IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 17, 2020, accepted September 28, 2020, date of publication October 1, 2020, date of current version October 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3028088

An Infrared and Visible Image Fusion Algorithm Based on LSWT-NSST

LI JUNWU[1], (Graduate Student Member, IEEE), BINHUA LI[1,2], AND YAOXI JIANG[1]

[1]Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China
[2]Key Laboratory of Applications of Computer Technologies of the Yunnan Province, Kunming University of Science and Technology, Kunming 650500, China

Corresponding author: Binhua Li (lbh@kust.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 11673009.

ABSTRACT Regarding the problems of image distortion, edge blurring, Gibbs phenomena in the traditional wavelet transform algorithm and the loss of subtle features in the Non-Subsampled Shearlet Transform (NSST), and considering the physical characteristics of infrared and visible images, an infrared and visible image fusion algorithm based on the Lifting Stationary Wavelet Transform (LSWT) and Non-Subsampled Shearlet Transform is proposed in this paper. First, since LSWT can quickly calculate and has all advantages of traditional WT, it is utilized to decompose infrared and visible images to obtain low-frequency coefficients and multi-scale and multi-directional high-frequency coefficients, respectively. Second, NSST multi-scale decomposition is used to extract the target features and detailed features of the image from the high and low-frequency sub-bands to obtain new high and low-frequency sub-bands. Third, according to the physical characteristics that low and high-frequency coefficients represent, different fusion rules are designed. Discrete Cosine Transform (DCT) and Local Spatial Frequency (LSF) are introduced in the low-frequency sub-band, and LSF adaptive weighted fusion rules are used in the DCT domain. The fusion strategy improves the regional contrast in the high-frequency sub-band with the spectral characteristics of human vision. Finally, the Inverse Lifting Stationary Wavelet Transform (ILSWT) is used to reconstruct the fusion coefficients to obtain the final fused images. To verify the advantages of the proposed algorithm in this paper, the classic and advanced 9 IR and VI fusion algorithms are selected for subjective and objective comparison. In the objective evaluation, a comprehensive ranking index is designed based on 9 classical indicators. Simulation experiments with 10 IR and VI fusion algorithms prove that the proposed algorithm has better performance and flexibility. The results show that the proposed algorithm in this paper fuses the images with clear edges, prominent targets, and good visual perception, and it outperforms state-of-the-art image fusion algorithms.

INDEX TERMS Lifting Stationary Wavelet Transform (LSWT), Non-Subsampled Shearlet Transform (NSST), Discrete Cosine Transform (DCT), Local Spatial Frequency (LSF), regional contrast, infrared and visible image fusion.

I. INTRODUCTION

Image fusion is a technology that performs registration on images obtained by different sensors on the same target; then, it utilizes certain algorithms to remove redundant information and integrate complementary information to generate more suitable fusion images for human visual perception [1]. Recently, image fusion technology obtains a boost and plays a key role in image segmentation and computer vision [2]–[4]. With the development of image fusion technology and reduction of hardware costs, higher reliability and comprehensiveness are expected for image fusion. Therefore, infrared and visible image fusion technology has received extensive research by scholars [5].

The infrared sensor and visible sensor can collect complementary information of the same scene. Infrared sensors catch rich thermal radiation information, which can clearly unveil hidden target outlines but cannot catch detailed information. The visible sensor characterizes the object through the spectral reflection, which yields fusion results that are more closely consistent with the human visual system. However, the image quality is limited by the environment and depth

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

of field conditions, especially at night and in low-visibility conditions. The fusion of infrared and visible images provides more comprehensive information with high resolution of the visible-light target and clear hidden infrared light target. The technology overcomes the limitation of single sensor-information acquisition and improves the visual effect of the image. Thus, image fusion technology has always been a research hot-spot and plays a key role in the fields of target tracking and detection, medical imaging, military reconnaissance, remote sensing, face recognition, and space exploration [6]–[12].

Currently image fusion algorithms are mainly divided into pixel-based and region-based methods. According to the transformation range, image fusion algorithms are divided into transformation-based and spatial-domain-based methods. Multi-scale fusion based on the transform domain is the mainstream framework, whose core idea is to map the source image to multiple transform domains. Classical multi-scale transform algorithms are: Wavelet Transform (WT) [13], Curvelet Transform (CT) [14], Non-Sampled Shearlet Transform (NSST) [15], Sparse Representation (SR) [16], Non-Subsampled Contourlet Transform (NSCT) [17], multi-resolution singular-value decomposition [18], etc. Image filtering technology based on the spatial domain is an important theory, which processes a single pixel or area pixel of the source image. Typical methods are: non-local mean filtering [19], guided filtering [20], global filtering [21], and bilateral filtering. The key of region-based infrared and visible image fusion technology is to extract distinctive features in the infrared light area, which can be achieved by image segmentation or saliency detection, and regions with strong infrared rays are effectively obtained [22], [23].

In recent years, deep learning has rapidly developed in various fields and been widely studied and applied in image fusion [24], [25]. By learning the weight parameters and loss function in the training layer and verification layer, this method can obtain rich image information with good results. Literature [26] uses a convolutional sparse representation method to extract the features of the detailed layer and use it for image fusion of infrared and visible-light; it has better and more malleable fusion effect than the traditional sparse method. Literature [27] applied CNN to image fusion for the first time, and proposed a multi-focus image fusion algorithm with a deep learning framework. The deep convolutional neural network is trained to extract the focused or de-focused area of the source image, and the fused image is generated through fusion decision post-processing, which is more robust. Literature [28] uses CNN to address two key issues of image fusion: activity level measurement and weight distribution. Then, it fully consider infrared and visible-light imaging methods and local similarity post-processing strategies to adjust the fusion decision map and obtains a good fusion effect. The deep learning method requires high hardware resources and time-consuming calculations. However, with the rapid development of GPU hardware technology, parallel computing and accelerated computing capabilities have been greatly improved, and hardware costs have been greatly reduced. Traditional image fusion technology has certain advantages in handling certain types of problems. Therefore, how to effectively integrate deep learning with traditional algorithms is also a research hotspot.

Recently, deep learning has rapidly developed in various fields and been widely applied in image fusion [24], [25]. Rich image information with good effect is obtained by learning weight parameters and loss functions from the training and verification layers. However, the structures of deep learning networks are usually complicated. When there are many training layers, a lot of training and learning is very time-consuming and introduces serious phenomenon of over-fitting. Moreover, it has high requirements on hardware resources, which are not suitable for the popularization of minicomputers, poor applicability and poor real-time performance.

Fusion methods based on transform domain usually have better fusion performance, but their basic functions and decomposition scales are relatively fixed. Based on this fusion framework, selecting the optimal basis function to better express the source image and design effective fusion rules for the decomposed sub-bands to improve the fusion quality of the image is a challenging research point, and its complexity is also high. The fusion method based on the spatial domain can avoid the transformation and inverse transformation process of the transform domain method. The initial fusion decision map is usually obtained by solving the image's activity degree, but the final fusion decision map requires subsequent optimization processing. This algorithm has a small amount of calculation and is simple and easy to implement, but its fusion performance is poor. The limitation of the image fusion method based on deep learning is that a large number of images need to be trained, and it is difficult to obtain real data of these. If simulating data is used to classify pixels to obtain a fusion decision map, noise is often introduced. Therefore, it is necessary to use an end-to-end unsupervised deep network to complete image fusion and obtain high-quality fused images. However, the existing training data does not have a standard reference image, which also brings specific difficulties in end-to-end image fusion. The classification of image fusion methods is not absolute. With the continuous development of related technologies, a variety of technical methods have shown a noticeable trend of cross-fusion.

Stationary Wavelet Transform (SWT) as an improved wavelet can effectively preserve the image texture and edge information. However, this algorithm can only represent the details of the image in three directions (horizontal, vertical, and diagonal), and it is weak for continuous regions. The algorithm performance is poor especially when the source image has complex details and continuous curves. NSCT has multi-scale direction anisotropy and shift invariance, which can effectively remove Gibbs effects [29]. However, NSCT has a complicated structure and a high computation cost.

NSST has strong flexibility and multi-directionality and is more efficient than NSCT. It can also well preserve the edges and curves of the image, but it is weak for subtle features of the image.

These multi-scale fusion methods for image reconstruction have the disadvantages of large computation burden, high complexity, poor real-time performance and high requirements on memory space. Lifting Stationary Wavelet Transform (LSWT) [30] has all advantages of the traditional wavelet and better performance than SWT. It also has fast computation speed, low memory requirements, and significant local characteristics in frequency and spatial domains. The shift invariance can effectively reduce the distortion of the image, but LSWT has poor performance for continuous curves.

Image fusion strategies greatly affect the quality of the fusion image. Multi-scale transform is used to decompose the source image into low and high-frequency coefficients. The low-frequency part gathers the main energy and represents approximate information of the source image. The high-frequency part is the representation of the edge and contour details of the source image. Traditional image fusion strategies mostly obtain high and low-frequency fusion decision maps by filtering the source image or performing calculations on the decomposition coefficients, and they have achieved good fusion results. However, this method ignores the physical characteristics of the high and low-frequency sub-bands, which causes the loss of details and reduces the effect of the fused image.

Discrete Cosine Transform (DCT) can focus on the key features of the source image on a small part of the DCT coefficients, which can concentrate information and compact energy during image processing [31], [32]. Local Spatial Frequency (LSF) can effectively reflect the regional characteristics of the source image, and is often used as the key parameter and key index of the image fusion algorithm [33].

Inspired by the above discussion, and by integrating the advantages of the image multi-scale frequency domain transform and the characteristics of LSWT and NSST, this paper proposes an infrared and visible image fusion algorithm based on LSWT-NSST. The NSST algorithm is used to obtain the continuous curve and edge to compensate for the deficiency of LSWT; LSWT algorithm is used to get subtle image features to make up for the deficiency of NSST. Different image fusion rules are designed according to the physical characteristics of infrared and visible-light, and the representation of high and low-frequency sub-bands. In low-frequency sub-bands, LSF adaptive weighted fusion rules are used in the DCT domain. In high-frequency sub-bands, a fusion strategy of improving regional contrast is adopted according to the spectral characteristics of human vision. A comprehensive ranking algorithm is designed based on 9 classic indicators for objective evaluation, which greatly enhances the performance of the overall evaluation of the fusion image and decreases subjective recognition.

Compared with the 9 advanced fusion algorithms, the main contributions of the proposed algorithm based on LSWT-NSST are as follows:

(1) LSWT and NSST are classical algorithms and have optimal fusion quality. By combining the advantages of two algorithms, the effect of image fusion is greatly improved, and the efficiency based on LSWT improves the running efficiency while ensuring image quality.

(2) By combining the physical characteristics of infrared and visible-light and the representation characteristics of high and low-frequency sub-bands, we design different image fusion rules in this paper. In the low-frequency part, the LSF adaptive weighted fusion rule is employed in the DCT domain, which greatly improves the target and detail characteristics of the fused image. In the high-frequency part, combining with the visual characteristics of the human eye, an improved regional contrast fusion strategy is proposed, which is more suitable for human vision, especially in image regions with high saliency such as edge contours.

(3) In this paper, seven classic indicators are selected, and a comprehensive ranking index is designed, which comprehensively considers the ranking indices of different algorithms in terms of image gray-scale, frequency, etc. Therefore, it greatly enhances the comprehensiveness of the distribution of image indicators. In addition, more consideration is given to the macro visual effect, which decreases the artificial subjective consciousness.

(4) The algorithm of this paper has improved the indicators of image decomposition, fusion rules and index evaluation. The improvements of the three aspects are combined to increase the performance of the proposed algorithm. The performance is superior, and the image fusion effect is perfect.

The remainder of this paper is organized as follows: The second part introduces the related knowledge background and theoretical algorithm of LSWT-NSST. The third part introduces the infrared and visible image fusion algorithm based on LSWT-NSST in detail. The fourth part shows the experimental configuration and simulation and subjectively and objectively analyzes the IR and VR image fusion effects. The final part is the conclusion of this work.

II. RELATED THEORIES

A. LSWT ALGORITHM

In image processing, the multi-scale decomposition based on the transform domain is more widely used and has stronger universality and stability than the model based on the spatial domain. The traditional wavelet as a classic algorithm of multi-scale transform in the field of image processing is a non-redundant decomposition algorithm and does not have shift invariance [34]. Lifting wavelet transform (LWT) overcomes the shortcomings of traditional wavelet, no longer relies on the traditional wavelet convolution operation, and can the construct Compactly Supported Biorthogonal Multi-wavelets in the spatial domain. For image decomposition, the high-frequency component of LWT uses a simple polynomial interpolation method, and the low-frequency component uses a scale function construction method to maintain some overall characteristics of the image. Therefore, the LWT algorithm is easy to implement and has a fast calculation speed. However, LWT does not have shift invariance, and the image fusion has Gibbs effect and serious distortion.

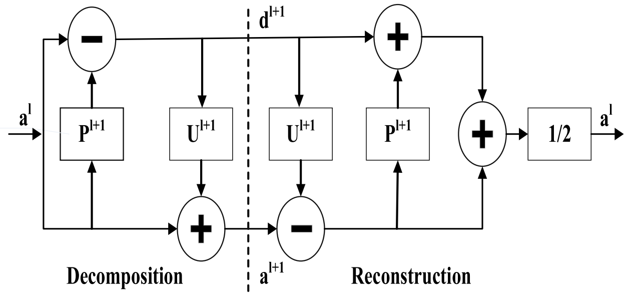


FIGURE 1. **Decomposition and reconstruction flow diagram of LSWT.**

To overcome the shortcomings of WT and second-generation LWT, this paper adopts LSWT as the multi-scale transformation algorithm. The filter extension is completed by canceling the parity-splitting steps, and the zero-filling operation of the corresponding filter coefficients of the LWT is canceled to achieve shift invariance of the LSWT. LSWT has the advantages of LWT, outstanding local characteristics in the spatial and frequency domains, and shift invariance, which can effectively avoid image distortion problems. The decomposition and reconstruction process is shown in Figure 1. where P and U represent the prediction operator and update operator, respectively. d^{l+1} and $a^{l+1}$are the low and high-frequency coefficients of input signal a^l at the $(l+1)(th)$ layer after decomposition by the LSWT algorithm. P^{l+1} and U^{l+1} are the prediction coefficients and update filter coefficients of LSWT, as defined in equations (1)-(2).

$$p^{l+1} = p_0, \underbrace{0, \cdots, 0}_{2^{l+1}-1}, p_1, \underbrace{0, \cdots, 0}_{2^{l+1}-1}, p_2, \cdots, p_{m-2}, \\ \times \underbrace{0, \cdots, 0}_{2^{l+1}-1}, p_{m-1} \tag{1}$$

$$u^{l+1} = u_0, \underbrace{0, \cdots, 0}_{2^{l+1}-1}, u_1, \underbrace{0, \cdots, 0}_{2^{l+1}-1}, u_2, \cdots, u_{n-2}, \\ \times \underbrace{0, \cdots, 0}_{2^{l+1}-1}, u_{n-1} \tag{2}$$

where $p_i(i = 0, 1, \cdots, m-1)$ and $u_j(j = 0, 1, \cdots, n-1)$ are the prediction coefficients and update filter coefficients of LWT, respectively; m and n are the numbers of coefficients of prediction operator P and update operator U, respectively.

B. NSST ALGORITHM

The Shearlet Transform inherits the advantages of WT and can realize the optimal sparse representation of the image, where multi-scale decomposition uses down-sampling pyramid filtering, and direction decomposition uses shear wave filtering by a shift window in the pseudo polar grid. The sub-sampling operation in the Multi-directional shear filter makes Shearlet Transform not have shift invariance and cause spectral aliasing.

To avoid this defect and retain the advantages of multiscale decomposition, Easley proposed the NSST algorithm [35]. NSST transform uses a non-subsampled Laplacian Pyramid (NLSP) for multiscale division. Using a two-dimensional convolution, NSST transforms the shear wave filter from a pseudo polar grid to a Cartesian system, which avoids sub-sampling and makes the NSST shift invariant. The decomposition process of NSST is shown in Figure 2.

The NSST image decomposition includes two main steps:

(1) Non-subsampled multiscale subdivision.

The source image is decomposed by the first layer of NLSP to obtain low-frequency coefficientsf_a^1 and high-frequency coefficientsf_d^1. The $(k+1)(th)$ layer NSLP decomposition is based on the low-frequency components of the $k(th)$layer. Therefore, after image f is decomposed by k layers of NLSP, one low-pass sub-band and k high-pass sub-bands are obtained.

(2) Localization of direction.

Shearlet filter is utilized to realize the direction localization of high-frequency coefficients. NLSP and Shearlet filters are used to make NSST algorithm multi-scale, multi-directional and shift invariant, which can effectively characterize the details of the source image and avoid Gibbs and eclipse phenomena.

C. DISCRETE COSINE TRANSFORM

DCT is a commonly used linear orthogonal transform in the field of image processing, whose outstanding advantage is the independent correlation of data, and it can concentrate the energy of the image in a few low-frequency components in the DCT domain. The DCT transform coefficients correspond to the low, mid, and high-frequency components of the image from the upper-left to the lower-right. The high-frequency coefficient is usually smaller than the low-frequency coefficient, so the energy is mainly concentrated in the low-frequency components.

$f(x, y)$ is a two-dimensional $M \times N$ image, and the definition of DCT is shown in equation (3).

$$F(u, v) = \frac{2}{\sqrt{M \times N}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y)C(u)C(v) \\ \times \cos\frac{(2x+1)u\pi}{2M} \cos\frac{(2y+1)v\pi}{2N}] \tag{3}$$

The definition of the two-dimensional $M \times N$ IDCT is shown in equations (4)-(6).

$$C(u) = \begin{cases} \sqrt{\dfrac{1}{M}}, & u = 0 \\ \sqrt{\dfrac{2}{M}}, & u \neq 0 \end{cases} \tag{4}$$

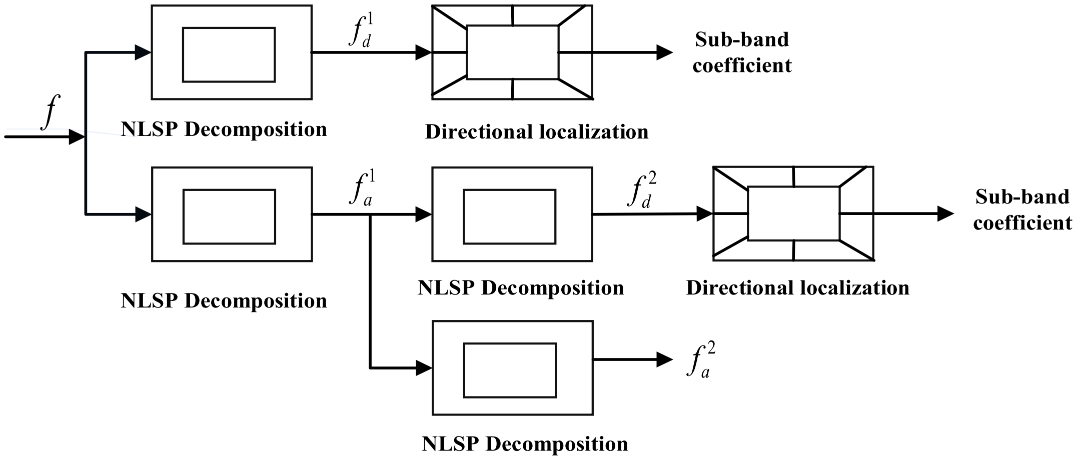


FIGURE 2. Decomposition block diagram of NSST.

$$C(v) = \begin{cases} \sqrt{\frac{1}{N}}, & v = 0 \\ \sqrt{\frac{2}{N}}, & v \neq 0 \end{cases} \tag{5}$$

$$f(x, y) = \frac{2}{\sqrt{M \times N}} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} [F(u, v)C(u)C(v) \times \cos \frac{(2x+1)u\pi}{2M} \cos \frac{(2y+1)v\pi}{2N}] \tag{6}$$

where $F(u, v)$ is the two-dimensional DCT transformation of the image. M and N are the width and height of the image, and $C(u)$ and $C(v)$ are the compensation coefficients. The low-frequency coefficient of DCT reflects the slow change of pixels, i.e., the image frame. The high-frequency coefficient reflects the rapid change of pixels, i.e., the image details.

III. INFRARED AND VISIBLE IMAGE FUSION ALGORITHM BASED ON LSWT-NSST

A. BASED ON THE COMBINED TRANSFORMATION THEORY OF LSWT and NSST

The key of infrared and visible image fusion is to effectively extract and fuse the complementary information of multi-source images. By integrating the advantages of LSWT and NSST algorithms, an infrared and visible image fusion algorithm based on LSWT-NSST is proposed. The multi-scale and multi-directional characteristics of the NSST algorithm cover the deficits of LSWT three-dimensional decomposition to retain more source image information through the redundancy of multi-scale decomposition. The high-frequency coefficients of LSWT decomposition are sparse, and its wavelet basis can fully reflect the texture characteristics of the source image in multiple directions and angles. The LSWT-NSST image fusion algorithm can compensate for the lack of subtle image features in NSST and greatly improve the efficiency. Its block diagram is shown in Figure 3.

The flowchart of the LSWT-NSST image fusion algorithm is as follows.

Step 1: Select two infrared light images (IR) and visible-light images (VR) with identical resolution.

Step 2: Perform the LSWT multi-scale decomposition of the infrared image and visible-light image to be fused to obtain a corresponding low-frequency sub-band (LL-Sub-IR and LL-Sub-VR) and multiple high-frequency sub-bands (LH-Sub-IR and LH-Sub-VR). The number of LSWT decomposition layers is set to 3.

Step 3: Perform the NSST multi-scale decomposition again on LL-Sub-IR, LL-Sub-VR, LH-Sub-IR and LH-Sub-VR, and obtain the new low-frequency sub-band images (LNLL-Sub-IR, LNLL-Sub-VR, LNHL-Sub-IR and LNHL-Sub-VR) and new multi-scale, multi-directional high-frequency sub-band images (LNLH-Sub-IR, LNLH-Sub-VR, LNHH-Sub-IR and LNHH-Sub-VR). The number of NSST decomposition levels is set to 3.

Step 4: Perform the DCT conversion on the newly obtained LNLL-Sub-IR, LNLL-Sub-VR, LNHL-Sub-IR and LNHL-Sub-VR. The key features of low-frequency sub-band images are concentrated on a small part of the coefficients in the DCT domain. The window scale of the DCT domain is set to $\omega = 4$.

Step 5: Calculate the LSF value of each low-frequency sub-band coefficient in the DCT domain. The calculation results are: LNLL-Sub-IR-DLV, LNLL-Sub-VR-DLV, LNHL-Sub-IR-DLV and LNHL-Sub-VR-DLV. Through the calculation, the regional characteristics of the low-frequency sub-band in the DCT domain can be further enhanced. The window scale of the LSF domain is set to $\omega = 3$.

Step 6: According to the characteristics of infrared and visible-light images, design the fusion rules of high and low-frequency coefficients. The low-frequency sub-band image adopts the LSF value adaptive weighted fusion rule in the DCT domain; the high-frequency sub-band image adopts a fusion strategy based on improving the regional contrast. Details are showns in section III-B (INFRARED AND VISIBLE IMAGE FUSION RULES).

Step 7: The LNLL-Sub-IR-DLV, LNLL-Sub-VR-DLV, LNHL-Sub-IR-DLV and LNHL-Sub-VR-DLV low-frequency

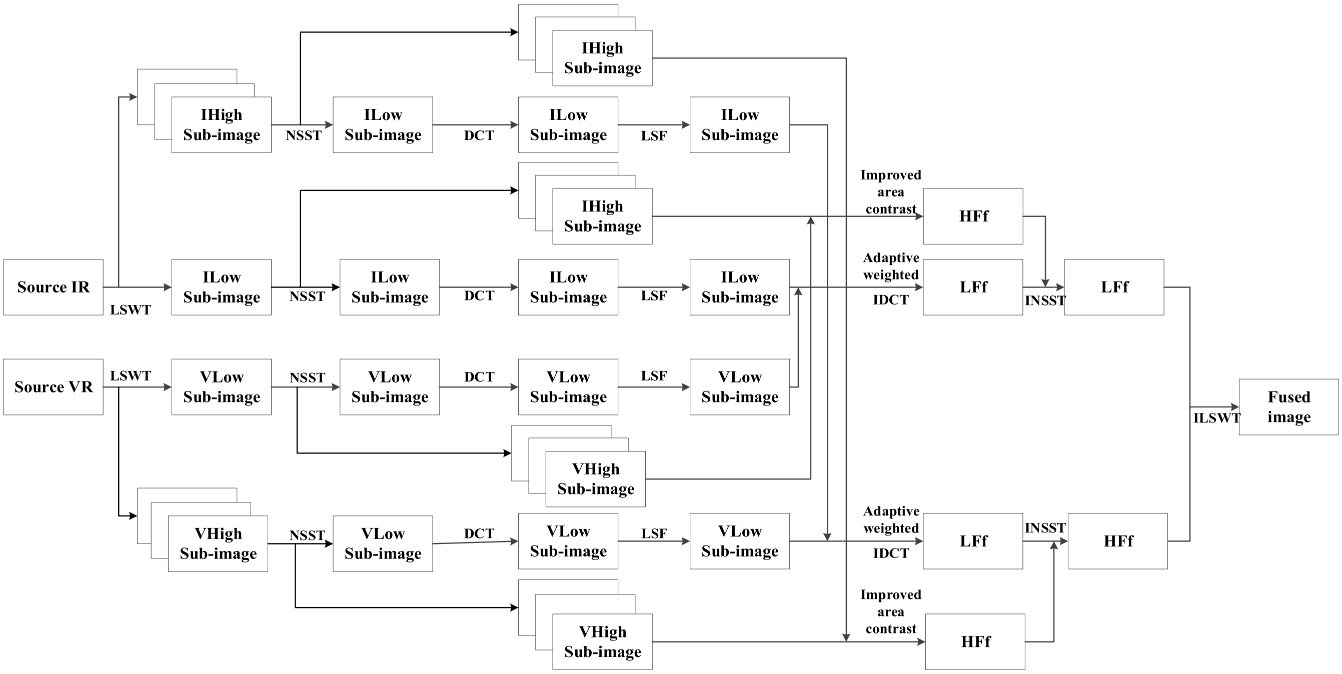


FIGURE 3. Block diagram of the LSWT-NSST image fusion algorithm.

sub-bands and LNLH-Sub-IR, LNLH-Sub-VR, LNHH-Sub-IR and LNHH-Sub-VR high-frequency sub-bands follow the high and low-frequency coefficient fusion rules designed in Step 6. Perform separate fusion operations to obtain the corresponding high and low-frequency fusion sub-images.

Step 8: Perform the IDCT inverse transformation on the low-frequency coefficients after the fusion of LNLL-Sub-IR-DLV, LNLL-Sub-VR-DLV, LNHL-Sub-IR-DLV and LNHL-Sub-VR-DLV to transform the energy to the NSST domain and obtain the low-frequency sub-band of the NSST domain.

Step 9: Perform the inverse NSST transform on the low-frequency sub-band obtained in the NSST domain to convert the fusion coefficient to the LSWT domain, and obtain the low-frequency coefficient of the LSWT domain.

Step 10: Perform the inverse NSST transform on the high-frequency coefficients after the fusion of LNLH-Sub-IR, LNLH-Sub-VR, LNHH-Sub-IR and LNHH-Sub-VR to convert the fusion coefficients to the LSWT domain and obtain the high-frequency coefficients of the LSWT domain.

Step 11: Perform ILSWT on the new high and low-frequency sub-bands to obtain the final fused image F.

B. INFRARED AND VISIBLE IMAGE FUSION RULES

Designing reasonable fusion rules as the key technology of the image fusion algorithm is very important for the fusion effect, since it determines the pixels or coefficients to select and merge into the final image. The image fusion rules require accurate and comprehensive important details and salient features of the source image, and the fusion of images to maximize adaptability to the human visual perception.

The physical properties of infrared and visible images are quite different. Infrared light imaging uses the infrared rays reflected by the target or the thermal radiation generated by the target to detect the target, and the infrared image sensor senses the temperature information of the measured object. The higher the object's temperature, the stronger the infrared spectrum signal, the brighter the infrared image, and the clearer the target. The infrared light image reflects the temperature characteristics of the target, its anti-environmental interference ability is strong, and it has good target detection ability. However, its target resolution is low, the background is fuzzy, and the ability to express details is relatively low. The visible light image sensor mainly uses the visible light information reflected by the object to image the object and can receive the visible spectrum information around the target scene. The visible light image reflects the background and contour information of the image. It has a high spatial resolution, obvious contrast, rich edge and structure texture information, and can better describe scene information. However, the image quality is poor in low visibility and poor lighting conditions.

The image's low-frequency coefficient is also called the approximate sub-band, which contains the main information of the infrared and visible-light images, represents the approximate components and average characteristics of the source image, and concentrates most of the energy of the infrared and visible-light images. The image's high-frequency coefficient is also called the detail sub-band, which describes the detailed information of the infrared and visible-light image, which directly affects the resolution and clarity of the fused image.

Traditional image fusion rules usually adopt the largest absolute value coefficient or weighted average. The former has high fusion efficiency but ignores the average characteristics of the image, which causes image distortion. The latter can reduce the information loss, but lacks the sharpness of the image.

Therefore, combining the characteristics of infrared and visible-light images and the characteristics of their high and low-frequency coefficients to design reasonable fusion rules can fully extract the spectral information of the visible light image and the thermal target information of the infrared image to improve the quality of image fusion effectively.

1) LOW-FREQUENCY COEFFICIENT FUSION RULES

After the source image has been orthogonally decomposed by LSWT and NSST algorithms, the corresponding low-frequency sub-bands and a series of high-frequency sub-bands of various scales and directions are obtained. For non-linear characteristics of the human visual system, a single pixel is closely related to its neighbor pixels. A single-point pixel cannot represent any information and only makes sense to associate with its neighbor set of pixels and be perceived by human vision. Since the low-frequency sub-band concentrates most of the energy of infrared and visible- light images, the key to low-frequency coefficient fusion is to maximize the preservation of important information of the source image, that is, the extraction and preservation of the features of essential regions of the image.

DCT can concentrate information into key features according to the frequency energy. LSF is a commonly used image region representation method, which can effectively select and extract the optimal image features. It has a strong ability to represent regional details and consists of the Local Row Frequency (LRF) and Local Column Frequency (LCF). Inspired by the literature [41], this paper performs the DCT conversion on low-frequency coefficients to obtain the regional characteristics of the DCT domain. Calculating the LSF feature matrix of DCT coefficients can further identify the low-frequency DCT coefficients and key features of the enhanced DCT domain. The formula to calculate the LSF value of the DCT coefficient is shown in equations (7)-(9).

$$LRF = \sqrt{\frac{1}{\omega^2}\sum_{i=1}^{\omega}\sum_{j=2}^{\omega}[DCT(i,j) - DCT(i,j-1)]^2} \tag{7}$$

$$LCF = \sqrt{\frac{1}{\omega^2}\sum_{i=2}^{\omega}\sum_{j=1}^{\omega}[DCT(i,j) - DCT(i-1,j)]^2} \tag{8}$$

$$LSF = \sqrt{LRF^2 + LCF^2} \tag{9}$$

where ω is the window size of LSF, and $DCT(i,j)$ is the DCT coefficient at position(i,j). LRF and LCF are local row frequency and local column frequency, respectively.

According to the difference between infrared and visible-light imaging systems and their sensitivity to local intensity and details, a quantitatively similar matching degree of the image region was designed. The similarity of the LSF feature region is defined in equation (10).

$$S_{A,B}(i,j) = \frac{2 * LSF_{A-ij}(i,j) * LSF_{B-ij}(i,j)}{LSF_{A-ij}(i,j)^2 + LSF_{B-ij}(i,j)^2} \tag{10}$$

Based on the above discussion, the low-frequency coefficient fusion rule is designed as follows.

If $S_{A,B}(i,j) < T$, then

$$FC_{ij} = \begin{cases} C_{A-ij}, & LSF_{A-\mathrm{ij}} > LSF_{B-ij} \\ C_{B-ij}, & LSF_{A-\mathrm{ij}} <= LSF_{B-ij} \end{cases} \tag{11}$$

If $S_{A,B}(i,j) >= T$, then

$$FC_{ij} = \omega(i,j) * C_{A-ij}(i,j) + (1 - \omega(i,j)) * C_{B-ij}(i,j) \tag{12}$$

$$\omega(i,j) = \begin{cases} \frac{1}{2} + \frac{1}{2} * [\frac{1 - S_{A,B}(i,j)}{1 - T}], \\ \quad LSF_A(i,j) > LSF_B(i,j) \\ \frac{1}{2} - \frac{1}{2} * [\frac{1 - S_{A,B}(i,j)}{1 - T}], \\ \quad LSF_A(i,j) <= LSF_B(i,j) \end{cases} \tag{13}$$

where (i,j) is the position of the DCT coefficients; FC_{ij} is the fused DCT coefficients; $C_{A-ij}(i,j)$ and $C_{B-ij}(i,j)$ are the DCT coefficients of A and B, respectively; $\omega(i,j)$ is the fusion weight; $S_{A,B}(i,j)$ is the similarity of the regional pixels of A and B. A larger $S_{A,B}(i,j)$ indicates that A and B are more similar to each other. Matching threshold T is set to 0.85 in this paper.

2) HIGH-FREQUENCY COEFFICIENT FUSION RULES

The multi-scale and multi-directional high-frequency coefficients obtained by the orthogonal decomposition by LSWT and NSST represent the characteristics of the source image, such as the edge and contour. The high-frequency sub-images decomposed by LSWT only contain three directions: horizontal, vertical and diagonal. The multi-directional NSST decomposition can overcome the problem of insufficient direction of LSWT and further extract the multi-scale and multi-directional details of high-frequency sub-bands. A large absolute coefficient value of the high-frequency represents a sudden change of pixels, and a region with a large sudden change contains rich details. If there is noise in the source image, this method will introduce artificial noise and reduce the quality of the fusion image.

According to the multi-resolution selection mechanism of human eye imaging, human vision is highly sensitive to the local contrast of the image. The contrast of the local direction is used to design the high-frequency coefficient fusion rule, which can fully consider the characteristics of infrared and visible-light. The definition is shown in equation (14).

$$CR^{k,m}(i,j) = \frac{|C^{k,m}(i,j)|}{C^0(i,j)} \tag{14}$$

where $C^0(i,j)$ is the low-frequency coefficient; $C^{k,m}(i,j)$ is the high-frequency coefficient of the scale $k(th)$ and direction

$m(th)$ after the NSST decomposition. Since the contrast of a single pixel has no meaning to the image, if the contrast of a single pixel is directly used in the fusion rule, the similarity between pixels will be dissevered, and noise will be introduced. Therefore, the concept of local region contrast is introduced, which is defined in equations (15)-(16).

$$\overline{C^0}(i,j) = \frac{1}{M \times N} \sum_{l=-(M/2-1)}^{(M/2)-1} \sum_{r=-(N/2-1)}^{(N/2)-1} \times C^0(i+l, j+r) \tag{15}$$

$$CR^{k,m}(i,j) = \frac{|C^{k,m}(i,j)|}{\overline{C^0}(i,j)} \tag{16}$$

where $\overline{C^0}(i,j)$ is the low-frequency region mean coefficient; M and N are the width and height of the image area block.

The significance of the coefficients varies in the high-frequency region. $C^{k,m}(i,j)$ is calculated by the region average method, which repels the fact that the human eye has a higher degree of attention to the significant regions. An improved local area contrast is designed as defined by equations (17)-(19).

$$D^{k,m}(i,j) = \sqrt{\begin{array}{l}[I^{k,m}(i,j) - I^{k,m}(i+1,j)]^2 \\ +[I^{k,m}(i,j) - I^{k,m}(i,j+1)]^2\end{array}} \tag{17}$$

$$E^{k,m} = \sum_{l=-(M-1)/2}^{(M-1)/2} \sum_{r=-(N-1)/2}^{(N-1)/2} |D^{k,m}(i+l, j+r)|^2 \tag{18}$$

$$\overline{CR}^{k,m}(i,j) = E(C^{k,m}(i,j))CR^{k,m}(i,j) \tag{19}$$

where $I_{i,j}$ is the high-frequency coefficient at location (i,j). $D^{k,m}(i,j)$ is the gradient at (i,j) pixel, whose scale and direction are k and m respectively. E is the regional gradient energy, which reflects both change degree and edge sharpness of the image. By combining with the area contrast, the high-frequency area with high saliency is given greater weight, and the visual result is more suitable for human vision.

Based on the above analysis, the high-frequency fusion rule is shown in equation (20).

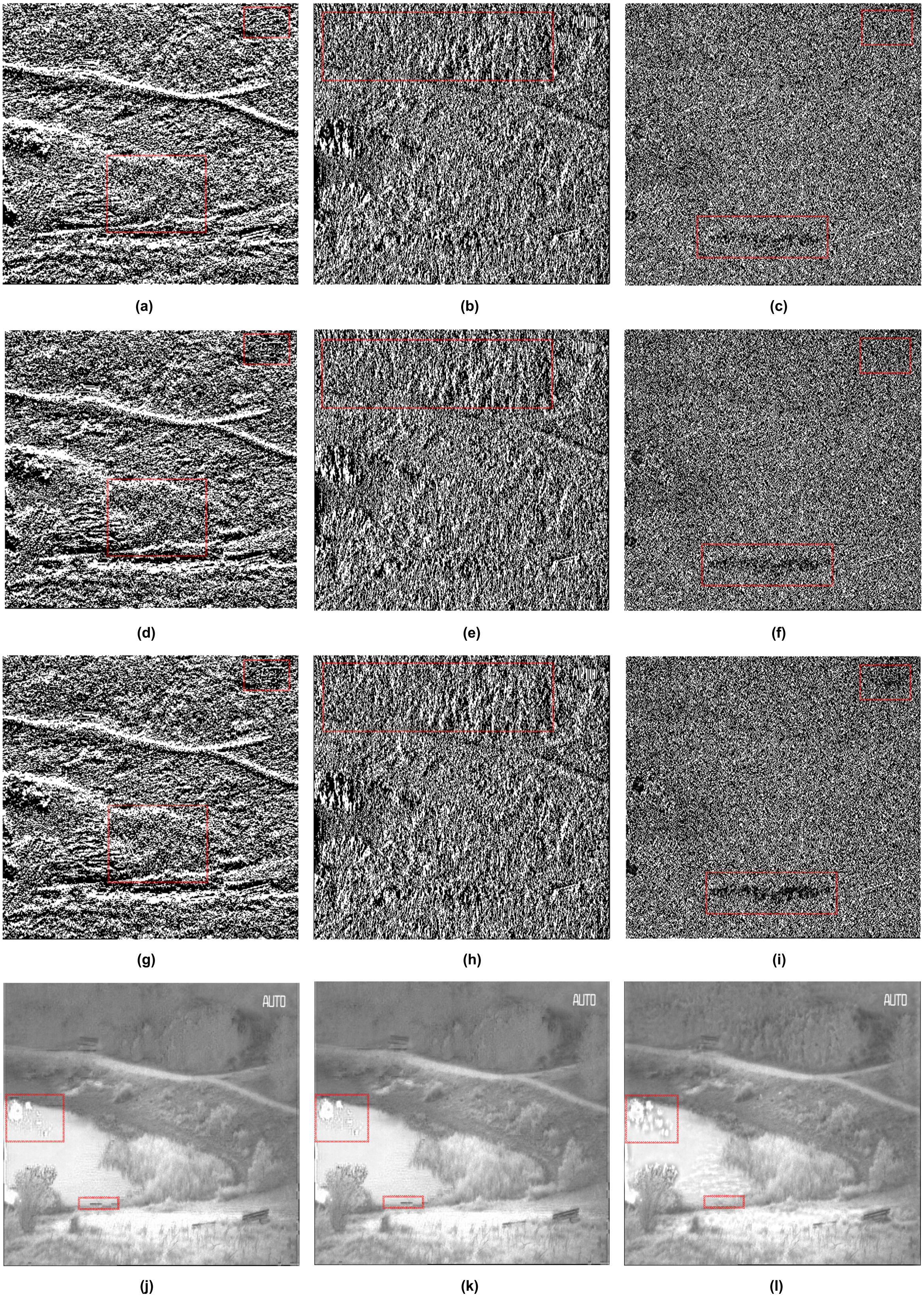
$$F^{k,m}(i,j) = \begin{cases} A^{k,m}(i,j), & \overline{CR}_A^{k,m}(i,j) > \overline{CR}_B^{k,m}(i,j) \\ B^{k,m}(i,j), & \overline{CR}_A^{k,m}(i,j) <= \overline{CR}_B^{k,m}(i,j) \end{cases} \tag{20}$$

where $F^{k,m}(i,j)$ is the high-frequency coefficient of fused image F in the $k(th)$ scale and $m(th)$ direction.

To verify the effectiveness of the improved regional contrast fusion strategy for high-frequency coefficients, the "Lake" image pair in the fourth part is selected as the verification data set (Due to the need to control the length of the text, only the "Lake" image team is compared and analyzed, and the algorithm comparison results of other data sets are similar). The high-frequency coefficients of LSWT multi-scale decomposition are subjectively analyzed by three fusion strategies of directional contrast, regional contrast and improved regional contrast from the horizontal, vertical and diagonal directions. Then, we use these three fusion strategies to reconstruct the final fusion image. These three types of reconstructed images are compared and analyzed to calculate the objective evaluation index values of image fusion in the fourth part of the article. For the simplicity of the analysis, the fusion strategies of directional contrast, regional contrast and improved regional contrast are abbreviated as DCS, ACS and IACS algorithms, respectively. The experimental results are shown in Figure 4 and Table 1.

The first, second, and third rows of Figure 4 indicate that the high-frequency sub-bands of the LSWT decomposition of the "Lake" infrared and visible-light high-frequency fusion image are reconstructed in the diagonal, horizontal and vertical directions using the DCS, ACS, and IACS fusion strategies. The fourth row represents the final fusion image using the DCS, ACS and IACS fusion strategies. In the horizontal direction, the red box marks in Figures (a), (d), and (g) show that the targets at the positions of water plants and “auto” in Figure (g) are more prominent than those in Figures (a) and (d). In the vertical direction, the red box marks in Figures (b), (e) and (h) show that the contour of the woods in the upper left corner of Figure (h) is clearer than that in Figures (b) and (e). In the diagonal direction, the red box marks in Figures (c), (f), (i) show that the contrast between water plants and auto text in Figure (i) is higher than that in Figures (f) and (i). For the final fusion image, the red box marks in Figures (j), (k), (l) show that the sky reflection of Figure (l) has a clearer target contour than Figures (j) and (k). The effect is also more natural. After careful observation, there are block artifacts in Figures (j) and (k). At the edge of the river bank, the fusion effect of Figures (j) and (k) is not very good, and there is a small black block in the fusion images. From a subjective viewpoint, all three integration strategies of DCS, ACS and IACS have achieved good results. After improving the regional contrast, the IACS fusion strategy is better than the DCS and ACS fusion quality.

The bold items in Table 1 represent the maximum of these three algorithms for ten evaluation indicators. A larger value indicates a better fusion effect. Table 1 shows that the IACS algorithm ranks first in AV, MI, SD, EN, AG and Q_{CB}, and SF ranks second. The difference between its SF index and ACS algorithm is 0.531, which is not large. The reason is that the single-pixel contrast of DCS splits the local correlation of the image, and the regional mean of ACS smooths the local features, which will introduce noise. SF reflects the spatial activity of the image, and noise will increase SF. For example, the SF values of the DCS and ACS algorithms are 12.6594 and 12.6812, respectively. Q_{CB} and MS-SSIM are visual evaluation indicators, which lead the ACS algorithm and DCS algorithm by 5.57, 6.05 and 6.29, 8.14 percentage points respectively. It can also be verified from Figures (j), (k) and (l). A comprehensive analysis of objective evaluation indicators shows that IACS has the better overall evaluation index and the best fusion effect.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

FIGURE 4. Fused high-frequency sub-bands and fused images using different fusion strategies in the fourth pair of IR and VI. (a) - (c) High-frequency fusion images in horizontal, vertical, and diagonal directions adopt DCS. (d) - (f) High-frequency fusion images in horizontal, vertical, and diagonal directions adopt ACS. (g) - (i) High-frequency fusion images in horizontal, vertical, and diagonal directions adopt IACS. (j) - (l) Fused images using three fusion strategy.

TABLE 1. Fusion quality indices with different fusion strategies for the fourth pair of IR and VI.

	AV	MI	SD	SF	EN	AG	Q_{CB}	Q_{ABF}	MS-SSIM	CRI
DCS	145.1663	3.2617	34.6998	12.6594	6.5698	4.0322	0.5581	0.5211	0.8328	0.4074
ACS	150.7068	3.2404	40.7634	**12.6812**	6.6258	4.1026	0.5629	0.5537	0.8513	0.6667
IACS	**160.8361**	**3.3141**	**42.8586**	12.1502	**7.4170**	**4.2925**	**0.6186**	**0.6051**	**0.9142**	**0.9259**

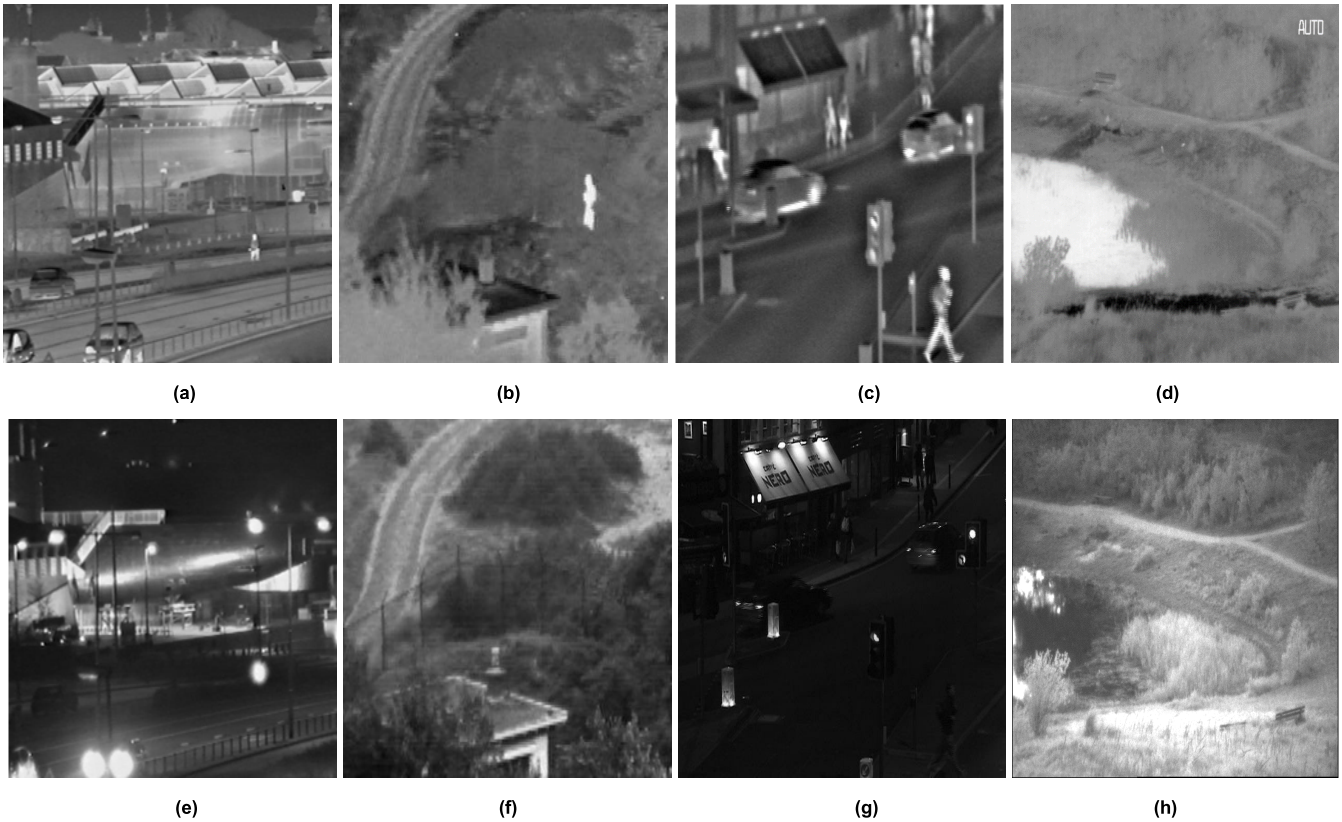


FIGURE 5. Four sets of source infrared and visible images (a) (e) Kayak's IR and VI. (b) (f) UNCamp 's IR and VI. (c) (g) Road 's IR and VI. (d) (h) Lake 's IR and VI.

The multi-scale decomposition method conforms to the multi-resolution physiological mechanism of human vision. In the high-frequency coefficients, the human eye has a high degree of recognition of the high salient area or high mutation area of the image. Thus, combining the characteristics of infrared and visible-light, this paper proposes an improved regional contrast fusion strategy for high-frequency images. Through the aforementioned subjective vision and objective evaluation analysis, the algorithm fusion quality ranking is obtained: IACS> ACS> DCS. In general, the images fused by the IACS algorithm have the best fusion effect of prominent infrared targets, clear contours and rich visible-light scene details.

IV. EXPERIMENTAL SIMULATION AND RESULT ANALYSIS

A. EXPERIMENTAL CONDITIONS AND SETTINGS

The algorithm experimental environment is as follows: The host is configured with Intel(R) Core(TM) i7. The main frequency is 1.99 GHz and the memory is 8 GB. The experimental simulation platform is MATLAB R2019b.

To objectively evaluate the performance of the fusion algorithm, this paper selects experimental materials from the infrared and visible image libraries. Four classic infrared and visible images were selected for fusion experiments. "Kayak," "UNCamp," "Bristol Queen's Road" and "Lake" are shown in Figure 5. To verify the effectiveness of the proposed algorithm, the selected four groups of images have different sizes from small to large: 256 × 256, 280 × 360, 496 × 632 and 576 × 768; the clear regions of the selected four groups of images must have different positions. In order to more comprehensively verify the advantages of the algorithm proposed in this paper, 21 pairs images in the literature [47] is selected for the experimental analysis of the classic infrared and visible-light image sets. The purpose of the fusion algorithm is to make the fusion image clearly and objectively reflect the real scene and conform to human vision.

To verify the accuracy of the algorithm in this paper, the LSWT-NSST algorithm proposed in this paper is compared with the traditional classical algorithm and recently proposed advanced algorithm as follows: Non-Subsampled Shearlet Transform (NSST) [36], Non-Subsampled Contourlet Transform (NSCT) [37], Cross Bilateral Filtering (CBF) [38], Guided Filtering (GF) [39], Latent Low-Rank Representation (LLRR) [40], Discrete Cosine and Local Spatial Frequency (DCLSF) [41], Saliency Detection (SD) [22] and Deep Convolutional Neural Networks (CNN1 and CNN2) [27], [28]. For experimental parameters setting, the open-source-code algorithm uses its fixed parameters, and non-open-source-code algorithm is simulated according to its detailed description in the references.

As classic algorithms for multi-scale image fusion, NSST and NSCT have better results than other traditional frequency domain methods. As classic algorithms for image fusion based on the spatial domain, CBF and GF have achieved good results. In particular, CBF smooths images through a nonlinear combination of neighborhood pixel values and has strong edge preservation. For LLRR, DCLSF and SD, which are improved multi-scale fusion algorithms proposed in recent years in the field of image fusion such as multi-focus, infrared and visible-light, the fusion effect has been greatly improved in both subjective and objective aspects. CNN1 and CNN2 are two deep learning image fusion algorithms proposed in recent years, where a deep convolutional neural network architecture is trained to extract the features of the source image in depth. The fused image has clear targets, rich detailed information, more obvious advantages and greatly improved performance than the traditional image fusion methods. Therefore, this paper selects these 9 algorithms as performance comparison algorithms, which are feasible for evaluating the advantages of our algorithm.

B. MAIN PARAMETER SETTINGS

The rationality of the image fusion algorithm parameter setting determines the level of algorithm fusion performance directly. Generally, the more layers decomposed by the multi-scale algorithm, the richer the detailed information expressed by the image will be. However, the increase in the number of decomposition layers often leads to a sudden increase in the amount of calculation. Therefore, in using the LSWT algorithm and the NSST algorithm to perform the first-layer and second-layer multi-scale decomposition, we must also consider the fusion performance and the amount of calculation to select the parameters. When using the LSWT algorithm to carry out the first-level decomposition, the literature [30] and [44] are referred to, and the decomposition scale of LSWT is determined to be 3 through experimental analysis. When using the NSST algorithm for the second-level decomposition, reference was made to the literature [15], [26] and [33], and the number of multi-scale filter directions corresponding to the high-frequency was determined to be [4 4 4] through experimental analysis. The corresponding multi-scale filter direction number is [3 3 3]. In the experiment, we found that with the increase of the LSWT and NSST algorithm's decomposition scale, their fusion effect will be improved, and the calculation will be increased, which is very time-consuming. After careful consideration, we select the decomposition level of LSWT and NSST algorithms to be 3. There will be a small amount of high-frequency information in the low-frequency sub-band of the LSWT domain, and a small amount of low-frequency information will also be contained in the high-frequency sub-band of the LSWT domain, so the scale setting is not appropriate to be too large when performing the second layer decomposition. The selection of the DCT window and the LSF window refers to the literature [38], and the settings are the same. For the selection of the fusion threshold of the low-frequency sub-band LSF domain and the high-frequency sub-band contrast area, because the adjustable range of these two parameters is small and the setting is easy, the general setting can be used.

The parameter setting is vital for the image fusion algorithm. Based on the above analysis, we reasonably set the parameters of the algorithm proposed in this paper to obtain a better fusion effect. The specific parameter settings are as follows: the LSWT decomposition level is set to 3; the NSST decomposition level is set to 3. Considering the difference in image high and low-frequency representative features, for LSWT multi-scale decomposition coefficients, the number of multi-scale filter directions corresponding to high frequencies is set to [4 4 4]; its low frequency corresponding multi-scale filter direction number is set to [3 3 3]. After the three-layer decomposition, the high-frequency coefficients of the LSWT domain have insufficient directivity. The filter in the high-frequency direction of NSST can compensate for the insufficiency of the high-frequency decomposition direction of LSWT and obtain more edge and contour information. After the three-layer decomposition of the low-frequency coefficients in the LSWT domain, the low-frequency sub-band contains a lot of detailed information and its low-frequency coefficients have a good fusion effect, but the low-frequency sub-band contains some high-frequency information, so better details and edge features in the low-frequency sub-band filtered by the NSST low-frequency direction filter can be obtained. The DCT window size is set to 4×4, which greatly affects the image fusion performance. The LSF window is set to 3×3, in case that it is too large to produce large redundant information. The fusion threshold is set to $\mathrm{T} = 0.85$ for LSF low-frequency sub-band. The high-frequency sub-band region contrast is set to 3×3.

C. EVALUATION OF THE EFFECT OF IMAGE FUSION

The rationality of the image fusion algorithm determines the quality of the fusion image, while the fusion quality is another important index. The fusion image has variable measurement standards because of application purposes or scene. Therefore, the comprehensive use of multiple evaluation criteria can better determine on the fusion effect.

Currently evaluation methods are divided into subjective and objective methods.

1) SUBJECTIVE EVALUATION

The differences between source image and fused image are directly evaluated by a subjective method with human vision system and mainly reflected by the image registration and clarity. The subjective method is suitable for images with significant differences, and it is relatively simple and intuitive, so it is an important method to determine the performance of the fusion images. However, this method is susceptible to the subject's knowledge level and subjective consciousness, which makes it difficult to set the standard scale and has great one-sidedness. Therefore, the comprehensive use of subjective and objective methods can make more accurate judgment on the fusion effect.

2) OBJECTIVE EVALUATION

The objective evaluation method can quantify the effect of image quality and effectively reduce the effect of human subjective factors. A quantifiable performance parameter is used to determine the pros and cons of each image fusion algorithm. Ten objective evaluation indicators are used in this paper: Average Value (AV), Mutual Information (MI), Standard Deviation (SD), Spatial Frequency (SF), Information entropy (EN) [41], Average Gradient (AG) [42], Q_{CB} [43], Edge strength coefficient(Q_{ABF}) [45], Multi-scale structural similarity(MS-SSIM) [46] and Comprehensive Ranking Index (CRI).

a: AVERAGE VALUE

AV is the average brightness of the fused image. Larger AV implies a brighter image, whose mean is defined in equation (21).

$$\mu = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} F(i,j) \tag{21}$$

where $F(i,j)$ is the pixel value of the fusion image at position(i,j); M and N are the width and height of the image.

b: STANDARD DEVIATION

SD is the degree of dispersion between the single pixel and the average pixel of the image. Larger SD implies that the image has higher contrast, wider gray value distribution, and more image information. The definition of standard deviation is shown in equation (22).

$$\sigma = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (F(i,j) - \mu)^2} \tag{22}$$

where $F(i,j)$ and μ are the gray value and mean value at (i,j) of the fusion image, respectively.

c: INFORMATION ENTROPY

EN is used to calculate the information richness and reflects the amount of information in the fused image. A larger EN indicates that the fused image has richer information and higher quality. The definition of the entropy value is shown in equation (23).

$$H = -\sum_{i=0}^{L-1} P_i \log_2(P_i) \tag{23}$$

where L is the total gray level of the image. P_i is the ratio of the number of pixels that the gray value is i to the total number of pixels of the image.

d: SPATIAL FREQUENCY

SF is the overall activity of the fusion image in the spatial domain. A larger SF corresponds to more image texture and edge information and higher quality of the fusion image. It is mainly composed of spatial Row Frequency (RF) and spatial Column Frequency (CF) and is defined as shown in equations (24)-(26).

$$RF = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=2}^{N} [F(i,j) - F(i,j-1)]^2} \tag{24}$$

$$CF = \sqrt{\frac{1}{M \times N} \sum_{j=1}^{M} \sum_{i=2}^{N} [F(i,j) - F(i-1,j)]^2} \tag{25}$$

$$SF = \sqrt{RF^2 + CF^2} \tag{26}$$

e: MUTUAL INFORMATION

MI is the degree of correlation information between the source image and the fused image. Larger MI implies stronger correlation, higher retention of the source image, and lower image distortion. The definition of mutual information is shown in equation (27).

$$MI = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \sum_{k=0}^{L-1} P_{AB,F} \log_2 \frac{P_{AB,F}(i,j,k)}{P_{AB,F}(i,j)P_F(k)} \tag{27}$$

where L is the total number of gray levels of the image. P_{AB} and P_F are the normalized histogram of the source image AB and fusion image F. $P_{AB,F}$ is the combined gray histogram after the normalization of the source image and fusion image.

f: AVERAGE GRADIENT

AG is the degree of image detail and texture changes. Higher AG implies that the fused image has more prominent texture and detail changes and contains more content. The definition of the average gradient of the image is shown in equation (28).

$$AG = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \sqrt{(\Delta I_x^2 + \Delta I_y^2)/2} \tag{28}$$

where $M$and $N$are the width and height of the image. $\Delta I_x$and ΔI_y are the differences between directions x and y, respectively.

g: Q_{CB}

Q_{CB} [43] is an evaluation index based on human visual perception. This method uses the Contrast Sensitivity Function (CSF) to calculate the local contrast of each image. Suppose that the input source image is I_A and I_B, and the fusion image is I_F, then its definition is shown in equations (29)-(30).

$$Q_C(x, y) = \lambda_A(x, y)Q_{AF}(x, y) + \lambda_B(x, y)Q_{BF}(x, y) \tag{29}$$

$$Q_{CB} = \overline{Q_C(x, y)} \tag{30}$$

where λ_A and λ_B are the salient maps of the source image I_A and I_B, respectively. $Q_{AF} \in [0, 1]$ and $Q_{BF} \in [0, 1]$ are the fidelity of information from source images I_A and I_B to fusion image I_F, respectively. Q_{CB} is the average value of the entire fusion quality map Q_C.

h: EDGE INTENSITY COEFFICIENT

Q_{ABF} [45] quantifies the amount of information retained on the edge of the image. It reflects the amount of edge information obtained from the source image of the fused image. $Q_{ABF} \in [0, 1]$, the closer the Q_{ABF}value is to 1, the more abundant the edge information of the source image is retained in the fusion image, and the better the fusion quality of the image. The definition is shown in equation (31).

$$Q_{ABF} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(Q_{AF}(i, j) \times \omega_A(i, j) + Q_{BF}(i, j) \times \omega_B(i, j))}{\sum_{i=1}^{M}\sum_{j}^{N}(\omega_A(i, j) + \omega_B(i, j))} \tag{31}$$

where (i, j) is the pixel position, M and N are the size of the image. Q_{AF} and Q_{BF} represent the edge strength of the source image A and B and the fused image F respectively. $\omega_A(i, j)$ and $\omega_B(i, j)$ represent the quantization weights of Q_{AF} and Q_{BF} respectively.

i: MULTI-SCALE STRUCTURAL SIMILARITY

MS-SSIM [46] is an indicator based on the human visual system. It is suitable for extracting structural information from the scene, and its measurement can better approximate the visual perception of better image quality. MS-SSIM can make SSIM measure the structural similarity between multi-scale images. The larger the value is, the better the fusion image effect. The definition is shown in equation (32).

$$MS-SSIM(A, B) = [l_S(A, B)]^{\alpha S} \prod_{i=1}^{S} [c_i(A, B)]^{\beta_i}[s_i(A, B)]^{\gamma_i} \tag{32}$$

where l is the brightness comparison between images A and B, c is the image contrast, s is the image structure, α, β, and $\gamma$are the relative importance of adjusting image brightness, contrast and structure, and S is the image scale.

j: COMPREHENSIVE RANKING INDEX

To evaluate the image fusion effectiveness of the proposed algorithm as a whole, a Comprehensive Ranking Index (CRI) was designed based on the above seven objective evaluation indices. The brightness, contrast, amount of information and details of the fusion image are comprehensively considered. The design idea is as follows:

a. Calculate the ranking of each fusion algorithm in this index in sequence. The value is S_{ai}^{j}, $i = 1, 2, \cdots, Z$, $j = 1, 2, \cdots, R$.
b. Calculate the single index score. If the ranking is S_{ai}, the single score is $Z - S_{ai}^{j} + 1$.
c. Calculate the comprehensive index score and weighted sum.
d. Normalize the index score. The definition is shown in equation (33).

$$CRI = \frac{\sum_{i=1}^{Z}\sum_{j=1}^{R} S_{ai}^{j} \times (Z - S_{ai}^{j} + 1)}{R \times Z} \tag{33}$$

where R is the number of indicators, and Z is the number of algorithms.

3) EXPERIMENTAL RESULTS AND DISCUSSION

The experiment verifies the effectiveness of the algorithm in this paper through the individual experimental analysis of four groups of classic infrared and visible-light images and overall fusion experiment of 21 pairs of classic infrared and visible-light image sets. In order to better compare and analyze the performance of the algorithms, we marked the top three indicators in bold. The value in bold blue indicates that it ranks first in the indicator and the fusion quality is the best, the value in bold green indicates that it ranks second in the indicator and the fusion quality is second, and the value in bold red indicates that it ranks third in the indicator and the fusion quality is third.

(1) The first set of source image pairs ‘‘Kayak,’’ the resolution is 256 × 256, which is the night view of the city including pedestrians, vehicles and buildings and other visible and infrared image scenes. The detailed information of IR and VI images is extracted to determine the performance of the fusion algorithm. The source image and fused images by all algorithms are shown in Figures 6 (a)-(k). From (c), (d) and (h) of Figure 6, the global contrast of the image is low. Figures (c), (d), (e), (g) have dotted artificial noise at the lights. In Figures (f) and (j), the bright light of the car lights disappeared, and the extracted infrared light information is too much, which causes the distorted visual effect. The brightness of the street lamps in Figures (c), (d), (e), (f), (g), (i) and (j) is obviously darker, and there are noises or distortions around the street lamps in several figures. In contrast, CNN2 based on deep learning and the algorithm based on LSWT-NSST proposed in this

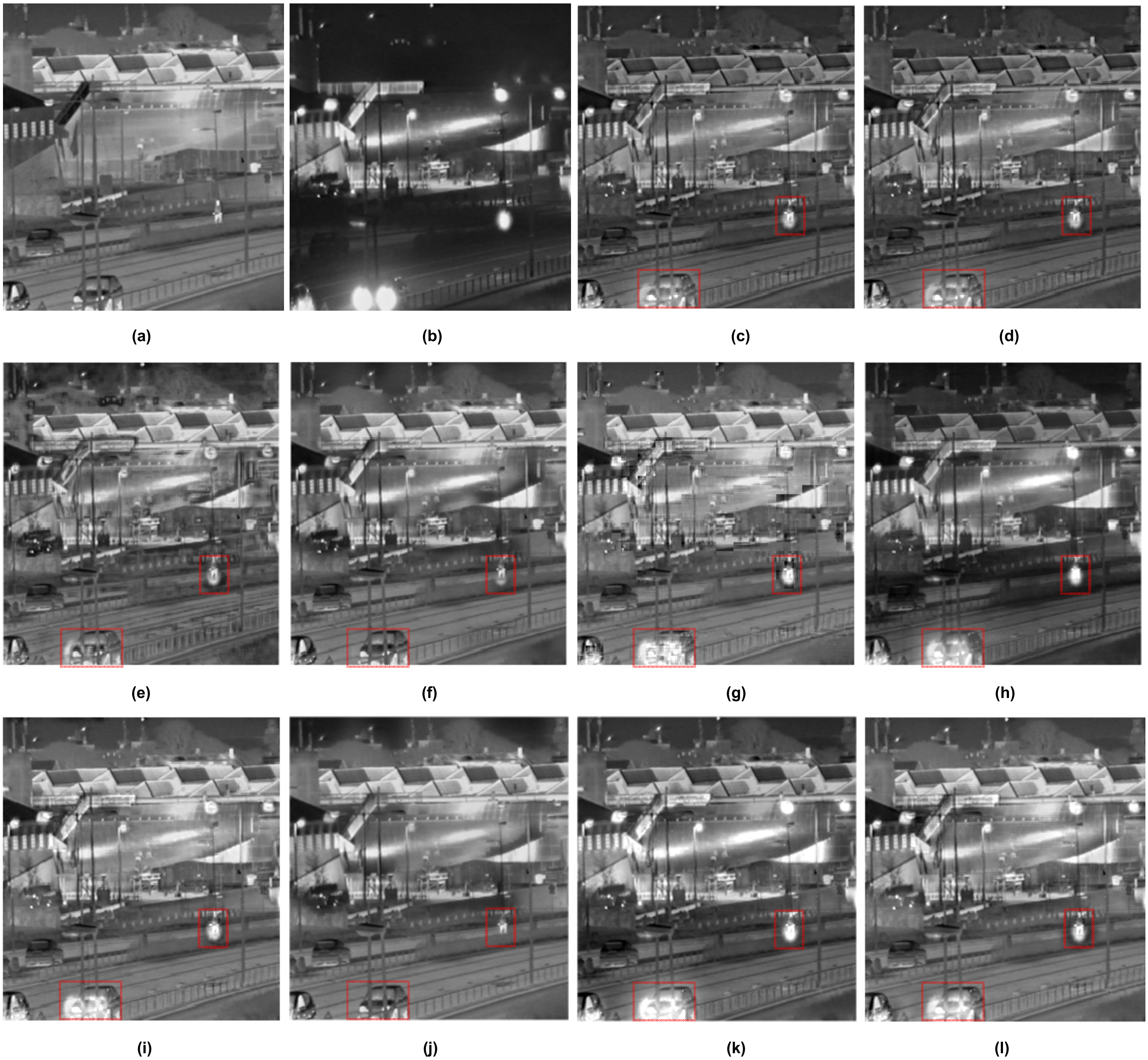


FIGURE 6. Source images and fused images using different methods in the first pair of IR and VI. (a) IR. (b) VI. (c) NSCT. (d) NSST. (e) CBF. (f) GF. (g) SD. (h) LLRR. (i) DCLSF. (j) CNN1. (k) CNN2. (l) Proposed method.

paper have close fusion effects from a visual viewpoint, and the discrimination is not obvious, but they are better than the other 8 fusion algorithms. Thus, these two algorithms can effectively extract the important features of infrared and visible images, and the resulting fusion image is clearer and has better visual effects than other methods as shown in Figures (k) and (l).

Table 2 lists the fusion quality indicators of the first pair of image fusion methods. From the perspective of these 10 objective evaluation indicators, the proposed algorithm in this paper ranks first for indicators AV, SF, EN and CRI, second for indicators SD and AG, and third for indicators MI, Q_{CB}, Q_{ABF} and MS-SSIM. In particular, it is 0.8889 on CRI, which leads the DCLSF, CNN1 and CNN2 algorithms by 14.45, 16.67 and 26.67 percentage points, respectively. The fusion frameworks CNN1 and CNN2 based on deep learning have certain advantages for certain indicators such as MI, SD and MS-SSIM, but the proposed algorithm index in this paper ranks higher as a whole, which indicates that the algorithm in this paper is optimal.

(2) The second set of source images is ''Bristol Queen's road'' with a resolution of 280 × 360. The person sheltered by trees in the IR image is clearly visible, but the details of the surrounding environment are blurred. The details of trees and fences in VI are very clear, and the contrast is higher, which is more suitable for human vision, but the person is not visible.

TABLE 2. Fusion quality indices with different methods for the first pair of IR and VI.

	AV	MI	SD	SF	EN	AG	Q_{CB}	Q_{ABF}	MS-SSIM	CRI
NSCT	88.3930	2.4955	38.7532	21.5267	7.1507	8.4679	0.6099	0.5846	0.9082	0.3667
NSST	88.3911	2.4867	38.5777	21.5799	7.1457	8.4901	0.6102	0.5798	0.9046	0.3222
CBF	104.3131	2.7478	42.5129	**22.2221**	7.3150	**9.1542**	0.5856	0.5834	0.7486	0.4778
GF	108.2681	3.5709	43.4467	21.0702	7.3125	8.3090	**0.6842**	**0.6617**	0.8459	0.5889
SD	90.6047	2.6567	**46.6809**	21.1479	7.3333	8.0864	0.6091	0.5733	**0.9423**	0.4556
LLRR	102.2520	2.8130	42.8210	16.4928	7.1803	6.5338	0.5966	0.5207	0.9225	0.3111
DCLSF	**117.4935**	**3.7170**	44.7271	**21.9256**	7.3676	**8.7413**	0.6491	0.5969	0.8813	**0.7222**
CNN1	**109.0612**	**4.6781**	44.1338	19.8632	**7.3935**	7.8875	**0.6737**	**0.6485**	0.8321	0.6222
CNN2	104.0445	2.8878	**51.3706**	21.8337	**7.4010**	8.6919	0.6248	0.6338	**0.9482**	**0.7444**
Proposed	**117.5004**	**3.5997**	**47.2411**	**23.8033**	**7.4050**	**8.8376**	**0.6679**	**0.6362**	**0.9228**	**0.8889**

By extracting the details of IR and VI images, we determine the performance of the algorithm. The source image and fused images by all algorithms are shown in Figures 7 (a)-(l). From the perspective of visual effects, the characters in Figures (f), (g), (i), (j), (k) and (l) are brighter and more prominent. The characters in Figures (c), (d) and (h) are relatively dim and there is a hole in the middle of the target person's body in Figure (e), which fails to well retain the prominent person target in the infrared image. In terms of details, there are noise and artifacts in the grass in the lower left corner of the building in Figures (c), (d), (e), (f), (g), (h), (i), (j) and (k). Especially in Figures (g), (i) and (j), the grass background is connected as a whole, and it is difficult to identify. The grass background in Figure (l) has a clear contour, which is better than other algorithms. The infrared white objects on both sides of the road in Figure (j) disappear, and there is a lot of noise in Figure (g). In general, the detail information in Figure (l) is clearer than that in other images and suitable for human eyes to observe.

Table 3 lists the fusion quality indicators of the second pair of image fusion methods. From the perspective of these 10 objective evaluation indicators, the algorithm proposed in this article ranks first for indicators AG, SF and CRI, second for indicators MI and AG, third for indicators SD EN, Q_{ABF} and MS-SSIM. Although it ranks fourth for indicator Q_{CB}, the difference with the first is only 0.011, which is very small. CRI is 0.8556, which leads the second to fourth by 7.78, 23.34 and 25.56 percentage points, respectively. The fusion frameworks CNN1 and CNN2 based on deep learning have certain advantages for certain indicators such as MI, SD, EN, Q_{ABF} and MS-SSIM. The visual effects of CNN2 and the proposed algorithm are better than other algorithms. From the overall evaluation and analysis, the algorithm proposed in this paper has the best effect.

(3) The third pair of source image pairs "Bristol Queen's road" has a resolution of 496 × 632. The image is a street view of the city, including visible and infrared image scenes of pedestrians, vehicles and public facilities. Detailed information is extracted from IR and VI images to determine the performance of the fusion algorithm. The source image and fused image by all algorithms are shown in Figures 8 (a)-(l). In Figure 8 (c)-(f), the fusion image has artifacts. The text on the "Advertising Board" on the Figures (d) and (e) is blurred compared to other algorithms, and artificial noise has been introduced, in particular, Figure (e) has serious noise. The overall brightness and contrast of Figures (c), (d), (h) and (j) are darker. There is a serious artifact between the car and the person on the left of Figure (j), and the car and the person are connected together. From a visual viewpoint, Figures (i), (k) and (l) have the best overall effect. However, after careful observation, the color of the eaves on the right side of the text baffle in Figure (k) is too dark, and the visual effect is not good. The two windows in the upper left corner of Figure (i) are less clear than Figure (l). From this group of experiments, the brightness of the image fused by the algorithm in this paper is closer to the source visible-light image and the target object is closer to the source infrared light image.

Table 4 lists the fusion quality indicators of the third pair of image fusion methods. From the perspective of these 10 objective evaluation indicators, the proposed algorithm in this paper ranks first for indicators AV, EN, MS-SSIM and CRI, second for indicators SF, AG and Q_{CB}, and third for indicators MI, SD and Q_{ABF}. The CRI is 0.8333, which leads the second to fourth by 8.89, 14.44 and 27.77 percentage points, respectively. The CBF algorithm has much larger SF and AG indicators than other algorithms because it introduces a lot of artificial noise, as shown in the figure. Fusion frameworks CNN1 and CNN2 based on deep learning have certain advantages for certain indicators such as SD, Q_{CB} and Q_{ABF}. The effect of most fusion algorithms is not subjectively ideal. However, the proposed algorithm in this paper, CNN2, and DCLSF algorithms have clear contours, high contrast and better results than other algorithms.

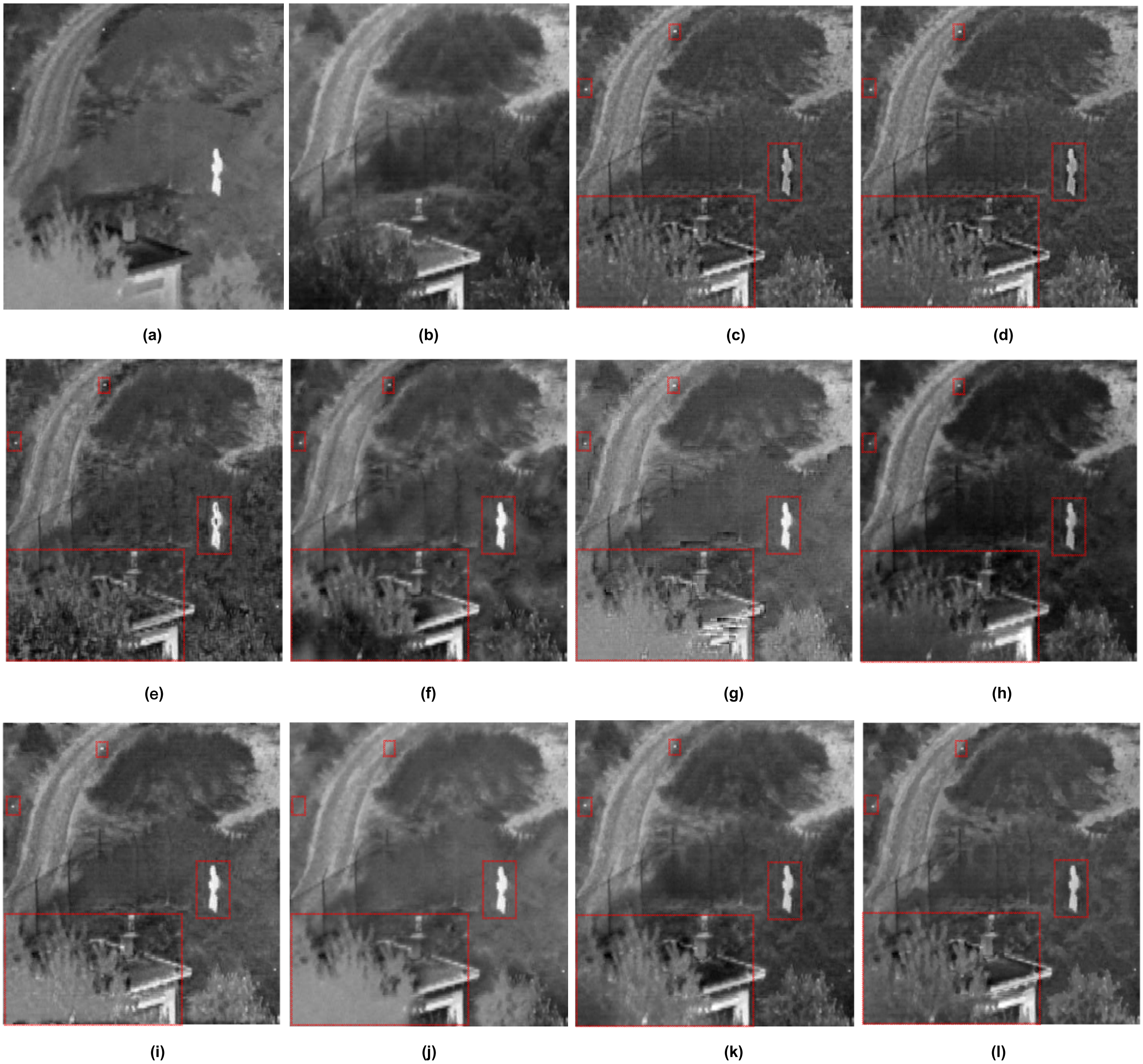


FIGURE 7. Source images and fused images using different methods in the second pair of IR and VI. (a) IR. (b) VI. (c) NSCT. (d) NSST. (e) CBF. (f) GF. (g) SD. (h) LLRR. (i) DCLSF. (j) CNN1. (k) CNN2. (l) Proposed method.

(4) The fourth pair of source image pair "Lake" has a resolution of 576 × 768. The image is a scene of a wild lake, including visible and infrared light scenes of lakes, trees, chairs and roads. By extracting the detailed information of IR and VI images, the performance of the fusion algorithm is determined. The source image and fused image by all algorithms are shown in Figures 9 (a)-(l). From Figures 9 (i), (j), (k) and (l), the brightness contrast of the image is higher, and the other images are darker. The object in the upper left corner of the chair in Figures (c)-(h) is blurry, and the details are not clear. In Figure (e), there are dark shadows in the pond at the left end of the lower chair and many irregular noises on the lake surface, and it has more noise overall. The lake surface in Figures (e), (f), (g) and (h) is darker. The reflections of the clouds on the left of Figures (i) and (j) are severely distorted. Figures (c), (d), (h), (i), (j) have serious fusion distortion at the upper left corner of the woods. Figures (k) and (l) have higher brightness than other figures, and the target details are clear. After careful observation, the near-focus parts of Figure (l) is more prominent than that in Figure (k), but the far-focus parts of Figures (k) and (f) are more prominent than that in Figure (l). Comparing these 10 algorithms, there is always a small part of the fused picture with noise, artifacts or distortion.

TABLE 3. Fusion quality indices with different methods for the second pair of IR and VI.

	AV	MI	SD	SF	EN	AG	Q_{CB}	Q_{ABF}	MS-SSIM	CRI
NSCT	91.0888	1.4956	24.4381	9.0898	6.3754	3.9259	0.5724	0.4520	0.9121	0.4222
NSST	91.0916	1.4856	24.3596	9.1367	6.3723	3.9473	**0.5783**	0.4477	0.9079	0.4222
CBF	89.8759	1.5302	27.3199	**10.5158**	6.5349	**4.6215**	0.5182	0.3858	0.7429	0.4000
GF	88.2701	1.8143	29.7353	9.0280	6.6969	3.7963	**0.5852**	**0.5128**	0.7965	0.5222
SD	90.9216	1.5177	28.9372	9.7413	6.6195	4.0938	0.5574	0.4517	**0.9327**	0.5333
LLRR	102.5323	1.7440	29.4864	7.3546	6.5392	3.1794	0.5157	0.4256	0.9017	0.3444
DCLSF	**110.8829**	**2.2782**	30.3310	9.6295	6.7691	4.1837	0.5305	0.4403	0.8767	0.6000
CNN1	**106.5087**	**4.0421**	**31.2053**	8.4331	**6.8447**	3.4860	0.5417	**0.5448**	0.7730	0.6222
CNN2	89.1381	1.9841	**36.8084**	**9.7424**	**7.0531**	**4.1969**	**0.5773**	0.5067	**0.9348**	**0.7778**
Proposed	**110.9159**	**2.3344**	**30.9534**	**10.8366**	**6.8171**	**4.3689**	0.5742	**0.5091**	**0.9136**	**0.8556**

TABLE 4. Fusion quality indices with different methods for the third pair of IR and VI.

	AV	MI	SD	SF	EN	AG	Q_{CB}	Q_{ABF}	MS-SSIM	CRI
NSCT	51.8982	1.9899	23.0690	12.2088	6.0054	3.0543	0.4927	0.5286	0.9082	0.2778
NSST	51.8985	1.9957	22.9760	12.2071	6.0018	3.0502	0.4917	0.5238	0.9065	0.2222
CBF	65.1589	1.5141	34.6737	**14.9873**	6.7027	**5.1341**	0.4473	0.4638	0.6986	0.3667
GF	71.7971	2.7138	35.5671	12.7406	6.6993	3.5244	**0.5509**	0.6365	0.8884	0.5556
SD	52.3405	1.9567	30.3458	11.6122	6.4136	3.2976	0.4604	0.5943	0.9579	0.3000
LLRR	60.4812	1.9780	27.4851	9.1842	6.0297	2.4135	0.4756	0.5037	0.9041	0.1444
DCLSF	**83.4059**	**4.3561**	36.0190	**12.6852**	**6.7716**	**3.5841**	0.5478	**0.6512**	**0.9726**	**0.7444**
CNN1	33.2465	**3.9758**	**41.3248**	12.2048	5.7575	2.7326	**0.6841**	0.5109	0.6888	0.3667
CNN2	**77.9765**	2.6521	**37.0487**	12.6188	**6.7090**	3.5297	0.5319	**0.6627**	**0.9745**	**0.6889**
Proposed	**83.4156**	**3.9010**	**36.1346**	**13.3541**	**6.7891**	**3.6053**	**0.5765**	**0.6391**	**0.9754**	**0.8333**

The comprehensive analysis shows that compared with other algorithms, the infrared and visible images fused by the algorithm in this paper have rich scenes, clear target contours, and strong contrast, but there are distortions in some parts, so the visual effect must be further improved.

Table 5 lists the fusion quality indicators of the fourth pair of image fusion methods. From the perspective of these 10 objective evaluation indicators, the proposed algorithm in this paper ranks first for indicators AV, EN and CRI, second for indicator AG, third for indicator SF, Q_{ABF} and MS-SSIM, and fourth for indicators MI, SD and Q_{CB}. In particular, the comprehensive index CRI is 0.8222, which leads the second to fourth by 5.55, 10.00 and 15.55 percentage points, respectively, and is nearly 20 percentage points ahead of other algorithms. The fusion frameworks CNN1 and CNN2 based on deep learning have certain advantages for certain indicators such as MI, SD and MS-SSIM. The overall evaluation of this algorithm is better than other algorithms in terms of fusion quality.

(5) Comprehensive Experiment

Select twenty one sets of classic infrared and visible-light images for experimental verification.

(6) Comparison of calculation efficiency

Figure 10 shows eight pairs of images randomly selected from the classic 21 pairs of infrared and visible-light images, and images fused using the algorithm proposed in this paper. From the fused images in Figure 10, it can be seen that the image fused using the algorithm in this paper has clear targets and outlines, contains more detailed information and contains less noise, and the overall fusion effect is good.

Table 6 uses different fusion algorithms to calculate each evaluation index's average value for 21 classic infrared and visible-light image sets. Table 7 shows the EN objective evaluation index table for 21 pairs of classic infrared and visible-light images using different fusion algorithms. Figure 11 is a graph of EN objective evaluation indicators drawn using different fusion algorithms for 21 pairs of classic infrared and visible-light images.

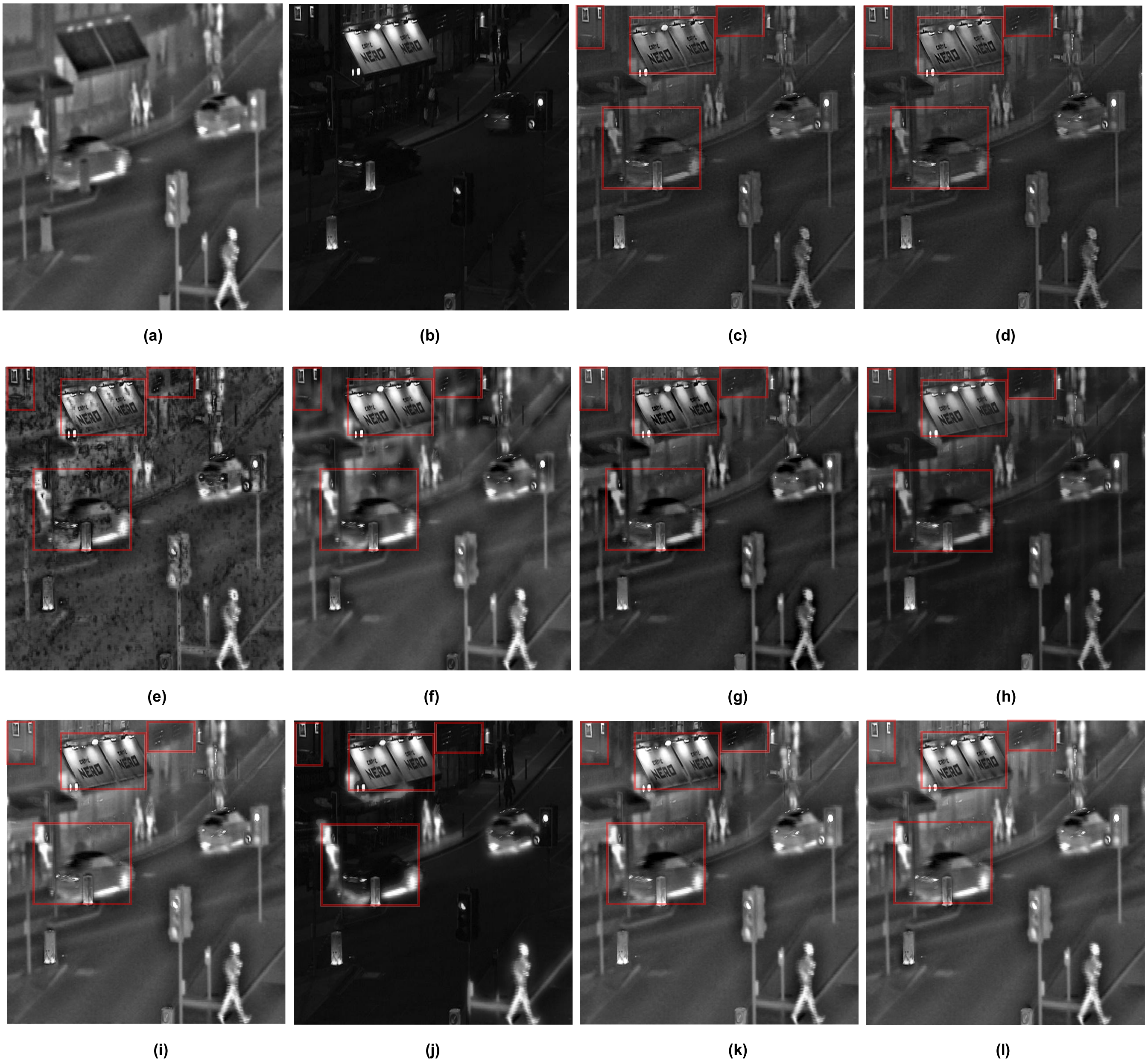
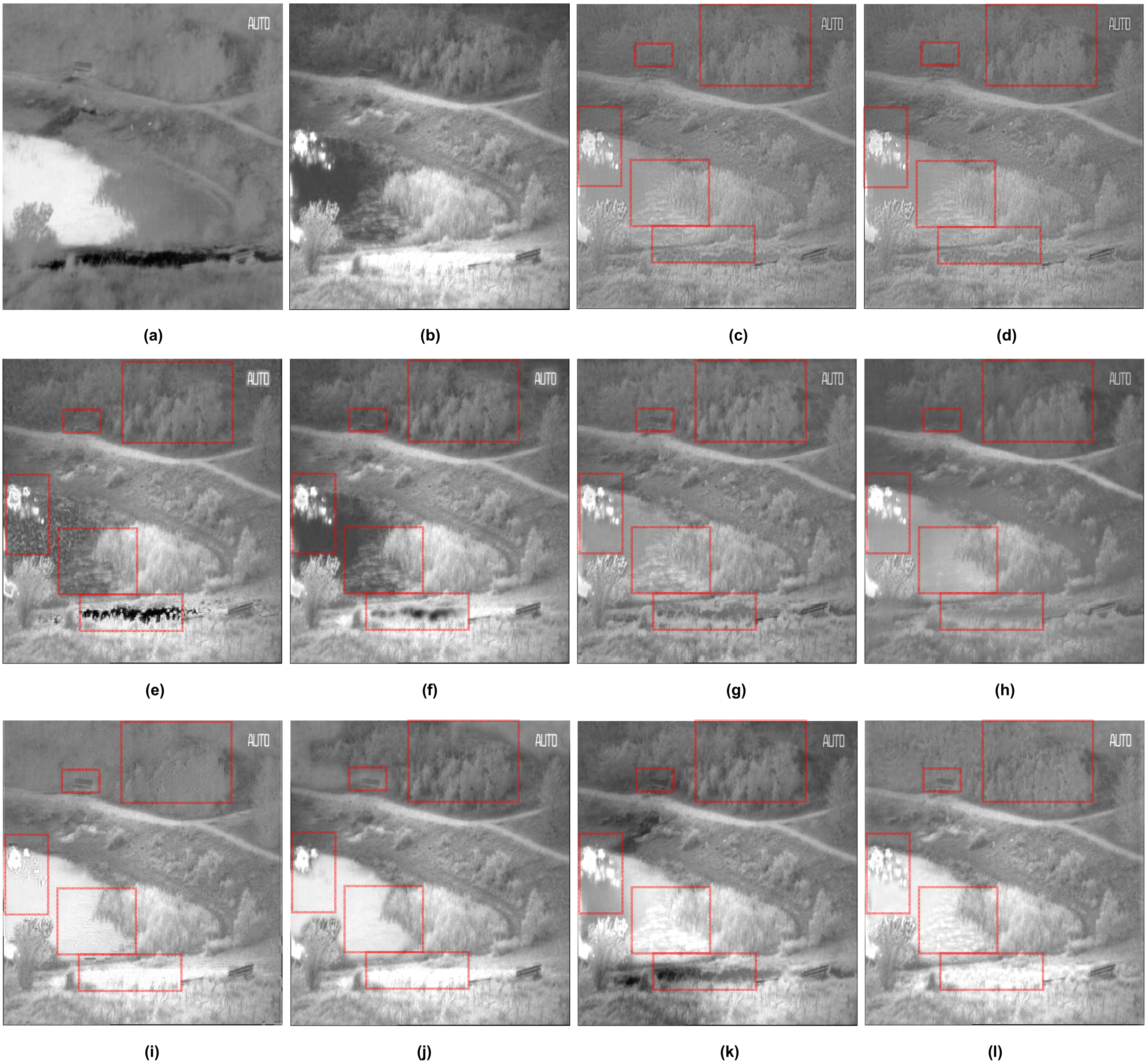


FIGURE 8. Source images and fused images using different methods in the third pair of IR and VI. (a) IR. (b) VI. (c) NSCT. (d) NSST. (e) CBF. (f) GF. (g) SD. (h) LLRR. (i) DCLSF. (j) CNN1. (k) CNN2. (l) Proposed method.

It can be seen from Table 6 that the fusion image index of the algorithm proposed in this paper is in the top three of the ten evaluation indexes. It is ranked first in indicators MV and CRI, second in indicators SF, EN, AG and Q_{CB}, and third in indicators MI, SD, Q_{ABF} and MS-SSIM. Especially in the CRI index, it leads the second to fourth places by 4.45, 15.56, and 24.45 percentage points. It shows that the algorithm proposed in this paper has a better overall effect on 21 pairs of image sets. It can be seen from Table 7 that the EN index of the fused image of our proposed algorithm is in the top three overall on the 21 pairs of image sets, but it is ranked fifth on image13 and image17, which is differs from the first place by 0.5229 and 0.3885 respectively. The gap is not big. It can also be seen from Figure 11 that the EN index of the image fused by the algorithm proposed in this paper is shown in the red pentagon, which is at the forefront of the 21 pairs of image sets as a whole. The EN index reflects the average information and texture richness of the image. It can be concluded that the fusion image of the algorithm proposed in this paper contains more information and rich texture information than other algorithms, and the overall fusion effect is the best.



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

FIGURE 9. Source images and fused images using different methods in the fourth pair of IR and VI. (a) IR. (b) VI. (c) NSCT. (d) NSST. (e) CBF. (f) GF. (g) SD. (h) LLRR. (i) DCLSF. (j) CNN1. (k) CNN2. (l) Proposed method.

(6) Comparison of calculation efficiency

The running time index cannot be ignored for the objective evaluation of the image fusion method. In this paper, four pairs of IR and VI images of different sizes in Figure 5 are used as examples of the calculation cost analysis, and the average value is run 10 times respectively. Its running efficiency is shown in Table 8. To conveniently compare and analyze the algorithms, the longest and shortest running times of the algorithm are bolded. The bold red value indicates the longest running time among all methods, and the bold blue value indicates the shortest running time among all methods.

Table 8 shows that when the size of the images increases, the average running time of each algorithm increases. To facilitate comparative analysis, the maximum and minimum average fusion times of various algorithms of the four groups of image sets are displayed in bold red and bold blue, respectively.

The GF algorithm has the shortest running time for the four image sets; the largest-size "Lake" image pair algorithm run for only 1.0273 seconds; the CNN1 algorithm has the longest running time on the four image sets; the running time of the largest-size "Lake" image pair algorithm was 208.2961 seconds. In the "Kayak" image pair, the running

TABLE 5. Fusion quality indices with different methods for the fourth pair of IR and VI.

	AV	MI	SD	SF	EN	AG	Q_{CB}	Q_{ABF}	MS-SSIM	CRI
NSCT	135.1179	1.6994	26.2381	11.7575	6.6275	4.0140	0.5453	0.5169	0.9037	0.3667
NSST	135.1186	1.6981	26.2107	11.7677	6.6264	4.0206	0.5439	0.5103	0.9005	0.3333
CBF	137.6852	**3.6092**	40.0125	**13.6533**	7.1999	**4.6937**	0.6012	0.5556	0.8431	0.6444
GF	137.6996	3.2521	**44.1348**	**12.2395**	**7.3738**	**4.2910**	**0.7779**	**0.6946**	0.8736	**0.7667**
SD	134.7636	1.5420	28.8305	11.2155	6.7620	3.8314	0.6013	0.5085	**0.9348**	0.3444
LLRR	143.9259	1.8145	27.4903	7.1417	6.6864	2.4692	0.5056	0.3858	0.8320	0.2556
DCLSF	**160.7758**	**3.4366**	42.4097	11.6815	7.0671	4.1080	0.5532	0.5355	0.8538	0.5778
CNN1	**154.7611**	**6.3248**	**47.1403**	11.5576	7.3043	3.8759	**0.6681**	**0.6384**	0.8061	0.6667
CNN2	141.4378	2.0700	**47.9781**	11.4933	**7.4159**	4.1364	**0.6372**	0.5953	**0.9358**	**0.7222**
Proposed	**160.8361**	3.3141	42.8586	**12.1502**	**7.4170**	**4.2925**	0.6186	**0.6051**	**0.9142**	**0.8222**

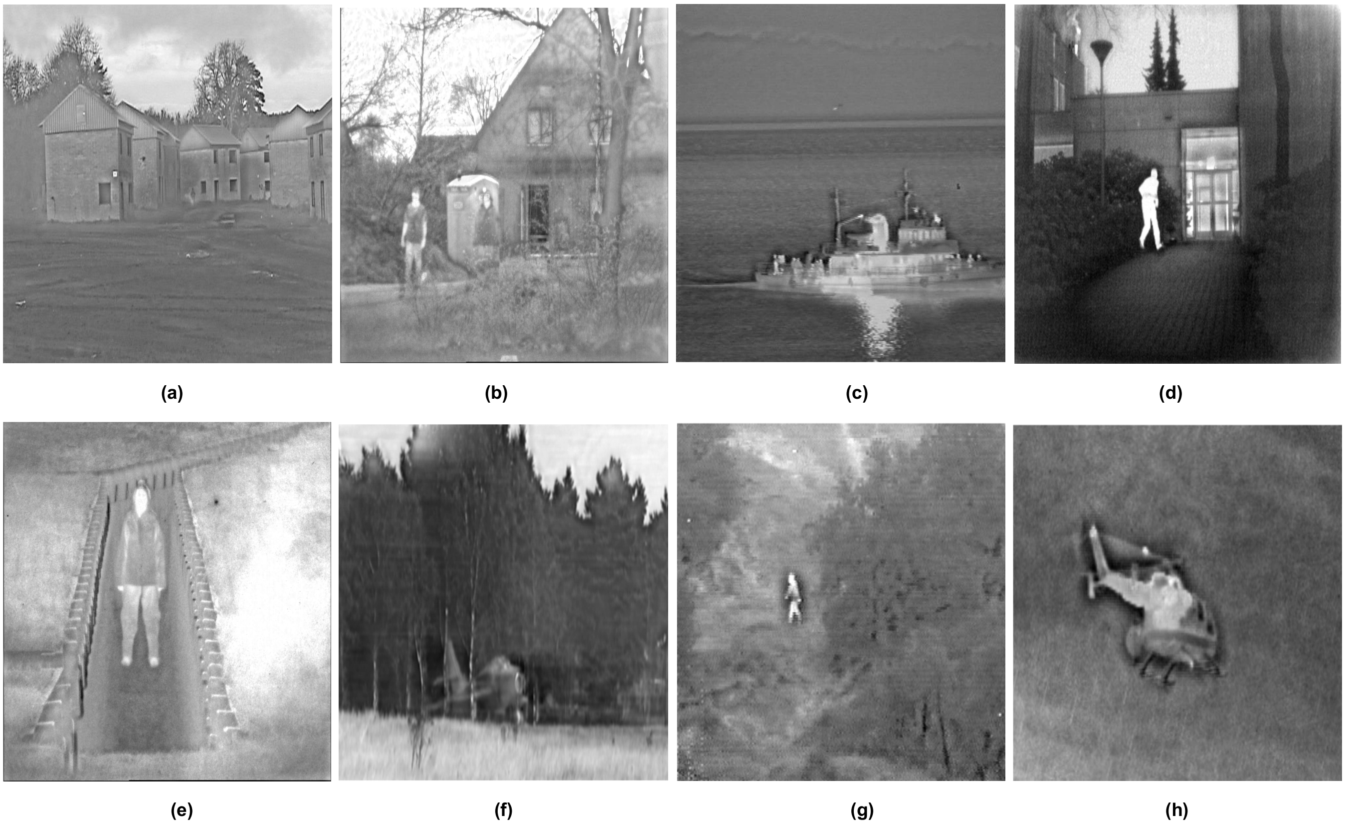

(a) (b) (c) (d)
(e) (f) (g) (h)

FIGURE 10. 8 pairs of fusion images using the proposed method from 21 pairs of infrared and visible-light images.

time of this algorithm is only shorter than that of CNN1 and higher than those of the other 8 algorithms. In the "UNCamp" image pair, the running time of the algorithm in this paper is shorter than the deep learning CNN1 and CNN2 algorithms and longer than the other 7 algorithms. In the "Road" image pair, the efficiency of the algorithm in this paper has been greatly improved and is lower than those of CNN1, LLRR, CNN2 and DCLSF algorithms. In the "Lake" image pair, the efficiency of this algorithm continues to improve, which is far lower than the 124.987 seconds for CNN1 algorithm and 100.0626 seconds for LLRR algorithm. When the image size continues to increase, the running efficiency of the algorithm in this paper will continue to improve, which has obvious advantages compared to multi-scale decomposition algorithms and deep learning.

However, even in large-scale image collections, the running time of the algorithm in this paper is only medium and much longer than the CBF, GF, SD and NSST algorithms.

TABLE 6. The average values of fusion quality indexes with different methods for 21 pairs fused images.

average	AV	MI	SD	SF	EN	AG	Q_{CB}	Q_{ABF}	MS-SSIM	CRI
NSCT	111.1183	1.7394	24.3786	10.9040	6.3199	4.0809	0.5079	0.5161	0.9158	0.3667
NSST	111.1185	1.7352	24.3360	10.9297	6.3184	4.0912	0.5085	0.5133	0.9131	0.3556
CBF	112.8376	2.4389	35.9078	**13.5377**	6.8569	**5.2872**	0.4933	0.4537	0.7087	0.5333
GFF	111.5111	**4.6585**	40.3927	10.9511	**6.9309**	4.1213	**0.6119**	**0.6092**	0.8124	**0.7111**
SD	111.1774	1.6573	28.2268	10.3089	6.5251	3.8814	0.5289	0.5093	**0.9418**	0.3889
LRR	120.1054	1.8363	25.8377	7.2536	6.3567	2.7251	0.48867	0.4138	0.8757	0.2667
SDLSF	**142.2701**	3.2486	37.2841	10.8954	6.7548	4.1353	0.4888	0.5343	0.8876	0.5667
CNN1	**132.9738**	**5.6550**	**42.2639**	10.4719	6.7582	3.7782	0.5411	**0.6085**	0.8032	0.6222
CNN2	122.6935	2.6643	**45.5192**	**11.2419**	**7.0654**	**4.2499**	**0.5619**	0.5909	**0.9437**	**0.8222**
Proposed	**142.2942**	**3.5159**	**41.0414**	**11.4180**	**7.0288**	**4.2793**	**0.5621**	**0.5944**	**0.9252**	**0.8667**

TABLE 7. The EN values for 21 fused images which obtained by fusion methods.

Methods	NSCT	NSST	CBF	GFF	SD	LRR	SDLSF	CNN1	CNN2	Proposed
image1	6.3754	6.3723	6.5349	6.6969	6.6195	6.5392	6.7691	**6.8447**	**7.0531**	**6.8171**
image2	6.0054	6.0018	6.7027	6.6993	6.4136	6.0297	**6.7716**	5.7575	**6.7090**	**6.7891**
image3	5.0687	5.0706	**6.4763**	**6.5419**	5.2264	5.0760	5.8786	6.0161	6.3728	**6.9344**
image4	6.4165	6.4156	**7.0708**	**7.1137**	6.6887	6.3750	6.7818	6.7376	6.9670	**6.9926**
image5	6.0081	6.0072	**6.5628**	6.3295	6.1182	6.0848	6.2136	6.3074	**6.6163**	**6.5377**
image6	6.5792	6.5763	7.0517	**7.2684**	6.7685	6.6473	6.8664	6.8521	**7.3352**	**7.1062**
image7	6.2799	6.2749	6.5221	6.5345	6.5052	6.4572	**6.8098**	**6.7977**	6.6542	**6.8593**
image8	6.8443	6.8428	7.3627	7.5237	7.0677	6.8550	7.1916	**7.5713**	**7.5477**	**7.5591**
image9	6.8239	6.8189	7.3540	**7.4746**	7.0785	6.8152	7.1555	**7.4723**	7.4693	**7.4947**
image10	6.7682	6.7672	7.0406	**7.2801**	6.9640	6.6068	6.7167	7.0136	**7.4753**	**7.1101**
image11	5.1044	5.1013	**5.7602**	4.8898	5.3011	5.2809	5.6070	4.9172	**5.7856**	**5.7891**
image12	6.6650	6.6640	6.7770	6.9896	6.8425	6.7016	**7.2023**	**7.1977**	7.1012	**7.2033**
image13	6.4274	6.4282	6.8778	**7.1144**	6.6129	6.5547	6.7381	**7.2300**	**7.3113**	6.8084
image14	6.6275	6.6264	7.1999	**7.3738**	6.7620	6.6864	7.0671	7.3043	**7.4159**	**7.4170**
image15	6.9078	6.9022	7.4231	**7.5345**	7.0737	6.8926	7.1718	7.1825	**7.5425**	**7.4324**
image16	6.5652	6.5638	6.9569	7.2032	6.8392	6.6399	7.1467	**7.3036**	**7.4114**	**7.2496**
image17	5.6215	5.6211	**6.7636**	**6.9347**	5.9323	5.7627	6.6707	6.6918	**7.0656**	6.6771
image18	5.6915	5.6972	5.9449	6.0011	5.8353	5.7537	**6.7215**	6.6633	**6.8165**	**6.8473**
image19	6.3202	6.3171	6.5205	6.9087	6.5571	6.2539	6.7278	**6.9609**	**7.1106**	**6.9961**
image20	6.9594	6.9586	**7.7084**	**7.7217**	7.0533	6.924	6.5052	5.8328	7.1018	**7.5519**
image21	6.6602	6.6592	7.3834	**7.4162**	6.7673	6.5548	7.1388	7.2678	**7.5121**	**7.5629**

The reason is that CBF and GF algorithms are spatial fusion algorithms, and their processing is pixel-based and does not undergo multi-scale decomposition, so the time is shorter. Although SD and NSST are multi-scale decomposition algorithms, SD only performs two-level decomposition, which is far lower than the decomposition scale of the algorithm in this paper. The LSWT-NSST algorithm is based on NSST multi-scale decomposition; to obtain a better fusion effect, the decomposition scale is larger than that based on the traditional NSST algorithm, so the running time will be longer than that of NSST algorithm. NSCT, LLRR and DCLSF are multi-scale decomposition algorithms and relatively complicated. The LLRR and DCLSF algorithms introduce many mathematical calculations to improve the fusion rules. The LSWT algorithm in this algorithm runs faster than the SWT algorithm of the DCLSF algorithm. Because we introduce NSST into the LSWT algorithm, the performance is mediocre in small-size image sets, but the efficiency is significantly improved in large or extra-large image sets. CNN1 and CNN2 are recently popular deep learning fusion algorithms. The image fusion quality of this algorithm is high, but it runs slowly on the CPU, as shown for the "Lake" image set.

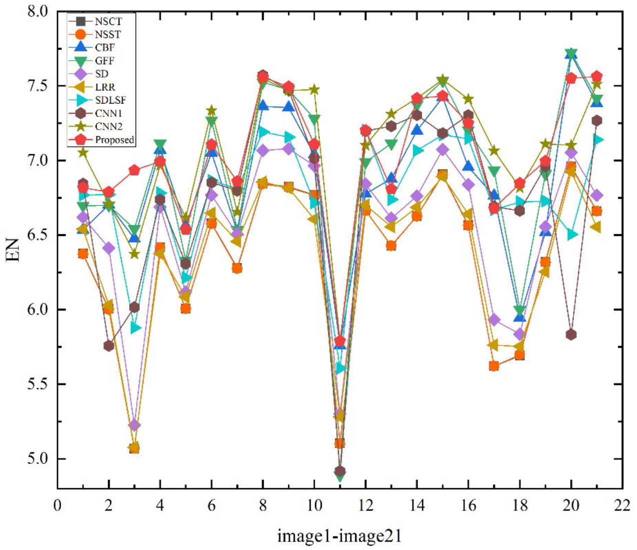


FIGURE 11. Plotting EN for 21 fused images obtained by the fusion methods experimentally compared.

TABLE 8. Average run time of different algorithms.

	Kayak(s)	UNCamp(s)	Road(s)	Lake(s)
NSCT	11.2950	17.4435	55.6885	79.8185
NSST	5.6036	9.7059	22.2129	43.2628
CBF	2.9292	3.9528	12.7105	18.0333
GF	**0.2672**	**0.4869**	**0.7597**	**1.0273**
SD	0.4118	0.6459	1.1057	1.3998
LLRR	17.6652	22.1356	108.7894	183.3717
DCLSF	16.5526	24.9758	80.9468	113.6166
CNN1	**48.1771**	**61.6193**	**158.2881**	**208.2961**
CNN2	25.6059	31.9061	83.4937	95.0167
Proposed	26.1499	31.3239	62.5079	83.3091

However, with the development of GPU and other hardware acceleration technologies and the reduction of costs, the running efficiency of deep learning will continue to accelerate.

The algorithm in this paper takes advantage of the fast calculation speed, saved memory, reduced storage space, and easy realization of inverse transformation of LSWT, which improves the running efficiency of the algorithm. It not only guarantees the quality of the fusion image but also reduces the running time of the algorithm. Because the running efficiency of the algorithm in this paper is medium, the running speed of the algorithm needs to be further improved.

V. CONCLUSION

This paper proposes an infrared and visible image fusion algorithm based on LSWT and NSST, which effectively utilizes the complementary advantages of LSWT and NSST multi-scale decomposition. First, the LSWT algorithm is used to perform a multi-scale decomposition of the IR and VI images to obtain low and high-frequency sub-band coefficients. Second, the high and low-frequency sub-bands of the LSWT domain are used in the multi-scale decomposition of NSST to further extract the target features and detailed features of the source image. The NSST algorithm can compensate for the insufficiency of the LSWT algorithm in decomposing the continuous curves and edges of the image; the LSWT algorithm can compensate for the disadvantage of the NSST algorithm in decomposing subtle features of the image. Through the high and low-frequency coefficients of the NSST domain, the target features of the low-frequency sub-bands of the LSWT domain and detailed features of the high-frequency sub-bands of the LSWT domain can be enhanced. Third, by combining the characteristics of IR and VI images and the characteristics of high and low-frequency coefficient representation, we design different fusion strategies for image fusion rules. The low-frequency part introduces the DCT algorithm and LSF features; then, adaptive weighted fusion rules are designed by LSF to enhance the regional characteristics of the DCT. The high-frequency part combines the imaging mechanism of human vision to design an improved regional contrast fusion strategy. Finally, IDCT, INSST and ILSWT algorithms are used to generate the final fusion image.

This paper conducts individual fusion experiments on four groups of classic infrared and visible-light images and conducts overall fusion experiments on 21 pairs of classic infrared and visible-light image sets. Nine classic and advanced image fusion algorithms are selected to compare fused images' subjective and objective effects with the algorithm proposed in this paper. Based on the objective evaluation, nine classic evaluation indices are selected, and a comprehensive ranking index is designed, which realizes the comprehensive consideration of image brightness, chroma, contrast, etc. The experimental results were comprehensively analyzed from the subjective and objective aspects. In terms of visual perception, the fusion image target with clear edges and high contrast in this paper is prominent. The ten objective evaluation indices are also higher than other algorithms, and the running efficiency is moderate. In summary, both subjective vision and objective evaluation show that the algorithm fusion image in this paper has the best effect and high quality and is an effective IR and VI image fusion algorithm.

Although the algorithm in this paper has a better fusion effect and a higher evaluation index than other algorithms, it faces some limitations in the running time of the algorithm, which reduces its performance of the algorithm to a certain extent. With the continuous development of deep learning technology, deep learning has achieved remarkable results in CV fields such as target detection, image recognition and image noise reduction, and the application in image fusion will also become more popular. The traditional image fusion method has certain advantages in some fields. In future research, the authors will combine advanced deep network models and traditional image processing algorithms to further extract multi-dimensional features of the image and perform unsupervised end-to-end image fusion. In terms of running efficiency, the suitable GPU acceleration technology or FPGU real-time processing technology is adopted to further improve the running efficiency and real-time performance of the algorithm. By improving the algorithm, we hope to obtain better fusion results.

ACKNOWLEDGMENT

The authors would like to thanks Prof. K. Ji from Yunnan Observatory, Chinese Academy of Sciences, for their valuable comments and suggestions for this study.

REFERENCES

[1] J. Zhao and H. Li, "An image fusion algorithm based on multi-resolution decomposition for functional magnetic resonance images," *Neurosci. Lett.*, vol. 487, no. 1, pp. 73–77, Jan. 2011.

[2] S. Wang and Y. Zhao, "A novel patch-based multi-exposure image fusion using super-pixel segmentation," *IEEE Access*, vol. 8, pp. 39034–39045, Feb. 2020.

[3] B.-L. Jian, W.-L. Chu, Y.-C. Li, and H.-T. Yau, "Multifocus image fusion using a sparse and low-rank matrix decomposition for Aviator's night vision goggle," *Appl. Sci.*, vol. 10, no. 6, p. 2178, Mar. 2020.

[4] Y. Kinoshita and H. Kiya, "Scene segmentation-based luminance adjustment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4101–4116, Aug. 2019.

[5] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion," *Inf. Fusion*, vol. 33, no. 33, pp. 100–112, Jan. 2017.

[6] X. Liu, Q. Liu, and Y. Wang, "Remote sensing image fusion based on two-stream fusion network," *Inf. Fusion*, vol. 55, pp. 1–15, Mar. 2020.

[7] X. Qian, S. Lin, G. Cheng, X. Yao, H. Ren, and W. Wang, "Object detection in remote sensing images based on improved bounding box regression and multi-level features fusion," *Remote Sens.*, vol. 12, no. 1, p. 143, Jan. 2020.

[8] M. Eslami and A. Mohammadzadeh, "Developing a spectral-based strategy for urban object detection from airborne hyperspectral TIR and visible data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 5, pp. 1808–1816, May 2016.

[9] X. Lu, J. Zhang, T. Li, and G. Zhang, "Synergetic classification of long-wave infrared hyperspectral and visible images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3546–3557, Jul. 2015.

[10] A. C. Muller and S. Narayanan, "Cognitively-engineered multisensor image fusion for military applications," *Inf. Fusion*, vol. 10, no. 2, pp. 137–149, Apr. 2009.

[11] X. Qian, L. Han, and Y. Cheng, "An object tracking method based on local matting for night fusion image," *Infr. Phys. Technol.*, vol. 67, pp. 455–461, Nov. 2014.

[12] L. Li, H. Ge, Y. Tong, and Y. Zhang, "Face recognition using Gabor-based feature extraction and feature space transformation fusion method for single image per person problem," *Neural Process. Lett.*, vol. 47, no. 3, pp. 1197–1217, Jun. 2018.

[13] P. Chai, X. Luo, and Z. Zhang, "Image fusion using quaternion wavelet transform and multiple features," *IEEE Access*, vol. 5, pp. 6724–6734, Mar. 2017.

[14] Q. Li, J. Du, and L. Xu, "Visible and infrared video fusion using uniform discrete curvelet transform and spatial-temporal information," *Chin. J. Electron.*, vol. 24, no. 4, pp. 761–766, Oct. 2015.

[15] S. Liu, J. Wang, Y. Lu, H. Li, J. Zhao, and Z. Zhu, "Multi-focus image fusion based on adaptive dual-channel spiking cortical model in non-subsampled shearlet domain," *IEEE Access*, vol. 7, pp. 56367–56388, Feb. 2019.

[16] Q. Zhang, Y. Liu, R. S. Blum, J. Han, and D. Tao, "Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review," *Inf. Fusion*, vol. 40, pp. 57–75, Mar. 2018.

[17] Z. Wang, J. Xu, X. Jiang, and X. Yan, "Infrared and visible image fusion via hybrid decomposition of NSCT and morphological sequential toggle operator," *Optik*, vol. 201, Jan. 2020, Art. no. 163497.

[18] G. Zhang, B. Xu, K. Zhang, J. Hou, T. Xie, X. Li, and F. Liu, "Research on a noise reduction method based on multi-resolution singular value decomposition," *Appl. Sci.*, vol. 10, no. 4, p. 1409, Feb. 2020.

[19] S. Bonny, Y. J. Chanu, and K. M. Singh, "Speckle reduction of ultrasound medical images using bhattacharyya distance in modified non-local mean filter," *Signal, Image Video Process.*, vol. 13, no. 2, pp. 299–305, Mar. 2019.

[20] W. Li, L. Jia, and J. Du, "Multi-modal sensor medical image fusion based on multiple salient features with guided image filter," *IEEE Access*, vol. 7, pp. 173019–173033, Dec. 2019.

[21] H. Talebi and P. Milanfar, "Global image denoising," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 755–768, Feb. 2014.

[22] D. P. Bavirisetti and R. Dhuli, "Two-scale image fusion of visible and infrared images using saliency detection," *Infr. Phys. Technol.*, vol. 76, pp. 52–64, May 2016.

[23] J. Han, E. J. Pauwels, and P. de Zeeuw, "Fast saliency-aware multi-modality image fusion," *Neurocomputing*, vol. 111, pp. 70–80, Jul. 2013.

[24] A. Azarang, H. E. Manoochehri, and N. Kehtarnavaz, "Convolutional autoencoder-based multispectral image fusion," *IEEE Access*, vol. 7, pp. 35673–35683, Mar. 2019.

[25] R. Hou, D. Zhou, R. Nie, D. Liu, L. Xiong, Y. Guo, and C. Yu, "VIF-net: An unsupervised framework for infrared and visible image fusion," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 640–651, Jan. 2020.

[26] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.

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

[28] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 16, no. 03, May 2018, Art. no. 1850018.

[29] T. Chu, Y. Tan, Q. Liu, and B. Bai, "Novel fusion method for SAR and optical images based on non-subsampled shearlet transform," *Int. J. Remote Sens.*, vol. 41, no. 12, pp. 4590–4604, Jun. 2020.

[30] Y. Chai, H. F. Li, and J. F. Qu, "Image fusion scheme using a novel dual-channel PCNN in lifting stationary wavelet domain," *Opt. Commun.*, vol. 283, no. 19, pp. 3591–3602, Oct. 2010.

[31] L.-Y. Hsu and H.-T. Hu, "Blind image watermarking via exploitation of inter-block prediction and visibility threshold in DCT domain," *J. Vis. Commun. Image Represent.*, vol. 32, pp. 130–143, Oct. 2015.

[32] L. Yu, Q. Han, X. Niu, S. M. Yiu, J. Fang, and Y. Zhang, "An improved parameter estimation scheme for image modification detection based on DCT coefficient analysis," *Forensic Sci. Int.*, vol. 259, pp. 200–209, Feb. 2016.

[33] X. Jin, R. Nie, D. Zhou, Q. Wang, and K. He, "Multifocus color image fusion based on NSST and PCNN," *J. Sensors*, vol. 2016, pp. 1–12, Nov. 2016.

[34] X. Jin, Q. Jiang, S. Yao, D. Zhou, R. Nie, J. Hai, and K. He, "A survey of infrared and visual image fusion methods," *Infr. Phys. Technol.*, vol. 85, pp. 478–501, Sep. 2017.

[35] G. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Appl. Comput. Harmon. Anal.*, vol. 25, no. 1, pp. 25–46, Jul. 2008.

[36] W. Kong, "Technique for gray-scale visual light and infrared image fusion based on non-subsampled shearlet transform," *Infr. Phys. Technol.*, vol. 63, pp. 110–118, Mar. 2014.

[37] J. Adu, J. Gan, Y. Wang, and J. Huang, "Image fusion based on nonsubsampled contourlet transform for infrared and visible light image," *Infr. Phys. Technol.*, vol. 61, pp. 94–100, Nov. 2013.

[38] B. K. Shreyamsha Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal, Image Video Process.*, vol. 9, no. 5, pp. 1193–1204, Jul. 2015.

[39] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[40] H. Li and X.-J. Wu, "Infrared and visible image fusion using latent low-rank representation," 2018, *arXiv:1804.08992*. [Online]. Available: http://arxiv.org/abs/1804.08992

[41] X. Jin, Q. Jiang, S. Yao, D. Zhou, R. Nie, S.-J. Lee, and K. He, "Infrared and visual image fusion method based on discrete cosine transform and local spatial frequency in discrete stationary wavelet transform domain," *Infr. Phys. Technol.*, vol. 88, pp. 1–12, Jan. 2018.

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

[43] 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.

[44] H. Li, S. Wei, and Y. Chai, ''Multifocus image fusion scheme based on feature contrast in the lifting stationary wavelet domain,'' *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 39, Feb. 2012.

[45] G. Piella and H. Heijmans, ''A new quality metric for image fusion,'' in *Proc. Int. Conf. Image Process.*, vol. 3, Sep. 2003, pp. III-173.

[46] Z. Wang, E. P. Simoncelli, and A. C. Bovik, ''Multiscale structural similarity for image quality assessment,'' in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, 2003, pp. 1398–1402.

[47] H. Li, X.-J. Wu, and J. Kittler, ''Infrared and visible image fusion using a deep learning framework,'' in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 2705–2710.

LI JUNWU (Graduate Student Member, IEEE) received the master's degree in traffic information engineering and control from the Automation and Electrical Engineering College, Lanzhou Jiaotong University, in 2019. He is currently pursuing the Ph.D. degree in computer science and technology with the Faculty of Information Engineering and Automation, Kunming University of Science and Technology.

His current research interests include deep learning, image processing, and image fusion.

BINHUA LI received the B.S. degree in industrial automation from the Jiangxi Institute of Technology (currently Nanchang University), in 1984, the M.S. degree in electrical engineering from Chongqing University, in 1987, and the Ph.D. degree in astrophysics from the Graduate School of Chinese Academy of Sciences (Yunnan Observatory), in 2002.

He is currently a Professor with the Faculty of Information Engineering and Automation, Kunming University of Science and Technology. His research interests include imaging technology and image processing, FPGA design and modern electronic systems, and astronomical telescope control and data acquisition.

YAOXI JIANG received the B.S. degree from Northwestern Polytecnical University, in 1998, and the M.S. degree from Kunming University of Science and Technology, in 2005. She was a Visiting Scholar with the University of California at San Diego, in 2015. She is currently pursuing the Ph.D. degree with the Kunming University of Science and Technology.

Her research interests include smart power systems and deep learning.