3,dD) | (4,dM,0,dD) | (4,dM,3,dD) |
| DWT | 3.4401s | 3.8668s | 2.1999s | 2.2343s |
| DTCWT | 7.1562s | 6.7786s | 3.2352s | 3.4209s |
| CVT | 5.4144s | 5.6333s | 3.1194s | 3.19071s |
| NSCT | 32.9456s | 19.1362s | 29.1812s | 14.3071s |
| NSST | 36.4241s | 21.6502s | 32.7587s | 16.6800s |
| FFST | 24.7700s | 6.5863s | 7.2075s | 7.1569s |

B. DIFFERENT ASPECTS OF THE ALGORITHM PERFORMANCE

Ablation study are conducted to prove the influence of R on fusion efficiency, the fusion effectiveness of high/low-frequency fusion rule.

1) THE INFLUENCE OF R ON FUSION EFFICIENCY

After source images are decomposed by subsampled MST, such as DWT, CVT and DTCWT, if most of the low-frequency sub-band patches don’t contain weak information patches, the thresholding step will take a little more time while keeping the same results. Therefore the efficiency improvement with the threshold is mainly based on non-subsampled MST, such as NSCT, NSST and FFST.

In order to verify the efficiency improvement, the runtime of MST(L,dM,3,dD) and MST(L,dM,0,dD) are compared. The 'Disk', 'Balloon' and 'Leaf' source images with different fuzzy region are used to evaluate the computational efficiency with or without the variance threshold. The results of run time are severally showed in Table 2, Table 3 and Table 4. The bold fonts of these three tables mean that at the same MST and the same decomposition level, the fusion time of the algorithm proposed in the paper greatly reduces the fusion time.

Based on the same down-sampled MSTs at the same decomposition level, from Table 2, except for DTCWT(L,dM,3,dD) and DTCWT(L,dM,0,dD), the calculation time of MST(L,dM,3,dD) is slightly more than that of MST(L,dM,0,dD), but the difference is less than 0.5 seconds. The reason is that the low frequency sub-band patches based on DTCWT three-level decomposition contain a small number of weak information patches. From Table 3 and Table 4, most of the calculation time of MST(L,dM,3,dD) is less than that of MST(L,dM,0,dD). Based on the same non-subsampled MST at the same decomposition level, the computation time of MST(L,dM,3,dD) is much smaller than that of MST(L,dM,0,dD).

In terms of computational complexity, The computational complexity of MST(L,dM,3,dD) is $O(TN^3 + T'N^2XY + XY + 1 + K'(1 + HK^2 + 256^2 + 1))$, where $T$ is the number of high frequency fusion rule.

VOLUME 9, 2021
frequency sub-bands, $K'$ is the number of patches that need to be directional dictionaries fused, $k = \|a_{km}\|^0$.

The computational complexities of the compared algorithm MST(L,M,R,A), ASR, MST(L,Ms,R,D) and MST(L,dM,0,dD) are $O(T) + O(1)$, $O(K(1+HK^2+256^2+1))$, $O(K(1+HK^2+256^2+1)+T)$ and $O(TN^3+TN^2XY+XY+1 + K(1 + HK^2 + 256^2 + 1))$, respectively.

In short, MST(L,M,R,A) has the lowest complexity; the time complexity of MST(L,dM,3,dD) is less than that of ASR; if MST is down-sampled wavelet, the time complexity of MST(L,dM,0,dD) is slightly larger than that of MST(L,Ms,R,D) and MST(L,dM,0,dD); if MST is non-sampled wavelet, the time complexity of MST(L,dM,3,dD) is far less than that of MST(L,Ms,R,D) and MST(L,dM,0,dD).

In summary, as long as the low-frequency sub-band patches contain weak information patches, the proposed algorithm MST(L,dM,3,dD) can improve the computational efficiency. The low computational complexity of the proposed algorithm in this paper is verified.

2) FUSION PERFORMANCE OF HIGH-FREQUENCY FUSION RULE

Based on the same low-frequency fusion strategy, different high-frequency fusion strategies are compared. The comparison of MST(L,M,R,D) with MST(L,M,R,D) and the comparison of MST(L,M,dM,3,dD) with MST(L,M,dM,3,dD) are used to show the effects of 'absolute max' (M) and direction contrast (dM). Visually, Figure.10 shows the results of a group of fused high frequency sub-bands. The first two images are the high-frequency information of the source images, and the last two images are the high-frequency fusion results of 'M' and 'dM' method respectively. From Table 5-10, the evaluation indexes of MST(L,M,R,D) is 93.75% higher than that of MST(L,M,R,D), and the evaluation indexes of MST(L,M,dM,3,dD) is 93.75% higher than that of MST(L,M,dM,3,dD). It can be obtained that direction contrast is extremely effective for the high-frequency fusion. The high-frequency sub-band mainly contains rich edge characterization and contour information, and the direction contrast can extract more information about contour and edge from the source images. All in all, the high-frequency fusion rule based on direction contrast is helpful to extract more effective information from the high-frequency sub-bands of the source images. The necessity of directionality for high-frequency fusion is verified.

3) FUSION PERFORMANCE OF LOW-FREQUENCY FUSION RULE

Based on the same high-frequency fusion rule, different low frequency fusion strategies are compared. The comparison of MST(L,M,R,D) with MST(L,M,dM,3,dD) and the comparison of MST(L,M,R,D) with MST(L,M,dM,3,dD) are used to show the difference effects of single dictionary and directional dictionaries-based. In Figure.11, the first two images are the low-frequency information of the source images, and the last two images are the low-frequency fusion results of 'D' and 'dD' method respectively. From Table 5-10, the evaluation indexes of MST(L,M,dM,3,dD) is 82.81% higher than that of MST(L,M,R,D), and the evaluation indexes of MST(L,M,dM,3,dD) is 84.38% higher than
that of MST(L,dM,R,D). It can be seen that the low-frequency sub-band coefficients chosen by fusion rule based on directional dictionaries are very effective. The low-frequency sub-band is an approximation of the source image at low resolution, including a large amount of basic information of the source image. Using the low frequency fusion rule based on directional dictionaries, the low frequency information can be better extracted, which greatly improves the quality of the fused image. The necessity of directionality for low-pass fusion is verified.

**C. ALGORITHM FUSION EFFECTIVENESS**

Based on the same MST and the same decomposition level, the subjective vision (Figure.12-17) and evaluation metrics (Table 5-10) of the six experiments are compared to obtain comprehensive comparison results. The bold fonts of Table 5-10 mean better fusion performance. In order to avoid lengthy experimental discussion, the influence of
decomposition level is only discussed in the down-sampled CVT and non-subsampled FFST. Other MSTs are analyzed only at the four level decomposition.

1) EXPERIMENTS BASED ON DWT
In visual effects, the blurred regions of the two source images (Figure.12(a) and (b)) in Figure.12(c)-(i) become more clear. With the help of quantitative assessments Table 5, the evaluation metric values of fusion result obtained by DWT(4,dM,3,dD) has four optimal values and two sub-optimal value. We can see from Table 5 that the DWT(4,dM,3,dD) method gets the first place for $Q_{CB}$, $Q_{G}$, $Q_{Y}$ and PSNR and has clear advantages over the other methods. The results of subjective and objective evaluations indicate that DWT(4,dM,3,dD) method improved the clarity of fused image, and the superiority of the proposed algorithm is verified. The fusion strategy (dM,3,dD) improves the DWT-based multi-focus image fusion results.

2) EXPERIMENTS BASED ON DTCWT
Considering visual comparison, Figure.13(c)-(i) all combine the effective information of the source images Lab A and Lab B. However, except Figure.13(h) obtained by the method proposed in this paper, there is shadow around the head of the person in Figure.13(c)-(i). The better fusion results in Figure.13(h) verify that the fusion strategy DTCWT(4,dM,3,dD) can better extract useful information from the source images. While, it can be seen from Table 6 the evaluation metrics of fusion result by DTCWT(4,dM,3,dD) has six optimal values and one sub-optimal value. The comprehensive evaluations further point out the effectiveness of (dM,3,dD) algorithm in DTCWT-based multi-focus image fusion.

3) EXPERIMENTS BASED ON CVT
Compared with the source images Figure.14(a) and Figure.14(b), the fused image Figure.14(c)-(o) successfully saves the focused portion of the two source images. However, after careful observation, it is found that the results based on three-level decomposition of CVT in Figure.14(c)-(h) have obvious fusion unevenness on the rightmost side of the large clock. Referring further to Table 7, the CVT(dM,3,dD) method has three optimal values ($Q_{CB}$, $Q_{G}$, PSNR) corresponding high visual fidelity, better description of edges and

![Figure 14](image-url)

![Figure 15](image-url)

![Table 9](image-url)
corners and low noise. The (dM,3,dD) fusion strategy has its obvious advantages for CVT-based multi-focus image fusion.

4) EXPERIMENTS BASED ON NSCT
In terms of visual effects, Figure.15(c)-(i) have better visual effect than the source images. By means of Table 8, the NSCT(4,dM,3,dD) method has seven optimal values and one sub-optimal value. The obvious superiority of NSCT(4,dM,3,dD) verifies that this method can obtain the largest amount of information from the source images. The combination of NSCT and (dM,3,dD) led to better NSCT-based multi-focus fusion results.

5) EXPERIMENTS BASED ON NSST
It can be seen from Figure.16 that Figure.16(c)-(i) extract useful information of the source images, so that the fused images are all clear. In virtue of quantitative assessments Table 9, the NSST(4,dM,3,dD) have six optimal values and one sub-optimal value. The NSST(4,dM,3,dD) method has a better extraction of the feature information of the source images, such as edges and corners, retaining the most structural information of the source images. Both the fusion results of Figure.16 and the evaluation indexes of Table 9 verify that the (dM,3,dD) fusion strategy can effectively extract the structural information of the source images in NSST-based multi-focus image fusion.

6) EXPERIMENTS BASED ON FFST
It can be found that the fusion result by FFST(L,dM,3,dD) has almost no artifacts and uneven fusion of the clothes and hair of the girl, and the contrast between the trunk and the girl is strong. With the help of Table 10, The FFST(L,dM,3,dD) method have six optimal values and one sub-optimal value. From Figure.17 and Table 10, the combination of FFST and (dM,3,dD) fusion strategy helps to improve the image quality of FFST-based multi-focus image fusion.

The comparison of MST(L,M,R,A), ASR, and MST(L,Ms, R,D) with our MST(L,dM,3,dD) can verify the fusion algorithm proposed in this paper is superior to MST-based method, SR-based method, and their simple combination.
V. CONCLUSION

This paper proposes a novel general multi-focus image fusion method based on multi-scale and multi-direction. In this method, the fusion rule based on direction contrast is used for the high frequency sub-band, and the direction weight is defined according to the high frequency MST sub-band direction. The low-frequency MST sub-band patches are classified according to the variance. The weak information patches are merged by weighted average rule. The strong information patches are merged based on directional dictionaries. Six MSTs (DWT, DTCWT, CVT, NSCT, NSST, FFST) are included in the experiment. Seven fusion rules are compared. The experiment results show that the algorithm proposed in this paper is best in subjective visual effects and comprehensive evaluation indexes. The full use of directionality can better extract the detail, contour and edge information of the source images. When a suitable variance threshold is used, the calculation efficiency has also been greatly improved, especially for the non-subsampled MSTs.

REFERENCES

[1] Y. Yang, Z. Nie, S. Huang, P. Lin, and J. Wu, “Multilevel features convolutional neural network for multifocus image fusion,” *IEEE Trans. Comput. Imag.*, vol. 5, no. 2, pp. 262–273, Jun. 2019.

[2] H. Zhai and Y. Zhuang, “Multifocus image fusion method based on a convolutional neural network,” *J. Electron. Imag.*, vol. 28, pp. 1–11, Mar. 2019.

MST(L,Ms,R,D). The evaluation indexes of MST(L,dM,3,dD) is 93.75% higher than that of MST(L,M,R,A), which indicates that the fusion algorithm proposed in this paper improves the multi-focus fusion result based on MST; the evaluation indexes of jMST(L,dM,3,dD) is 71.88% higher than that of ASR algorithm proposed in [19], which indicates that the fusion algorithm proposed in this paper is superior to the SR-based fusion algorithm ASR; the evaluation indexes of MST(L,dM,3,dD) is 96.88% higher than that of MST(L,Ms,R,D) proposed in [22], [23], which indicates that the multi-focus image fusion results by the proposed algorithm in this paper are better than algorithm combining MST-based and SR-based proposed in [22], [23]. So, the general multi-focus fusion method proposed in this paper is superior to the traditional MST-based method, the SR-based method ASR and the algorithm combining MST-based and SR-based. It is verified that a good multi-scale fusion method should consider the directivity of high/low frequency sub-bands.

VI. CONCLUSION

This paper proposes a novel general multi-focus image fusion method based on multi-scale and multi-direction. In this method, the fusion rule based on direction contrast is used for the high frequency sub-band, and the direction weight is defined according to the high frequency MST sub-band direction. The low-frequency MST sub-band patches are classified according to the variance. The weak information patches are merged by weighted average rule. The strong information patches are merged based on directional dictionaries. Six MSTs (DWT, DTCWT, CVT, NSCT, NSST, FFST) are included in the experiment. Seven fusion rules are compared. The experiment results show that the algorithm proposed in this paper is best in subjective visual effects and comprehensive evaluation indexes. The full use of directionality can better extract the detail, contour and edge information of the source images. When a suitable variance threshold is used, the calculation efficiency has also been greatly improved, especially for the non-subsampled MSTs.

REFERENCES

[1] Y. Yang, Z. Nie, S. Huang, P. Lin, and J. Wu, “Multilevel features convolutional neural network for multifocus image fusion,” *IEEE Trans. Comput. Imag.*, vol. 5, no. 2, pp. 262–273, Jun. 2019.

[2] H. Zhai and Y. Zhuang, “Multifocus image fusion method based on a convolutional neural network,” *J. Electron. Imag.*, vol. 28, pp. 1–11, Mar. 2019.
[3] R. Lai, Y. Li, J. Guan, and A. Xiong, “Multi-scale visual attention deep convolutional neural network for multi-focus image fusion,” IEEE Access, vol. 7, pp. 114385–114399, Aug. 2019.

[4] S. Maqsood, U. Javed, M. M. Riaz, M. Muzammil, F. Muhammad, and R. G. Baraniuk, “The dual-tree complex wavelet transform,” IEEE Trans. Signal Process., vol. 67, no. 11, pp. 2855–2869, Nov. 2019.

[5] W. Liu and Z. Wang, “A novel multi-focus image fusion method using multiscale shearing non-local guided averaging filter,” Signal Process., vol. 166, Jan. 2020, Art. no. 107252.

[6] H. Li, B. S. Manjunath, and S. Mitra, “Multisensor image fusion using the wavelet transform,” Graph. Models Image Process., vol. 57, no. 3, pp. 235–245, Mar. 1995.

[7] C. E. Heil and D. F. Walnut, Continuous and Discrete Wavelet Transforms, ser. Applied and Numerical Harmonic Analysis, B. F. Rajan, Ed. Boston, MA: Birkhäuser, 1995.

[8] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, “The dual-tree complex wavelet transform,” IEEE Signal Process. Mag., vol. 22, no. 6, pp. 123–131, Nov. 2005.

[9] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, “Remote sensing image fusion using the curvelet transform,” Inf. Fusion, vol. 8, no. 2, pp. 143–156, Apr. 2007.

[10] T. Li and Y. Wang, “Biological image fusion using a NSCT based variable-weight method,” Inf. Fusion, vol. 12, no. 2, pp. 85–92, Apr. 2011.

[11] B. Wang and S. Li, “Multi-focus image fusion based on non-subsampled shearlet transform,” JET Image Process., vol. 7, no. 6, pp. 633–639, Mar. 2013.

[12] W. Yajie, P. Quanbo, W. Yanyan, and Y. Zhoufeng, “A fusion method for visible light and infrared images based on FFST and compressed sensing,” in Proc. 29th Chin. Control Decis. Conf. (CCDC), May 2017, pp. 5184–5188.

[13] G. Pajares and J. Manuel de la Cruz, “A wavelet-based image fusion tutorial,” Pattern Recognit., vol. 37, no. 9, pp. 1855–1872, Sep. 2004.

[14] Y. Wen, X. Yang, T. Celik, O. Sushkova, and M. K. Albertini, “Multifocus image fusion using convolutional neural network,” Multimedia Tools Appl., vol. 79, pp. 34531–34543, May 2020.

[15] Q. Zhang, G. Li, Y. Cao, and J. Han, “Multi-focus image fusion based on negative-sparse representation and patch-level consistency rectification,” Pattern Recognit., vol. 104, Aug. 2020, Art. no. 107325.

[16] S. Maqsood and U. Javed, “Multi-modal medical image fusion based on two-scale image decomposition and sparse representation,” J. Biomed. Signal Process. Control, vol. 57, Mar. 2020, Art. no. 101810.

[17] X. Ma, S. Hu, S. Liu, J. Fang, and S. Xu, “Multi-focus image fusion based on joint sparse representation and optimum theory,” Signal Process., Image Commun., vol. 78, pp. 125–134, Oct. 2019.

[18] 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.

[19] Y. Liu and Z. Wang, “Simultaneous image fusion and denoising with adaptive sparse representation,” IET Image Process., vol. 9, no. 5, pp. 347–357, May 2015.

[20] B. Wei, X. Yang, T. Celik, O. Sushkova, and M. K. Albertini, “A wavelet-based image fusion tutorial,” Pattern Recognit., vol. 37, no. 9, pp. 1855–1872, Sep. 2004.

[21] B. Wei, X. Yang, T. Celik, O. Sushkova, and M. K. Albertini, “Multifocus image fusion using convolutional neural network,” Multimedia Tools Appl., vol. 79, pp. 34531–34543, May 2020.

[22] C. E. Heil and D. F. Walnut, Continuous and Discrete Wavelet Transforms, ser. Applied and Numerical Harmonic Analysis, B. F. Rajan, Ed. Boston, MA: Birkhäuser, 1995.

[23] C. E. Heil and D. F. Walnut, Continuous and Discrete Wavelet Transforms, ser. Applied and Numerical Harmonic Analysis, B. F. Rajan, Ed. Boston, MA: Birkhäuser, 1995.

[24] B. Wei, X. Yang, T. Celik, O. Sushkova, and M. K. Albertini, “Multifocus image fusion using convolutional neural network,” Multimedia Tools Appl., vol. 79, pp. 34531–34543, May 2020.

[25] G. Pajares and J. Manuel de la Cruz, “A wavelet-based image fusion tutorial,” Pattern Recognit., vol. 37, no. 9, pp. 1855–1872, Sep. 2004.

[26] B. Wei, X. Yang, T. Celik, O. Sushkova, and M. K. Albertini, “Multifocus image fusion using convolutional neural network,” Multimedia Tools Appl., vol. 79, pp. 34531–34543, May 2020.

[27] B. Wei, X. Yang, T. Celik, O. Sushkova, and M. K. Albertini, “Multifocus image fusion using convolutional neural network,” Multimedia Tools Appl., vol. 79, pp. 34531–34543, May 2020.

[28] R. Dian, S. Li, L. Fang, and Q. Wei, “Multispectral and hyperspectral image fusion with spatial-spectral sparse representation,” Inf. Fusion, vol. 49, pp. 262–270, Sep. 2019.

[29] Y. Zhang, X. Z. Bai, and T. Wang, “Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure,” Inf. Fusion, vol. 35, pp. 81–101, May 2017.

[30] P. Jagalingam and A. V. Hegde, “A review of quality metrics for fused images,” Aquatic Procedia, vol. 4, pp. 133–142, 2015.

[31] S. Aymaz and C. Köse, “A novel image decomposition-based hybrid technique with super-resolution method for multi-focus image fusion,” Inf. Fusion, vol. 45, pp. 113–127, Jan. 2019.

[32] M. Hossny, S. Nahavandi, and D. Creighton, “Comments on ‘Information measure for performance of image fusion,’” Electron. Lett., vol. 44, pp. 1066–1067, Aug. 2008.

[33] Y. Chen and R. S. Blum, “A new automated quality assessment algorithm for image fusion,” Image Vis. Comput., vol. 27, no. 10, pp. 1421–1432, Sep. 2009.

[34] C. S. Xydeas and V. S. Petrovic, “Objective image fusion performance measure,” Electron. Lett., vol. 36, no. 4, pp. 308–309, Mar. 2000.

[35] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, “Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 1, pp. 94–109, Jan. 2012.

[36] C. Yang, J.-Q. Zhang, X.-R. Wang, and X. Liu, “A novel similarity based quality metric for image fusion,” Inf. Fusion, vol. 9, no. 2, pp. 156–160, Apr. 2008.

[37] M. Deshmukh and U. Bhosale, “Image fusion and image quality assessment of fused images,” Int. J. Image Process., vol. 4, no. 5, pp. 484–508, Apr. 2010.

**BINGZHE WEI** received the B.S. degree from Pingdingshan University, Henan, China, in 2014, and the M.S. and Ph.D. degrees in computational mathematics from the Xi’an University of Architecture and Technology, in 2017. She is currently pursuing the Ph.D. degree with the School of Mathematics and Statistics, Xidian University. Her research interests include image fusion and image denoising.

**XIANGCHU FENG** received the B.S. degree in computational mathematics from Xi’an Jiaotong University, Xi’an, China, in 1984, and the M.S. and Ph.D. degrees in applied mathematics from Xidian University, Xi’an, in 1989 and 1999, respectively. He is currently a Professor with the School of Mathematics and Statistics, Xidian University. His research interests include numerical analysis, wavelets, and partial differential equations for image processing.

**WEIWEI WANG** received the B.S., M.S., and Ph.D. degrees from Xidian University, Xi’an, China, in 1993, 1998, and 2001, respectively. She is currently a Professor with the School of Mathematics and Statistics, Xidian University. Her research interests include pattern classification, subspace clustering, sparse representation, low-rank representation, and their applications in image processing.

***