International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.4 (2016), pp.37-46
http://dx.doi.org/10.14257/ijsip.2016.9.4.04

Fusion of Infrared and Visible Image Based on Compressed Sensing and Nonsubsampled Shearlet Transform

WANG Xin[1,2], MENG Jian[2] and LIU Fu[1*1]

1. College of Communication Engineering, Jilin University, Changchun 130022, China ;
2. School of Computer Science& Engineering, Changchun University of Technology, Changchun 130012, China;
E-mail:wangxin315@ccut.edu.cn(WANG Xin)
**Corresponding author E-mail:liufu@jlu.edu.cn(LIU Fu)*

Abstract

In order to solve storage and computation cost problems for the traditional whole sampling image fusion algorithms, a new method of infrared and visible light image fusion is put forward based on compressed sensing (CS) theory. Nonsubsampled shearlet transform (NSST) is introduced as the sparse transform. Compressed sensing is applied to fuse the high frequency subbands decomposed by NSST. The high frequency coefficients are compressed for measured values which are fused by the rules of spatial frequency weighting. Regional energy together with regional standard deviation guides the fusion of the low frequency subband. Finally, the fused image is gained through inverse NSST. The experimental results show that high-quality fused images can be obtained with only one layer NSST. The fused image quality is better than the several traditional fusion algorithms based on compressed sensing.

***Keywords**: Image Fusion, Compressed Sensing, Non-subsampled Shearlet Transform, Spatial Frequency, Minimal total variation*

1. Introduction

Fusion of multi-sensors can increase the credibility of images and the reliability of the imaging system while reducing the requirements for single sensor performance. Fused image can show the related information about the targets or scenes in a more accurate and comprehensive way, which facilitates direct observation and post- processing using computer technology. Fusion of Infrared and visible light images is a very important branch of image fusion. Infrared sensors create images based on the thermal radiation theory which can highlight the target area in the scene but the problem is that it cannot provide enough details of the scenes. Visible light sensors, however, can provide detailed information about the scene via reflected lights from the object. Therefore, fusion of Infrared and visible light not only possesses the good target characteristics of infrared images, but also preserves the details of visible lights.

Pixel-level image fusion method has been used widely because of its high fidelity. However, with the development of sensor technology, the size of the images is increasing, which brings a great challenge to the pixel-level fusion in storage and calculation. Compressive Sensing (CS) is a new kind of signal sampling theory, which breaks the limitation of Nyquist Sampling Theory.

For the sparse signals, it's able to sample signal at a rate much lower than the Nyquist sampling rate and recover the original signal by the reconstruction algorithm. Compressed Sensing theory can reconstruct the original image by using undersampling information,

---

[*] Corresponding Author

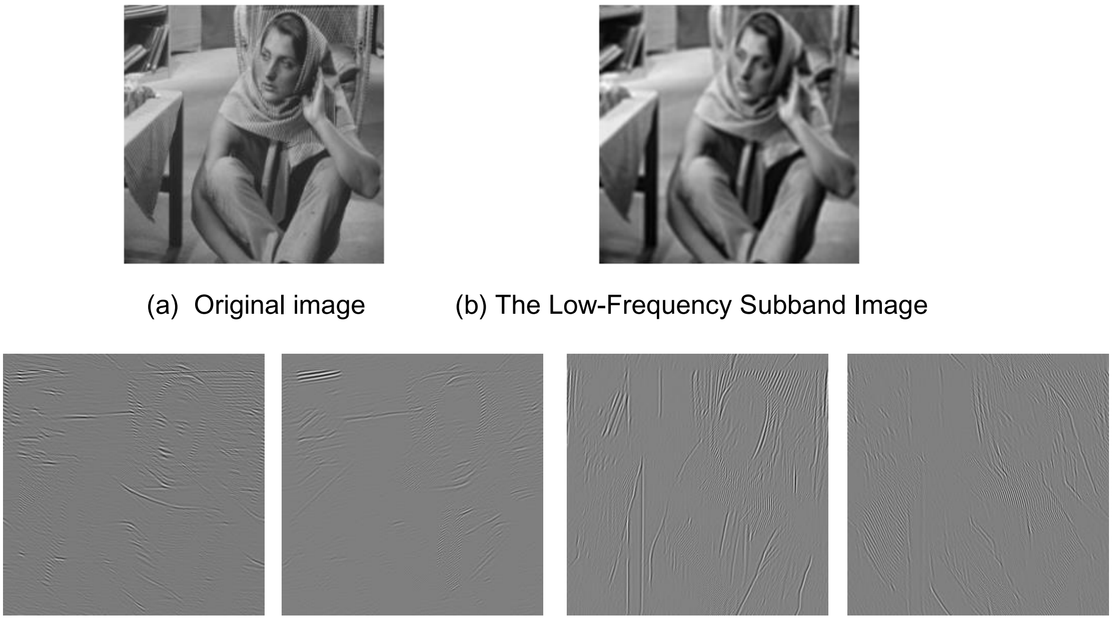
ISSN: 2005-4254 IJSIP

which can greatly improve the fusion efficiency and alleviate the pressure of storage. Fusion algorithms based on Compressed Sensing has been proposed in succession, which has become the research focus at present. Ref. [1-2] studied the image fusion method under the undersampling modes of star, double star, single radioactive, double radioactive with different sparse basis. According to the characteristics of the coefficients distribution with different sparse bases, the sampling mode is designed to improve the fusion quality with the equal sample rate; Ref. [3-4] took the wavelet as the sparse basis and the corresponding fusion algorithm is studied; Ref. [5-6] studied the fusion method under the sparse basis of the trained dictionary. The sparsity and noise robustness are increased through dictionary learning method, which also improves the quality of the fused image with noise pollution.

Discrete Cosine Transform（DCT）,Wavelet transform and trained dictionary(eg. Kernel Singular Value Decomposition (KSVD) dictionary) *et.al.*. are commonly used as image sparse representation in conventional CS fusion algorithms. Because DCT transform does not have good time-frequency analysis performance and wavelet can only capture the limited direction information, the reconstructed image quality is poor; although trained dictionary can achieve better fused image quality, the computation complexity is higher. Nonsubsampled shearlet transform (NSST) is thus introduced as the sparse transform in this paper. The image fusion approach based on CS is applied to the decomposed high frequency subband coefficients. The new fusion rules are developed respectively to the low frequency and the measured high frequency subands. Finally, the inverse transform of NSST is used to get the final fused image. The experimental results show that high-quality fused images can be obtained with only one layer NSST. The fused image quality is better than the several traditional fusion algorithms based on compressed sensing.

2. Sparse Transformation

NSST transform [7] possesses direction sensitivity, translation invariance and many other advantages. It is the real 2D image sparse representation. The process of multi-scale decomposition by NSST is the same as NSCT, which is realized mainly by Nonsubsampled pyramid filter bank (NSP). After N-level NonSubsampled pyramid decomposition, one low frequency sub-band image and N numbers of bandpass subbands images with the same size but different scales are obtained. Unlike NSCT transformation, NSST abandons the original direction filter during the decomposition process, applying shear filter to realize the direction localization of the bandpass subbands. The result of one layer NSST decomposition is shown in Figure 1. After one layer decomposition, a low-frequency sub-band and a band-pass sub-band are obtained. Set the direction localized series of the band-pass sub-band is 2, and then the decomposition directions become 4, namely 4 high frequency subbands. Let l_k be the direction series of the kth layer decomposition and after k-layer NSST, one low frequency subband and $\sum_k 2^{l_k}$ high frequency subbands are obtained. All of these subbands possess the same size as the original image.



(a) Original image (b) The Low-Frequency Subband Image

(c) Four High Frequency Subbands Images

Figure 1. One layer NSST of Image Barbara

Since there is no down sampling during the decomposition process of NSST, translation invariance is maintained. NSST can select direction number of decomposition according to the actual need, which overcomes the limited direction number of NSCT and can better express the image edges and detail characteristics. Compared with the algorithmic complexity, unlike NSCT, NSST doesn't need synthesize inversely direction filters during inverse transformation. Therefore the operation efficiency is much higher. Based on the above analysis, NSST is selected to decompose infrared image A and visible light image B to obtain high and low frequency subband coefficients , $\left\{H^{A}_{j,k}, L_{A}\right\}$ and $\left\{H^{B}_{j,k}, L^{B}\right\}$, where L is the low frequency subband and $H_{j,k}$ is the high frequency subband of jth scale on the direction of k.

3. Image Fusion Algorithm Based on NSST and CS

3.1 Fusion Scheme

After NSST,the source images are decomposed into low-frequency subband and high- frequency subbands. High-frequency subbands are considered to be sparse, but the low-frequency subband is the approximation of the original image at different scales,which is not sparse. If the low-frequency coefficients and the high-frequency coefficients are measured together, the correlation among coefficients of low frequency approximation components is broken, resulting in poor reconstruction results. Therefore, we only measure the high-frequency coefficients and low frequency coefficients are fused directly, which improves the reconstruction accuracy and the effect of the fused image. The specific fusion steps are as follows:

Step1: The coefficients $\left\{H^{A}_{j,k}, L_{A}\right\}$ and $\left\{H^{B}_{j,k}, L^{B}\right\}$corresponding to the two source images A and B are obtained by NSST respectively.

Step2: Fusing the low-frequency coefficients L_A and L_B directly to obtain the fused low-frequency coefficients L_F.

Step3: The high frequency coefficients $H_{j,k}^{A}$ and $H_{j,k}^{B}$ are transformed by DFT and measured respectively to get the corresponding measurements X and Y.

Step4: Fuse X and Y to get the measured value Z.

Step5: After inverse DFT and reconstruction with minimal total variation (TV) algorithm to obtain the fused high frequency coefficients $H_{j,k}^{F}$.

Step6: Inverse NSST transformation to get the final fused image.

The fusion diagram is shown as in Figure 2.

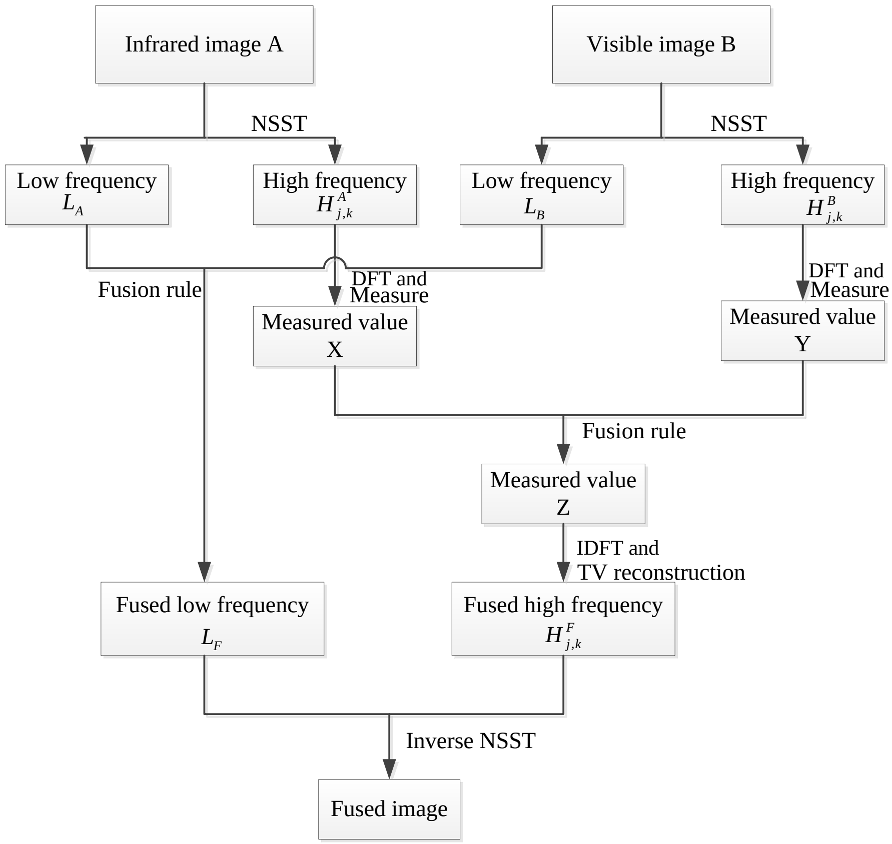


Figure 2. Image Fusion Block Diagram Based on CS

3.2 The Low-Frequency Subband Fusion Rule

Image Fusion based on multi-scale transform focuses more on high frequency sub-bands fusion rules. For low frequency sub-band, "direct average method" is commonly adopted. Simple average method will weaken infrared target, decrease contrast of scenes and cause the loss of some useful source images information. For infrared and visible light image fusion, the primary purpose is to detect the hot target in the scene. Therefore, a good fusion algorithm should be able to transfer as much infrared heat target information as possible to the fused image. In the background area, algorithm should be selected according to the detail richness degree of source images to transfer as much spectral information of the source images as possible to the fused image. Based on this idea, a strategy is proposed based on regional standard deviation and regional energy to jointly guide fusion of low-frequency sub-band coefficients. Specific steps are shown as follows:

(1) Let $L_I(\mathrm{x,y})$ be the NSST low frequency coefficient located at (x,y) of image I (I denotes infrared image A or visible image B). According to formula (1) and (2), local energy and local standard deviation are calculated centered at pixel (x,y). M*N represents the size of local region and is selected to be 3*3 in this paper.

$$E_I(x,y)=\sum_{m\in M}\sum_{n\in N}L_I{}^2(x+m,y+n) \tag{1}$$

$$SD_I(x,y)=\sqrt{\frac{1}{M*N}\sum_{m\in M}\sum_{n\in N}(L_I(x+m,y+n)-\overline{C})^2} \tag{2}$$

$$\overline{C}=\frac{1}{M*N}\sum_{m\in M}\sum_{n\in N}L_I(x+m,y+n) \tag{3}$$

(2) According to formula (4) and (5), ratios of local energy and local standard variation can be calculated for infrared image A and visible light image B.

$$R_E(x,y)=E_A(x,y)/E_B(x,y) \tag{4}$$

$$R_S(x,y)=SD_A(x,y)/\mathrm{SD}_B(x,y) \tag{5}$$

R_E represents the degree of variance between infrared local energy and visible light local energy. If $R_E>1$, then high-energy feature of image A is more significant than that of image B[8]. The standard deviation of image reflects the gray distribution of the image. The higher standard deviation indicates grayscale variation more drastically and the more abundant spectral information. So R_S represents relative abundance degree of spectral information of the two source images. If $R_S>1$, then the spectral information of image A is more abundant than that of image B.

(3) If $R_E\geq T$ ($T=mean[\mathrm{R}_E(x,y)]+\mathrm{k*std}[\mathrm{R}_E(\mathrm{x,y})],\mathrm{k}\in(0,1)$, $mean[\bullet]$ denotes average value and $std[\bullet]$ represents standard deviation), the fused low frequency subband coefficient at (x,y) is selected from the infrared image. If $R_E<T$, then the fused low frequency coefficient should be chosen based on the richness of the region details R_S. If $R_S<1$, the spectral information of visible light image is richer than that of infrared image, the fused coefficient should be chosen from the visible light region.

(4) If the above conditions are all not satisfied, both the details and energy features of the two sources are not so obvious. The adaptive weighted fusion rule is adopted. All the specific expressions described above are as follows:

$$L_F(x,y)=\begin{cases} L_A(x,y) & R_E(x,y)\geq T \\ L_B(x,y) & R_E(x,y)<T \; and \; R_S(x,y)<1 \\ \omega_1*L_A(x,y)+\omega_2*L_B(x,y) & else \end{cases} \tag{6}$$

In which,

$$k_1(x,y)=\frac{E_A(x,y)}{E_A(x,y)+E_B(x,y)}\text{；}\ k_2(x,y)=1-k_1(x,y)$$
$$\lambda_1(x,y)=\frac{SD_A(x,y)}{SD_A(x,y)+SD_B(x,y)}\text{；}\ \lambda_2(x,y)=1-\lambda_1(x,y)\text{；}$$
$$\omega_1(x,y)=\frac{k_1(x,y)*\lambda_1(x,y)}{k_1(x,y)*\lambda_1(x,y)+k_2(x,y)*\lambda_2(x,y)}\text{；}\ \omega_2(x,y)=1-\omega_1(x,y) \tag{7}$$

3.3 The High-Frequency Subbands Fusion Rule

Three major steps of compressed sensing include: image sparse, measurement and reconstruction. In this paper, the high frequency subbands $H_{j,k}^A$ and $H_{j,k}^B$ are firstly transformed by DFT (Discrete Fourier Transform) respectively and get the corresponding results denoted as $D_{j,k}^A$ and $D_{j,k}^B$. Then star matrix as shown in Figure 3 is selected for measurement.

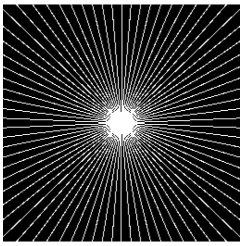


Figure 3. Star Sampling Pattern

In the star matrix, the pixels of white region whose values are “1” are selected as sampling points. After measurement, the new fusion rule of spatial frequency(SF) weighted is adopted here for the reason that the bigger SF is , the richer detail of the image；Furthermore, SF is not sensitive to the noise. The spatial frequency SF of Image $I(i,j)$ is：

$$R_F=\sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j)-I(i,j-1)]^2} \tag{8}$$
$$C_F=\sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j)-I(i-1,j)]^2} \tag{9}$$
$$SF=\sqrt{(R_F)^2+(C_F)^2} \tag{10}$$

Let measurement matrix be ϕ :

$$X=(x_1,x_2,...,x_M)^T=\Phi D_{j,k}^A \tag{11}$$
$$Y=(y_1,\mathrm{y}_2,...,\mathrm{y}_M)^T=\Phi D_{j,k}^B \tag{12}$$

X and Y represent the measurement of the high frequency subbands $D_{j,k}^A$ and $D_{j,k}^B$

respectively. Spatial frequencies of X and Y are calculated respectively based on formula 10 and written as： SF_x and SF_y . Let $Z=(z_1,z_2,...,z_M)^T$ be the measurement after fusion, then：

$$z_i = w_x x_i + w_y y_i, i = 1,2,...,M \tag{13}$$

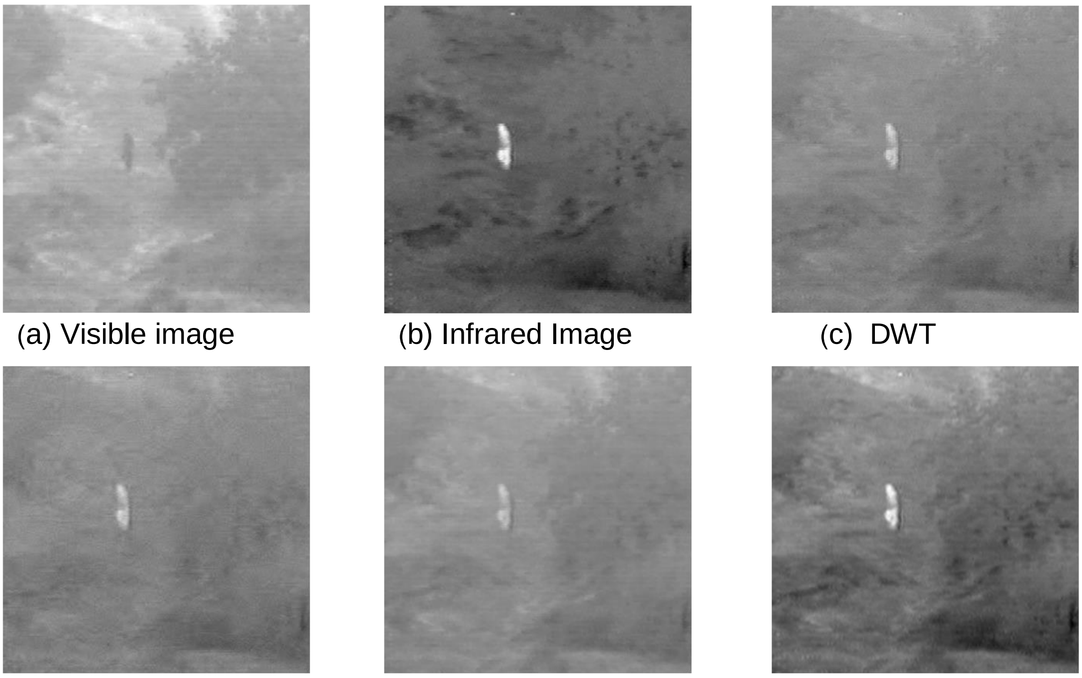
In which :

$$w_x = SF_x/(SF_x + SF_y);\, w_y = SF_y/(SF_x + SF_y) \tag{14}$$

The minimal total variation (TV) algorithm is chosen to reconstruct Z. For the reconstruction of two-dimensional signal, the minimal TV can achieve the good reconstruction results and can better preserve image details [9].The final fused high-frequency subbands are obtained by inverse DFT to Z. Finally the fused high and low frequency subbands are inversely transformed by NSST to obtain the final fused image.

4. Experimental Results

In order to verify the performance of the algorithm, compare the proposed method with the fusion method based on DWT[10], compressed sensing fusion algorithm based on wavelet sparse transform[3] and discrete Fourier transform sparse transfrom[1] respectively (denoted as DWT, DWT + CS, DFT + CS).One-level NSST decomposition with two direction series are chosen, the same level as DWT in Ref.[9]and Ref.[3].The weighted factors are $\alpha = \beta$ =0.5 in Ref.[3].The sampling rates are all 40% in Ref.[3] and Ref.[1],as well as in this paper. The two groups of infrared and visible light source images after registration are respectively named as: "Trees" and "UNcamp" .All the images' size is 256 *256. Fusion experimental results are as shown in Figure.4 – Figure. 5.



(a) Visible image (b) Infrared Image (c) DWT

(d) DWT+CS (e) DFT+CS (f) Proposed method

Figure 4. Fusion Results on Trees Image Set

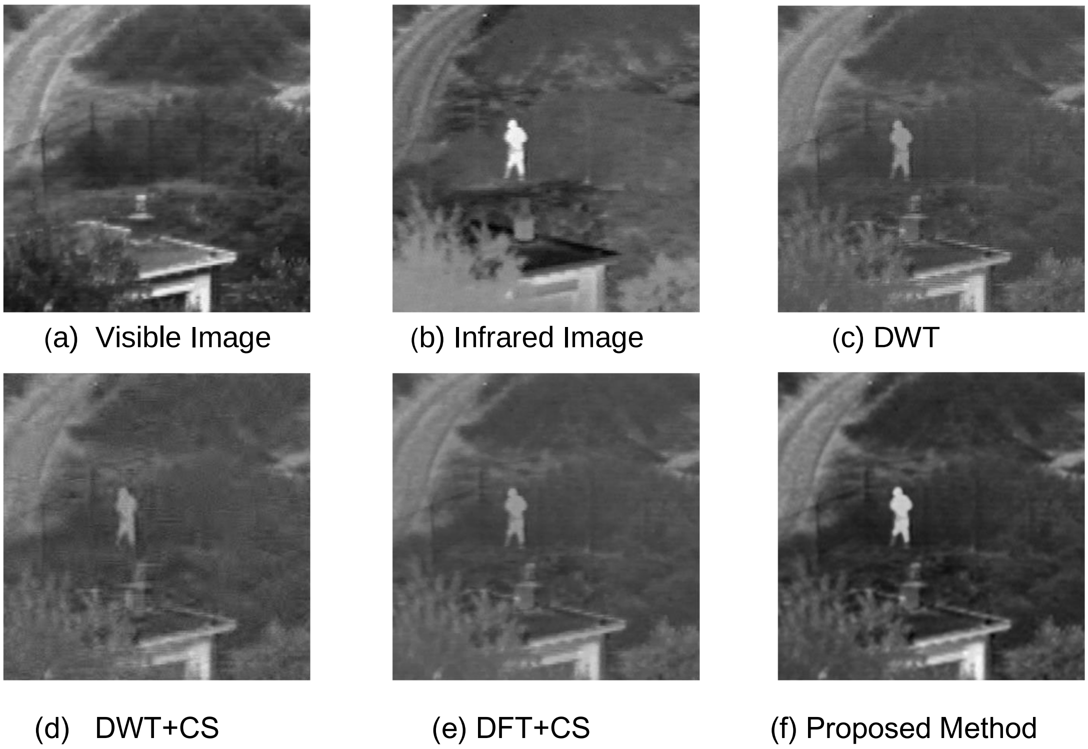


Figure 5. Fusion Results on UNcamp Image Set

Based on the results of the contrast experiments, we can see that details of the fused images obtained from DWT, DWT + CS and DFT+CS are all very vague. For example , the edges of the trees shown in Figure 4(c)-(e); Roads, fences, roof contour and the trees behind of the house all have different degrees of blur in Figure 5(c)-(e). Moreover, the target (person) of the fused image is almost "drowned" in the background area, seen from the fusion results of the three methods. While the proposed method in this paper applies only one-layer NSST which leads to better image fusion effect than the other three algorithms. The target area is highlighted in the fused image and reserves more detail information.

Only point target objects can be well expressed by DWT algorithm, which results in blur edge and contour in the fused image. DWT + CS algorithm captures edges poorly because the algorithm adopts wavelet as sparse matrix; Measurement by random matrix can lead to reflect the structure information incorrectly. While the star matrix is used in this paper to ensure the measured values associated with structural information. Simple weighted average rule for low low-frequency fusion leads to weakened infrared target information by DWT and DWT + CS algorithms. DFT + CS algorithm applies DFT (Discrete 'Fourier transform) to the whole image for sparse and star matrix mainly chooses the low-frequency components located in the center of the image after DFT. And high frequency coefficients are only sampled a few points. All these reasons result in the phenomenon of blurred edge and contour. In this paper, compressed sensing is only used for high frequency sub-bands, and the sampled high frequency coefficients are far more than DFT+CS, which results in a better fusion image quality than the algorithm.

Information entropy (E), standard deviation (SD), mutual information (MI) and edge retention ($Q^{AB/F}$) [11] are used to measure the fused image quality. The objective evaluation results are shown in Table 1. The four evaluation indexes of the algorithm proposed in this paper are of the maximum value for each group images. Objective evaluation of the experimental results has further confirmed the superiority of the algorithm.

Table 1. Objective Evaluation of Different Algorithms

Image set	Method	E	SD	MI	$Q^{AB/F}$
Trees	DWT	6.1809	17.6906	0.7062	0.3344
	DWT+CS	5.9327	15.0240	0.6458	0.2389
	DFT+CS	5.9224	15.0164	0.6900	0.3481
	New	**6.9069**	**24.2645**	**0.9566**	**0.4658**
UNcamp	DWT	6.3052	24.3118	0.7090	0.2882
	DWT+CS	6.2614	23.1550	0.6372	0.2038
	DFT+CS	6.1600	22.6317	0.8210	0.2954
	New	**7.0317**	**34.1352**	**0.8553**	**0.4016**

5. Conclusion

Fusion of infrared and visible images can achieve complementary advantages of the two sensors. NSST has the characteristics of multidirectional, translation invariance and good real-time. In this paper, NSST is introduced to CS fusion. A new compressed sensing fusion scheme is proposed to achieve fusion of high-frequency coefficients. Use the star observation matrix, which makes the measured values associated with the structural information; spatial frequency weighted fusion rule is adopted rather than “choose maximum” can avoid introducing noise in the fusion process; Minimum TV algorithm can ensure the reconstruction quality. The use of local energy feature together with local detail feature fuses the low-frequency subband directly, which effectively highlights the target area. Simulation results show that the proposed fusion algorithm achieves better image fusion effect than several other conventional compressed sensing algorithms.

Acknowledgement

This work is supported by National Natural Science Foundation of China， and by Jilin Province Natural Science Foundation under Grant NO.201215127.

References

[1] X. LI and S.-Y. Qin, “Efficient fusion for infrared and visible images based on compressive sensing principle [J]”, IET Image Processing, vol. 5, no. 2, **(2011)**, pp. 141-147.
[2] B. Kang and W. P. Zhu, “Fusion framework for multifocus images based on compressed sensing[J]”, IET Image Processing， vol. 7, no. 4, pp. **(2013)**， pp. 290—299.
[3] W. Xu, J. G. Wen and Y. ZH. Chen, “Fusion of remote sensing image with compressed sensing based on wavelet sparse basis[C]”, 2014 Sixth International Conference on Measuring Technology and Mechatronics Automation, Hunan, China, January 10 - 11 **(2014)**.
[4] X. Sh. Huang， Q. F. Dai and Y. Q. Cao, “Compressive sensing image fusion algorithm based on wavelet sparse basis [J]”, Application Research of Computers， vol. 29, no, **(2012)**, pp. 3581—3583.
[5] S. Li , H. Yin and L. Fang, “Remote sensing image fusion via sparse Representations over learned dictionaries[J]”, JEEE Transactions on Geoscience and Remote Sensing， vol. 51, no. 9, **(2013)**， pp. 4779-4789.
[6] M. Ding, L. Wei and B. Wang, “Research on fusion method for infrared and visible images via compressive sensing[J]”, Infrared Phys. Technol, **(2013)**, pp. 57: 56 -67.
[7] G. Easley, D.Labate and W. Q. Lim, “Sparse directional image representation using the discrete shearlet transforms[J]”, Appl. Comput. Harm. Anal, vol. 25, no. 1, **(2008)**, pp. 25-46.
[8] L. hongyan, N. Ziyu, L. Ke and L.Meng, “Image weighted fusion algorithm based on regional energy

and layers correlation [J]", Journal of Harbin University of Science and Technology, vol. 19, no. 5, **(2014)**, pp. 99-102.
[9] E. Candès, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements[J]", Communications on Pure and applied Mathematics, vol. 59, no. 8, **(2006), pp.**1207-1233.
[10] Gui Hong-qu, Da Li-zhang, Ping Fan-yan. Medical image fusion by wavelet transform modulus maxima[J]. OPTICS EXPRESS, 2001, 9(4): 184-190.
[11] Xydes C S, Petrovi V. Objective image fusion performance measure[J]. Electronics Letters, 2000, 36 (4): 308-309.

