A Improved Infrard And Visible Light Image Fusion Algorithm

XIAO Zhong-jie
Network Center
Qinghai University For Nationalities
Xining, China,810007
E-mail: xiaozhongjie0951@163.com

Abstract—**On the basis of Nonsubsampled Contourlet Transform (NSCT) research, this paper proposed an improved image fusion algorithm based on the infrared and visible light image fusion. First of all, the NSCT was used to decompose the infrared image and the visible light image. Then the pixel feature energy weighted fusion rule and the neighborhood variance feature information fusion rule were used to get the low-frequency and high-frequency coefficients. Finally, the inverse NSCT was used to reconstruct the image to get the fused image. The experiment results show that this algorithm could effectively improve the image fusion quality.**

Keywords- image fusion；（Nonsubsampled Contourlet Transform, NSCT）；reconstruction；infrared image；visible light image

I. INTRODUCTION

Image fusion is the process of combining the images or image sequences based on the same scene captured by two or more sensors in the same time or different time into a single image. The infrared images captured by the infrared sensor focus on the target in the scene. The visible image could better describe the background information. It could get panoramic description by fusing the infrared image and the visible image with the same scene together. The resulting image will be more informative than any of the input images [1].

There are two kinds of image fusion algorithm. They are the space domain based algorithm and the transform domain based algorithm. In the space domain algorithm, there are weighted average method, principal component analysis and etc. The multi-scale analysis method is a kind of transform domain based algorithm. In the multi-scale analysis method, the wavelet transform are widely used in the image fusion because of its multi-resolution and time-frequency localization characteristics [2]. But the wavelet transform could not recognize the line singularity and surface singularity inherent in natural images. And its capture directionality information is restricted [3]. So the fused image usually generate details blur. The NSCT [4] not only effectively capture the geometric characteristics in the images, but also have the multi-scale and time-frequency localization characteristics. It also has the multi-directional characteristic. On the basis of this, the image energy is more concentrated after the sparse representation. After the NSCT decomposition, every sub-band images have the same size. It is convenience for the following fusion calculation. In the

discrete domain, the NSCT adopted the filter bank to realize the image multi-scale and multi-directional decomposition. The computing speed is improved.

According to the NSCT translation invariant characteristics and perfect description capability for the image geometric characteristics especially the edge direction information, this study proposed an improved infrared image and visible light image fusion algorithm. It adopted the pixel feature energy weighted fusion rule for the low-frequency sub-band coefficients. It used the neighborhood variance feature information fusion rule for the high-frequency sub-band coefficients. The experiment results show that the improved algorithm has good fusion performance.

II. BASIC THEORY

Contourlet transform is a kind of two-dimensional image representation mode. It is consists of two steps. They are the Laplacian Pyramid [5-6] (LP) decomposition and Directional Filter Bank (DFB) filter. First of all, the LP is used to decompose the original image to the low-frequency sub-band and high-frequency sub-ban. The high-frequency sub-band is decomposed to several directional sub-bands. For the low-frequency sub-band, above process is repeated to do the multi-resolution and multi-direction decomposition. In the LP decomposition process, the interlaced sampling is used on the image. This process reduces redundancy of the transformation to a great extend. In the same time, it causes the result that the Contourlet transform has no translation invariance feature. This limits the application range. So M.N.Do proposed the Nonsubsampled Contourlet Transform (NSCT) to change this situation.

In the NSCT, the Nonsubsampled Pyramid Filter Bank (NSPFB) is used to do multi-scale decomposition. Then the Nonsubsampled Direcional Filter Bank (NSDFB) is used to do directional decomposition on each sub-band image. The processing procedure is shown in fig.1 [7]. The input original image and a two-dimensional filter are convoluted to get the low-frequency images. The difference of the original image and the low-frequency image is the high-frequency detail image. Then, the high-frequency detail image is filtered by the NSDFB to get the directional details. By the N-layer NSCT decomposition, the input image can be decomposed to $1+\sum_{j=1}^{N} 2^{l_j}$ sub-band images. The sub-band images have the same sizes with the original images. Where l_j are the directional decomposition levels on the scale j.
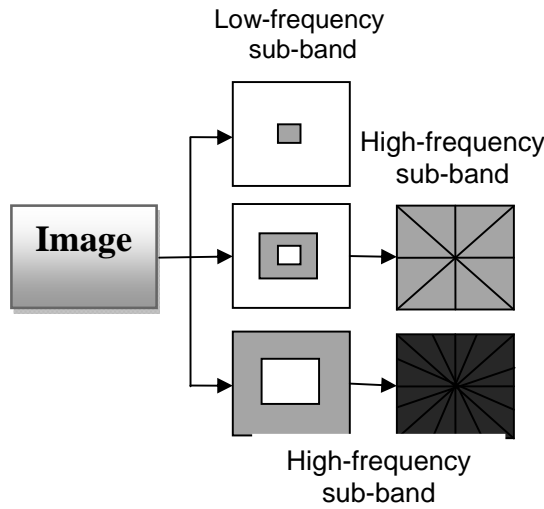
Low-frequency
sub-band

High-frequency
sub-band

Image

High-frequency
sub-band

Figure 1.   NSCT Block Diagram

III.   THE STEPS OF THIS FUSION ALGORITHM

The steps of this fusion algorithm are shown as follows [8]:

- Step1 Do multi levels NSCT transform on the original images A and B. The result is the NSCT coefficients of different scales. They are the low-frequency coefficients and the high-frequency coefficients.

- **Step2** Fuse the low-frequency efficient by the pixel feature energy weighted fusion rule. The results are the low-frequency coefficients.

- **Step3** Fuse the high-frequency efficient by the neighborhood variance feature information fusion rule. The results are the high-frequency efficient.

- **Step4** Reconstruct the low-frequency and high-frequency coefficients in each scale. The inverse NSCT is used to reconstruct image. The result is the fused image.

IV.   THE ALGORITHM IN THIS PAPER

A.   Low-frequency sub-band coefficients fusion

The low-frequency sub-band contains most of the energy. It is the image contour. The main aim of the infrared and visible light image fusion is abstracting the target. The target has large energy. So the decomposed low-frequency component is most energy of the transformed image. So the choice of the low-frequency coefficients fusion rule is vital for the infrared and visible image fusion. This study adopted pixel feature energy weighted fusion rule.

The detail information of the image is represented by the variance. The average brightness of the pixel is represented by the pixel mean. The brightness of the pixel in certain region is represented by window mean. The calculation formula of the variance, pixel mean and window mean is shown as follows:

$$\sigma(p) = \sum_{m \in K, N \in N} [S(x, y) - U(p)]^2$$

$$U(p) = \frac{1}{(M \times N)} \sum_{m \in M, n \in N} [S(x+m, y+n)]$$

$$U_R(p) = \frac{1}{(M \times N)} \sum_{m \in M, n \in N} [U(x+m, y+n)]$$

Where, $\sigma(p)$, $U(p)$ and $U_R(p)$ denote the local invariance centering around the point, pixel mean and window mean respectively. M and N denote the window size. m and n range between K and L. The physical significances of the corresponding pixels between the infrared image and the visible image are different. If only calculates pixel window energy without considering the regional characteristics, it would lose the difference between the target contour and the background contour. So it is necessary to consider the relevance between the image pixel energy and the located region feature. On this basis, a novel region feature based pixel energy is defined as follows:

$$E(x, y) = U^2(p) + K^s(p) \times \frac{(C(p) - U(p))^2}{\sigma^2} \times C^2(p)$$

Where $K^S(p) = \frac{[U_R^S(p)]^2}{([U_R^I(p)]^2 + [U_R^S(p)]^2)}$, $S \in \{I, V\}$.

I and V represent the infrared and visible light image. $U^2(p)$ denotes the main pixel energy. In the $K^s(p) \times \frac{(C(p) - U(p))^2}{\sigma^2} \times C^2(p)$, K is the adjustment factor. $[U_R^S(p)]^2$ represents the region features by the energy. It adjusts the ratio of the details to the overall energy by the pixel location. $\frac{(C(p) - U(p))^2}{\sigma^2}$ denotes the contribution rate of the detail to overall energy. When the pixel is in the target area, the pixel energy in the infrared image is comparatively concentrated. It has large energy in this region. While in the visible light image, it has small region energy. The weighting K helps to retrain the details in the visible image. When the pixel is in the background regions, the region energy variation is relatively small. K changes little. The visible image has abundance details. So it has large detail energy. In a word, the feature energy compromises the merits of infrared and visible image advantages. It considers the overall features to the energy. It avoids the limitation of the mechanistic energy weighting.

$$C_{j,0}^F(x, y) = \omega_1 C_{j,0}^I + \omega_2 C_{j,0}^V$$

$$\omega_1 = \frac{E_{j,0}^I(x,y)}{E_{j,0}^I(x,y) + E_{j,0}^V(x,y)}$$

$$\omega_1 = \frac{E_{j,0}^V(x,y)}{E_{j,0}^I(x,y) + E_{j,0}^V(x,y)}$$

B. High-frequency coefficient fusion

The high-frequency got by the NSCT decomposition represents the details of the image. The larger the region variance, the clearer the corresponding region of the original image will become. So the choice of the high-frequency coefficients could be guided by the neighborhood variance feature information. The neighborhood variance of certain coefficient can be defined as follows:

$$STD_l(i,j) = \sum_{(m,n) \in \omega} (|H_l(m,n)| - \bar{H}_l)^2$$

Where, $STD_l(i,j)$, $l = A, B$ is the neighborhood high-frequency coefficient variance. $H_l(m,n)$, $l = A, B$ is the high-frequency coefficient of the image l in point (m,n). $\bar{H}_l$, $l = A, B$ is the mean of the absolute value of high-frequency coefficient of image l in region ω. And there is $i - k \le m \le i + k$, $j - k \le n \le i + k$, $\omega = 2 * k + 1$, $k = 1, 2, \ldots$.

The original image with larger neighborhood high-frequency coefficient variance is chosen as the high-frequency coefficient of the fused image.

$$H_F(i,j,k) = \begin{cases} H_A(i,j,k) & SDT_A(i,j) \ge SDT_B(i,j) \\ H_B(i,j,k) & SDT_A(i,j) < SDT_B(i,j) \end{cases}$$

Where $H_F(i,j,k)$, $l = A, B, F$ is the high-frequency coefficients of the image in the k–layer decomposition point (i,j).

C. Image decompostion

Every directional sub figure in each scale could be processed by the above method to finish the image fusion. Then the inverse NSCT is used on the fused sub-band coefficient to get the reconstructed fusion image F.

$$H_F^{i,j,k} \cdot C_{j,0}^F = F$$

### V. EXPERIMENT RESULT AND ANALYSIS

The experiment adopted the strictly registered images infrared and visible images. In fig.2, (a) and (b) are the infrared image and visible image respectively. The size of them is 256×256. In order to evaluate the effect of this algorithm, this paper compared this algorithm with the Laplacian pyramid method (LP method), Wavelet method (WT method), traditional Contourlet fusion method

(Contourlet method). In the LP method, the number of pyramid decomposition layers is 2. The fusion rule is that the sub-image on the top layer was fused by the average method, and the sub-images on other layer were fused by selecting larger pixel values method. In the WT method, the maximum value was taken in the high-frequency layer and the average value was taken in the low-frequency layer. The decomposition directions in each layer of Contourlet method and this algorithm are 1, 2, 2.

The infrared image, visible light image and the fused results fused by different algorithm are shown in fig 2. Fig 2(a) is infrared image. In this image, the targets are clear and other scenes are fuzzy. Fig 2(b) is visible light image. It is in low light. So the targets are blurring. But the background is clear. Fig2 (c)-(f) are the fused results fused by the LP method, WT method, Contourlet method and the method in this paper.

Assessing from the subjective standpoint, for the visual effect, the fused image got by this algorithm better obtains the scene feature information in the visible image, and the target information in the infrared image, as well as the edge details. Although the fused images got by the LP method and the WT method reserves the main scene, the edges are blurry. The fused image got by the Contourlet method has clearer edges than the LP and WT method, but its background is not as clear as it in the fused image got by this algorithm. In fig.2, the targets in the fused image by this method are clear than them in other three fused images.

(a) Infrared image   (b) Visible image

(c) LP method   (d) WT method
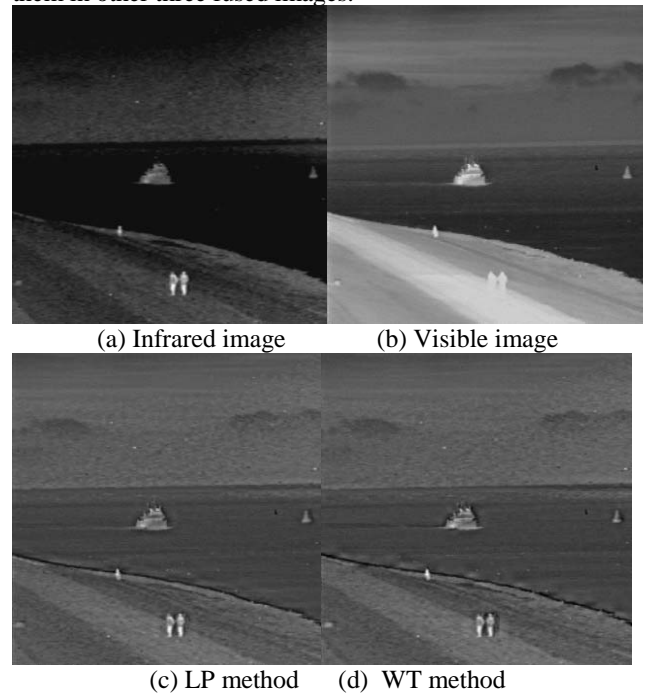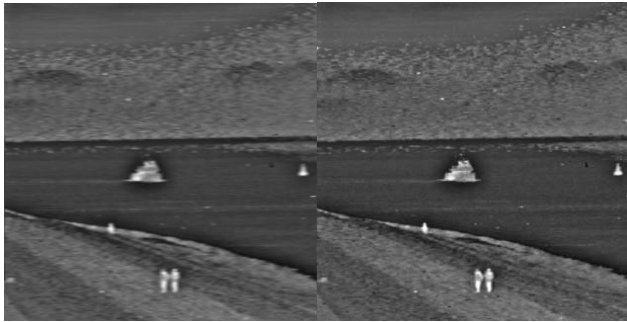
(e) Contourlet method   (f) Method in this paper

Fig.2 Visible image, infrared image, fused images using different methods

Assessing from the objective standpoint, entropy (EN), mutual information (MI) and average gradient (AvG) are used as the object evaluation factors of the image fusion effect. The entropy is one important factor of the image information rich degree. The mutual information is used to calculate how much information is transferred to the fusion result. The average gradient is used to describe the clarity of the image. The larger the three values, the better the fusion results. The table 1 shows that the entropy, mutual information and average gradient are all higher than other three methods. So the method in this paper is better than other three methods.

Tab. 1   Objective evaluation of fusion results

Method	EN	MI	AvG
LP	6.8235	2.3842	5.5781
WT	6.8324	2.4928	5.6356
Contourlet	6.8539	2.4245	5.7658
This Method	7.0627	2.6639	5.8936

VI. CONCLUSION

The complementary or redundancy information supplied by the infrared and visible images could effectively improve signal to noise ratio and obtain reliable image information. This paper proposed a improved NSCT fusion method based on the infrared and visible light images characteristics and fusion requirement. This paper improved the high-frequency coefficient and low-frequency coefficient fusion rules. The low-frequency sub-band images adopted the pixel feature energy weighted fusion rule. The high-frequency sub-band images adopted the neighborhood variance feature information fusion rule. The fusion experiment results show that this algorithm has good robustness. It could effectively extract edges and texture information. The fused images have abundance scene information and clear target. So this algorithm is an effective infrared and visible image fusion method.

REFERENCES

[1] ZHOU Feng-fei,CHEN Wei-dong,LI Liang-fu. Fusion of IR and visible images using region growing [J]. Journal of Applied Optic,2007,28(6)：737-741.

[2] WANG Hong-mei,ZHANG Ke,LI Yan-jun. Image fusion algorithm based on wavelet transform [J]. Infrared and Laser Engineering,2005,34(3)：328-332.

[3] Cheng Yinglei, Zhao Rongchun, Wan Bing. An optimal algorithm of multisensor image fusion based on wavelet transform[C] // Proceedings of7th International Conference on Signal Processing, Beijing,2004: 1049-1051.

[4] da Cunha A L,Zhou J P,Do M N. The nonsubsamp- led contourlet transform : theory,design,and applications [J]. IEEE Transactions on Image Processing(S1057-7149),2006,15(10)：3089-3101.

[5] Do M N, Vetterli M. Framing pyramids[J].IEEE Trans.Signal Proc., 2003,51(9): 2329-2342.

[6] Bamberger R H, Smith M J T. A filter bank for the directional decomposition of images: Theory and design[J].IEEETrans. Signal Proc., 1992,40(4): 882-893.

[7] Huang Qingqing; Ji Yuan; Yang Jian. Improved fusion method for infrared and visible remote sensing imagery using NSCT. Industrial Electronics and Applications (ICIEA), 2011 6th IEEE Conference. 2011: 1012 – 1015.

[8] Songfeng Yin, Liangcai Cao, Qiaofeng Tan, Guofan Jin. Infrared and Visible Image Fusion based on NSCT and Fuzzy Logic. Mechatronics and Automation (ICMA), 2010 International Conference. 2010: 671 - 675