




NSMT: A Novel Non-subsampled Morphological Transform Fusion Algorithm for Infrared–Visible Images

Peng Hu^{1,2}  · Chenjun Wang^{1,2} · Dequan Li^{1,2} · Xin Zhao^{1,2}

Received: 17 March 2023 / Revised: 16 September 2023 / Accepted: 16 September 2023 /
Published online: 9 October 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Bright-dark components and edge details are the most important complementary information between infrared and visible images. To extract and fuse them efficiently, a novel non-subsampled morphological fusion algorithm is proposed in this paper. The algorithm uses non-subsampled pyramid (*NSP*) as the spatial-frequency splitter to decompose the source image to get a series of high-frequency detail images and one low-frequency background image. Then, a dual-channel multi-scale top–bottom hat (*MTBH*) decomposition is constructed to extract the bright-dark details from the low-frequency background. In addition, to extract the edge details with different directions from high-frequency images, a dual-channel multidirectional inner-outer edge (*MIOE*) decomposition is constructed. Through these decompositions, the bright-dark information and edge details present in the source images can be effectively extracted. Then, based on the distinct roles of the extracted information, the decomposed images are fused using diverse fusion strategies. Subsequently, the fused image is reconstructed using the appropriate inverse transforms corresponding to each decomposition. The experimental results demonstrate that the fusion images generated by this algorithm exhibit richer details and higher image contrast compared to those produced by state-of-the-art algorithms.

✉ Peng Hu
aust_hp@163.com
Chenjun Wang
cumt1279@163.com
Dequan Li
dqli@aust.edu.cn
Xin Zhao
zhaoxin@aust.edu.cn

¹ State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mines, Anhui University of Science and Technology, Huainan 232001, China

² School of Artificial Intelligence, Anhui University of Science and Technology, Huainan 232001, China

Keywords Image fusion · Multi-scale decomposition · Morphological decomposition · Non-subsampled pyramid

1 Introduction

As commonly known, infrared sensors have the capability to accurately capture the infrared radiation emitted by objects, independent of lighting conditions [16]. Therefore, infrared images can highlight targets with high thermal radiation well (such as the people in the infrared image in Fig. 1) [11]. Unfortunately, the spatial resolution of infrared images is insufficient for displaying detailed textures clearly. Conversely, visible images can provide clear textures and rich details, such as edges or lines. However, in situations with inadequate illumination or complex scenes, visible images often struggle to display targets effectively (such as the people in the visible image in Fig. 1) [26]. Hence, relying solely on a single infrared or visible image cannot offer comprehensive information about a specific scene in numerous computer vision applications. In dealing with this challenge, image fusion technology emerges as an effective solution capable of synthesizing the information from both infrared and visible source images [41]. As a result, infrared–visible image fusion technology plays a crucial role in various fields, including object recognition [35], remote sensing [37], surveillance [15], and more.

To acquire high-quality infrared–visible fusion images, a large number of fusion algorithms have been proposed. Among them, the fusion methods based on multi-scale transformation have garnered significant attention and witnessed rapid development [7, 8, 17]. For instance, the effectiveness of wavelet-based multi-scale transform algorithms has been substantiated through numerous experiments and applications, rendering them persistently studied and utilized to this day [9, 32, 40, 45]. However, a limitation of these methods is their inability to effectively represent two-dimensional singular information such as edges or lines [6]. Consequently, one has to propose multidirectional wavelet [2–4], curvelet [1], contourlet [23], shearlet [14], and other



Fig. 1 A pair of typical infrared–visible images. The left one is an infrared image, and the right one is a visible image

methods [5, 29] to solve the above problem. Regrettably, these approaches often necessitate designing complex filter banks, and their exceptional fusion performance comes at the expense of real-time processing capability.

With the development of technology, hybrid multi-scale methods have emerged as a solution to overcome the performance limitations of individual algorithms [12, 13, 42]. For example, Zhou et al. have proposed a novel fusion method based on hybrid multi-scale decomposition. In their algorithm, the fused image can produce enhanced visual effects by incorporating infrared spectral information [12]. Furthermore, with the widespread adoption of deep learning, a plethora of deep-learning-based fusion methods have been proposed in recent years [19–21, 27]. For example, in Ref. [18], a fusion framework for infrared–visible images based on deep learning is presented, while Ref. [20] combines a residual network with zero-phase analysis for the fusion of infrared–visible images.

These deep-learning-based fusion methods are innovative and have achieved notable achievements. But, according to Ref. [39], extensive fusion experiments suggest that the improvement in fusion performance brought about by deep-learning technology is quite limited. In most cases, these methods do not exhibit significant advantages compared to traditional classical algorithms. Additionally, there are two prominent drawbacks associated with deep-learning-based methods. Firstly, their fusion performance heavily relies on the availability of training datasets. Unfortunately, constructing large-scale datasets is often challenging in many fields, and ground-truth fusion images are frequently unavailable. Secondly, their generalization capability is severely constrained. Fusion models trained on datasets of multi-focus images usually show poor performance when confronted with other image types. As a result, it is crucial to develop a general, high-performance fusion method for infrared–visible images.

In general, the key to obtaining high-quality fusion images lies in extracting and effectively fusing the complementary information present in the source images. As is commonly acknowledged, the primary sources of complementary information between infrared and visible images are texture details (such as edges and lines) and bright-dark information. Therefore, fully extracting and efficiently fusing these elements is crucial for achieving optimal fusion outcomes. To address this objective, this paper proposes a novel fusion method called non-sampled morphological transformation (NSMT). Firstly, the *NSP* decomposition is applied to decompose the source image into one low-frequency background image (*LF*) and a series of high-frequency detail images (*HF*) with different resolutions. Secondly, a dual-channel multi-scale morphological decomposition called *MTBH* is constructed to extract the bright-dark details from the *LF* image. Simultaneously, a dual-channel multi-direction morphological edge decomposition called *MIOE* is constructed to decompose the *HF* images to fully extract the edges with various directions. Subsequently, the edge images including smooth images and detailed images are fused by the approach of choosing the absolute maximum and window-gradient maximum, respectively. For the bright-dark information from the background, the rule of “choosing brighter from the bright and darker from the dark” is used to enhance image contrast. Finally, the fusion image is reconstructed by corresponding inverse transformation.

The rest of this paper is organized as follows: in Sect. 2, we will present and elaborate on the details of the proposed algorithm, covering three key aspects: decomposition, fusion, and reconstruction. Section 3 demonstrates the fusion performance of our method through a series of fusion experiments. Finally, Sect. 4 provides the main conclusions drawn from this research.

2 Proposed Algorithm

The flowchart of our *NSMT* is shown in Fig. 2, which consists of three processes, i.e., decomposition, fusion, and reconstruction. In Fig. 2, the *iNSP*, *iMTBH*, and *iMIOE* are the inverse transformations of *NSP*, *MTBH*, and *MIOE*, respectively.

2.1 Decomposition

The decomposition process of *NSMT* is shown in Fig. 3. First, the source image $f(x, y)$ is decomposed to obtain K high-frequency detail images HF_i and one low-frequency

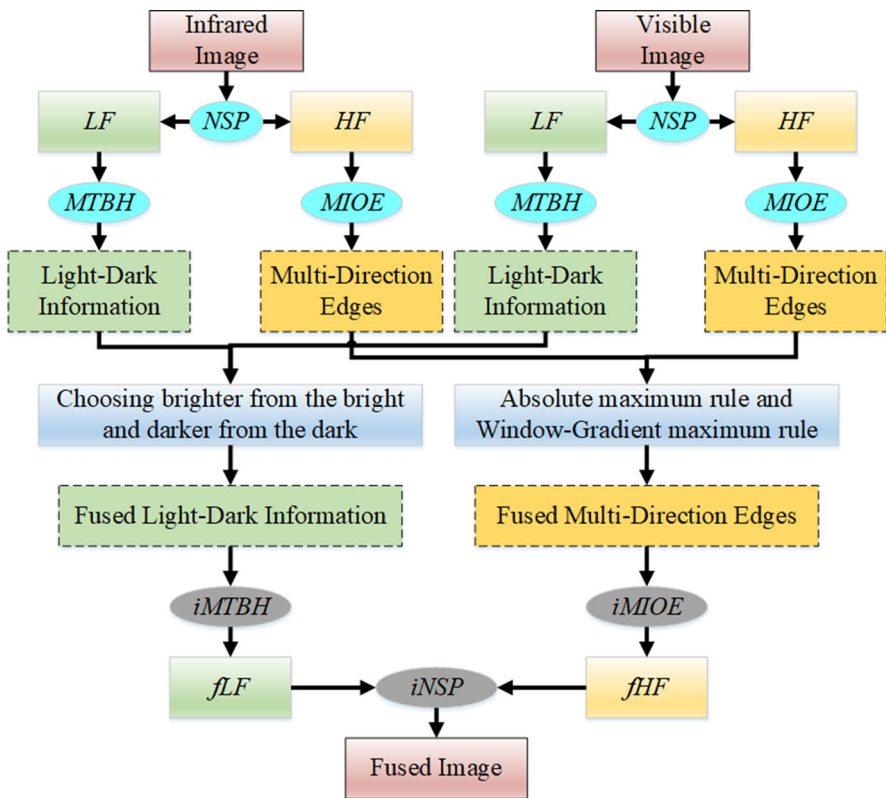


Fig. 2 The flowchart of our *NSMT*

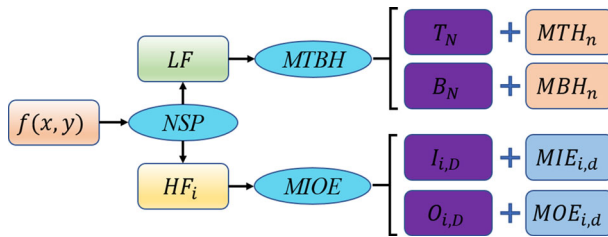


Fig. 3 Decomposition schema of our NSMT

background image LF by K -level NSP . Second, the LF image is decomposed to obtain the bright-dark smooth image T_N - B_N and the bright-dark detail images TH_n - BH_n by N -level $MTBH$. For the high-frequency detail image HF_i , we can get the smooth inner-outer edge images $I_{i,D}$ - $O_{i,D}$ and multi-direction inner-outer edge detail images $IE_{i,d}$ - $OE_{i,d}$ by D -direction $MIOE$.

2.1.1 NSP Decomposition

To obtain shift-invariant property, Zhou et al. realized the NSP decomposition by constructing dual-channel 2-D non-subsampled filter banks [43]. When the decomposition filter banks $\{H_0(z), G_0(z)\}$ and the reconstruction filter banks $\{H_1(z), G_1(z)\}$ satisfy Bezout equation, such as Eq. (1), the decomposition will satisfy the condition of complete reconstruction.

$$H_0(z) \cdot H_1(z) + G_0(z) \cdot G_1(z) = 1 \tag{1}$$

Moreover, to realize multi-scale decomposition, each stage needs to sample the previous filter according to the sampling matrix $D = 2I$ (I represent the second-order unit matrix). The decomposition process is shown in Eq. (2).

$$\begin{cases} f_0(x, y) = F \\ f_i(x, y) = H_0(z^{2^{i-1}I}) * f_{i-1}(x, y) \\ h_i(x, y) = G_0(z^{2^{i-1}I}) * f_{i-1}(x, y) \end{cases} \tag{2}$$

where F represents the source image; $H_0(*)$ and $G_0(*)$ represent the low-pass and high-pass filters after up-sampled, respectively; $f_i(x, y)$ and $h_i(x, y)$ represent the low- and high-frequency image in the i -th layer.

An image can be decomposed into $K + 1$ sub-images (including one low-frequency image and K high-frequency images) by K -level NSP decomposition. Taking image “Kaptein” as an example, the decomposed results by 4-level NSP as shown in Fig. 4.

According to the decomposed results, it is easy to find that the edges and lines are mainly concentrated in high-frequency images, and their fineness increases with the layer. In addition, the details in the LF image are very blurred, in which the main information is the bright-dark background information. This result provides practical

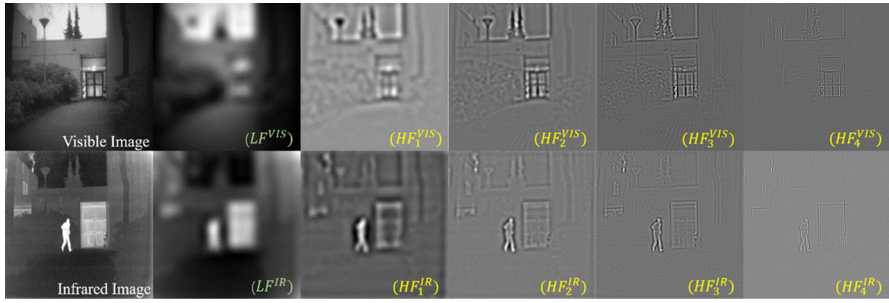


Fig. 4 The decomposed results of the “Kaptein” image by 4-level NSP

feasibility for extracting multi-direction edges from HF_i and bright-dark information from LF .

For reconstruction, high-frequency and low-frequency reconstructed filters are used to filter high- and low-frequency images, respectively. The reconstruction of the i -th layer is shown in Eq. (3).

$$f_i(x, y) = H_1(z^{2^i I})f_{i+1}(x, y) + G_1(z^{2^i I})h_{i+1} \tag{3}$$

2.1.2 MTBH Decomposition

Morphological top-hat (TH) and bottom-hat (BH) can be defined based on morphological $Open$ and $Close$. Their definitions are shown in Eq. (4). The TH can extract patches that are brighter than neighbors, namely bright details. And, the function of BH is to extract dark details [44].

$$\begin{cases} TH(F) = F - Open(F) \\ BH(F) = Close(F) - F \end{cases} \tag{4}$$

Due to the TH and BH decomposition at a single scale cannot extract the bright-dark details adequately. Multi-scale $Open$ and $Close$ are defined, respectively, in Eq. (5).

$$\begin{cases} OPen_n(F) = F \circ nS = (F \ominus nS) \oplus nS \\ Close_n(F) = F \cdot nS = (F \oplus nS) - \ominus nS \end{cases} \tag{5}$$

where $n = 1, 2, \dots, N$ is the scale factor, which is an integer and determines the size of structuring elements, \oplus and \ominus represent $Dilate$ and $Erode$, respectively. In this paper, the shape of the structuring element S is “disk.” The nS represents the n -th structuring element whose size is determined by $2n + 1$. So, based on the pyramid idea, the multi-scale top-hat MTH decomposition can be defined such as Eq. (6).

$$\begin{cases} T_0(x, y) = F \\ T_n(x, y) = \text{Open}_n(T_{n-1}(x, y)) \\ \text{MTH}_n(x, y) = T_{n-1}(x, y) - T_n(x, y) \end{cases} \quad (6)$$

where F is a source image, $\text{MTH}_n(x, y)$, $n = 1, 2, \dots, N$ represent a set of bright detail images. In addition, the last layer image is a smooth bright background image, which is represented by $T_N(x, y)$. By a similar process, we can obtain the decomposition of multi-scale bottom-hat MBH shown in Eq. (7).

$$\begin{cases} B_0(x, y) = F \\ B_n(x, y) = \text{Close}_n(B_{n-1}(x, y)) \\ \text{MBH}_n(x, y) = B_n(x, y) - B_{n-1}(x, y) \end{cases} \quad (7)$$

In Eq. (7), $\text{MBH}_n(x, y)$ represents the dark detail image and the smooth dark background image is $B_N(x, y)$.

Figure 5 shows the decomposed results of the LF image from Fig. 4 by 5-level MTBH . As can be seen from Fig. 5, the high-layer image contains bright-dark details, and the fineness of bright-dark information increases with the number of layers. This is mainly because MTBH only compares with the pixels around the structuring element when extracting the bright-dark information. Therefore, bright-dark details will appear in high-layer images, while large areas with the bright-dark feature will be captured by low-layer images or directly retained in smooth images as backgrounds. Taking the pedestrian in the infrared image as an example, the pedestrian area has not been extracted in the high-layer bright detail image MTH_4^{IR} or MTH_5^{IR} . However, in the low-layer image MTH_1^{IR} , the pedestrian area has been successfully extracted. Moreover, the sky in the visible image is retained as the background in the smooth

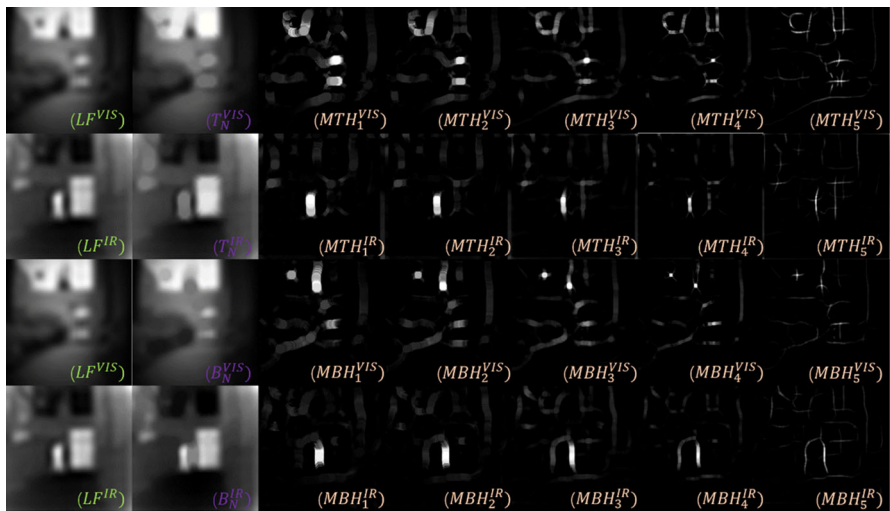


Fig. 5 The decomposed results of LF image by 5-levels MTBH

image T_N^{VIS} (which is the background information after removing the bright-dark details) due to its too large area. In addition, it should be noted that in the dark detail image $MBH_n(x, y)$, even the dark information is displayed with bright pixels.

According to Eq. (6) and (7), the source image F can be perfectly reconstructed by MTH or MBH decomposition, respectively, as shown in Eq. (8).

$$\begin{cases} F = T_0(x, y) = T_N(x, y) + \sum_{n=1}^N MTH_n(x, y) \\ F = B_0(x, y) = B_N(x, y) - \sum_{n=1}^N MBH_n(x, y) \end{cases} \quad (8)$$

2.1.3 MIOE Decomposition

Considering the directionality of edges, a series of 7-pixels “line” structuring element S_d with different directions are used to construct multi-direction inner-edge (MIE) and outer-edge (MOE) to extract the edge details in different directions. The definition is shown in Eq. (9).

$$\begin{cases} MIE_d = F - F \ominus S_d \\ MOE_d = F \oplus S_d - F \end{cases} \quad (9)$$

where the d represents the direction number; the F is the source image. Moreover, the inner-outer edge smooth image I_D and O_D are shown in Eq. (10).

$$\begin{cases} I_D = \frac{1}{D} \sum_{d=1}^D (F \ominus S_d) \\ O_D = \frac{1}{D} \sum_{d=1}^D (F \oplus S_d) \end{cases} \quad (10)$$

Taking the HF_3 as the sample, the 4-direction decomposition is shown in Fig. 6. We can easily find that the edges in the decomposed images have significant directionality from $MIE_{3,1}$ to $MIE_{3,4}$ (or from $MOE_{3,1}$ to $MOE_{3,4}$). It should be noted that the fineness of these multi-direction edge images is consistent, and the difference is only in the direction of the edges.

According to Eq. (9) and (10), the reconstruction process can be easily deduced from MIE or MOE , respectively, as shown in Eq. (11).

$$\begin{cases} IE^R = \frac{1}{D} \sum_{d=1}^D (I_D + MIE_d) \\ OE^R = \frac{1}{D} \sum_{d=1}^D (O_D - MOE_d) \end{cases} \quad (11)$$

where D represents the total number of directions, and IE^R and OE^R represent reconstructed images based on MIE and MOE , respectively.



Fig. 6 The 4-directions edge detail images of HF_3

2.2 Fusion

To make full use of the extracted information, our *NSMT* adopts different fusion rules for the decomposed results from *MTBH* and *MIOE*.

2.2.1 Fusion for *MTBH*

There is a fact that high contrast is more conducive to highlighting image details. So, for the fusion of bright-dark information (including the smooth layer and detailed layer), the rule of choosing the brighter from the bright and darker from the dark is proposed to improve the image contrast. The technique of fusion can be formulated as Eq. (12).

$$\begin{cases} fT = \max(T^{VIS}, T^{IR}) \\ fB = \min(B^{VIS}, B^{IR}) \end{cases} \quad (12)$$

where fT and fB represent the fused bright and dark images; T and B represent the extracted bright and dark images.

2.2.2 Fusion for MIOE

For the edge smooth image $I_{i,D}$ and $O_{i,D}$, the rule of choosing absolute maximum is adopted such as Eq. (13).

$$fIO_{i,D}(x, y) = \begin{cases} IO_{i,D}^{VIS}(x, y), & |IO_{i,D}^{VIS}(x, y)| > |IO_{i,D}^{IR}(x, y)| \\ IO_{i,D}^{IR}(x, y), & \text{otherwise} \end{cases} \quad (13)$$

where $fIO_{i,D}(x, y)$ represents the fused inner-outer edge smooth image; $IO_{i,D}^{VIS}(x, y)$ and $IO_{i,D}^{IR}(x, y)$ represent the extracted infrared and visible inner-outer edge smooth image.

In addition, the window-gradient maximum rule is applied to fuse the inner-outer edge detailed image $MIE_{i,d}$ and $MOE_{i,d}$. In our NSMT, the size of the window is 3×3 , and the fusion rule is shown in Eq. (14). The $WG(x, y)$ is the window gradient of a specific pixel (x, y) .

$$fMIOE_{i,d}(x, y) = \begin{cases} MIOE_{i,d}^{VIS}(x, y), & WG_{VIS}(x, y) > WG_{IR}(x, y) \\ MIOE_{i,d}^{IR}(x, y), & \text{otherwise} \end{cases} \quad (14)$$

where $WG(x, y) = \sqrt{G_x^2 + G_y^2}$, in which the G_x and G_y represent horizontal and vertical gradients, and their definitions are shown in Eq. (15).

$$\begin{cases} G_x = \sum_{x=1}^3 \sum_{y=1}^3 [f(x, y) - f(x, y-1)] \\ G_y = \sum_{x=1}^3 \sum_{y=1}^3 [f(x, y) - f(x-1, y)] \end{cases} \quad (15)$$

2.3 Reconstruction

2.3.1 Reconstruction of Background Image fLF

In the process of reconstructing the LF image, we hope to ensure the integrity of bright-dark information and flexibly adjust the image contrast by using the proportion of bright-dark components. So, the smooth images are synthesized by the average method, and the bright-dark detail images by the weighted method. According to Eq. (8), the reconstruction process of fLF can be expressed as shown in Eq. (16),

$$fLF = (fT_N + fB_N) / \left(2.0 + \alpha \sum_{n=1}^N fMTH_n - (1 - \alpha) \sum_{n=1}^N fMBH_n \right) \quad (16)$$

where fLF is the low-frequency background image LF of the fused image; α is a weight factor to adjust the proportion of bright-dark details. In this paper, $\alpha = 0.5$ without any prior knowledge.

2.3.2 Reconstruction of Edge Detail Images fHF_i

As we all know, the inner and outer edges are also equally important. So, the average method is used to synthesize the fHF_i as shown in Eq. (17),

$$fHF_i = \frac{fIE_i^R + fOE_i^R}{2} \quad (17)$$

where the fIE_i^R and fOE_i^R represent the reconstructed inner and outer fused image according to Eq. (11).

2.3.3 Reconstruction of Fusion Image

According to Eq. (3), the $f_0(x, y)$ can be reconstructed by iterating fLF and fHF_i layer by layer. Then, the fusion image F can be obtained directly according to $F = f_0(x, y)$.

3 Experiments and Discussions

To explore the fusion performance of our *NSMT*, some state-of-the-art methods are selected as comparison objects such as *FPDE* [5], *HMSD* [42], *JSR* [24], *LATLRR* [22], *MSVD* [30], *VSMWLS* [28], and some deep-learning-based methods (including *DLF* [18], *DenseFuse* [19], and *ResNet* [20]). In this paper, about our *NSMT*, there are five sub-images, i.e., one low-frequency background image LF and four high-frequency detail images HF_i ($i = 1, 2, 3, 4$) by 4-level *NSP* decomposition. In addition, to extract bright-dark information as much as possible from the LF image, the number of decomposition levels of *MTBH* is set to 5 ($N = 5$). For the HF_i image, the numbers of directions of *MIOE* from $i = 1$ to 4 are set to 2, 4, 8, 16 (namely $D = 2, 4, 8, 16$), respectively.

Ten pairs of typical infrared–visible images have been selected as fusion objects from the public database, and all images have been pre-registered. These images are often selected as fusion objects in many infrared–visible image fusion studies because they are very representative and contain common objects such as buildings, trees, roads, pedestrians, and grasslands. The source images and fusion results are shown in Fig. 7. The first two rows are original infrared and visible source images, the last row is the fusion results of our *NSMT*, and the rest nine rows correspond to the fusion results of nine comparison methods, namely *DLF*, *DenseFuse*, *FPDE*, *HMSD*, *JSR*, *LATLRR*, *MSVD*, *ResNet*, and *VSMWLS*. According to the fusion results, we can find that all methods can fuse infrared and visible images successfully.

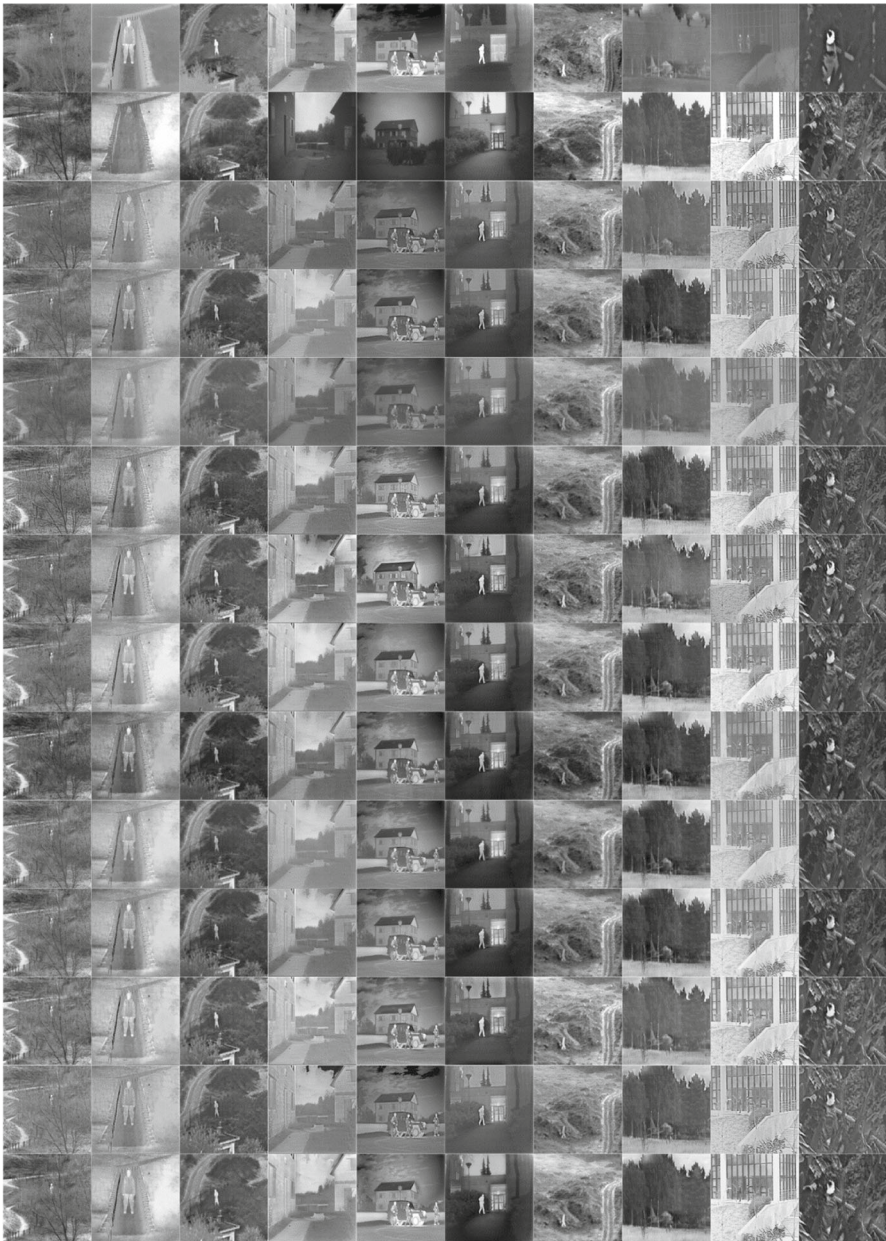


Fig. 7 Fusion results of ten pairs of infrared–visible images. From top to bottom: infrared images, visible images, fusion results of *DLF*, *DenseFuse*, *FPDE*, *HMSD*, *JSR*, *LATLRR*, *MSVD*, *ResNet*, *VSMWLS* our *NSMT*

3.1 Subjective Evaluation

Due to the limitation of page size, it is very difficult to judge which one is the best or worst according to the above figure. To show their difference more intuitively, taking the first column images as an example, the regions in red rectangles were selected as observation objects as shown in Fig. 8.

According to selected regions, it is not difficult to find that the results of our *NSMT* and *FPDE* are better than that of other methods in terms of sharpness, and they can show clearer outline of the branches. As we all know, the edge details are mainly from the visible image. Therefore, when the visible image is taken as the reference, the fusion result of *NSMT* can better retain the information from the source image. There are two main reasons for the above phenomenon. On the one hand, the *NSMT* extracts the bright-dark information from the low-frequency background and fuses them with the rule of “choosing brighter from the bright and darker from the dark,” which can effectively preserve the bright-dark distribution from the source images. On the other hand, the multidirectional edges extracted from the high-frequency images can effectively maintain the edge integrity with the rules of absolute maximum and window-gradient maximum. The combination of the above two aspects makes the fusion image generated from *NSMT* have an excellent visual effect.

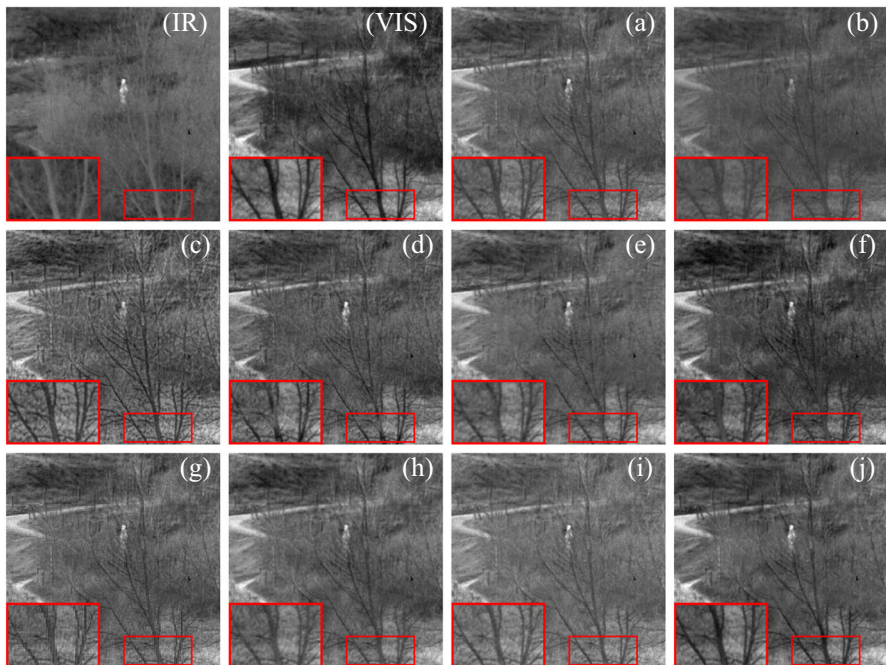


Fig. 8 The fusion images from first column. From (a) to (j) are the fusion results of *DLF*, *DenseFuse*, *FPDE*, *HMSD*, *JSR*, *LATLRR*, *MSVD*, *ResNet*, *VSMWLS*, and our *NSMT*

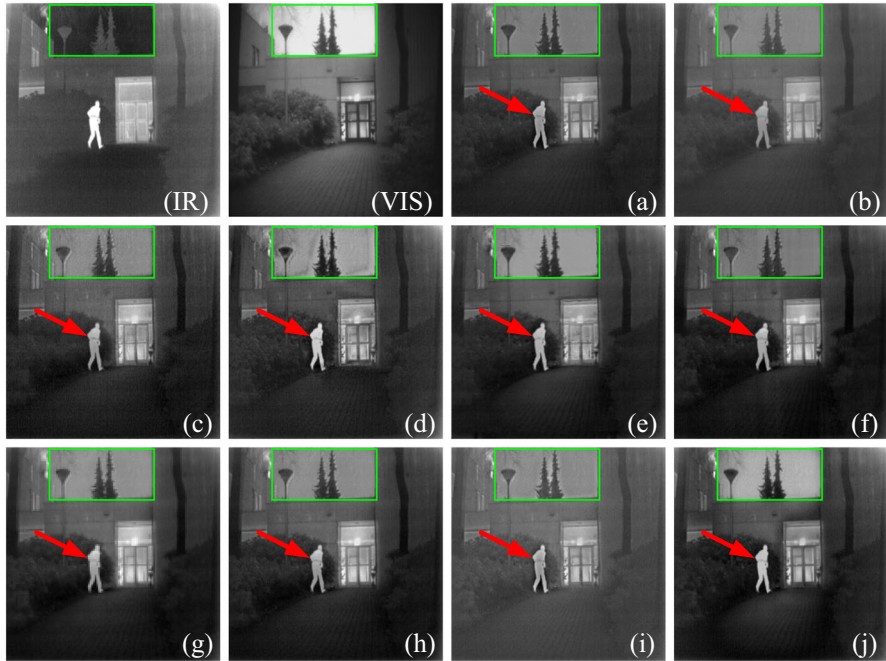


Fig. 9 The fusion images from sixth column. From (a) to (j) are the fusion results of *DLF*, *DenseFuse*, *FPDE*, *HMSD*, *JSR*, *LATLRR*, *MSVD*, *ResNet*, *VSMWLS* our *NSMT*

In order to further explore the performance of the *NSMT*, the sixth column in Fig. 7 is taken as the observation object, as shown in Fig. 9. First, as far as the pedestrian indicated by the red arrow is concerned, the pedestrian in the fusion images of *HMSD* and *NSMT* has higher brightness than that of others. However, when we focus on the green rectangular regions, *NSMT* has obvious advantages over *HMSD* in terms of sharpness and edge integrity. In addition, *NSMT* shows better performance than other methods in terms of contrast in this region. An excellent contrast can make the results generated by *NSMT* have higher sharpness and more notable targets. These results further verify the excellent performance of *NSMT* in fusing infrared and visible images.

3.2 Objective Evaluation

In practice, an ideal fusion image is often associated with specific tasks, so it is not always easy to find a universal quality evaluation method. Although the subjective evaluation can judge the fusion quality from the human perspective. But, the quality difference among fusion results is very small in most cases. So, it is difficult to give uniform and accurate results due to differences in individual mental state, visual acuity, psychological factors, and so on. To circumvent the above deficiency, many objective evaluation methods and metrics are proposed. However, there is no universally accepted method to objectively evaluate image fusion results.

In this paper, to comprehensively evaluate our *NSMT* from different perspectives, we choose eight commonly used and typical objective metrics, namely standard deviation (*SD*), information entropy (*IE*) [34], mutual information (*MI*) [33], Q_0 [36], Q_w [31], $Q^{AB/F}$ [38], Q_E [31] and *VIF* [10]. Among them, the *SD* can evaluate the image contrast from a statistical viewpoint. The *IE* can measure the information amount of the fused image from the viewpoint of information theory. The *MI* can measure the information amount that is transferred from source images to the fused image. The Q_0 can evaluate the distortion degree of a fused image, which combines three factors of image distortion related to the human visual system, namely loss of correlation, contrast distortion, and luminance distortion. The Q_w is used to evaluate the salience of information from source images. The $Q^{AB/F}$ can measure the amount of edge details transferred from source images into the fused image. The Q_E was designed by modifying Q_w , which contains edge information and visual information, simultaneously. The *VIF* can measure the visual information fidelity of the fusion image relative to the source images.

The evaluation results of all metrics based on ten pairs of typical infrared–visible images are shown in Fig. 10. It can be found from the above figure that the *NSMT* has generated the best results in most cases on the other six metrics except Q_0 and *VIF*. According to *SD*, we can find that our *NSMT* is significantly ahead of other algorithms in the ninth pair of images. This is because the brightness difference between the source images of the ninth pair is very obvious. About the ninth pair of images, we can see that the infrared image is very dark on the whole, while the overall brightness of the visible image is high. Therefore, it can be judged that our *NSMT* can make good use of the bright–dark components to synthesize high-contrast fused images when the brightness difference between the source images is large. Additionally, we note that in terms of *MI*, the fusion results of *NSMT* have significant advantages in most images. This demonstrates that the *NSMT* can retain the information from the source image better than other algorithms. This conclusion can also be drawn from the performance of *NSMT* on metric $Q^{AB/F}$. Moreover, the superiority of *NSMT* in metric Q_E further proves its excellent edge extraction ability. As it should be, for Q_0 and *VIF*, our algorithm also shows good fusion performance, although not obvious.

To show the fusion performance of our algorithm as a whole, the average results of all images for each metric are shown in Table 1, in which the best results have been marked in bold values. According to this table, we can intuitively see that our algorithm is ahead of these state-of-the-art algorithms in almost all metrics (not including Q_0 and *VIF*), especially in image contrast (by *SD*), and mutual information (by *MI*). Although the performance of *NSMT* is mediocre in metrics Q_0 and *VIF*, it still shows satisfactory competitiveness.

3.3 Comparison of Fusion Time

In this section, the fusion processing speed of our method will be shown by fusing the ten pairs of images from Fig. 7. To ensure a more reasonable comparison of fusion speeds across all methods, we exclude the worst and best fusion results. The mean fusion processing time of the remaining eight results is then calculated and presented

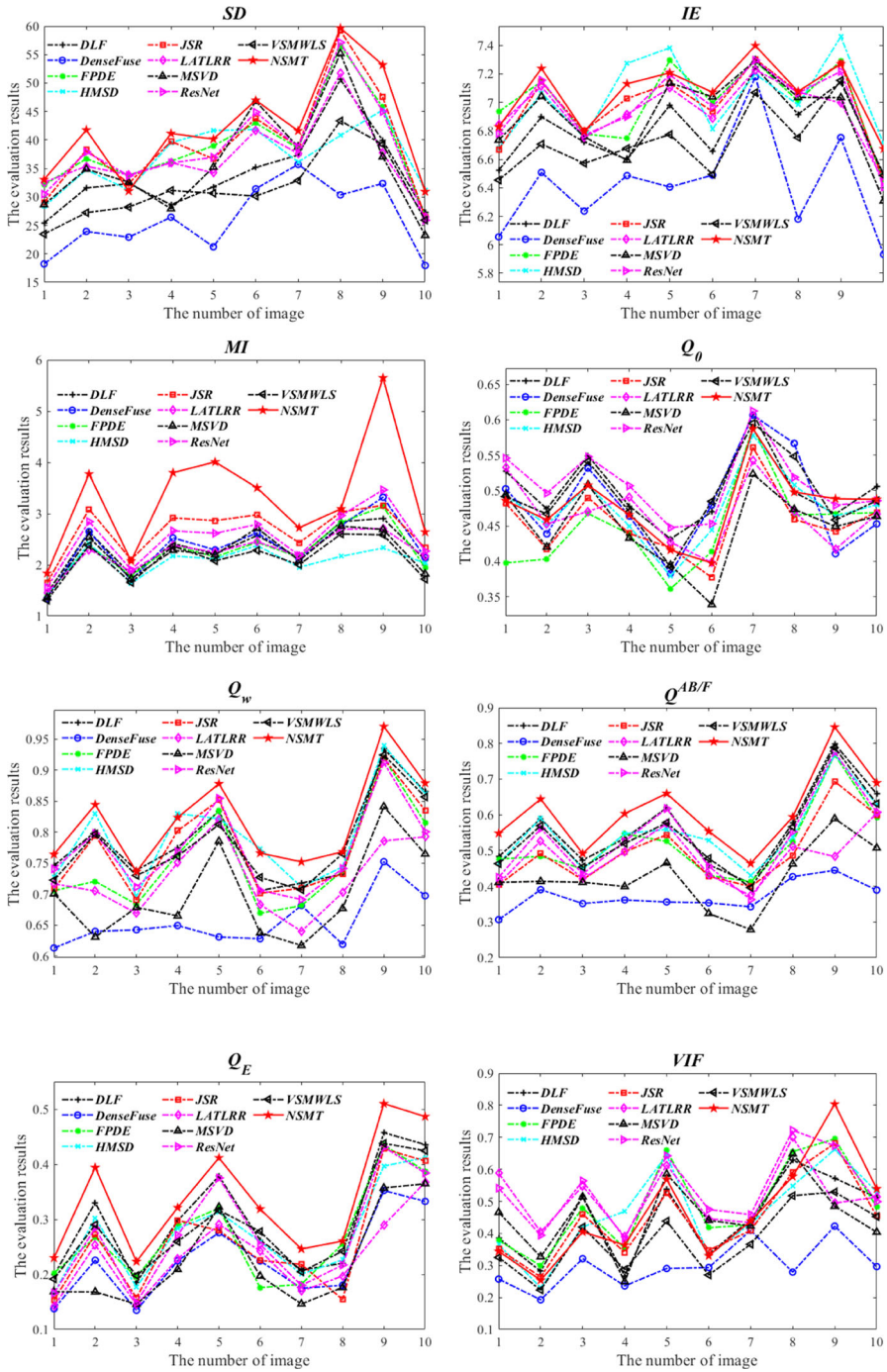
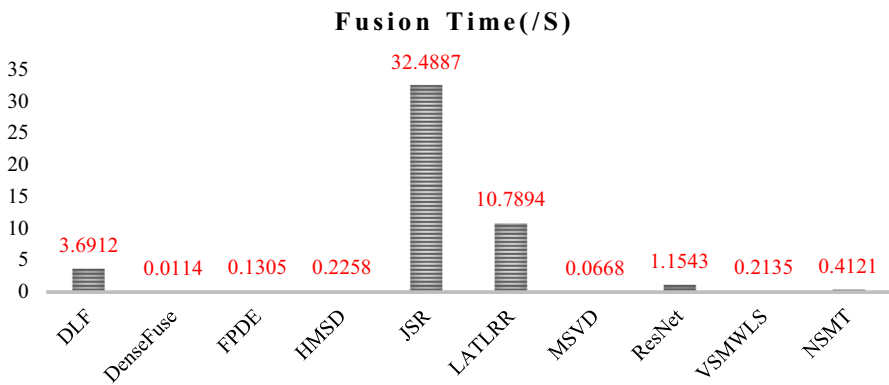


Fig. 10 Objective comparisons of the eight metrics

Table 1 The average results of each metric

	SD	IE	MI	Q_0	Q_w	$Q^{AB/F}$	Q_E	VIF
<i>DLF</i>	34.0134	6.8175	2.3082	0.5010	0.7852	0.5601	0.2974	0.4480
<i>DenseFuse</i>	26.0659	6.4221	2.3467	0.4837	0.6556	0.3724	0.2263	0.2999
<i>FPDE</i>	38.8473	6.9924	2.2623	0.4474	0.7534	0.5210	0.2709	0.4847
<i>HMSD</i>	37.0820	7.0319	2.0606	0.4756	0.7940	0.5503	0.2737	0.4631
<i>JSR</i>	39.2066	6.9805	2.6548	0.4558	0.7749	0.4962	0.2594	0.4425
<i>LATLRR</i>	36.7409	6.9359	2.2455	0.4669	0.7268	0.4832	0.2325	0.5131
<i>MSVD</i>	36.0400	6.8973	2.2177	0.4499	0.7000	0.4263	0.2256	0.4543
<i>ResNet</i>	38.6827	6.9756	2.5255	0.5094	0.7727	0.5318	0.2712	0.5354
<i>VSMWLS</i>	31.2337	6.7139	2.0985	0.4935	0.7806	0.5445	0.2843	0.3829
<i>NSMT</i>	41.9540	7.0686	3.3103	0.4793	0.8185	0.6094	0.3403	0.4624

Bold values indicate the best results

**Fig. 11** The fusion processing time (unit is second) of all methods

in Fig. 11. The experimental platform employed in this study is Windows 10, with an AMD 3900 × CPU and 32GB of memory. As depicted in Fig. 11, it is evident that *DenseFuse* exhibits the fastest fusion speed, while *JSR* has the slowest. Although our *NSMT* does not have a significant advantage over *DenseFuse*, its fusion speed remains competitive.

4 Fusion of Noisy Image

It is well known that infrared thermal radiation imaging is easily polluted by environmental thermal noise. Ref [25] points out that morphological operation has certain denoising ability. To test this conjecture, the white Gaussian noise with different degrees is added to infrared source images. In this paper, the mean value of white Gaussian noise is 0, and the variance is 5, 10, 15, and 20, respectively. In general,

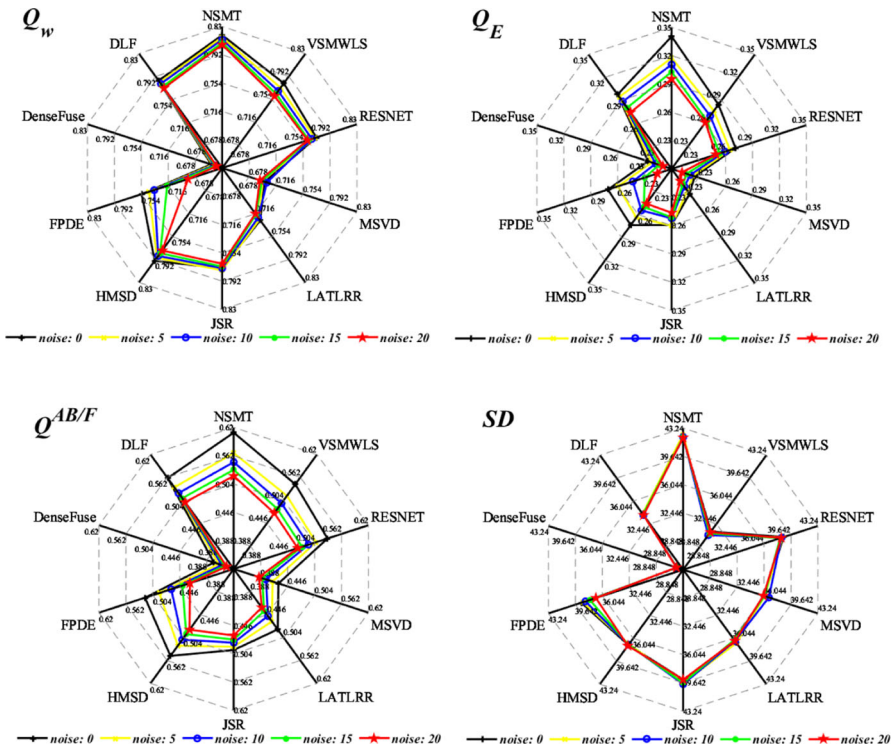


Fig. 12 Objective comparisons of the four metrics for noisy image fusion

noise will blur image details, such as edges or lines. In addition, the addition of Gaussian white noise will affect the brightness distribution of the image to a certain extent, thus changing the contrast of the original image. Therefore, for the evaluation of noisy image fusion, four evaluation metrics are selected, namely Q_w , Q_E , $Q^{AB/F}$, and SD . The evaluation results of four metrics based on ten pairs of typical infrared–visible images are shown in Fig. 12. The axis of the radar chart represents the fusion algorithms.

According to Fig. 12, we can draw some very useful conclusions. First, the noise in the infrared image indeed degrades the image quality of fusion results. Second, the noise has a significant impact on salient features (such as edges and lines). And, the salient features decreases as the noise level increases, which can be drawn from the metrics Q_w , Q_E , and $Q^{AB/F}$. Third, the noise has a relatively small effect on contrast (according to SD), especially when the noise intensity is not high, this effect can be ignored. However, when the noise increases to a certain level, the image contrast will decrease significantly. Even so, the results of SD show that the contrast of the fused images from *NSMT*, *DLF*, *DenseFuse*, *HMSD*, *LATLRR*, and *RESNET* are almost unaffected by noise. Fourth and most importantly, for image fusion with noise, our *NSMT* performs the best among all algorithms regardless of any metric, and its performance is significantly superior to other algorithms. In summary, our *NSMT* can

not only successfully fuse noisy images but also effectively reduce the impact of noise on the fusion results and produce very satisfactory results.

5 Conclusions

In this paper, we propose a new image fusion algorithm called *NSMT*. Firstly, the *NSP* is used to decompose the source image to obtain the low- and high-frequency components, and then the multi-scale *MTBH* and multi-direction *MIOE* are constructed to furtherly extract the bright-dark and edge details. In the decomposition process, the *MTBH* decomposition can efficiently extract the bright-dark details from the low-frequency background image. Meanwhile, for the high-frequency detail images, the *MIOE* decomposition can extract the edge details in different directions well. In the fusion process, the rule of “choosing brighter from the bright and darker from the dark” can effectively fuse the bright-dark information to improve the image contrast and show better visual effects. Moreover, the rules of choosing absolute maximum and window-gradient maximum can retain and highlight the multi-direction edge from the high-frequency detail images well. According to the fusion experiments, we find that the proposed method has absolute superiority in fusion quality compared with those state-of-the-art methods. Our *NSMT* can not only generate fused images with high contrast but also preserve the information from source images well. In addition, our *NSMT* has shown its excellent performance in extracting and retaining the edge information of visible images from both subjective and objective aspects. In summary, it can be concluded that the proposed method in this paper is a very competitive fusion algorithm for infrared–visible images.

Acknowledgements We sincerely thank the reviewers and editors for carefully checking our manuscript. This work is supported by the Scientific Research Foundation of the Education Department of Anhui Province (No. 2022AH050801), Scientific Research Fund for Young Teachers of Anhui University of Science and Technology (No. QNZD2021-02), Anhui Provincial Natural Science Foundation (No. 2208085ME128), Scientific Research Fund of Anhui University of Science and Technology (No. 13210679), Huainan Science and Technology Planning Project (No. 2021005).

Data availability All experimental images can be obtained through the public dataset “Toet A. TNO Image fusion dataset” https://figshare.com/articles/TN_Image_Fusion_Dataset/1008029. In addition, the experimental fusion images will be made available upon reasonable request for academic use and within the limitations of the provided informed consent by the corresponding author upon acceptance.

Declarations

Conflict of interest We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

1. M. Arif, G. Wang, Fast curvelet transform through genetic algorithm for multimodal medical image fusion. *Soft. Comput.* **24**(3), 1815–1836 (2020)

2. A. Averbuch, P. Neittaanmäki, V. Zheludev, M. Salhov, J. Hauser, Image inpainting using directional wavelet packets originating from polynomial splines. *Signal Process. Image Communication* **97**, 116334 (2021)
3. A. Averbuch, P. Neittaanmäki, V. Zheludev, M. Salhov, J. Hauser, Coupling BM3D with directional wavelet packets for image Denoising. *arXiv preprint arXiv:2008.11595* (2020)
4. R.H. Bamberger, M.J. Smith, A filter bank for the directional decomposition of images: theory and design. *IEEE Trans. Signal Process.* **40**(4), 882–893 (1992)
5. D. P. Bavisetti, G. Xiao, G. Liu, Multi-sensor image fusion based on fourth order partial differential equations. In *2017 20th International conference on information fusion (Fusion)* (2017). pp. 1–9
6. E. Candes, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms. *Multisc. Model. Simulat.* **5**(3), 861–899 (2006)
7. J. Chen, X. Li, L. Luo, X. Mei, J. Ma, Infrared and visible image fusion based on target-enhanced multiscale transform decomposition. *Inf. Sci.* **508**, 64–78 (2020)
8. B. Cheng, L. Jin, G. Li, Infrared and visual image fusion using LNSST and an adaptive dual-channel PCNN with triple-linking strength. *Neurocomputing* **310**, 135–147 (2018)
9. T. Deepika, Analysis and comparison of different wavelet transform methods using benchmarks for image fusion. *arXiv preprint arXiv:2007.11488* (2020)
10. Y. Han, Y. Cai, Y. Cao, X. Xu, A new image fusion performance metric based on visual information fidelity. *Information Fusion* **14**(2), 127–135 (2013)
11. P. Hu, F. Yang, H. Wei, L. Ji, X. Wang, Research on constructing difference-features to guide the fusion of dual-modal infrared images. *Infrared Phys. Technol.* **102**, 102994 (2019)
12. P. Hu, F. Yang, L. Ji, Z. Li, H. Wei, An efficient fusion algorithm based on hybrid multiscale decomposition for infrared-visible and multi-type images. *Infrared Phys. Technol.* **112**, 103601 (2021)
13. M. Kumar, N. Ranjan, B. Chourasia, Hybrid methods of contourlet transform and particle swarm optimization for multimodal medical image fusion. In *2021 international conference on artificial intelligence and smart systems (ICAIS)* (2021), pp. 945–951
14. G. Kutyniok, D. Labate, Construction of regular and irregular shearlet frames. *J. Wavelet Theory Appl* **1**(1), 1–12 (2007)
15. H.J. Kwon, S.H. Lee, Visible and near-infrared image acquisition and fusion for night surveillance. *Chemosensors* **9**(4), 75 (2021)
16. S. Li, X. Kang, L. Fang, J. Hu, H. Yin, Pixel-level image fusion: a survey of the state of the art. *Inform. Fusion* **33**, 100–112 (2017)
17. H. Li, L. Liu, W. Huang, C. Yue, An improved fusion algorithm for infrared and visible images based on multi-scale transform. *Infrared Phys. Technol.* **74**, 28–37 (2016)
18. H. Li, X. J. Wu, J. Kittler, Infrared and visible image fusion using a deep learning framework. In *2018 24th international conference on pattern recognition (ICPR)* (2018), pp. 2705–2710
19. H. Li, X.J. Wu, DenseFuse: a fusion approach to infrared and visible images. *IEEE Trans. Image Process.* **28**(5), 2614–2623 (2018)
20. H. Li, X.J. Wu, T.S. Durrani, Infrared and visible image fusion with ResNet and zero-phase component analysis. *Infrared Phys. Technol.* **102**, 103039 (2019)
21. J. Li, H. Huo, C. Li, R. Wang, Q. Feng, AttentionFGAN: Infrared and visible image fusion using attention-based generative adversarial networks. *IEEE Trans. Multimedia* **23**, 1383–1396 (2020)
22. H. Li, X. J. Wu, Infrared and visible image fusion using latent low-rank representation. *arXiv preprint arXiv:1804.08992* (2018)
23. H. Liu, G.F. Xiao, Y.L. Tan, C.J. Ouyang, Multi-source remote sensing image registration based on contourlet transform and multiple feature fusion. *Int. J. Autom. Comput.* **16**, 575–588 (2019)
24. C.H. Liu, Y. Qi, W.R. Ding, Infrared and visible image fusion method based on saliency detection in sparse domain. *Infrared Phys. Technol.* **83**, 94–102 (2017)
25. J. Liu, K. Tian, H. Xiong, Y. Zheng, Fast denoising of multi-channel transcranial magnetic stimulation signal based on improved generalized mathematical morphological filtering. *Biomed. Signal Process. Control* **72**, 103348 (2022)
26. J. Ma, Y. Ma, C. Li, Infrared and visible image fusion methods and applications: a survey. *Information fusion* **45**, 153–178 (2019)
27. J. Ma, W. Yu, P. Liang, C. Li, J. Jiang, FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inform Fusion* **48**, 11–26 (2019)
28. J. Ma, Z. Zhou, B. Wang, H. Zong, Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys. Technol.* **82**, 8–17 (2017)

29. S. Minghui, L. Lu, P. Yuanxi, J. Tian, L. Jun, Infrared & visible images fusion based on redundant directional lifting-based wavelet and saliency detection. *Infrared Phys. Technol.* **101**, 45–55 (2019)
30. V.P.S. Naidu, Image fusion technique using multi-resolution singular value decomposition. *Def. Sci. J.* **61**(5), 479 (2011)
31. G. Piella, H. Heijmans, A new quality metric for image fusion. In *proceedings 2003 international conference on image processing (Cat. No. 03CH37429)* (2003) Vol. 3, pp. III-173
32. G. Qi, M. Zheng, Z. Zhu, R. Yuan, A DT-CWT-based infrared-visible image fusion method for smart city. *Int. J. Simul. Process Model.* **14**(6), 559–570 (2019)
33. G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion. *Electron. Lett.* **38**(7), 1 (2002)
34. J.W. Roberts, J.A. Van Aardt, F.B. Ahmed, Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *J. Appl. Remote Sens.* **2**(1), 023522 (2008)
35. X. Wang, K. Zhang, J. Yan, M. Xing, D. Yang, Infrared image complexity metric for automatic target recognition based on neural network and traditional approach fusion. *Arab. J. Sci. Eng.* **45**, 3245–3255 (2020)
36. Z. Wang, A.C. Bovik, A universal image quality index. *IEEE Signal Process. Lett.* **9**(3), 81–84 (2002)
37. H. Xu, D. Xu, S. Chen, W. Ma, Z. Shi, Rapid determination of soil class based on visible-near infrared, mid-infrared spectroscopy and data fusion. *Remote Sens.* **12**(9), 1512 (2020)
38. C.S. Xydeas, V. Petrovic, Objective image fusion performance measure. *Electron. Lett.* **36**(4), 308–309 (2000)
39. X. Zhang, P. Ye, G. Xiao, VIFB: A visible and infrared image fusion benchmark. In *proceedings of the IEEE/CVF Conference on computer vision and pattern recognition workshops* (2020) pp. 104–105
40. C. Zhao, Y. Huang, Infrared and visible image fusion method based on rolling guidance filter and NSST. *Int. J. Wavelets Multiresolut. Inf. Process.* **17**(06), 1950045 (2019)
41. J. Zhou, W. Li, P. Zhang, J. Luo, S. Li, J. Zhao, Infrared and visible image fusion method based on NSST and guided filtering. *Optoelectr. Sci. Mater.* **11606**, 82–88 (2020)
42. Z. Zhou, B. Wang, S. Li, M. Dong, Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters. *Inform. Fusion* **30**, 15–26 (2016)
43. J. Zhou, A. L. Cunha, M. N. Do, Nonsubsampled contourlet transform: construction and application in enhancement. In *IEEE international conference on image processing 2005* (2005) 1: I-469
44. P. Zhu, Z. Huang, A fusion method for infrared-visible image and infrared-polarization image based on multi-scale center-surround top-hat transform. *Opt. Rev.* **24**, 370–382 (2017)
45. K. Zhuo, Y. HaiTao, Z. FengJie, L. Yang, J. Qi, W. JinYu, Research on multi-focal image fusion based on wavelet transform. *J. Phys. Conf. Ser.* **1994**(1), 012018 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.