Region-based fusion of infrared and visible images using nonsubsampled contourlet transform

Baolong Guo (郭宝龙), Qiang Zhang (张 强), and Ye Hou (侯 叶)

ICIE Institute, School of Electromechanical Engineering, Xidian University, Xi'an 710071

Received October 9, 2007

With the nonsubsampled contourlet transform (NSCT), a novel region-segmentation-based fusion algorithm for infrared (IR) and visible images is presented. The IR image is segmented according to the physical features of the target. The source images are decomposed by the NSCT, and then, different fusion rules for the target regions and the background regions are employed to merge the NSCT coefficients respectively. Finally, the fused image is obtained by applying the inverse NSCT. Experimental results show that the proposed algorithm outperforms the pixel-based methods, including the traditional wavelet-based method and NSCT-based method.

OCIS codes: 100.0100, 350.2660, 110.3080.

Image fusion aims at synthesizing information from multiple source images to obtain a more accurate, complete and reliable description of the same scene. The fusion of the infrared (IR) image and the visible image is an increasingly important topic and is being employed in many fields such as night vision, video surveillance, and so on.

During the last decade, a number of fusion algorithms have been proposed, and the fusion methods based on the multiscale transform (MST) are the most typical. The commonly used MST tools include the Laplacian pyramid[1,2] and the wavelet transform (DWT)[3]. In general, due to the perfect properties of the DWT such as multiresolution, spatial and frequency localization, and direction, the DWT-based methods are superior to the pyramid-based methods[4]. However, the DWT also has some limitations such as limited directions and non-optimal-sparse representation of images. Thus, some artifacts are easily introduced into the fused images using the DWT-based methods, which will reduce the quality of the resultant image consequently[5]. In 2006, Cunha *et al.* proposed a novel multiscale geometric analysis (MGA) tool, namely, the nonsubsampled contourlet transform (NSCT)[6]. The NSCT is not only with multiscale, localization, and multi-direction, but also with properties of shift-invariance and the same size between each subband image and the original image. Therefore, the NSCT is more suitable for image fusion.

In addition, most of the above fusion algorithms are mainly pixel-based methods. However, for most image fusion applications, it seems more meaningful to combine objects rather than pixels[7]. Therefore, some region-based fusion algorithms[8−12] have been proposed in recent years. In this paper, we present a novel region-based fusion algorithm (NSCT_RG) for IR and visible images using the NSCT. Segmentation is firstly performed on the IR image and, consequently, the NSCT coefficients from the target regions and the background regions are merged separately. Finally the fused image is obtained by performing inverse NSCT.

The NSCT is a shift-invariant version of the contourlet transform[13]. To achieve shift-invariance, the NSCT eliminates the downsamplers and the upsamplers during the decomposition and the reconstruction of the image, and then it is built upon the nonsubsampled pyramid filter banks (NSPFBs) and the nonsubsampled directional filter banks (NSDFBs). Figure 1 displays the construction of the NSCT.

The NSPFB and NSDFB, employed by the NSCT, are both two-channel nonsubsampled filter banks (NSFBs). And both of them satisfy the Bezout identity, which guarantees the perfect reconstruction. To achieve the multiscale decomposition, the two-channel NSPFB is iteratively used. Such expansion is conceptually similar to the one-dimensional (1D) 'à trous' wavelet transform. To achieve finer direction decomposition, the NSDFB is also iteratively used. For example, to achieve the four-channel direction decomposition, the image is firstly filtered by the original fan filters; Secondly, the filtered subband images are respectively filtered by the upsampled filters, in which the sampling matrix $\mathbf{D}$ is the quincunx matrix, i.e., $\mathbf{D} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. Then the four-channel direction decomposition is obtained.

The building block two-channel NSFBs in the NSPFB and the NSDFB are invertible; therefore, the NSCT is clearly invertible. As well, the NSCT can satisfy the anisotropic scaling law, which is a key property in establishing the expansively nonlinear approximation behavior. This property can be ensured by doubling the number of directions in the NSDFB expansion at every other

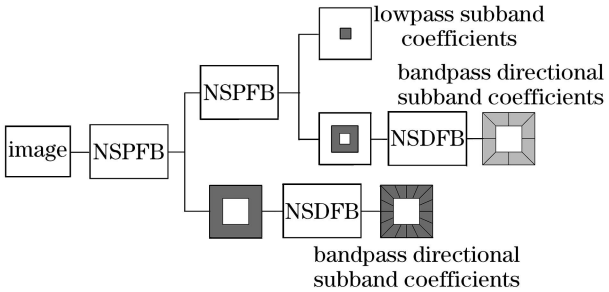

Fig. 1. Nonsubsampled contourlet transform.

1671-7694/2008/050338-04 © 2008 Chinese Optics Letters

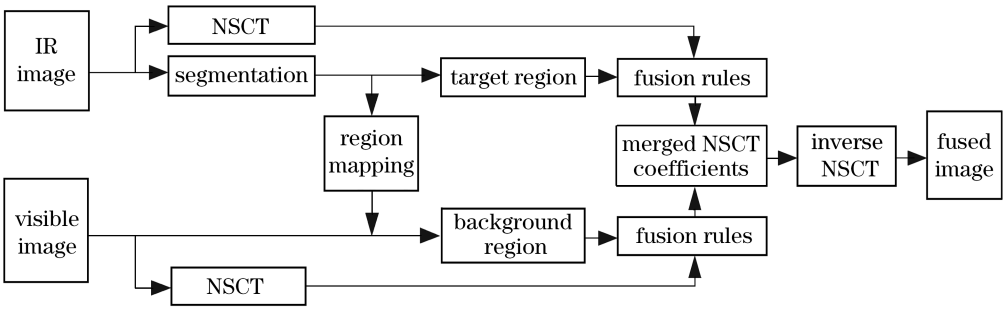

Fig. 2. Block diagram of the proposed image fusion algorithm.

scale. After J level NSCT decomposition, one lowpass subband image and $\sum_{j=1}^{J} 2^{l_j}$ bandpass directional subband images are obtained, all of which have the same size as the input image. And the l_j denotes the direction decomposition levels in the NSDFB at the jth scale.

Figure 2 shows the block diagram of the proposed NSCT_RG algorithm. To be convenient, we suppose the source IR and visible images have been registered before image fusion.

During the fusion process, the first step is to choose the IR image as object image to perform segmentation by region merging. In the segmented IR image, it is easy to find the target regions, which have higher contrast compared with the neighboring background[8]. Then a confidence measure[12] for each candidate region is computed, and candidate regions with high confidence are selected as target regions. The regions are also mapped to the visible image.

The second step is to employ different fusion rules for the target regions and the background regions to merge the source NSCT coefficients. For the target regions, a special fusion rule should be employed to preserve the full target information as much as possible[8]. Therefore, the corresponding fusion rule for the target region $R_{\rm T}$ can be written as

$$C^{\rm F}(m,n) = C^{\rm Iir}(m,n), \qquad (m,n) \in R_{\rm T}, \tag{1}$$

where $C(m,n)$ includes all subband coefficients of the NSCT.

Compared with the target regions, the background regions are with abundant detail information. To extract more information from the source images as much as possible, we firstly introduce the structure similarity (SSIM)[14], which is defined as:

$$S_{\rm SSIM}({\rm Iir},{\rm Ivi}) = \frac{\left(2\overline{\rm Iir}\cdot\overline{\rm Ivi}+C_1\right)}{\left(\overline{\rm Iir}^2+\overline{\rm Ivi}^2+C_1\right)}\cdot\frac{(2\sigma_{\rm Iir,Ivi}+C_2)}{(\sigma_{\rm Iir}^2+\sigma_{\rm Ivi}^2+C_2)}, \tag{2}$$

where C_1 and C_2 are small constants. $\overline{\rm Iir}$ and $\overline{\rm Ivi}$ denote the mean of the IR image Iir and the visible image Ivi, respectively, $\sigma_{\rm Iir}$ and $\sigma_{\rm Ivi}$ denote the corresponding variance, and $\sigma_{\rm Iir,Ivi}$ denotes the covariance. Here, we employ the SSIM to measure the corresponding region similarity between the source images. And a threshold α is also introduced in this paper.

In case $S_{\rm SSIM}({\rm Iir},{\rm Ivi}) \geq \alpha$, the two regions of the IR image and the visible image are with higher similarity and more redundant information exists between the source images[15]. Then the fusion rules are written as

$$C_{j_0}^{\rm F}(m,n) = \left(C_{j_0}^{\rm Iir}(m,n) + C_{j_0}^{\rm Ivi}(m,n)\right)/2, \ (m,n) \in R_{\rm B}, \tag{3}$$

$$C_{j,r}^{\rm F}(m,n) = \begin{cases} C_{j,r}^{\rm Iir}(m,n), & E_{j,r}^{\rm Iir}(m,n) > E_{j,r}^{\rm Ivi}(m,n) \\ & \& \quad (m,n) \in R_{\rm B} \\ C_{j,r}^{\rm Ivi}(m,n), & E_{j,r}^{\rm Iir}(m,n) \leq E_{j,r}^{\rm Ivi}(m,n) \\ & \& \quad (m,n) \in R_{\rm B} \end{cases}, \tag{4}$$

where $C_{j_0}(m,n)$ denotes the lowpass subband coefficients, $C_{j,r}(m,n)$ denotes the bandpass directional subband coefficients at the jth scale and on the rth direction in the NSCT domain. $R_{\rm B}$ denotes the background region. $E_{j,r}(m,n)$ denotes the local area energy of the NSCT coefficients and is defined as (the size of the local area $M_1 \times N_1$ may be 3×3 or 5×5, ect.)

$$E_{j,r}(m,n) = \sum_{x=-(M_1-1)/2}^{(M_1-1)/2} \sum_{y=-(N_1-1)/2}^{(N_1-1)/2} |C_{j,r}(m+x,n+y)|^2. \tag{5}$$

In case $S_{\rm SSIM}({\rm Iir},{\rm Ivi}) < \alpha$, the two regions are with less similarity and more conflictive information exists between the source images[15]. To select the regions as the fused regions properly, a region salience measure should be defined. Here, to make best use of the direction information of the NSCT, we use the normalized Shannon entropy[11] as the region salience measure, which is defined as

$$S(R) = \frac{1}{|R|} \sum_{\forall j,r,(m,n)\in R} |C_{j,r}(m,n)|^2 \log |C_{j,r}(m,n)|^2, \tag{6}$$

with the convention $0\log 0 = 0$, where $|R|$ is the size of the region R. Then the fusion rule can be written as

$$C^{\rm F}(m,n) = \begin{cases} C^{\rm Iir}(m,n), & S^{\rm Iir}(R) > S^{\rm Ivi}(R) \\ & \& \quad (m,n) \in R_{\rm B} \\ C^{\rm Ivi}(m,n), & S^{\rm Iir}(R) \leq S^{\rm Ivi}(R) \\ & \& \quad (m,n) \in R_{\rm B} \end{cases}. \tag{7}$$

Table 1. Performance of Different Fusion Methods

	Method	AVE_PL	DWFT_PL	NSCT_PL	NSCT_RG
Experiment 1	Q(Iir,F)	0.1673	0.1966	0.2010	0.1743
	Q(Ivi,F)	0.1962	0.2722	0.2802	0.3318
	Q(Iir,Ivi,F)	0.3635	0.4688	0.4811	0.5061
	L(Iir,Ivi,F)	0.6071	0.3841	0.3811	0.3747
	N(Iir,Ivi,F)	0.0620	0.3413	0.3226	0.2429
Experiment 2	Q(Iir,F)	0.2485	0.2939	0.2982	0.2497
	Q(Ivi,F)	0.2314	0.2565	0.2591	0.3484
	Q(Iir,Ivi,F)	0.4798	0.5504	0.5573	0.5980
	L(Iir,Ivi,F)	0.4816	0.3065	0.3027	0.2782
	N(Iir,Ivi,F)	0.0820	0.3290	0.3237	0.2018

After merging the NSCT coefficients, the final step is to perform the inverse NSCT on the merged coefficients, and then the fused image is obtained.

Two sets of IR and visible images of the same scene have been used to demonstrate the effectiveness of the proposed fusion algorithm. For comparison, the fusion is also performed using other three pixel-based methods including the averaging method (AVE_PL), the discrete wavelet frame transform based method (DWFT_PL) and the NSCT based method (NSCT_PL). In the DWFT_PL method and the NSCT_PL method, the lowpass subband coefficients and the bandpass subband coefficients are simply merged by the 'averaging' scheme and the 'absolute-value-maximum' scheme respectively. Figures 3 and 4 show the fusion results.

Visual comparison indicates that our proposed method is far superior to the other three methods. In Fig. 3, it is clear that the fence detail from the visible image is far better transferred into the fused image obtained by the proposed method and the human figure is much brighter in Fig. 3(f) than that in the fused images obtained by the other methods. In Fig. 4, similar conclusions can also be obtained.

The gradient-based representation metric[16] is used as the evaluation criteria for quantitative assessment of the fusion performance. We use the symbols $Q(\mathrm{Iir},\mathrm{F})$ and $Q(\mathrm{Ivi},\mathrm{F})$ to represent the information contributions of individual input images Iir and Ivi toward the fused image F in the fusion process. And we use the symbols $Q(\mathrm{Iir},\mathrm{Ivi},\mathrm{F})$, $L(\mathrm{Iir},\mathrm{Ivi},\mathrm{F})$ and $N(\mathrm{Iir},\mathrm{Ivi},\mathrm{F})$ to represent the total information transferred from the two inputs, the information lost during the fusion process and the artificial information introduced into the fused image, respectively. Larger $Q(\mathrm{Iir},\mathrm{F})$, $Q(\mathrm{Ivi},\mathrm{F})$ and $Q(\mathrm{Iir},\mathrm{Ivi},\mathrm{F})$ values indicate that more information has been extracted from the input images and then injected into the fused image. Smaller $L(\mathrm{Iir},\mathrm{Ivi},\mathrm{F})$ and $N(\mathrm{Iir},\mathrm{Ivi},\mathrm{F})$ values indicate that less useful information has been lost and less false information has been introduced during the fusion process.

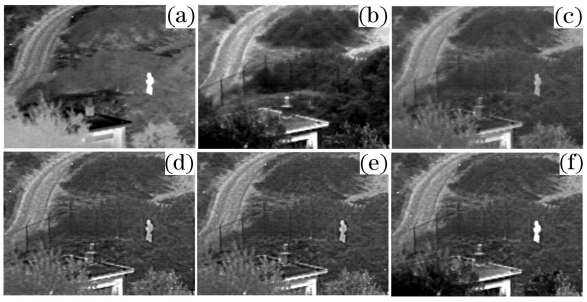

Fig. 3. Test images and fusion results in experiment 1. (a) IR image; (b) visible image; (c) fused image using the AVE_PL method; (d) fused image using the DWFT_PL method; (e) fused image using the NSCT_PL method; (f) fused image using the proposed NSCT_RG method.

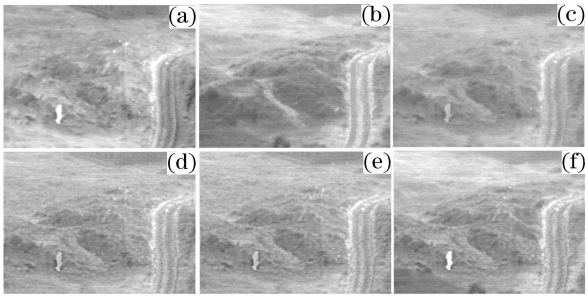

Fig. 4. Test images and fusion results in experiment 2. (a) IR image; (b) visible image; (c) fused image using the AVE_PL method; (d) fused image using the DWFT_PL method; (e) fused image using the NSCT_PL method; (f) fused image using the proposed NSCT_RG method.

Table 1 gives the quantitative results. According to Table 1, we observe that, due to the fact that the averaging operation is directly performed on the source images, few artifacts are introduced into the fused image obtained by the AVE_PL method, but much more useful information is also lost during the fusion process. The NSCT-based methods, especially the proposed NSCT_RG method, significantly outperform the DWFT-based method in terms of higher values of $Q(\mathrm{Iir},\mathrm{F})$, $Q(\mathrm{Ivi},\mathrm{F})$ and $Q(\mathrm{Iir},\mathrm{Ivi},\mathrm{F})$, together with lower values of $L(\mathrm{Iir},\mathrm{Ivi},\mathrm{F})$ and $N(\mathrm{Iir},\mathrm{Ivi},\mathrm{F})$, which greatly agrees with what have been discussed above. However, the improved performance is with a cost of increased computational complexity of the fusion process.

This work was jointly supported by the National Natural Science Foundation of China (No. 60572152) and the National 863 Program of China (No. 2006AA01Z127). The original IR and visible images, which are available on line at www.imagefusion.org, are kindly supplied by Alexander Toet of the TNO Human Factors Research Institute. Q. Zhang is the author to whom the corre-

spondence should be addressed, his e-mail address is zhangqiang134@163.com.

References

1. Q. Miao and B. Wang, Acta Opt. Sin. (in Chinese) **27,** 1605 (2007).
2. P. Zhao and Z. Pu, Acta Opt. Sin. (in Chinese) **27,** 40 (2007).
3. G. Pajares and J. M. de la Cruz, Pattern Recognition **37,** 1885 (2004).
4. S. Li, J. T. Kwork, and Y. N. Wang, Pattern Recognition Lett. **23,** 985 (2002).
5. Q. Zhang and B. Guo, J. Jilin Univ. (Eng. and Technol. Edn.) (in Chinese) **37,** 458 (2007).
6. A. L. da Cunha, J. Zhou, and M. N. Do, IEEE Trans. Image Processing **15,** 3089 (2006).
7. G. Piella, Information Fusion **4,** 259 (2003).
8. C. Liu, Z. Jing, G. Xiao, and B. Yang, Chin. Opt. Lett. **5,** 274 (2007).
9. Z. Li, Z. Jing, and S. Sun, Chin. Opt. Lett. **2,** 578 (2004).
10. H. Wang, Z. Jing, and J. Li, Chin. J. Lasers (in Chinese) **32,** 351 (2005).
11. N. Cvejic, D. Bull, and N. Canagarajah, IEEE Sensors J. **7,** 743 (2007).
12. Z. Wang, Z. Qin, and Y. Liu, J. Zhejiang Univ. Sci. A **8,** 56 (2007).
13. M. N. Do and M. Vetterli, IEEE Trans. Image Processing **14,** 2091 (2005).
14. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, IEEE Trans. Image Processing **13,** 600 (2004).
15. C. Yang and J. Zhang, J. Xidian Univ. (in Chinese) **33,** 871 (2006).
16. V. Petrovic and C. Xydeas, in *Proceedings of the Tenth IEEE International Conference on Computer Vision* **2,** 1866 (2005).