Tech Science Press

Infrared and Visible Image Fusion Based on NSST and RDN

Peizhou Yan[1], Jiancheng Zou[2,*], Zhengzheng Li[1] and Xin Yang[3]

[1]College of Electrical and Control Engineering, North China University of Technology, Beijing, 100043, China
[2]College of Sciences, North China University of Technology, Beijing, 100043, China
[3]Department of Computer Science, Middle Tennessee State University, Murfreesboro, 37132, USA
*Corresponding Author: Jiancheng Zou. Email: zjc@ncut.edu.cn

Abstract: Within the application of driving assistance systems, the detection of driver's facial features in the cab for a spectrum of luminosities is mission critical. One method that addresses this concern is infrared and visible image fusion. Its purpose is to generate an aggregate image which can granularly and systematically illustrate scene details in a range of lighting conditions. Our study introduces a novel approach to this method with marked improvements. We utilize non-subsampled shearlet transform (NSST) to obtain the low and high frequency sub-bands of infrared and visible imagery. For the low frequency sub-band fusion, we incorporate the local average energy and standard deviation. In the high frequency sub-band, a residual dense network is applied for multiscale feature extraction to generate high frequency sub-band feature maps. We then employ the maximum weighted average algorithm to achieve high frequency sub-band fusion. Finally, we transform the fused low frequency and high frequency sub-bands by inverse NSST. The results of the experiment and application in real world driving scenarios proved that this method showed excellent performance when objectively compared to the indexing from the other contemporary, industry standard algorithms. In particular, the subjective visual effect, fine texture, and scene were fully expressed, the target's edge distinct was pronounced, and the detailed information of the source image was exhaustively captured.

1 Introduction

Infrared image is a kind of radiation image which distinguishes the target from the background according to the radiation difference. It has the ability to capture the thermal radiation of the object, but it cannot reflect the real scene information. Visible image is a reflection image with multiple high-frequency components, which can show the scene details under certain lighting conditions. However, it is difficult to capture all the useful information in the environment of low light intensity and heavy fog at night. They have strong complementarity in reflecting scene information.

Image fusion is to analyze and extract the complementary information from two or more images of the same target to generate a composite image which can accurately describe the scene information. Image fusion is of great significance for image recognition and understanding. It has been widely used in computer vision, remote sensing [1], super-resolution [2], target detection, and recognition and other fields. In recent years, more and more driving assistance systems have been applied in the automotive field. As an important part, the driver's face information is usually collected in real time to monitor the driver's fatigue condition. But the light changes in the cab are very obvious, and it is usually in a low illumination, which means that it is difficult to obtain a clear driver's face image from a single camera all day long. However, the fusion technology of infrared and visible images has made it possible.

The image fusion algorithm includes two directions: the traditional fusion algorithms based on model driven and the fusion algorithms based on neural network driven by data.

1.1 Fusion Algorithms Based on Model Driven

The traditional fusion algorithms are based on the models such as multi-scale decomposition [3], sparse representation [4,5], and guided filter [6].

Son [7] used Block Matching 3D (BM3D) [8] to pre-denoise the visible image and then fused the denoised visible image with the near-infrared image. The problem of step-by-step processing was that the near-infrared image information was not used in noise reduction, which would lead to poor noise reduction effect and more information loss.

Multi-scale fusion [9] was used to calculate the wavelet scale map on each layer. Then, wavelet coefficients were de-noising based on a scale map to suppress high-frequency noise. Finally, after noise reduction, the coefficients were fused and transmitted upward as the visible high frequency information of the next layer. However, it was easy to produce halos and other defects in the edge area.

Based on the model of the retinex, Jung et al. [10] fused the near-infrared information in the dark area with large visible noise, so as to reduce the noise of the fusion image. However, the method of improving SNR by integrating near infrared brightness would also lead to color inaccuracy.

Jee et al. [11] transformed the fusion operation into near-infrared assisted visible noise reduction and superimposed near-infrared texture. The essence of the denoising algorithm was guided filtering [12], which also has the problem of color inaccuracy.

Inspired by the nonlocal mean de-noising filter, Zhang et al. [13] proposed a group sparse representation model. The structural group matrix was constructed by using nonlocal similarity, and the singular value decomposition and iteration of the structure group matrix was carried out to obtain the sub dictionary corresponding to the structure group.

Zhang et al. [14] applied group sparsity theory to remote sensing image fusion, and obtained sub dictionary and sparse coefficient through SVD within the group. After that, the general component fusion framework was used for fusion, and the experimental results effectively retained the source image information, but the effect of vegetation fusion was not ideal.

Li et al. [15] proposed a nonlocal sparse K-SVD dictionary learning image fusion method. The structure group matrix was constructed by using nonlocal similarity. The sparse coefficient was calculated by learning dictionary of the structure group matrix. It has shown good results in many kinds of image fusions. However, due to the high dimension of the matrix obtained by vectorization of the structure group, the complexity of dictionary learning increases significantly.

1.2 Fusion Algorithms Based on Neural Network

In recent years, deep learning technology has shown remarkable effects in image processing applications. Due to the lack of high-quality training data set, there is not much research on the fusion of visible and infrared images based on deep learning [16].

Jung et al. [17] produced simulated noise images by adding Gaussian noise to visible images. But in the actual scene of low illumination, noise in visible image is more complex, and intrinsic noise such as device noise accounts for a large proportion. As a result, it was not effective for the fusion of real data.

On the other hand, the neural-network-based fusion of visible and near-infrared images requires high registration accuracy, and most of the existing data sets are unable to meet the requirements. Using scale invariant feature transform (SIFT) method, Brown et al. [18] provided a visible-infrared image dataset for registration, but the registration accuracy still could not meet the requirements for training the deep learning model.

Tang et al. [19] proposed an end-to-end convolutional neural network for the fusion of near-infrared and visible images. The fused image could combine the advantages of the signal-to-noise ratio of a NIR image and the color of a visible image.

Li et al. [20] combined convolutional layers, fusion layer and dense block in the encoding network, and the output of each layer was connected to all other layers. Different fusion strategies were applied to fuse the features extracted from source images. It showed excellent performance in both subjective and objective indicators.

Li et al. [21] designed AttentionFGAN to highlight the typical parts existing in infrared and visible images, in which multi-scale attention mechanism was introduced to both generator and discriminator of GAN. The multi-scale attention mechanism could not only capture comprehensive spatial information to help generator focus on the foreground target information of infrared image and background detail information of visible image, but also constrain the discriminators focus more on the attention regions rather than the whole input image. This method preserves the effective detail features of the source image to the greatest extent.

In the past research, several open-source datasets were used for training and testing, such as TNO [22–26] and INO [22], and there were rare studies on the fusion of driver's face image in driving scene. In our case, we captured the driver's facial images via a camera with infrared and colored lenses, and applied our method in the fusion task, which came out to be effective in fusing the driver's facial features.

2 System Architecture

We built a system based on non-subsampled shearlet transform (NSST) [27] and residual dense network (RDN) [28] named NSST-RDN, as shown in Fig. 1.

NSST is an improvement of shearlet transform. It inherits the characteristics of shearlet transform and avoids pseudo-Gibbs. Compared with the wavelet transform, it has the advantages of low complexity and high efficiency, which has been widely used in image segmentation, edge detection, recognition, and other fields.

RDN is a combination of residual network (ResNet) [29] and dense network (DenseNet). It proposes a more extreme dense connection strategy, that is, each layer will accept all previous layers as additional input. ResNet is a short-circuit connection between each layer and previous layers (generally 2-3 layers). and it is established by adding elements. DenseNet is to concatenate each layer with all previous layers in the channel dimension to reuse the features. It effectively alleviates the problem of gradient vanishing and encourages feature reuse, which means a reduction in the number of parameters. It can improve the flow of image information and gradient, and make the training process easier.
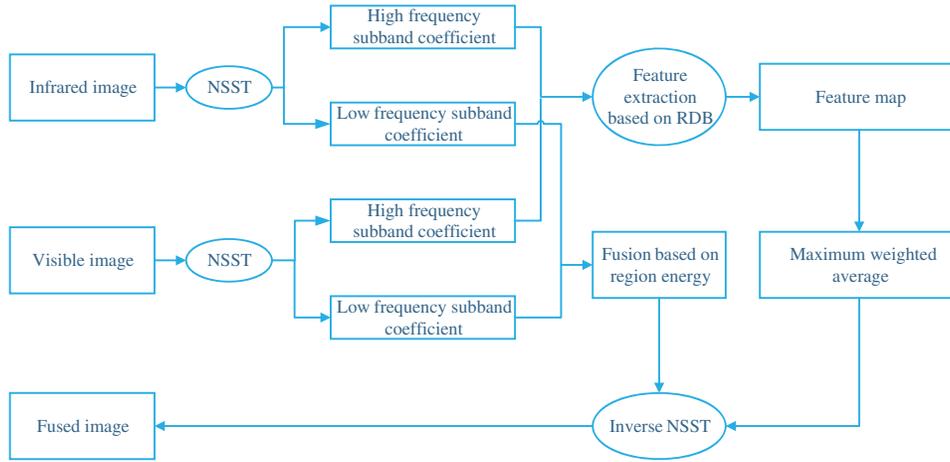
Figure 1: NSST-RDN system architecture image fusion

The steps are as follows.

Step 1. NSST

Step 1.1. The non-sub sampled Laplacian pyramid filter bank (NSLP) was used to achieve multi-scale decomposition, which ensures translation invariance and suppress pseudo-Gibbs [30].

Step 1.2. Shear filter (SF) was used to achieve direction localization. After N-level decomposition of the source image, one low-frequency sub-band image and N high-frequency sub-band images with the same size but different scales were obtained.

Step 2. The average energy in the neighborhood of low frequency sub-band coefficients was calculated. The regional energy eigenvalues of each sub-region were calculated. The low-frequency sub-band coefficients were calculated by the weighted strategy based on the regional energy eigenvalues.

Step 3. The local features were extracted by dense convolution layers. Residual dense block (RDB) allows the states of the previous RDB to be directly connected to all layers of the current RDB, forming a continuous memory (CM) mechanism.

Step 4. Local feature fusion was used to adaptively learn more effective features from previous and current local features, and stabilize the training of larger networks.

Step 5. The fused image is obtained by inverse NSST transform.

3 Fusion Algorithms

This section introduces the details of the fusion algorithm in this paper.

3.1 NSST

Shearlet is a system theory which combines geometry and multiresolution analysis by classical affine [31].

If $\psi \in L^2(R^2)$ meets the following conditions:

1. $\hat{\psi}(\zeta) = \hat{\psi}(\zeta_1, \zeta_2) = \hat{\psi}_1(\zeta_1)\hat{\psi}_2(\frac{\zeta_1}{\zeta_2})$, in which $\hat{\psi}()$ is the Fourier transformation of $\psi()$;

2. ψ_1 is a continuous wavelet, and $\hat{\psi}_1 \in C^\infty(R)$, $\text{supp}\hat{\psi}_1 \subset \left[-2, -\frac{1}{2}\right] \cup \left[\frac{1}{2}, 2\right]$;

3. $\hat{\psi}_2 \in C^\infty(R)$, and $\text{supp}\hat{\psi}_2 \subset [-1,1]$, $\hat{\psi}_1 > 0$, $\left\|\hat{\psi}_2\right\| = 1$;

then the system can be generated by ψ, $A_a = \begin{bmatrix} a & 0 \\ 0 & \sqrt{a} \end{bmatrix}$ and $B_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$:

$$\left\{ \psi_{ast}(x) = a^{-\frac{3}{4}}\psi(A_a^{-1}B_s^{-1}(x-t)), a \in B^+, s \in R, t \in B^2 \right\} \tag{1}$$

in which $\psi_{ast}(x)$ is called shearlet, A_a and B_s are called anisotropic dilation matrix and shear matrix.

The shearlet transform of function $f \in L^2(R^2)$ is defined as:

$$SH_\psi, f(a,s,t) = \langle f, \psi_{ast} \rangle \tag{2}$$

It can be seen that the shearlet transform is a function of scale (a), shear (s) and translation (t). Usually, the parameters are taken as follows:

$$A_a = A_0 = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, B_s = B_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, t \in Z^2$$

NSST is an improvement of shearlet transform. By using non-subsampled Laplacian filter banks, NSST obtains the multiscale decomposition of the original image. Then, the shear filter combination is used to decompose the sub-band images of different scales. The sub-band images with different scales and directions are obtained.

3.2 Low Frequency Sub-band Fusion

Low frequency sub-band reflects the outline and basic information of the image. The commonly used fusion methods for low frequency sub-bands are: weighted average method, absolute value method, standard deviation selection method, etc. In this paper, the method of combining local average energy with local standard deviation was applied [32]. The steps are as follows.

Step 1. The average of low frequency coefficient (C_L), the regional energy (E_L) and local standard deviation (D_L) were calculated:

$$C_L(x,y) = \frac{1}{M \times N} \sum_{(r=-(M-1)/2)}^{(M-1)/2} \sum_{(c=-(N-1)/2)}^{(N-1)/2} C_L(x+r,y+c) \tag{3}$$

$$E_L(x,y) = \frac{1}{M \times N} \sum_{(r=-(M-1)/2)}^{(M-1)/2} \sum_{(c=-(N-1)/2)}^{(N-1)/2} C_L^2(x+r,y+c)w(x+r,y+c) \tag{4}$$

$$D_L(x,y) = \sqrt{\frac{1}{(M-1)(N-1)} \sum_{(r=-(M-1)/2)}^{(M-1)/2} \sum_{(c=-(N-1)/2)}^{(N-1)/2} (C_L(x+r,y+c) + \overline{C}_L(x,y))^2} \tag{5}$$

The size of the window is $M \times N$, and the window function is $W = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 1 \end{bmatrix}$.

Step 2. According to E_L and D_L, regional features were constructed.

$$H_1(x,y) = \frac{E_L(x,y)}{(1 + D_L(x,y))} \tag{6}$$

$$H_2(x,y) = E_L(x,y) + D_L(x,y) \tag{7}$$

$$R_P = H_1^2 + H_2^2, (P = A, B) \tag{8}$$

In Eq. (6), "1" was used to prevent the invalidation when D_L was zero. In Eq. (8), the higher the value of R, the more information the region contains.

Step 3. The strategy for Low frequency sub-band fusion is:

$$F_L(x,y) = \begin{cases} C_L^A(x,y), H_{1a} > H_{1b}, H_{2a} > H_{2b} \\ C_L^B(x,y), H_{1b} > H_{1a}, H_{2b} > H_{2a} \\ q_1 \times C_L^A(x,y) + q_2 C_L^b(x,y), otherwise \end{cases} \tag{9}$$

in which q_1 and q_2 are the weights, and the following conditions were satisfied:

$$\begin{cases} q_1 = \dfrac{R_A}{R_A + R_B} + 0.0001 \\ q_2 = 1 - q_1 \end{cases}$$

3.3 RDN

RDN is based on deep learning, and was proposed for image super-resolution [32].

The main innovation of RDN is the structure of residual dense block (RDB), combining ResNet and DenseNet, as Fig. 2.
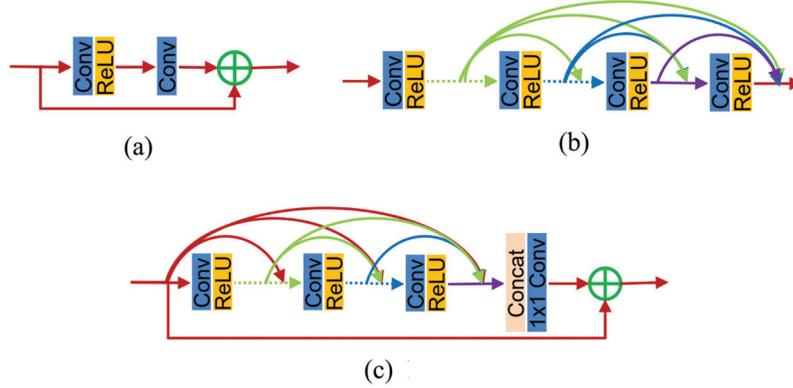
Figure 2: Comparison of network structures (a) Residual block (b) Dense block (c) Residual dense block

There are three parts in an RDB.

1. Contiguous Memory.

   The state of the previous RDB was passed to all the layers of the current RDB. For instance, for the adjacent RDB-c and RDB-d, the mathematical model is:

   $$F_{d,c} = \sigma\left(W_{d,c}[F_{d-1}, F_{d,1}, \ldots, F_{d,c-1}]\right) \tag{10}$$

   in which $F_{d,c}$ consists of feature maps. σ is the ReLU activation function. $W_{d,c}$ is the weights of Conv layer c without the bias.

2. Local feature fusion

   The feature number was reduced due to the direct connection between the preceding RDB and the current RDB. Furthermore, a 1×1 convolutional layer was introduced. Local feature fusion was formulated as:

$$F_{d,LF} = H_{LFF}^d \left(\left[F_{d-1}, F_{d,1}, \ldots, F_{d,c}, \ldots, F_{d,C} \right] \right) \tag{11}$$

in which H_{LFF}^d is the function of the 1×1 convolutional layer in the $d-$th RDB.

3. Local residual learning

On the basis of the above, the output of the $d-$th RDB is:

$$F_d = F_{d,c} + F_{d,LF} \tag{12}$$

With the increase of network depth, the features of each conv layer will be gradually graded. The hierarchical features can provide important details for image reconstruction because of their different receptive fields.

3.4 High Frequency Sub-band Fusion

High frequency sub-band images mainly contain the edge features and texture details. Therefore, the fusion rules of high frequency sub-band directly affect the resolution and clarity of the fused image.

In this paper, high frequency sub-band images of infrared and visible images were input into the trained RDN model to extract deep features, and then generate high frequency sub-band feature map. By using the maximum weighted average fusion strategy, the weight map is obtained.

The formulate is:

$$Fsub(i,j) = \max(\alpha_1 sub_A + \alpha_2 sub_B) \tag{13}$$

in which sub_A and sub_B are the high frequency sub-band images of infrared and visible images, α_1 and α_2 are the weights of sub_A and sub_B, respectively.

3.5 Image Reconstruction

According to the fusion rules of low frequency sub-band and high frequency sub-band mentioned above, the basic part $F_b(x,y)$ and detail part $F_d(x,y)$ were obtained. By using the inverse NSST, the image was reconstructed:

$$F(x,y) = F_b(x,y) + F_d(x,y) \tag{14}$$

4 Experimental Results

In order to verify the effectiveness of the method in this paper, the UN-Camp-images were taken as an example, and the results were compared with the following six methods: NSCT-PCNN [33], SR [34], NSST [35], NSST-PAPCNN [36], CNN [37], and NSST-CNN-APCNN [38].

There are two main aspects of image fusion quality evaluation, namely subjective evaluation, and objective index evaluation.

4.1 Subjective Evaluation

Subjective evaluation refers to the performance of the fused image details through visual perception.

The fusion results of the UN-Camp are as shown in Fig. 3. The source images of gray scale of visible image and infrared image are Figs. 3a and 3b, respectively. Figs. 3c–3i are the fused images of different fusion strategies.

- Fig. 3c: clear texture in the guardrail, but blurred trees.
- Fig. 3d: a fusion result with poor contrast.

- Figs. 3e and 3f: clear target but serious loss in texture.
- Fig. 3g: strong layering in gray scale, but discontinuous details in the left.
- Fig. 3h: clear texture structure and details.
- Fig. 3i: higher contrast than Fig. 3h.
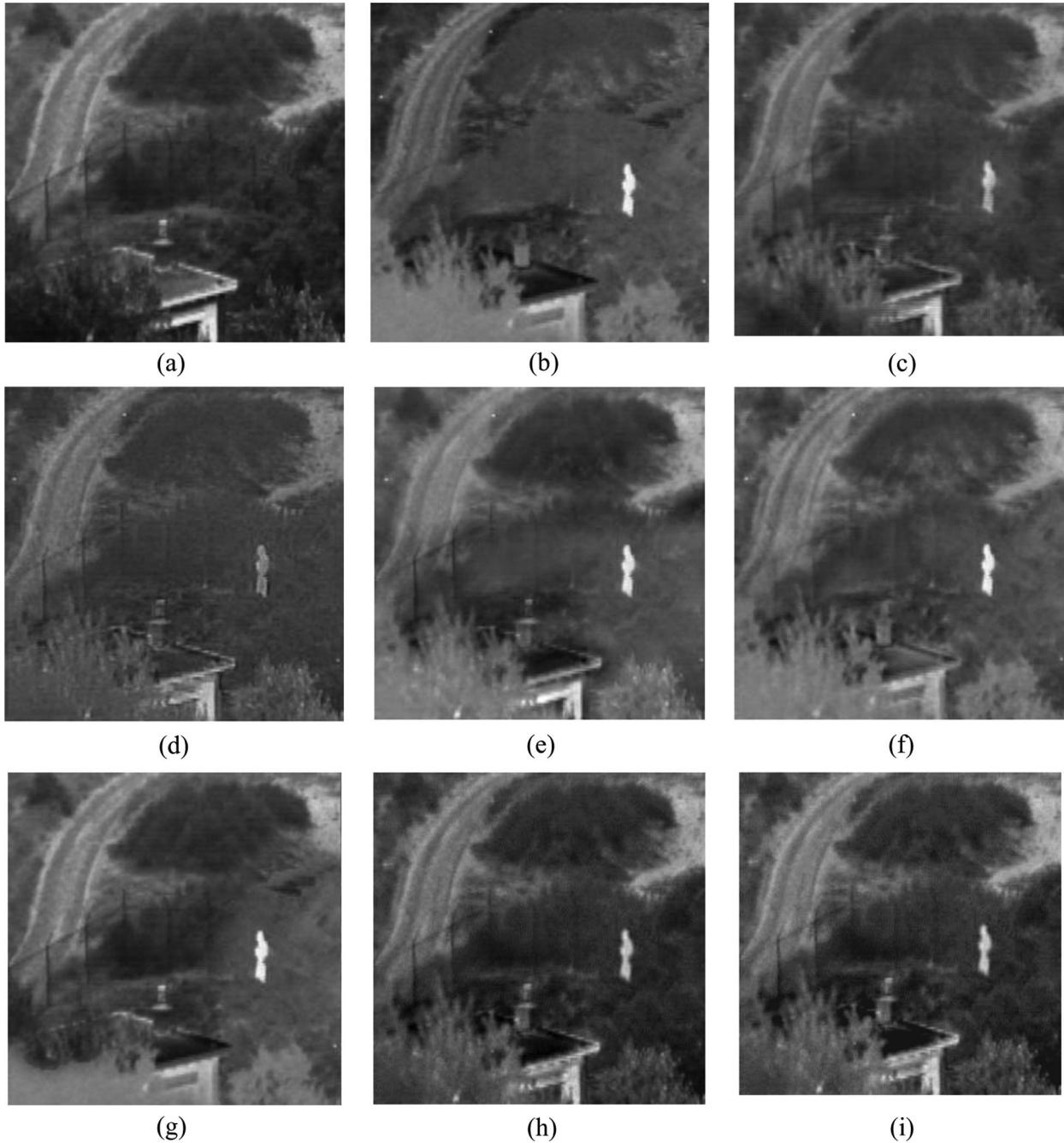
Figure 3: Results comparison of different fusion algorithms (a) Gray scale of visible image (b) Infrared image (c) NSCT-PCNN (d) SR (e) NSST (f) NSST-PAPCNN (g) CNN (h) NSST-CNN-APCNN (i) NSST-RDN

4.2 Objective Evaluation

In order to assess the performance of the algorithms more objectively, five indexes were selected to evaluate the quality of fused images: spatial frequency (SF), sharpness (SP), structural similarity (SSIM), Xydeas-Petrovic ($Q^{AB/F}$) and visual information fidelity for fusion (VIFF).

As shown in Fig. 4, our fusion method shows excellent performance in the five indicators than the other six algorithms.
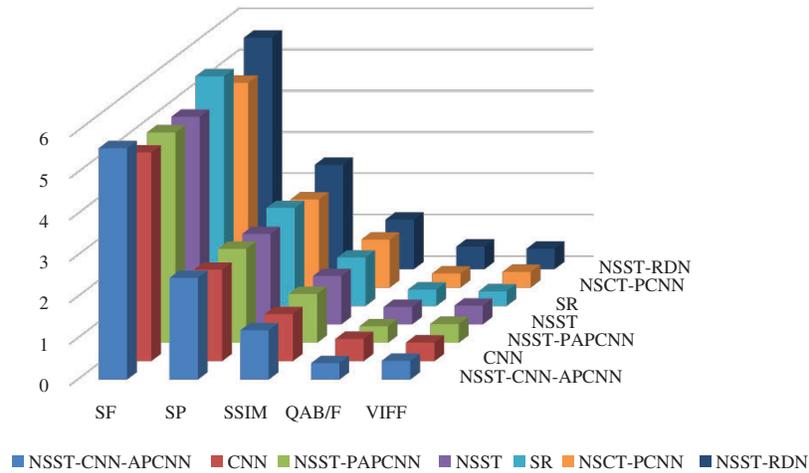
Figure 4: Objective evaluation indexes

5 Application

Visible image can reflect the texture and details of the scene. The infrared image is generated according to the thermal radiation of the object and is not affected by the external illumination. Therefore, visible image and infrared image fusion technology are particularly suitable for all-weather image processing and computer vision.

As an important part of the driving assistance system, the detection of driver's facial features is quite a critical issue. We took the vehicle cab as the application scenarios, and applied our method to the fusion of driver's facial images to capture the driver's facial images in both daytime and nighttime. The camera with a colored lens and an infrared lens was mounted above the dashboard [39], as shown in Fig. 5.

Figure 5: Camera in test vehicle

We took the driver's infrared images and visible images in the actual driving scene with low intensity of illumination intensity, and compared the results of different image fusion algorithms, which was similar to Section 4.

As can be seen from Figs. 6 and 7, the fused image generated from our algorithm was better than those from the other algorithms in texture, definition and indexes, and the details of the original images were well retained.
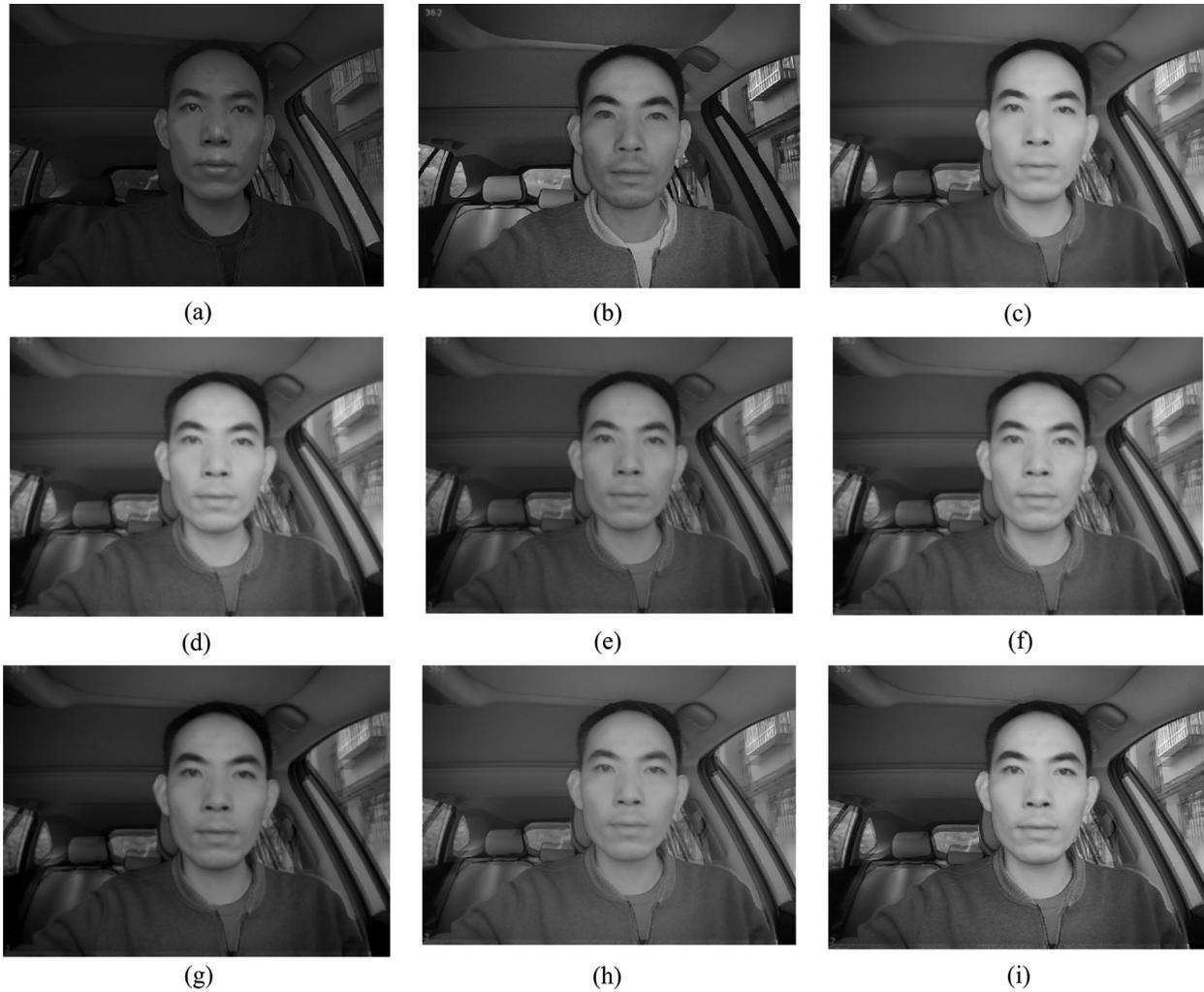
Figure 6: Fusion comparison in vehicle cab (a) Gray scale of visible image (b) Infrared image (c) NSCT-PCNN (d) SR (e) NSST (f) NSST-PAPCNN (g) CNN (h) NSST-CNN-APCNN (i) NSST-RDN
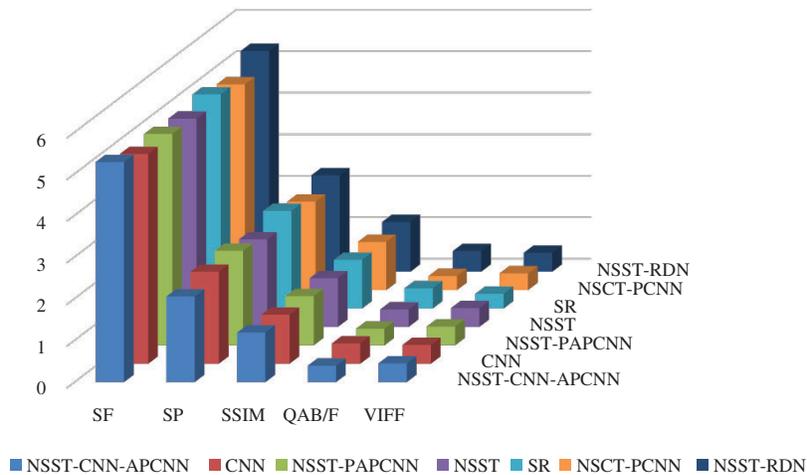
Figure 7: Objective evaluation of fusion comparison in vehicle cab

6 Conclusion

In this paper, on the basis of existing image processing algorithms, infrared image and visible image are decomposed into low frequency sub-band and high frequency sub-band by NSST transform. Low-frequency sub-band fusion was proceeded by combining local average energy with local standard deviation. RDN model was used to extract multi-scale features of high frequency sub-band, and then feature mapping was generated. The maximum weighted average fusion strategy was applied for high frequency sub-band fusion. Finally, the fusion image was obtained by inverse NSST transform. The results of the experiment and application in the driver's cab showed the fused image generated by our method has better visual effects. Moreover, it generally retains the largest amount of information in the source images.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] G. Ravikanth, K. V. N. Sunitha and B. E. Reddy, "Location related signals with satellite image fusion method using visual image integration method," *Computer Systems Science and Engineering*, vol. 35, no. 5, pp. 385–393, 2020.

[2] J. Zou, Z. Li, Z. Guo and D. Hong, "Super-resolution reconstruction of images based on microarray camera," *Computers, Materials & Continua*, vol. 60, no. 1, pp. 163–177, 2019.

[3] A. B. Hamza, Y. He, H. Krim and A. Willsky, "A multiscale approach to pixel-level image fusion," *Integrated Computer-Aided Engineering*, vol. 12, no. 2, pp. 135–146, 2005.

[4] Q. Zhang, Y. Liu, R. S. Blum, J. G. Han and D. C. Tao, "Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review," *Information Fusion*, vol. 40, no. 3, pp. 57–75, 2018.

[5] C. Cheng and D. Lin, "Image reconstruction based on compressed sensing measurement matrix optimization method," *Journal on Internet of Things*, vol. 2, no. 1, pp. 47–54, 2020.

[6] D. R. Zhu, L. Xu, F. B. Wang, T. Liu and Z. T. Chu, "Multi-focus image fusion algorithm based on fast finite shearlet transform and guided filter," *Laser & Optoelectronics Progress*, vol. 55, no. 1, pp. 011001, 2018.

[7]   C. H. Son, "Near-infrared fusion via a series of transfers for noise removal," *Signal Processing*, vol. 143, pp. 20–27, 2018.

[8]   K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[9]   H. N. Suand and C. K. Jung, "Multi-spectral fusion and denoising of RGB and NIR images using multi-scale wavelet analysis," in *Proc. ICPR*, Beijing, BJ, China, pp. 1779–1784, 2018.

[10] T. Y. Jung, D. M. Son and S. H. Lee, "Visible and NIR Image Blending for Night Vision," in *Proc. IPCV*, Las Vegas, LAS, USA, pp. 110–113, 2019.

[11] S. Jee and M. G. Kang, "Sensitivity improvement of extremely low light scenes with RGB-NIR multispectral filter array sensor," *Sensors*, vol. 19, no. 5, pp. 1256, 2019.

[12] K. M. He, J. Sun and X. O. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[13] J. Zhang, D. B. Zhao and W. Gao, "Group-based sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336–3351, 2014.

[14] X. Zhang, Y. J. Xue, S. Q. Tu, Y. M. Hu and X. F. Ning, "Remote sensing image fusion based on structural group sparse representation," *Journal of Image and Graphics*, vol. 21, no. 8, pp. 1106–1118, 2016.

[15] Y. Li, F. Y. Li, B. D. Bai and Q. Shen, "Image fusion via nonlocal sparse K-SVD dictionary learning," *Applied Optics*, vol. 55, no. 7, pp. 1814–1823, 2016.

[16] R. Chen, L. Pan, C. Li, Y. Zhou, A. Chen *et al.,* "An improved deep fusion CNN for image recognition," *Computers, Materials & Continua*, vol. 65, no. 2, pp. 1691–1706, 2020.

[17] C. K. Jung, K. L. Zhou and J. W. Feng, "Fusionnet: Multispectral fusion of RGB and NIR images using two stage convolutional neural networks," *IEEE Access*, vol. 8, pp. 23912–23919, 2020.

[18] M. Brown and S. Süsstrunk, "Multi-spectral SIFT for scene category recognition," in *Proc. CVPR*, Colorado Springs, CO, USA, pp. 20–25, 2011.

[19] C. Y. Tang, S. L. Pu, P. Z. Ye, F. Xiao and H. J. Feng, "Fusion of low-illuminance visible and near-infrared images based on convolutional neural networks," *Acta Optica Sinica*, vol. 40, no. 16, pp. 1610001, 2020.

[20] H. Li and X. Wu, "DenseFuse: A fusion approach to infrared and visible images," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2614–2623, 2019.

[21] J. Li, H. Huo, C. Li, R. Wang and Q. Feng, "AttentionFGAN: Infrared and visible image fusion using attention-based generative adversarial networks," *IEEE Transactions on Multimedia (Early Access)*, pp. 99, 2020.

[22] J. Y. Ma, P. W. Liang, W. Yu, C. Chen, X. J. Guo *et al.,* "Infrared and visible image fusion via detail preserving adversarial learning," *Information Fusion*, vol. 54, no. 11, pp. 85–98, 2020.

[23] J. Chen, X. J. Li, L. B. Luo, X. G. Mei and J. Y. Ma, "Infrared and visible image fusion based on target-enhanced multiscale transform decomposition," *Information Sciences*, vol. 508, no. 4, pp. 64–78, 2020.

[24] Z. X. Zhao, S. Xu, C. X. Zhang, J. M. Liu and J. S. Zhang, "Bayesian fusion for infrared and visible images," *Signal Processing*, vol. 177, no. 3, pp. 107734, 2020.

[25] H. B. Yan and Z. M. Li, "Infrared and visual image fusion based on multi-scale feature decomposition," *Optik*, vol. 203, no. 8, pp. 163900, 2020.

[26] H. Li, X. Wu and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *24th Int. Conf. on Pattern Recognition (ICPR)*, Beijing, China, pp. 2705–2710, 2018.

[27] H. Hermessi, O. Mourali and E. Zagrouba, "Multimodal image fusion based on non-subsampled shearlet transform and neuro-fuzzy," in *Int. Workshop on Representations, Analysis and Recognition of Shape and Motion from Imaging Data*. Tunisia: Sidi Bou Said Village, pp. 161–175, 2016.

[28] Y. L. Zhang, Y. P. Tian, Y. Kong, B. N. Zhong and Y. Fu, "Residual dense network for image super-resolution," in *Proc. CVPR*, Salt Lake City, USA, pp. 2472–2481, 2018.

[29] K. M. He, X. Y. Zhang, S. Q. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las Vegas, LAS, USA, pp. 770–778, 2016.

[30] Y. F. Feng, H. Yin, H. Q. Lu, K. Chen, L. Cao *et al.,* "Infrared and visible light image fusion method based on improved fully convolutional neural network," *Computer Engineering*, vol. 46, no. 8, pp. 243–249, 2020.

[31] X. B. Gao, W. Lu, D. C. Tao and X. L. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1409–1423, 2009.

[32] W. W. Zhu and J. F. Li, "Medical image fusion algorithm based on non-subsampled shearlet transform and feature synthesis," *Computer Systems & Applications*, vol. 28, no. 10, pp. 170–177, 2019.

[33] W. C. Hao and N. Jia, "Infrared and visible image fusion method based on adaptive PCNN in NSCT domain," *Journal of Xihua University: Natural Science Edition*, vol. 33, no. 3, pp. 11–14, 2014.

[34] S. W. Zhang, W. Li and X. J. Zhao, "A visible and infrared image fusion method based on sparse representation," *Electrooptics and Control*, vol. 24, no. 6, pp. 47–52, 2017.

[35] G. R. Gao, L. P. Xu and D. Z. Feng, "Multi-focus image fusion based on non-subsampled shearlet transform," *IET Image Processing*, vol. 7, no. 6, pp. 633–639, 2013.

[36] M. Yin, X. N. Liu and Y. Liu, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 1, pp. 49–64, 2019.

[37] L. Y. Mei, X. P. Guo and J. H. Zhang, "Multi focus image fusion using deep convolution neural network based on spatial pyramid pooling," *Journal of Yunnan University: Natural Science Edition*, vol. 41, no. 1, pp. 18–27, 2019.

[38] A. Y. Dong, Q. Z. Du, B. Su, W. B. Zhao and W. Yu, "Infrared and visible image fusion based on convolutional neural network," *Infrared Technology*, vol. 42, no. 7, pp. 660–669, 2020.

[39] Y. Y. Sun, P. Z. Yan, Z. Z. Li, J. C. Zou and D. Hong, "Driver fatigue detection system based on colored and infrared eye features fusion," *Computers, Materials & Continua*, vol. 63, no. 3, pp. 1563–1574, 2020.