Journal of Electronic Imaging

JElectronicImaging.org

Multisensor image fusion approach utilizing hybrid pre-enhancement and double nonsubsampled contourlet transform

Qiong Zhang
Xavier Maldague

Qiong Zhang, Xavier Maldague, "Multisensor image fusion approach utilizing hybrid pre-enhancement and double nonsubsampled contourlet transform," *J. Electron. Imaging* **26**(1), 010501 (2017), doi: 10.1117/1.JEI.26.1.010501.

Multisensor image fusion approach utilizing hybrid pre-enhancement and double nonsubsampled contourlet transform

Qiong Zhang* **and Xavier Maldague**
Laval University, Computer Vision and Systems Laboratory, 1065 Avenue de la Médecine, Québec City, G1V 0A6 Québec, Canada

Abstract. A multisensor image fusion approach established on the hybrid-domain image enhancement and double nonsubsampled contourlet transform (NSCT) is proposed. The hybrid-domain pre-enhancement algorithm can promote the contrast of the visible color image. Different fusion rules are, respectively, selected and applied to obtain fusion results. The double NSCT framework is introduced to obtain better fusion performance than the general single NSCT framework. Experimental outcomes in fused images and performance results demonstrate that the presented approach is apparently more advantageous. *© 2017 SPIE and IS&T* [DOI: 10.1117/1.JEI.26.1.010501]

Keywords: image fusion; hybrid image enhancement; nonsubsampled contourlet transform; double nonsubsampled contourlet transform; log-Gabor filter.

Paper 16100L received Feb. 11, 2016; accepted for publication Dec. 16, 2016; published online Jan. 9, 2017.

1 Introduction

Due to the restrictions on the characteristics of an infrared (IR) vision sensor, the acquired IR images suffer from low contrast, edge blurring, less texture detail information, etc., although thermal radiations emitted from objects are recorded in the scene. Similarly, a regular visible image sensor also has deficiencies, because it only reflects objects in a scene superficially without penetrating into the mechanisms of the targets, even though it provides more high-frequency information from the background, which benefits the objects localization and awareness. The IR-visible image fusion technology[1,2] provides a composite image with characteristics stemming from two categories of the images.

The nonsubsampled contourlet transform (NSCT), proposed by da Cunha et al.,[3] has been broadly utilized in image denoising, enhancement, fusion, and other fields. The NSCT is a multiscale, shift-invariant, linear-phase transform method, and its true capacity of two-dimensional transformation can decompose an image into directional subimages to capture the intrinsic geometrical structure, such as edges, textures, and rich details. In addition, it has been applied in image fusion issues and achieved better results than other traditional transform-based methods. Nevertheless, if only different fusion rules are merely developed under the generic framework of NSCT-based IR-visible image fusion issues, the performance of the fusion could not be improved much. Thus, the pre-enhancement of the source image(s) before fusion processing could inspire a better solution.

In this letter, we introduce a hybrid frequency-spatial domain image enhancement algorithm on the visible image to obtain a pre-enhanced RGB color image as being the new input visible source image for the following fusion procedure. The adaptive regional average energy fusion rule is applied to the low-frequency coefficients fusion issue, and an adaptive regional log-Gabor energy fusion rule is presented to be applied to the high-frequency coefficients fusion problem. More importantly, the double NSCT framework[4] is applied to better promote the fusion performance. The proposed fusion scheme, named hybrid-double-NSCT, shows significant advantages than traditional fusion methods in fused image quality and performance assessment.

2 Hybrid-Domain Image Enhancement

The traditional homomorphic image enhancement methods are simply attributed to the consideration of frequency domain. However, it is only able to enhance the details of the image and could not intensify the effect of the outlines well enough.

The mathematical morphological image processing methods consist of a set of operators that can transform images according to the size, shape, connectivity, etc. This methodology could usually achieve the smoothness for the outline of the image, e.g., using the open and close filtering methods to remove partial noise.

On the basis of the advantages of the above approaches in frequency and space domains, a hybrid-domain image enhancement approach can be presented[5] to enhance the IR image or visible image as a first step for the following fusion process.

The original image f can be processed by the gamma correction, log conversion, and Fourier transform at first, which can be given as

$$Z(u,v) = \mathbf{F}[\log f_\gamma(x,y)] = \mathbf{F}\{\log[cf(x,y)^\gamma]\}, \quad (1)$$

where c and γ are the parameters for the correction of the source image $f(x,y)$.

The homomorphic filter is able to eliminate the influence of nonuniform illumination without loss of the image details. The filter transfer function $H(u,v)$ can be applied to be multiplied by the above equation, such as

$$H_Z(u,v) = Z(u,v)H(u,v). \quad (2)$$

The image after homomorphic filtering can be resolved by the inverse Fourier transform and the exponential transform on $H_Z(u,v)$. Here, we present a filter function stemming from Ref. 5 in the form of

*Address all correspondence to: Qiong Zhang, E-mail: qiong.zhang.1@ulaval.ca

$$H(u,v) = (r_H - \alpha_1 r_L)[1 - e^{-c(D^2/D_0^2)}] + \frac{r_H + \alpha_2 r_L}{1 + \alpha_2}, \quad (3)$$

where α_1 and α_2 are parameters to adjust the high- and low-frequency regulation parameters r_{H} and r_{L}. Some other coefficients are defined as

$$D^2 = u^2 + v^2, \quad D_0 = \frac{1}{\beta}[(p_{\max} - \bar{p})^2 + (p_{\min} - \bar{p})^2], \quad (4)$$

where D_0 is the harmonic coefficient, β is a parameter for adjustment, $p_{\max}$, $p_{\min}$, and $\bar{p}$ denote the maximum, minimum, and average pixel gray values of the image, respectively.

The frequency-domain enhanced image $\hat{f}$ can be obtained by the inverse fast Fourier transform and inverse log transformations.

As a technique of nonlinear filtering in image processing, mathematical morphology can use a set of operators to transform the image according to its characteristics. Here, we use a combination of top-hat and bottom-hat transforms to enhance the image outline in the spatial domain and the final enhanced image can be obtained by

$$f_{\mathrm{enh}} = \hat{f} + k_{\mathrm{th}}\mathrm{Top-hat}(\hat{f}) - k_{\mathrm{bh}}\mathrm{Bot-hat}(\hat{f}), \quad (5)$$

where k_{th} and k_{bh} are parameters for adjustment of the weights of transforms. The top-hat method can smooth the target boundary and enhance the shadow details; the bottom-hat method can fill the empties and connect similar objects.

3 Nonsubsampled Contourlet Transform-Based Adaptive Fusion Rules

The NSCT,[3] as a fully shift-invariant, multiscale and multidirectional transform-domain technique, has been proven to be superior in the issue of image fusion. Here, in this section, the fusion rules are specifically considered for the NSCT-decomposed low- and high-frequency coefficients, respectively.

3.1 *Low-Frequency Coefficients Fusion*

In the low-frequency subband coefficients fusion, the adaptive regional average energy fusion rule is employed, because the low-frequency coefficients contain the most detailed information, which approximates the patterns of the original image, and the regional energy method can maintain as much as the low-frequency information.

The regional energy of the low-frequency coefficient images A and B can be written as

$$E_{j0}^A(x,y) = \sum_{m=-M}^{M}\sum_{n=-N}^{N} w_0(m,n)[C_{j0}^A(x+m,y+n)]^2,$$
$$E_{j0}^B(x,y) = \sum_{m=-M}^{M}\sum_{n=-N}^{N} w_0(m,n)[C_{j0}^B(x+m,y+n)]^2, \quad (6)$$

in which, $w_0(m,n)$ is chosen as a 3×3 window mask, $C_{j0}^A(x,y)$ and $C_{j0}^B(x,y)$ denote the low-frequency coefficients decomposed by the two source images, respectively.

The weight factors can be defined to be

$$W_{j_0}^A(x,y) = \frac{E_{j_0}^A(x,y)}{E_{j_0}^A(x,y) + k_0 E_{j_0}^B(x,y)},$$
$$W_{j_0}^B(x,y) = 1 - W_{j_0}^A(x,y), \quad (7)$$

where k_0 is a factor for the adjustment of the regional energy weight.

Thus, the low-frequency coefficient after fusion will be

$$C_{j_0}^F(x,y) = W_{j_0}^A(x,y)C_{j_0}^A(x,y) + W_{j_0}^B(x,y)C_{j_0}^B(x,y). \quad (8)$$

3.2 *High-Frequency Coefficients Fusion*

As is known, the high-frequency subband coefficients represent the edges and curves of the source images. In general, it is considered that the coefficients with larger absolute values usually have better local or edge features. However, the noisy information might also be regarded as the high-frequency coefficients, which probably cause calculation inaccuracy for sharpness values, and thus influence the fusion results and performance.

The traditional maximum absolute selection fusion rule does not take into consideration the impact from surrounding pixels. In addition, human eyes are more sensitive to texture or local features with detailed information but not sensitive to a single pixel value. Log-Gabor filter,[6,7] with optimal spatial orientation and wide frequency information, is more suitable to represent local information from the image, which is more accord with the human visual system.

The log-Gabor filter under the polar coordinates can be defined as

$$G(f,\theta) = \exp\left\{-\frac{[\log(f/f_0)]^2}{2[\log(\sigma/f_0)]^2}\right\} \cdot \exp\left\{-\frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}\right\}, \quad (9)$$

where f_0 is the center frequency and θ_0 is the filtering direction. In addition, the bandwidth of the radial filter, B_f, is determined by the parameter σ, and the bandwidth of the filtering orientation, B_θ, is determined by the parameter σ_θ.

The scales of the log-Gabor wavelets can correspond to the scales of the NSCT-decomposed high-frequency coefficients, in the case of the high-frequency subbands having the same numbers of directions on each NSCT decomposition scale.

For any single source image, the signal response can be described by a convolution form as follows:

$$HG_{k,l}(x,y) = H_{k,l}(x,y) * G_{k,l}(x,y), \quad (10)$$

where $H_{k,l}(x,y)$ is the coefficient value located at (x,y) in high-frequency subimages at the k'th scale and l'th direction, and $G_{k,l}(x,y)$ is the corresponding log-Gabor filter value.

The corresponding log-Gabor energy can be defined as

$$EH_{k,l}(x,y) = \sqrt{\Re[HG_{k,l}(x,y)]^2 + \Im[HG_{k,l}(x,y)]^2}, \quad (11)$$

in which, $\Re[\cdot]$ and $\Im[\cdot]$ denote the real part and the imaginary part, respectively. Similarly, the adaptive regional average log-Gabor energy can be defined as

$$E^A_{k,l}(x,y) = \sum_{m=-M}^{M}\sum_{n=-N}^{N} w_0(m,n)[EH^A_{k,l}(x+m,y+n)]^2,$$
$$E^B_{k,l}(x,y) = \sum_{m=-M}^{M}\sum_{n=-N}^{N} w_0(m,n)[EH^B_{k,l}(x+m,y+n)]^2, \quad (12)$$

in which, w_0 can be chosen as the same window size as described in the above subsection.

Therefore, the high-frequency coefficients fusion rule can be a selection rule through

$$C^F_{k,l}(x,y) = \begin{cases} H^A_{k,l}(x,y), E^A_{k,l}(x,y) \geq E^B_{k,l}(x,y) \\ H^B_{k,l}(x,y), E^A_{k,l}(x,y) \end{cases}, \quad (13)$$

where $C^F_{k,l}(x,y)$ represents the corresponding fused high-frequency coefficient.

4 Double-Nonsubsampled Contourlet Transform Image Fusion Scheme

Based on the hybrid-domain enhancement and the adaptive fusion rules, we propose a hybrid-double-NSCT fusion scheme to achieve extraordinary calculation results from experimental implementations. The procedure can be described as in Fig. 1, where the input source data for the second-fusion are from the first-fusion result image and the pre-enhanced visible image.

5 Experimental Results and Analysis

Experiments have been carried out using the IR-visible datasets with the image dimensions of 768×1024. These datasets[8] were acquired by the Stereovision system in the field experiments conducted in Corte, France in November 2013. The vision system was comprised of two industrial multispectral visible/near-IR cameras, and it captured visible and IR images at the same time.

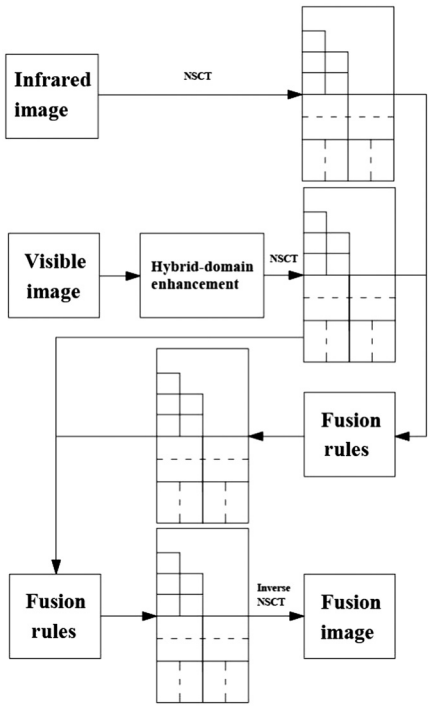

Fig. 1 The proposed hybrid-double-NSCT fusion framework.

In each dataset, there are two raw images on a same scene, including one IR image and one RGB visible image. Each time for the process of image fusion, the registered IR image in grayscale and the registered visible image in colorscale on the same scene are input into the fusion approach as source images at the same time.

The dataset used in this letter has very low noise, which can be considered as clean data source. To cater for the fusion rules and balance the computation speed and performance results, all of the NSCT decomposition and reconstruction were conducted under $2^2 - 2^2$ scales and directions.

Figure 2 displays the source and result images from experiments on a selected dataset, in which the original IR image, original visible RGB image, fused grayscale image, and fused RGB colorscale image are displayed, respectively. Figure 3 displays the low-frequency coefficients fused image, directional high-frequency coefficients fused images in two decomposition levels, in comparison with the final fused grayscale image.

It is obvious that the final fused grayscale and colorscale images exhibit all contours, details, and characteristics inherited from the source IR and visible images. In the fused images, the separation effect between the foreground human-checkerboard target and the background scene has been evidently intensified, the image contrast, luminance, and colors are better redistributed, and the image edges and details are significantly enhanced. The image quality of the objects, such as sky, clouds, mountains, trees, shadows, etc., is apparently augmented.

The performance of other different fusion methods,[8] such as wavelet transform (WT), contourlet transform (CT), and general NSCT-based schemes, is compared with the proposed approach in the aspects of standard deviation (SD),[9] Shannon entropy (H),[10] root-mean-square error (RMSE),[11] mutual information (MI),[12] and edge-based similarity metrics ($Q^{AB/F}$).[12]

As a rule, the NSCT-based fusion method has better performance in all of these measures compared with the

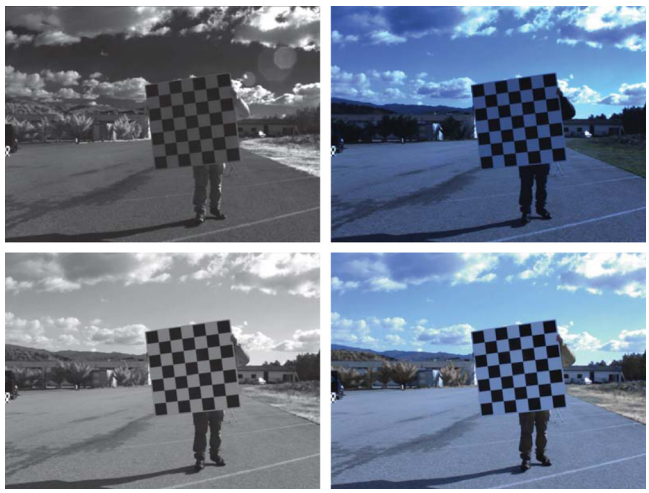

Fig. 2 Hybrid-double-NSCT fusion images: (a) IR image, (b) visible image, (c) fused image in grayscale, and (d) fused image in RGB colorscale.

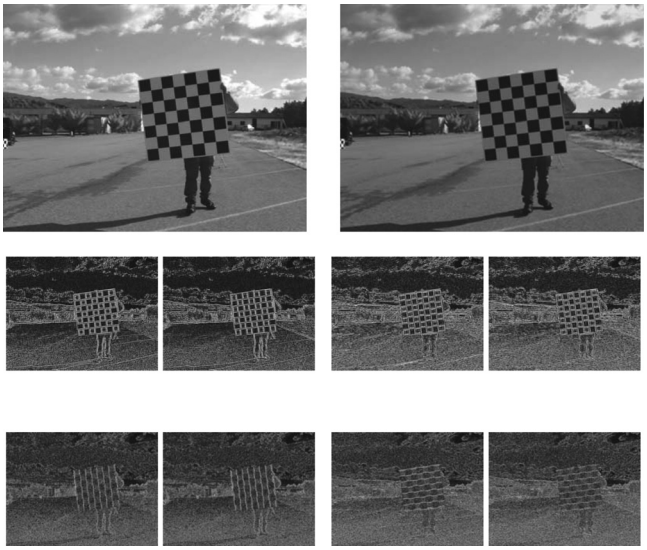

Fig. 3 Hybrid-double-NSCT fused coefficients images: (a) fused image in grayscale, (b) low-frequency coefficients fused image, (c1)–(c4) first-level high-frequency coefficients fused images, and (d1)–(d4) second-level high-frequency coefficients fused images.

Table 1 Performance of different fusion methods.

	SD	H	RMSE	MI	$Q^{AB/F}$
IR image	41.1175	6.9767	N/A	N/A	N/A
Visible image	53.2545	7.3542	N/A	N/A	N/A
WT	42.2954	7.0014	23.4291	5.1348	0.6964
CT	45.7620	7.2749	33.9656	5.4815	0.5706
NSCT	45.8144	7.2609	34.1048	6.1323	0.6978
Hybrid-double-NSCT	52.0676	7.4654	41.2115	9.1325	0.7532

WT- and CT-based fusion methods, and the CT-based method has better measures of performance than the WT-based method. In general, the performance of SD and H for the source visible image behaves better than for the source IR image, because the IR image has lower global grayscale intensity and contrast, whereas the visible image has better high-frequency information with better contrast quality. Thus, under the single decomposition-reconstruction framework, SD and H performance for the final fused image should, respectively, fall between that for the IR image and the visible image.

Table 1 exhibits the values of quantitative performance of the fusion methods for comparison. Compared with the WT-based method, the CT-based method evidently increases the performance in almost all indices except the edge similarity index $Q^{AB/F}$, which is caused by its filtering mechanism and directional selectivity. The NSCT-based approach behaves better performance than CT-based approach except the H, because NSCT has better capability in denoising and smoothing than CT.

However, it is noted that all of the performance metrics results of our proposed fusion approach evidently demonstrate its superiority compared with all the other approaches. In particular, the presented method obtains much higher SD and MI, due to the impact of double-NSCT structure and log-Gabor filtering, which ensures the contrast enhancement quality of the fused image obviously better than that of any other methods. Experiments on other datasets were also conducted, all of which prove that our proposed fusion algorithm achieves better fusion performance in image quality and computation results.

6 Conclusion

This letter presents a joint multisensor image fusion approach on the basis of hybrid-domain image pre-enhancement and double NSCT techniques. Compared with the conventional IR-visible image fusion schemes, such as WT-, CT-, and NSCT-based algorithms, the proposed scheme obtains evidently better fused image quality both in grayscale and colorscale and achieves comparably better performance results in mainstream measures, such as SD, H, RMSE, MI, and $Q^{AB/F}$. Furthermore, the exact qualification of the proposed approach for improving the image fusion quality and performance could be studied as a fresh research topic in the near future.

Acknowledgments

The authors would like to acknowledge partial funding support from the National Sciences and Engineering Research Council of Canada, Canada, and dataset support from Dr. Tom Toulouse at the University of Corsica, France.

References

1. G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Inf. Fusion* **4**(4), 259–280 (2003).
2. G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.* **37**(9), 1855–1872 (2004).
3. A. da Cunha, J. Zhou, and M. Do, "The nonsubsampled contourlet transform: theory, design, and applications," *IEEE Trans. Image Process.* **15**, 3089–3101 (2006).
4. Z. Chen, C. Zhang, and P. Wang, "High-quality fusion for visible and infrared images based on the double NSCT," in *7th Int. Congress on Image and Signal Processing (CISP'14)*, pp. 223–227 (2014).
5. Q. Zhang, J. Fleuret, and X. Maldague, "A hybrid frequency-spatial domain infrared image enhancement approach evaluated by fuzzy entropy," *Proc. SPIE* **9105**, 91050L (2014).
6. P. Yao et al., "Iris recognition algorithm using modified log-Gabor filters," in *18th Int. Conf. on Pattern Recognition (ICPR'06)*, Vol. **4**, pp. 461–464 (2006).
7. R. Raghavendra and C. Busch, "Novel image fusion scheme based on dependency measure for robust multispectral palmprint recognition," *Pattern Recognit.* **47**, 2205–2221 (2014).
8. Q. Zhang and X. Maldague, "An adaptive fusion approach for infrared and visible images based on NSCT and compressed sensing," *Infrared Phys. Technol.* **74**, 11–20 (2016).
9. S. Ghahramani, *Fundamentals of Probability*, 2nd ed., p. 438, Prentice Hall, New Jersey (2000).
10. C. E. Shannon, "A mathematical theory of communication," *The Bell Syst. Tech. J.* **27**(3), 379–423 (1948).
11. P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," *Aquatic Procedia* **4**, 133–142 (2015).
12. G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.* **38**, 313–315 (2002).