Multi-sensors Image Fusion via NSCT and GoogLeNet

LI Yangyu, *WANG Caiyun*[*], *YAO Chen*

College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P. R. China

(Received 17 June 2020; revised 10 September 2020; accepted 20 September 2020)

Abstract: In order to improve the detail preservation and target information integrity of different sensor fusion images, an image fusion method of different sensors based on non-subsampling contourlet transform (NSCT) and GoogLeNet neural network model is proposed. First, the different sensors images, i. e., infrared and visible images, are transformed by NSCT to obtain a low frequency sub-band and a series of high frequency sub-bands respectively. Then, the high frequency sub-bands are fused with the max regional energy selection strategy, the low frequency sub-bands are input into GoogLeNet neural network model to extract feature maps, and the fusion weight matrices are adaptively calculated from the feature maps. Next, the fused low frequency sub-band is obtained with weighted summation. Finally, the fused image is obtained by inverse NSCT. The experimental results demonstrate that the proposed method improves the image visual effect and achieves better performance in both edge retention and mutual information.

Key words: image fusion; non-subsampling contourlet transform; GoogLeNet neural network; infrared image; visible image

CLC number: TN911.73 **Document code:** A **Article ID:** 1005-1120(2020)S-0088-07

0 Introduction

Image fusion technology is an effective method to fuse the same scene image acquired by different sensors into an image with clear target and background[1]. The technology is widely used in the fields of remote sensing, military and medical image processing. Infrared and visible images are typical for different sensors. The principle of visible light sensor is to use reflected light to form images which have clear texture, but the effect is poor in night vision and complex weather conditions. So, it is easy to neglect target. On the contrary, infrared sensors form images based on the object's own thermal radiation. The detection effect of the target is better in the infrared images. However, due to the limitation of infrared imaging mechanism, the definition is low and edges are blurred. Therefore, with the use of the complementarity and redundancy of images obtained by different sensors, we can combine the rich background texture information in visible image and the target information in infrared image to generate fused images, which have good target information and clear background information.

Image fusion methods start from the simple weighted average fusion in the spatial domain and the method based on principal component analysis, to the subsequent image fusion based on multi-scale transformation and some other multi-source image fusion methods based on sparse representation (SR), saliency, deep learning and so on. The image fusion method using the theory of multi-scale transformation (MST) has gradually developed into a classic method due to its simple but effective and practical algorithm. Han et al.[2] and Wang et al.[3] researched the infrared and visible image fusion based on discrete wavelet transform (DWT). Lewis et al.[4] improved the DWT-based method and proposed a kind of fusion method based on dual-tree

*Corresponding author, E-mail address: wangcaiyun@nuaa.edu.cn.

How to cite this article: LI Yangyu, WANG Caiyun, YAO Chen. Multi-sensors image fusion via NSCT and googlenet[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2020, 37(S): 88-94.

http://dx.doi.org/10.16356/j.1005-1120.2020.S.011

complex wavelet transform (DTCWT). Because the non-subsampling contourlet transform (NSCT) can capture the directional feature and edge information, the NSCT-based fusion methods[5-7] became the research focus. The SR-based fusion methods are used in multi-focus image fusion. Yang et al.[8] and Liu et al.[9] applied this method to achieve good results in multi-focus image fusion. Saliency-based fusion method[10] gets the salient features of the source image, divides the image into the target saliency map and the non-saliency map, and fuses the two parts separately. Some other methods such as gradient transfer fusion (GTF)[11] and multi-scale singular value decomposition (MSVD)[12] are under research recently. Deep learning method[13] is a kind of weighted fusion method, and the weight matrix value depends on the feature maps, which are extracted from source images with the use of neural network.

In this paper, we present a novel and effective fusion method via NSCT and GoogLeNet neural network for infrared and visible image fusion.

1 Preliminaries

NSCT is proposed by Cunha et al.[14] in 2006, as shown in Fig.1. NSCT uses the non-subsampled pyramid decomposition (NSP) and the non-subsampled direction filter banks (NSDFB) to obtain the decomposition coefficients of different scales and directions of source images.

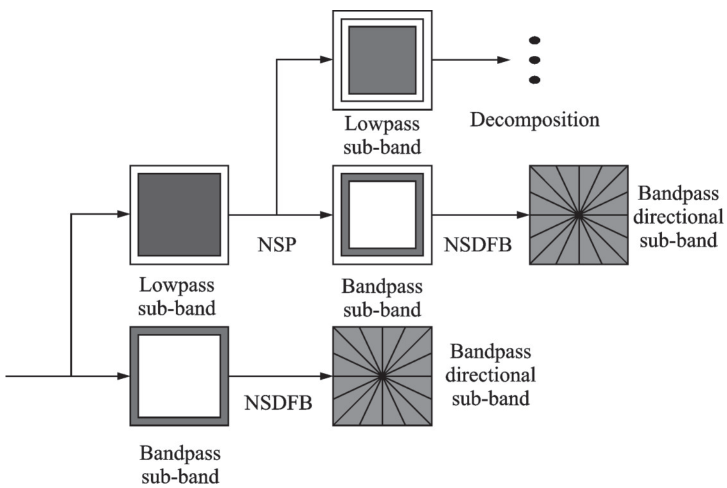

Fig.1 Decomposed structure of NSCT

NSCT is a kind of multiscale analysis tool, and multiscale analysis has been proved to be very effective for image fusion and other image processing tasks. The NSCT fusion methods can be summed up in three steps: (1) The low frequency and high frequency sub-bands in the transform domain are obtained by decomposing the source images; (2) Design fusion rules and fuse the low frequency and high frequency sub-bands respectively; (3) Obtain the fused image by corresponding inverse transform with the use of fused low frequency and high frequency sub-bands. The image fusion framework based on multiscale analysis involves two basic problems: The selection of multiscale decomposition method and the strategy for multiscale coefficient fusion. Most of the image features adopted by the fusion strategy are simple, while the fused image effect is not good enough. If we can get the features that can express the source image better, we can design more suitable fusion strategy.

Deep learning methods have good performance in image target recognition tasks, due to the excellent feature extraction and description capabilities at each layer of the neural network. The image features extracted by neural network are richer and more accurate representation of image information than previous features. Among the neural networks, GoogLeNet neural network[15] is a brand new neural network model proposed by Szegedy. It reduces the number of training parameters and the possibility of over-fitting when the training data set is limited, and also improves the classification accuracy. Compared with other network models, GoogLeNet neural network model has a depth of only 22 layers. The feature graph of each layer contains the contours and gradient features, and it is the key to high classification accuracy.

As analyzed above, the strategy for multi-scale coefficient fusion determines the quality of the fusion image, and the feature maps of GoogLeNet, which is trained on ImageNet, can express the source image precisely. We design fusion strategy of low frequency sub-bands with GoogLeNet feature maps. The fusion strategy is adaptive and comprehensive.

2 The Proposed Fusion Method

This paper presents a new method of infrared and visible image fusion, as shown in Fig.2.

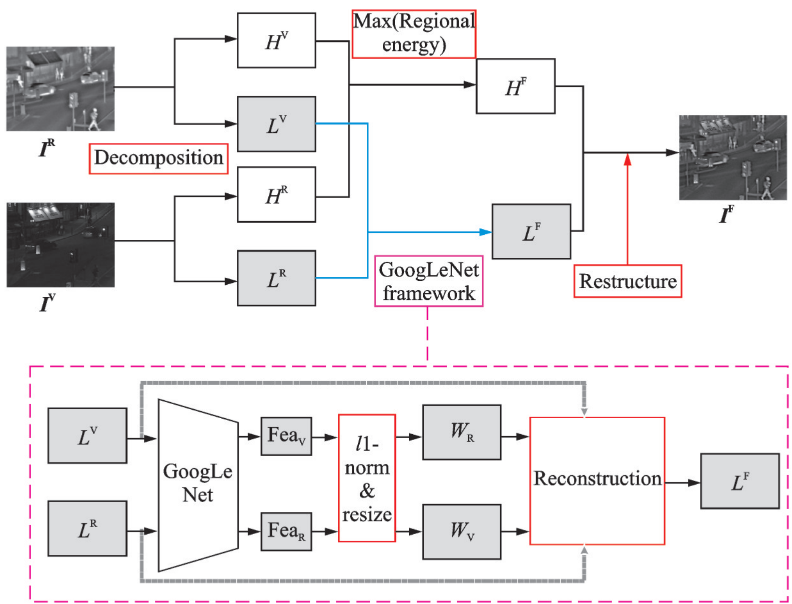

Fig.2 Framework of the proposed method

2.1 Decomposition with NSCT

Suppose that there are K preregistered source images, and in our paper we choose $K=2$. But the fusion strategy is the same for $K>2$. Here, we name the visible source image as $I_V(x, y)$ and the infrared source image as $I_R(x, y)$.

As mentioned above, NSCT is an effective decompose method. So in our paper, we use this method to decompose the source images.

The low frequency sub-band image $L^V(x, y)$ and $L^R(x, y)$ and high frequency sub-band images $H_{j,r}^V(x, y)$ and $H_{j,r}^V(x, y)$ can be obtained by Eq.(1) and Eq.(2).

$$I_V(x,y)=L_J^V(x,y)+\sum_{j=1}^{J}H_{j,r}^V(x,y) \tag{1}$$

$$I_R(x,y)=L_J^R(x,y)+H_{j,r}^R(x,y) \tag{2}$$

where j and r represent the scale and direction number of decomposition.

2.2 Adaptive fusion strategy with GoogLeNet

The low frequency sub-band images contain the basic contour information, which carries most of the energy of the source image. The processing of low frequency sub-band image will affect the final fusion result. In our paper, this part of the fusion strategy is designed as follow.

(1) Input the low frequency sub-band image $L_J^V(x, y)$ and $L_J^R(x, y)$ into the GoogLeNet network model, trained with the ImageNet set. With the use of feature extraction function of network model, feature graphs $\mathbf{Fea}_V^i$ and $\mathbf{Fea}_R^i$ of different depths of neural network are obtained, where $i=1, 2, \cdots, n$, $\mathbf{Fea}_V^i$ corresponds to visible image, and

$\mathbf{Fea}_{\mathrm{R}}^{i}$ corresponds to infrared image.

(2) Calculate the maximum value of pixel points corresponding to n feature graphs with $l1$-norm, then we can obtain feature map $\mathbf{Fea}_{\mathrm{R}}$ of infrared low frequency sub-band images and feature map $\mathbf{Fea}_{\mathrm{V}}$ of visible low frequency sub-band images.

(3) The feature map is extended to the size of the source image by the method of up-sampling interpolation. Next the feature matrix which can represent image contour feature information is obtained, and we would calculate the corresponding weight matrix $\boldsymbol{W}_{\mathrm{R}}$ and $\boldsymbol{W}_{\mathrm{V}}$ from feature matrix by Eqs.(3) and (4).

$$\boldsymbol{W}_{\mathrm{R}}(x,y)=\frac{\mathbf{Fea}_{\mathrm{R}}(x,y)}{\mathbf{Fea}_{\mathrm{R}}(x,y)+\mathbf{Fea}_{\mathrm{V}}(x,y)} \tag{3}$$

$$\boldsymbol{W}_{\mathrm{R}}(x,y)=\frac{\mathbf{Fea}_{\mathrm{R}}(x,y)}{\mathbf{Fea}_{\mathrm{R}}(x,y)+\mathbf{Fea}_{\mathrm{V}}(x,y)} \tag{4}$$

The fused low frequency sub-band image $\boldsymbol{L}_J^{\mathrm{F}}(x,y)$ is obtained by combining the low frequency sub-band image with the corresponding weight matrix by Eq.(5).

$$\boldsymbol{L}_J^{\mathrm{F}}(x,y)=\boldsymbol{L}_J^{\mathrm{R}}(x,y)\times \boldsymbol{W}_{\mathrm{R}}(x,y)+\boldsymbol{L}_J^{\mathrm{V}}(x,y)\times \boldsymbol{W}_{\mathrm{V}}(x,y) \tag{5}$$

2.3 Maximum of regional energy

The high frequency sub-band image energy is low but contains the main target energy, which affects the sharpness of fused images. In order to make the target area of fusion image clearer, high frequency sub-band images are fused with the rule of the maximum of regional energy. Taking the (3×3) region center at (x,y) in high frequency sub-band images $\boldsymbol{H}_{j,r}^{\mathrm{R}}(x,y)$ and $\boldsymbol{H}_{j,r}^{\mathrm{V}}(x,y)$, the regional energy of the two regions $\boldsymbol{E}_{j,r}^{\mathrm{R}}(x,y)$ and $\boldsymbol{E}_{j,r}^{\mathrm{V}}(x,y)$ is calculated by Eq.(6).

$$\boldsymbol{E}_{j,r}=\sum_{x'=-1}^{1}\sum_{y'=-1}^{1}\boldsymbol{H}_{j,r}(x+x',y+y')^2 \tag{6}$$

The high frequency sub-band images $\boldsymbol{H}_{j,r}^{\mathrm{F}}(x,y)$ is calculated by Eq.(7).

$$\boldsymbol{H}_{j,r}^{\mathrm{F}}(x,y)=\begin{cases}\boldsymbol{H}_{j,r}^{\mathrm{R}}(x,y) & \boldsymbol{E}_{j,r}^{\mathrm{R}}(x,y)\geqslant \boldsymbol{E}_{j,r}^{\mathrm{V}}(x,y)\\ \boldsymbol{H}_{j,r}^{\mathrm{V}}(x,y) & \text{Else}\end{cases} \tag{7}$$

2.4 Reconstruction

The final infrared and visible fusion image $\boldsymbol{I}_{\mathrm{F}}(x,y)$ is obtained by corresponding inverse NSCT of the fused low frequency sub-band image $\boldsymbol{L}_J^{\mathrm{F}}(x,y)$ and the fused high frequency sub-band images $\boldsymbol{H}_{j,r}^{\mathrm{F}}(x,y)$ of each layer and each direction.

2.5 Summary of the proposed method

The proposed image fusion method is shown in Table 1.

Table 1 NSCT-GoogLeNet image fusion algorithm steps

Input: Infrared image $\boldsymbol{I}_{\mathrm{R}}(x,y)$, visible image $\boldsymbol{I}_{\mathrm{V}}(x,y)$
Output: Fused image $\boldsymbol{I}_{\mathrm{F}}(x,y)$
(1) Image decomposition: Decompose $\boldsymbol{I}_{\mathrm{R}}(x,y)$ and $\boldsymbol{I}_V(x,y)$ to $\boldsymbol{L}_J^{\mathrm{R}}(x,y)$, $\boldsymbol{L}_J^{\mathrm{V}}(x,y)$ and $\boldsymbol{H}_{j,r}^{\mathrm{V}}(x,y)$, $\boldsymbol{H}_{j,r}^{\mathrm{V}}(x,y)$;
(2) Adaptive fusion strategy with GoogLeNet: Obtain GoogLeNet feature graphs Fea_V^i and Fea_R^i; calculate weight matrix $\boldsymbol{W}_{\mathrm{R}}$ and $\boldsymbol{W}_{\mathrm{V}}$; obtain $\boldsymbol{L}_J^{\mathrm{F}}(x,y)$;
(3) Maximum of regional energy: Obtain $\boldsymbol{H}_{j,r}^{\mathrm{F}}(x,y)$ with the rule of maximum of regional energy;
(4) Reconstruction: Obtain $\boldsymbol{I}_{\mathrm{F}}(x,y)$ with corresponding inverse NSCT.

3 Experimental Results

3.1 Experimental settings

In our experiment, the source infrared and visible images were collected from TNO Human Factors Research Institute[16]. Three pairs of infrared and visible images were selected, which are UN-camp, Sandpath and Band respectively.

For comparison, we selected several recent and classical fusion methods to perform the same experiment, including: DTCWT[4], NSCT[5], ASR[9], GTF[11] and MSVD[12]. The decomposition levels for DTCWT and NSCT are all 4, and the decomposition directions in NSCT are [2 3 3 4]. All the fusion algorithms are implemented in MATLAB R2018a on Intel Core i7-8750h @2.20 GHz CPU with 8 GB RAM.

3. 2 Subjective results

As shown in Figs.3—5, there are the fusion results of three pairs of infrared and visible images obtained by different fusion methods. The target information in the infrared image and the background information in the visible image are well retained in

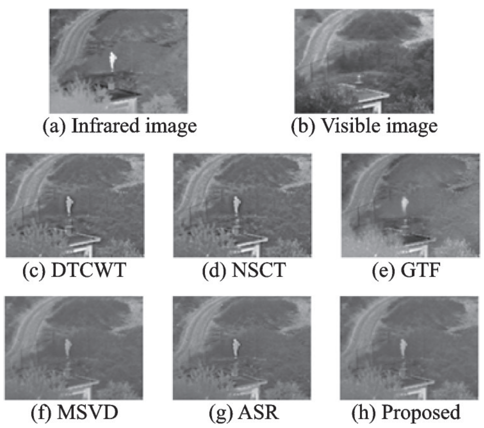

Fig.3 Result on “UNcamp” image

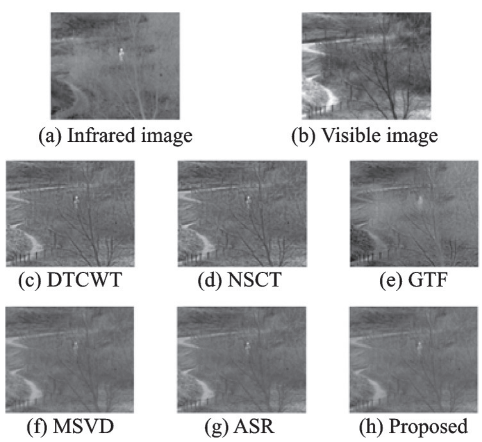

Fig.4 Result on “Sandpath” image

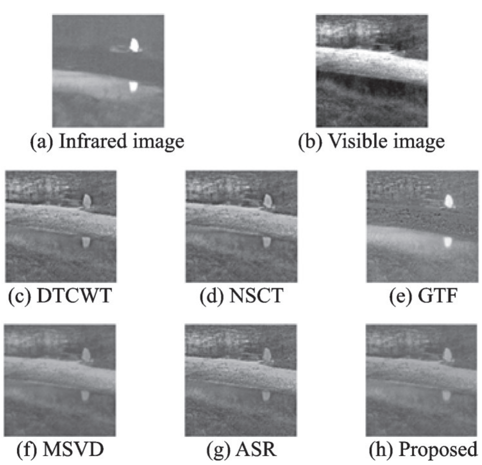

Fig.5 Result on “Band” image

the fusion image by all six methods. The target brightness in the GTF fusion image is the highest but ill-defined. The MSVD fusion images are blurrier than other methods. As we can see in the result, fusion images by the proposed method have clear target information and fine texture.

3. 3 Objective evaluation

Five objective evaluation indexes are used to compare and evaluate the fusion method. They are spatial frequency SF, information entropy IE, edge information retention Q_{abf}[17], weighted fusion index Q_w[18] and fusion runtime. It should be noted that a larger value means the better fusion performance, while runtime is just the opposite.

During the experiment, we add Gaussian noise with variance of 5, 10 and 25 to all three pairs of source images, and then calculate the average value of the evaluation results after multiple tests. The results are shown in Tables 2—4.

Table 2 Comparison of UNcamp fusion result evaluation

Method	DTCWT	NSCT	GTF	MSVD	ASR	Ours
SF	11.987	**12.204**	8.878	9.762	9.832	11.986
IE	6.507	6.567	6.681	6.265	6.345	**6.725**
Q_{abf}	0.447	0.489	0.408	0.326	0.431	**0.498**
Q_w	0.397	0.417	0.458	0.287	0.482	**0.489**
Runtime/s	0.192	1.142	0.671	**0.104**	85.066	1.750

Table 3 Comparison of sandpath fusion result evaluation

Method	DTCWT	NSCT	GTF	MSVD	ASR	Ours
SF	11.102	11.534	10.305	10.012	8.924	**11.586**
IE	6.401	6.443	6.502	6.148	6.159	**6.765**
Q_{abf}	0.484	**0.496**	0.492	0.336	0.413	0.491
Q_w	0.437	0.446	**0.514**	0.366	0.445	0.509
Runtime/s	0.273	2.923	4.261	**0.263**	240.88	1.652

Table 4 Comparison of band fusion result evaluation

Method	DTCWT	NSCT	GTF	MSVD	ASR	Ours
SF	23.188	23.225	21.815	9.288	22.779	**24.424**
IE	6.939	6.961	6.778	6.478	6.854	**6.965**
Q_{abf}	0.672	0.679	0.580	0.166	0.674	**0.688**
Q_w	0.483	0.494	0.476	0.185	0.445	**0.511**
Runtime/s	0.092	0.765	0.446	**0.079**	65.806	1.354

From above results, the following conclusions can be drawn.

(1) The IE of the proposed method is the high-

est, which indicates that the fusion image is the clearest and contains the most information in the image.

(2) The Q_{abf} and Q_w of the proposed method are the highest in the UNcamp and Band experiment, and slightly less than other methods in the Sandpath experiment. The proposed method is better than the other methods.

(3) The fusion time of the proposed method is the 5th in the six methods, but it's only slower less than 1 s compared with the fastest MSVD method. So it meets the requirement of real-time fusion.

4 Conclusions

An infrared and visible image fusion method based on NSCT and GoogLeNet neural network model is proposed. We use a multi-scale model NSCT to decompose the source image, the fuse low frequency sub-band image by the use of GoogLeNet, and the fuse high frequency sub-bands with the rule of the maximum of regional energy. The experimental results show that the proposed fusion method can not only highlight the target information in infrared images, but also retain the background details in visible images. It is superior to other fusion methods in terms of vision, SF, IE, Q_{abf} and Q_w. In addition, the proposed fusion method has a short running time, which can meet the real-time requirement.

References

[1] ARDESHIR G. Image fusion: Advances in the state of the art[J]. Information Fusion, 2006, 8(2): 114-118.

[2] HAN L, LI C F, PU X J. Image segmentation and stationary wavelet transform fuse infrared and visible image[J]. Journal of Chongqing University, 2003, 36 (6): 112-118.(in Chinese)

[3] WANG Y, WANG M Q. Study of infrared and visible image fusion technology based on wavelet transform[J]. Infrared, 2013,34(3): 12-14.

[4] LEWIS J J, ROBERT J O, STAVRI G N, et al. Pixel and region-based image fusion with complex wavelets [J]. Information Fusion, 2007, 8(2): 119-130.

[5] ZHANG Q, GUO B L. Multi-focus image fusion using the non-subsampled contourlet transform[J]. Signal Processing, 2009, 89(7): 1334-1346.

[6] CAI H Y, ZUO L R, ZU P, et al. Infrared and visible image fusion based on non-subsampled contourlet transform and intuitionistic fuzzy set[J]. Acta Photonica Sinica, 2018, 47(6): 225-234.

[7] ZHANG Y. Image fusion based on non-subsampled contourlet transform[J]. Computer Engineering and Applications, 2017, 53: 335-339.

[8] YANG B, LI S. Multi-focus image fusion and restoration with sparse representation[J]. IEEE Trans Instrum Meas, 2010, 59(4): 884-892.

[9] LIU Y, WANG Z F. Simultaneous image fusion and denosing with adaptive sparse representation[J]. IET Image Processing, 2015, 9(5): 347-357.

[10] BAVIRISETT I, DURGA P, DHUL I, et al. Two-scale image fusion of visible and infrared images using saliency detection[J]. Infrared Physics Technology, 2016, 76: 52-64.

[11] MA J Y, CHEN C, LI C. Infrared and visible image fusion via gradient transfer and total variation minimization[J]. Information Fusion, 2016, 31(9): 100-109.

[12] NAIDU V P S. Image fusion technique using multi-resolution singular value decomposition[J]. Information Fusion, 2011, 61(9): 479-484.

[13] LI H, XIAO J W, TARIQ S, et al. Infrared and visible image fusion with ResNet and zero-phase component analysis[J]. Infrared Physics & Technology, 2019(12): 103039.

[14] DA CUNHA A L, ZHOU J P, DO M H. The non-subsampled contourlet transform: Theory, design, and applications[J]. IEEE Transactions on Image Processing, 2006, 10(15): 3089-3101.

[15] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolution[C]//Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR).[S.l.]: IEEE Computer Society, 2015:1-9.

[16] TOET A. TNO image dataset[EB/OL]. (2014-10-08). https://doi. org/10.6084/m4. figshare. 1008029.v1.

[17] XYDEAS C S, PETROVIC V. Objective image fusion performance measure [J]. Electronics Letters, 2000, 36(4): 308-309.

[18] PIELLA G, HEIJMANS H. A new quality metric for image fusion[C]//Proceedings 2003 International Conference on Image Processing. Barcelona, Spain:

IEEE, 2003: 173-176.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 61301211) and the China Scholarship Council(No.201906835017).

Authors Mr. **LI Yangyu** received the B.E.degree in information engineering from Nanjing University of Aeronautics and Astronautics (NUAA) in 2016, and under postgraduate in NUAA since 2018. He is major in multi-sensors image processing and target recognition.

Prof. **WANG Caiyun** received the Ph.D. degree in engineering from Beijing University of Aeronautics and Astronautics (BUAA) in 2008. In June 2008, she joined the Academy of Astronautics, NUAA, Nanjing, China. Prof. WANG is engaged in the research of radar signal processing, target recognition and so on.

Author contributions Prof. **WANG Caiyun** revised and modified the manuscript. Mr. **LI Yangyu** designed the study, conducted the analysis and wrote the manuscript. Ms. **YAO Chen** assisted in simulation experiments. All authors commented on the manuscript draft.

Competing interests The authors declare no competing interests.

(Production Editor: LIU Yandong)

一种基于NSCT与GoogLeNet的多传感器图像融合算法

李阳雨，王彩云，姚 晨

（南京航空航天大学航天学院，南京211106，中国）

摘要：为提高多传感器融合图像的细节保持性与目标信息完整性，提出一种基于非下采样轮廓波变换(Non-Subsampling Contourlet transform, NSCT)与GoogLeNet神经网络模型相结合的异传感器图像融合算法。本文采用的异传感器图像为红外与可见光图像，首先将红外与可见光图像分别进行NSCT变换，分解得到一个低频子带系数和一系列多尺度、多方向的高频子带系数；然后将高频子带系数采用区域能量取大策略进行融合，将低频子带系数输入GoogLeNet神经网络模型中提取特征图，从特征图中自适应计算加权融合的权值矩阵参数，加权求和得到融合后的低频子带系数；最后经过NSCT逆变换得到融合图像。实验结果表明，该算法有效地提高了图像视觉效果，且在边缘保持度、互信息等客观指标上有明显提高。

关键词：图像融合；非下采样轮廓波变换；GoogLeNet神经网络；红外图像；可见光图像