Infrared and Visible image Fusion based on saliency detection and shearlet tranform

Pengyu Yan[1], Ping Zhang[1,], Chun Fei[2], Xiaowei Wang[1], Jiang Wu[1]*

1. School of Computer Optoelectronic Information, University of Electronic Science and Technology of China, Chengdu 611731, China;

2. School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China;

**E-mail: andyyim-ypy@outlook.com*

Abstract: Infrared and visible image fusion can integrate the target information of an infrared image and the spatial detail of a visible image to constitute a fused image, which has more complete and accurate description of the same scene than a single image. In this paper, a novel image fusion method using saliency detection based on non-subsampled shearlet transform (NSST) is proposed for infrared and visible image fusion. Firstly, the infrared target of infrared image is effectively highlighted using saliency detection based on NSST. Then, the different fusion rule is designed for the saliency regions and non-saliency regions. The final fuse image is obtained through inverse NSST. Experimental results demonstrate the proposed method has super to the traditional fusion method, which can highlight the infrared target of infrared image meanwhile reserve the scene and texture details of visible image.

Key words: image fusion; saliency detection; shearlet transform; infrared image; visible image

1. INTRODUCTION

Image fusion combines different information from two or more source images of the same scene into a new image. These source images may be acquired by different sensors or by different methods with one sensor. The fused image can provide more information than signal source image. Image fusion is widely applied, such as machine vision, target detection, medical diagnostics, remote sensing, etc. [1~3]. Infrared images have the good ability to detect the hidden targets while they have poor image quality because of low contrast and details. Visible images have high spatial resolution and abundant texture details while they are limited by the light conditions and cannot detect the hidden targets. So the fusion of infrared and visible images is very useful to improve the ability of target detection and scene description, which combine the both advantages of infrared and visible images. Typical applications such as night surveillance, fire alarm, military detection, etc. [4,5].

However, many image fusion algorithms are not suitable for infrared and visible image fusion because infrared and visible images have different characteristics. In order to effectively combine the target object of infrared image and the scene information of visible images, many image fusion algorithms based on saliency detection have been proposed [6~10]. Chen et al. proposed a steerable weight fusion algorithm based on visual saliency. Gaussian pyramids feature maps are computed to measure the saliency and salient regions are obtained through maximum entropy. Then different salient regions with different weights are fused to new image [6]. Zhao et al. used the weighted least squares filter to construct the multi-scale decomposition and obtained the salient map with grayscale contrast for each layer. Then the fusion result is reconstructed by synthesizing different scales with synthetic weights [7]. Han et al. proposed a biased fusion based on Markova random field saliency classification for the infrared and visible sequences [7]. Peng et al. used a new method to fuse images in non-subsampled contourlet transform domain based on region saliency analysis and the salient map is achieved by image segmentation and multi-scale contrast analysis [9]. Rosa et al. presented a salient map to define the level of activity for each coefficients of the dual-tree complex wavelet transform and derived the rules of fusion according to the saliency level [10].

2. SALIENCY DETECTION FOR INFRARED IMAGE

Infrared image only has brightness information, which is lack of color information of visible image. Therefore, the saliency detection of infrared image is easier to implement in the frequency domain. N. Imamoglu et al. has presented a saliency detection method based on wavelet transform, using the high frequency coefficients of wavelet decomposition to construct global and local characteristic maps. Their method has obtained excellent saliency detection effect for the visible image [11]. Inspired by their idea, we propose a new saliency detection method based on

non-subsampled shearlet transform (NSST). NSST has better direction decomposition characteristics than the wavelet transform. The framework is shown in Fig.1.

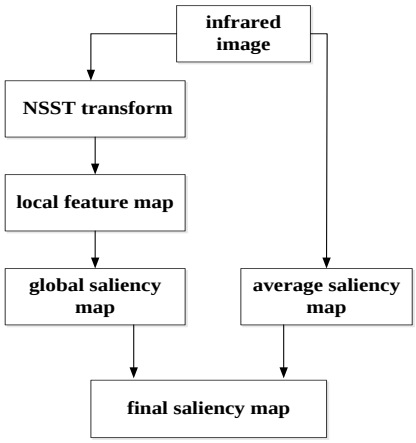


Figure.1 Saliency detection for infrared image based on NSST

Firstly, calculate the local feature map based on NSST. The original infrared image is implemented by NSST transform and the image is decomposed into low frequency subband coefficients $C_{l=0}(i,j)$ and a series of multi-scale multi-direction of high frequency subband coefficients $C_l^k(i,j)$. $l=\{1,2...,l_{\max}\}$ is the number of decomposed layer and $k=\{1,2...,k_{\max}\}$ is the number of decomposed direction. $l_{\max}$ and $k_{\max}$ are max decomposed layer and direction, (i,j) is the coordinate of decomposed coefficients.

Different from visible image, infrared image primarily describes the characterization of the infrared target in the scene because the infrared target and background information are quite different. In order to effectively highlight the infrared target, we use a series of high frequency subband coefficients of NSST transform to reconstruct the local feature map and ignore the low frequency subband coefficients and the first layer of high frequency subband coefficients.

$$F_{l'}(i,j)=(INSST\{C_{l'}^{k}(i,j)\})^2,\quad l'=\{2,...,l_{\max}\} \quad (1)$$

Secondly, global saliency maps is obtained based on probability density distribution method according to above local feature maps. Define the characteristic vector $\mathbf{f}(i,j)$ as all the feature maps of the position (i,j). Multi-dimensional Gaussian probability density distribution is defined as follows:

$$p(\mathbf{f}(i,j))=\frac{1}{(2\pi)^{n/2}|\Sigma|}\times e^{(-1/2(\mathbf{f}(i,j)-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{f}(i,j)-\boldsymbol{\mu}))} \quad (2)$$

$$\Sigma=E[(\mathbf{f}(i,j)-\boldsymbol{\mu})(\mathbf{f}(i,j)-\boldsymbol{\mu})^T] \quad (3)$$

$$SM_{Global}(i,j)=(\log(p(\mathbf{f}(i,j))^{-1}))^{1/2} \quad (4)$$

In order to eliminate noise effects, 3*3 Gaussian filter is used to get the final global saliency map.

$$SM'_{Global}(i,j)=SM_{Global}(i,j)*GF_{3\times3}(i,j) \quad (5)$$

Although the global salient map based on NSST can highlight the infrared target, it shows that the highlight part is focused on the edge region rather than the entire salient region. This phenomenon is called as inconsistencies highlighted salient region [12], which means there are high saliency values at the edge of region and low saliency values in the interior of region.

Achanta et al. proposed that the salient object of the image can be completely highlighted using the whole information of image [13]. So we use this idea to obtain the average saliency map of the infrared image. First of all, 3*3 gaussian filter is used to smooth the infrared image $f_{IR}(i,j)$ for eliminating noise effect.

$$f'_{IR}(i,j)=f_{IR}(i,j)*GF_{3\times3}(i,j) \quad (6)$$

Then, calculate the average salient map $S_{avg}(i,j)$ of the image, μ is the average of $f'_{IR}(i,j)$. And we combine the global salient map and the average salient map to obtain the final salient map.

$$SM_{avg}(i,j)=\left\|(f'_{IR}(i,j)-\mu)\right\|_2 \quad (7)$$

$$SM(i,j)=SM_{avg}^{\alpha}(i,j)\times SM_{global}^{\beta}(i,j)*GF_{5\times5}(i,j) \quad (8)$$

3 FUSION BASED ON SALIENCY DETECTION

According to the characteristics of infrared and visible image, we propose a fusion method using the saliency detection proposed above. The framework is shown in Fig. 2. Firstly, the infrared and visible image is converted into NSST transform domain, respectively. Then, saliency map is obtained using high frequency subband coefficients of the infrared image. The fusion rule is defined based on the saliency map. The infrared image is dominated in saliency region and the visible image is dominated in nonsaliency region. The fused image is obtained through inverse NSST transform. The details of the fusion rule are as follow.

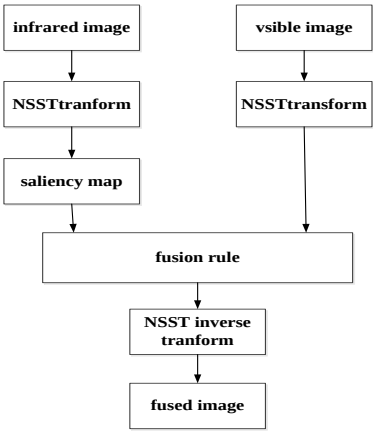


Figure.2 Infrared and visible image fusion based on saliency detection

Set the low frequency coefficient in the location (i, j) of infrared and visible image is $LC_{IR}(i, j)$ and $LC_{VS}(i, j)$, respectively. High frequency coefficient is $SC_{IR}^{l,\theta}(i, j)$ and $SC_{VS}^{l,\theta}(i, j)$, respectively. The scale is l and the direction is θ . The low frequency coefficient of the fused image is $LC_F(i, j)$ and the high frequency coefficient of that is $SC_F^{l,\theta}(i, j)$.

(1) Fusion rule of low frequency coefficients

We use Otsu threshold segmentation for salient map $SM(i, j)$ to obtain segmentation map $Seg(i, j)$ .The corresponding area $Seg(i, j) = 1$ is the infrared target area.

For low frequency coefficients, fusion rule is as follows:

$$LC_F(i, j) = \begin{cases} LC_{IR}(i, j) & if\ Seg(i, j) = 1 \\ LC_{VS}(i, j) & if\ Seg(i, j) = 0 \end{cases} \quad (9)$$

(2) Fusion rule of high frequency coefficients

The high frequency coefficient of NSST transform is the details description of source image such as edge and texture. So the fusion rule of high frequency coefficient should make the fused image to maintain the rich details as far as possible. We use regional energy of high frequency coefficients as the fusion rule. High regional energy value of high frequency coefficients usually is corresponding to the intensity change area of the source image. The regional energy is defined as follows:

$$E^{l,\theta}(x, y) = \sum_{(x,y)\in\phi} W(x, y) \left| SC^{l,\theta}(x, y) \right|^2 \quad (10)$$

$$W(x, y) = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (11)$$

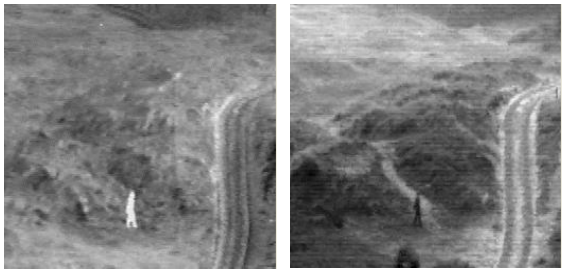
$$SC_F^{l,\theta}(x, y) = \begin{cases} SC_A^{l,\theta}(x, y) & E_A^{l,\theta}(x, y) > E_B(x, y) \\ SC_B^{l,\theta}(x, y) & E_A^{l,\theta}(x, y) < E_B(x, y) \end{cases}$$

4 EXPERMENTS

The proposed method is compared with other five common and excellent methods, including methods based on the DWT [14], NSCT [15], NSST [16] , DWT using Itti saliency detection [17] and NSCT using SR saliency [18]. Those methods are simple marked as DWT, NSCT, NSST, DWT_SD and NSCT_SD, respectively. The proposed method is marked as NSST_SD. The wavelet base is "db8" in DWT method. The decomposition layer number of DWT, NSCT and NSST is set to 3, direction decomposition direction number is set to (4, 4, 8), scale filter is "maxflat", direction filter is "pkva". The DWT, NSCT and NSST adopted the fusion rule is: low-frequency subband is average fusion, high frequency subband is absolute value maximum fusion. All methods are tested on three image databases, which are named ‘Uncamp’, ‘Duine’and ‘Trees’. They are downloaded from TNO Human Factors Research Institute. Image size is 256×256. Each source image is shown as in Fig. 3, 4 and 5, respectively. The fused image is shown as in Figure 6, 7 and 8, respectively.

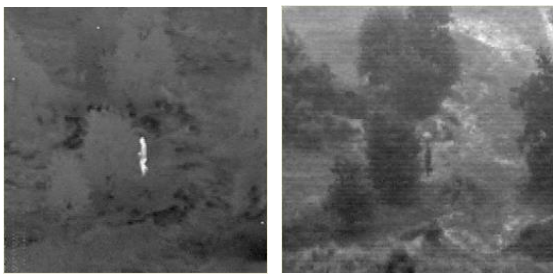
(a) Infrared image (b) Visible image

Figure.3 UN Camp images



(a) Infrared image (b) Visible image

Figure.4 Duine images



(a) Infrared image (b) Visible image

Figure.5 Tree images

For objective evaluation of the infrared and visible image fusion, we choose information entropy (IE), spatial frequency (SF) and average gradient (AG) to evaluate the fused image. IE evaluates the ability to contain the average information in the fused image, which is defined as formula (12). SF evaluates the ability to express small details in the fused images, which is defined as formula (13). AG evaluates the ability to measure the detail contrast and texure change in the fused images, which is defined as formula (14).

$$E = -\sum_{i=0}^{L-1} p_i \ln p_i \tag{12}$$

where p_i is the probability of gray value i, L is the gray level.

$$SF(x,y) = \sqrt{F_R^2(x,y) + F_C^2(x,y)}$$

$$F_R(x,y) = \sqrt{\frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=1}^{N} [I(x,y+1) - I(x,y)]^2} \tag{13}$$

$$F_C(x,y) = \sqrt{\frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=1}^{N} [I(x+1,y) - I(x,y)]^2}$$

$$AG = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{\left(\left(\frac{\partial f(i,j)}{\partial i}\right)^2 + \left(\frac{\partial f(i,j)}{\partial j}\right)^2\right) \Big/ 2} \tag{14}$$

Where $F_R(x,y)$, $F_C(x,y)$ are the row and column gradients, respectively, and $\partial f(i,j)/\partial i$ and $\partial f(i,j)/\partial j$ is the difference of i direction and j direction , respectively.

Moreover, we choose mutual information (MI) to evaluate the relationship between the source image and the fused image, , which is defined as formula (15). F is the fused image and A, B is the source image.

$$MI_F^{AB} = MI(A,F) + MI(B,F) \tag{15}$$

$$MI(A,F) = \sum_{i,j} (h_{AF}(i,j) \log_2 \frac{h_{AF}(i,j)}{h_A(i) h_F(j)} \tag{16}$$

$$MI(B,F) = \sum_{i,j} (h_{BF}(i,j) \log_2 \frac{h_{BF}(i,j)}{h_B(i) h_F(j)} \tag{17}$$

Where $h_A(i,j)$, $h_B(i,j)$, $h_F(i,j)$ is the normalized histogram of the source image A,B and the fused image F, respectively. $h_{AF}(i,j)$, $h_{BF}(i,j)$ is the joint normalized histogram of A and F,B and F, respectively.

(a)DWT (b)DWT_SD (c)NSCT

(d)NSCT_SD (e)NSST (f)NSST_SD

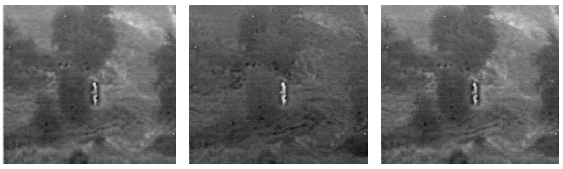
Figure.6 UN Camp fused image

(a)DWT (b)DWT_SD (c)NSCT

(d)NSCT_SD (e)NSST (f)NSST_SD

Figure.7 Duine fused image

(a)DWT (b)DWT_SD (c)NSCT



(d)NSCT_SD (e)NSST (f)NSST_SD

Figure.8 Tree fused image

Table.1 Objective evaluation

		DWT	DWT_SD	NSCT	NSCT_SD	NSST	NSST_SD
UN Camp	IE	6.26	6.27	6.43	7.14	6.58	**7. 16**
	SF	9.48	9.75	11.89	12.31	12.57	**12.71**
	AG	3.53	3.59	4.86	5.14	5.17	**5.23**
	MI	1.44	1.65	1.60	3.30	1.60	**3.34**
Duine	IE	5.78	5.80	5.88	6.40	5.88	**6.41**
	SF	6.07	6.17	6.89	7.00	7.04	**7.14**
	AG	2.34	2.37	3.02	3.10	3.05	**3.12**
	MI	1.40	1.48	1.41	2.41	1.42	**2.45**
Tree	IE	5.70	5.71	5.78	6.21	5.78	**6.22**
	SF	6.75	6.83	7.54	7.59	7.71	**7.78**
	AG	2.37	2.39	2.96	3.05	3.00	**3.10**
	MI	1.44	1.64	1.58	2.31	1.59	**2.39**

Experimental results further show that the performances of three methods based on the saliency detection are superior to the corresponding multi-scale transform fusion methods. From Table 1, it can be observed that IE, SF, AG and MI values of DWT_SD, NSCT_SD and NSST_SD methods are higher than the corresponding DWT, NSCT and NSST methods. However, accuracy of the saliency detection greatly affects the performance of the fused image for those four methods based on saliency detection. DWT_SD method uses L. Itti salient model, combining brightness, color, movement and direction. But the infrared image is lack of color information. So L. Itti model for infrared target detection is not accurate and effects the fusion results. In Fig. 6(b), Fig. 7(b) and Fig. 8(b), the infrared target are only partly highlighted. The qualities of the fused images based on NSCT_SD and NSST_SD methods are much better than those of NSCT and NSST methods, especially for the target contrast and texture details. For Tree image, we almost can't see the contour of trees and roads in Fig. 8 (c) and Fig. 8 (e), but the fused images in Fig. 8 (d) and 8 (f) maintain the contour of trees and roads at the same time to highlight the infrared target. NSST_SD method can obtain more obvious infrared target and more clear edge compared with NSCT_SD. The infrared target in Fig. 6 (d) is surrounded by some shadows but the infrared target contour in Fig. 6 (f) has no artifacts, remaining the integrity of the original information in the source infrared image, no artifacts. This is because the NSCT_SD method adopts spectral residual saliency detection so that the saliency area is too large. NSST_SD method uses the proposed saliency method superior to spectral residual to improve the performance of the fused image.

Moreover, Table 1 shows that all objective evaluation values of NSST_SD method are optimal. Based on the analysis above, it can be concluded that NSST_SD has good performance in terms of quantitative and visual evaluations.

5 CONCULUSION

Descriptions of infrared and visible images for targets and scenes have their pros and cons. The fused image can combine infrared target information of the infrared image and scene detail information of the visible image to describe fully the targets and scenes. We propose a fusion method based on saliency detection using the NSST, NSST_SD. The different fusion rules are designed according to the salient region and nonalienated region. The final fused image is obtained by inverse NSST. The experiment results show that the methods based on saliency detection outperform the traditional DWT, NSCT and NSST method. Furthermore, among the methods based on based on saliency detection, NSST_SD is superior to DWT and NSCT_SD in terms of visual analysis and quantitative evaluation. In the future, it is worthy to further investigate other methods of saliency detection that may affect fusion performance. And it is possible to apply the proposed method to other image fusion of different types of images.

6 REFERENCE

[1] A.Goshtasby, S.Nikolov. Image fusion: Advances in the state of the art. Information Fusion, 2007, 2(8):114-118.

[2] A. Anish, T. J. Jebaseeli. A Survey on multi-focus image fusion methods. International Journal of Advanced Research in Computer Engineering & Technology, 2012, 1(8): 319-324.

[3] A.P. James, B. V. Dasarathy. Medical image fusion: A survey of the state of the art. Information Fusion, 2014, 19:4-19.

[4] Y. Ma, J. Chen, C. Chen, F. Fan, J.Y. Ma, Infrared and visible image fusion using total variation model. Neurocomputing, 2016, 202: 12-19.

[5] W. Gan , X.H. Wu, W. Wu, X. M. Yang, C. Ren, X.H. He. Infrared and visible image fusion with the use of multi-scale edge-preserving decomposition and guided image filter, 2015, 72:37-51.

[6] Y. F. Chen, N. Sang, H. W. Wang. Algorithm for visible and infrared images fusion.Journal of Huazhong University of Science and Technology, 2013, 41(I): 112-115.

[7] J.F. Zhao, Q. Zhou, Y.T. Chen, et al. Fusion of visible and infrared images using saliency analysis and detail preserving based image decomposition [J]. Infrared Physics & Technology, 2013, 56: 93-99.
[8] J.G. Han, E. J. Pauwels, P. Zeeuw. Fast saliency-aware multi-modality image fusion. Neurocomputing, 2013, 111:70-80.
[9] H. Peng, J. F. Zhao, H.J. Feng. Dual band image fusion method based on region saliency. Journal of Zhejiang University, 2012, 46(11): 2109-2115.
[10] R. S. Rosa, J.A. García, J. F. Valdivia. From computational attention to image fusion. Pattern Recognition Letters , 2011,32:1778-1795
[11] N. Imamoglu, W. Lin, Y.M. Fang . A saliency detection model using low-level features based on wavelet transform. IEEE Transactions on Multimedia, 2013 ,15(1): 96 -105
[12] H.Y. Jing, Q. Han, X.M. Niu. Survey of Salient R egion Detection Algorithms.Interlligent computer and application,2014,4(1):37-39
[13] R. Achanta, S. Hemami, F. Estrada. Frequency-tuned salient region detection. IEEE Conference on Computer Vision and Pattern Recognition, 2009, 1597-1604
[14] G. Pajares, J.M. Cruz, A wavelet-based image fusion tutorial. Pattern Recognition, 2004, 37(9):1855-1872
[15] S. Yang, M. Wang, et al. Image fusion based on a new contourlet packet. Information Fusion, 2010, 11(2): 78–84.
[16] Q.G. Miao, C. Shi, et al. A novel algorithm of image fusion using shearlets. Optical Communication, 2011, 284: 1540–1547.
[17] X.Y. Ren, W. Liu, P. Duan.Image fusion based on wavelet transform and visual attention. MIPPR 2009: Multispectral Image Acquisition and Processing, 74941X.
[18] Y. Q. Li. A saliency-based multi-scale image fusion model. Journal of Jilin University,2013,51(2):285-288.