A NEW FUSION METHOD FOR REMOTE SENSING IMAGES BASED ON SALIENT REGION EXTRACTION

Libao Zhang and Jue Zhang*

College of Information Science and Technology, Beijing Normal University, Beijing 100875, China

ABSTRACT

The goal of the remote sensing image fusion is to inject the detail information extracted from panchromatic (PAN) images to multispectral (MS) images with minimized spectral distortion. However, different regions in the image may practically have different demands on the spatial and spectral resolution. In this paper, a new fusion method for remote sensing images based on salient region extraction is proposed. By introducing the hybrid visual saliency analysis, information in the PAN and MS image are automatically partitioned into two categories: salient and non-salient regions. Then, a sub-region fusion strategy is applied to fuse the non-salient and salient regions respectively. For non-salient regions, such as farmland and mountains, the wavelet transform is used in the process of spatial infusion to suppress spectral distortion. As for salient regions like residential areas, the windowed IHS transform is carried out for its merits of effective integration of spatial and spectrum information. Experimental results demonstrate that our proposal achieves a better balance between spatial injection and spectral maintenance in different regions.

Index Terms—Image fusion, remote sensing, saliency analysis, IHS transform, wavelet transform

1. INTRODUCTION

Remote sensing image fusion is an important branch of image fusion, aiming to integrate advantages of images from multiple complementary sources and provide new images with better depictions [1]. Over recent decades, various strategies have been proposed to address the issue. Brovey method [2] is a linear weighted method that mainly focuses on the enhancement of spatial details. It achieves a low computation complexity but suffers from severe spectral distortion. The principal component analysis (PCA) decomposition [3] is a typical component substitution model, eliminating the redundancy effectively. Intensity–hue–saturation (IHS) transform [4, 5] separates the spectral and intensity feature to conduct the intensity modulation. It performs well in the injection of spatial details and subsides spectral distortion to a certain extent. However, those methods aforementioned pervasively provide unsatisfied spectral maintenance. To cope with this limitation, multiresolution analysis is introduced as a new approach. Discrete wavelet transform (DWT) [6, 7] is one of classical ways to accomplish the multi-sources image fusion. Levis et al. [8] proposed a novel region-based fusion scheme using the complex wavelet (RB-CWT). Aiazzi et al. [9] presented a context-driven principle with oversampled multiresolution analysis. Recently, Sujoy et al. [10] proposed an algorithm in the gradient domain using the maximum gradient magnitude (MMGD). Li et al. [11] made use of the guided filtering based method to obtain fused images.

In some practical applications, there is a need for observers or follow-up treatments that different types of regions in remote sensing images differ in requirements for spectral features and spatial details. For instance, residential areas, buildings and airports need more spatial details for better descriptions of various ground targets and they are also more significant in images. Regions with less saliency, such as farmland and mountains are generally discriminated by spectral characteristic, thereby requiring undistorted spectral features. However, most existing methods execute a unified fusion processing of the whole image without distinguishing the demand of different regions. To solve the problem, a new fusion method for remote sensing image based on salient region extraction is introduced in this paper. The new method is proposed to:

1) Adaptively determine regions in different needs of spatial and spectral information by saliency analysis.
2) Achieve better sub-region depictions in fused images.

Experimental results reveal that our proposal has great potential for handling the complex needs of different regions in the fused image.

2. METHODOLOGY

In our method, PAN and MS images are automatically classified into salient and non-salient regions by hybrid visual saliency analysis. Then, with the sub-region fusion strategy, we fuse salient regions via the windowed IHS transform and merge non-salient regions by the discrete wavelet transform (DWT). Figure 1 shows the framework of the proposed method.

*Corresponding author: libaozhang@163.com. This research was supported by National Natural Science Foundation of China (61571050 and 61071103) and Beijing Natural Science Foundation (4162033).

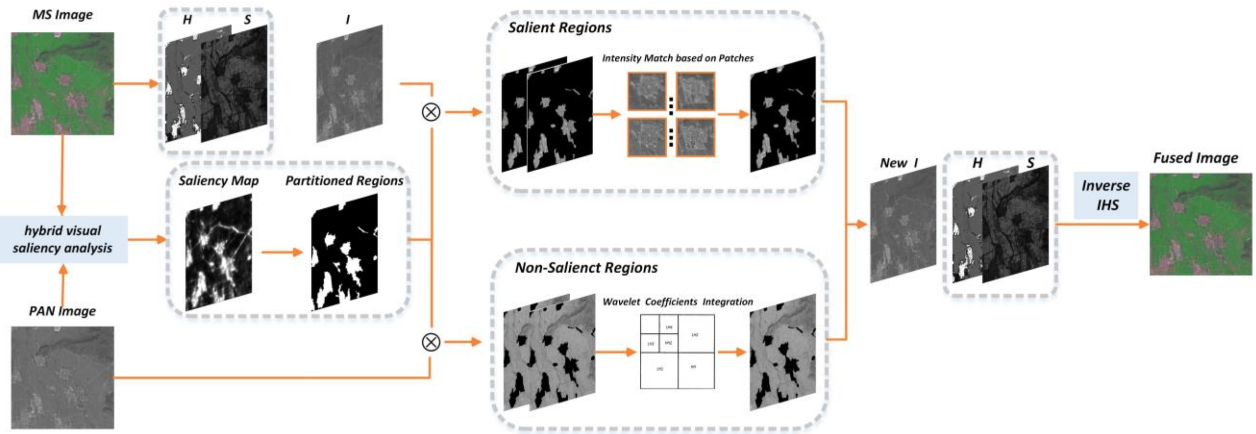


Fig.1. *The framework of the proposed model.*

2.1. Salient region extraction

In this section, the hybrid visual saliency analysis model (HVSA) is proposed to extract salient regions from remote sensing images. In the proposed model, features in both PAN image and MS images are jointly considered to fetch the saliency. Figure 2 shows the flowchart of salient region extraction.

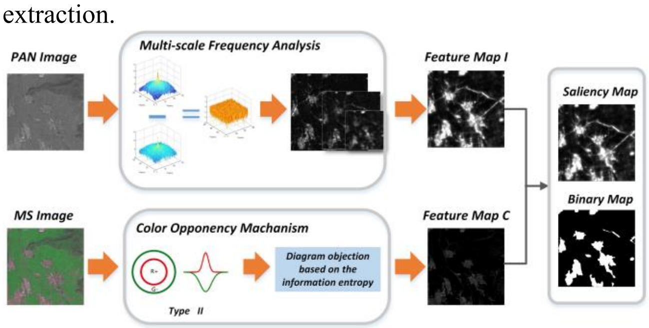


Fig.2. *The flowchart of salient region extraction.*

2.1.1. Intensity feature extraction

Intensity feature is a kind of significant saliency information in remote sensing images. In this paper, we propose an efficient intensity feature extraction algorithm based on multi-scale frequency analysis strategy. It includes three steps as follows:

Step 1: Multi-scale decomposing based on the adaptive direction prediction-based lifting wavelet transform (ADP-LWT) [12, 13]. The ADP-LWT can analyze the local spatial correlations in all directions and then chooses a direction of prediction and update in which the high-frequency energy is minimal (the local edge or texture tangent direction). So it can concentrate energy to low-frequency subbands more efficiently than traditional DWT.

Step 2: In each level low-frequency subband, the spectral residual (SR) [14, 15] model is used to generate an intensity feature map belonging to different scales. We can compute the amplitude $A_\delta(f)$ and phase spectrums $\varphi_\delta(f)$ of down-sampled images at different scales through the Fourier Transform $F[\cdot]$ as (1).

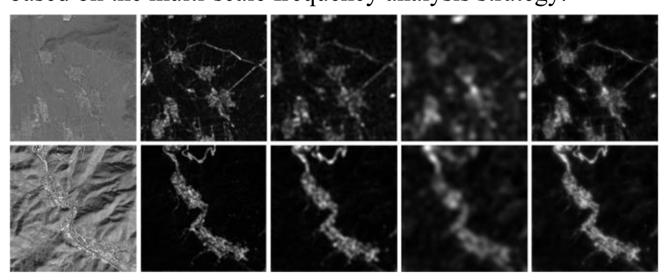
$$F_\delta(f) = F[I_\delta(x,y)] = |A_\delta(f)| \exp(j \cdot \phi_\delta(f)) \qquad (1)$$

The spectral residuals are defined as follows:

$$SR_\delta = log(A_\delta(f)) - log(A_\delta(f)) * h \qquad (2)$$

where h represents an average filter with the size of 3×3.

Step 3: All intensity feature maps belonging to different scales are fused to generate the final intensity feature map $\tilde{I}_\delta$. Figure 3 shows the intensity feature maps of different scales based on the multi-scale frequency analysis strategy.

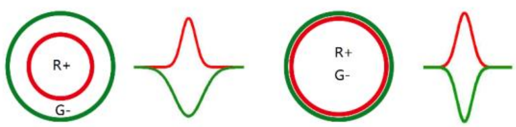


(a) (b) (c) (d) (e)

Fig.3. *Intensity feature maps based on the multi-scale frequency analysis strategy. (a) PAN images (b)* 512×512 *(c)* 256×256 *(d)* 128×128 *(e) Final intensity feature map.*

2.1.2. Color feature extraction

The color feature has been revealed to be an important cue for distinguishing diverse scenes in human visual attention mechanism [16]. According to biological experiments, the color feature is processed via the color opponent mechanism in human vision systems (HSV). The mechanism works as follows: In the center of the cells' receptive fields (RF), neurons are excited by one color and suppressed by another while the situation would be reverse in the surround [17]. Such opponency pervasively exists among R-G-B color contrast channels in HSV. The RF of cells in layers of retina can be divided into two groups: type I and type II. The type I RF has balanced opponent coefficients while type II has unbalanced ones [18]. Furthermore, researchers have attested that type II RF responds better to color-defined boundaries. Figure 4 shows the schematic diagram of the RF of type I and type II. Based on the RGB color space, traditional color opponency algorithms address color cues through the simulation of a three-layered neural network. For remote sensing images, color information theoretically corresponds to spectral bands in MS images, which are not always in one-to-one conformity with RGB channels. For example, the SPOT satellite system provides red, blue, far-infrared and near-infrared bands without blue band [19]. As a result, traditional color opponency algorithms based on RGB color space generate feature maps with inaccurate detection when applied to remote sensing images. Hence, we propose a new color opponent method based on the integration of the R-G opponent channel and the diagram objection via the information entropy.



Type I *Type II*

Fig.4. *Different types of receptive fields.*

First, we use the Gaussian filters with standard deviate σ for R , G channel. The outputs are denoted by $\tilde{R}$ and $\tilde{G}$ respectively. Subsequently, the R-G contrast channel is generated by imitating layers of the retina with the receptive field of type I.

$$\tilde{O}(x,y) = \tilde{R} + w \times \tilde{G} \quad w \in (-1,0) \tag{3}$$

where w denotes the unbalanced opponent coefficient.

$$H(k) = -\frac{l_k}{M}\sum_{i=1}^{k}\log\frac{l_k}{M} \tag{4}$$

Here, $H(k)$ denotes the information entropy and l_k is the number of pixels in $\tilde{O}(x,y)$ with the value of k . M is the total number of pixels in $\tilde{O}(x,y)$.

Finally, we can obtain the color feature map $\tilde{C}$ through the histogram objection.

$$\tilde{C}(x,y) = H(k) \quad as \quad \tilde{O}(x,y) = k \tag{5}$$

2.1.3. The generation of the saliency map

We introduce a weighted combination of feature maps to generate the final map via feature competition. The combination coefficients are adaptively decided by the significant contrast between salient regions and the background in feature maps. The detailed computational formulas are as follows:

$$S = w_i \times \Pi\left(I_\delta, n\right) + w_c \times \tilde{C} \tag{6}$$

$$w_i = \left[Max\left(\Pi\left(I_\delta, n\right)\right) - Mean\left(\Pi\left(I_\delta, n\right)\right)\right]^2 \tag{7}$$

$$w_c = \left[Max\left(\tilde{C}\right) - Mean\left(\tilde{C}\right)\right]^2 \tag{8}$$

where $\Pi(\cdot)$ denotes the across-scale fusion method. S represents the saliency map.

2.2. Sub-region fusion strategy

2.2.1. Fusion Approach for salient regions

Residential areas, buildings etc. with well-defined boundaries and abundant details are quite notable in remote sensing images, which can be obtained using the salient extraction method. The fusion framework based on the IHS transform is proved to be simple and of high spatial quality [20]. However, due to the steep difference between the intensity component in the MS image and estimated one, the IHS-based methods cannot keep the chromatic perception unchanged. In order to minimize spectral distortion of salient regions, we introduce the windowed IHS transform to reduce the difference during the estimation. Instead of the simple linear weighting based strategy, the windowed IHS algorithm matches the intensity component of the MS image with the PAN image to produce accurate approximations. We calculate the local mean value $\bar{P}_W$ in the PAN image in a sliding patch with the size of $n \times n$ to amend the intensity patch Q_W in the MS image after IHS transform. The new intensity patch is computed as follows:

$$T = Q_W \times \frac{\bar{P}_W}{\bar{Q}_W} \tag{9}$$

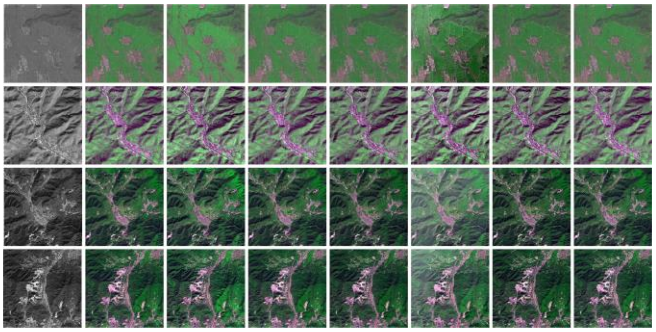
where P_W is the intensity patch of salient regions in PAN images and $\bar{Q}_W$ is the mean value of Q_W .

2.2.2 Fusion Approach for non-salient regions

In non-salient regions, the *I* component of the MS image and the intensity of the PAN image are both decomposed by the wavelet transform. In wavelet-based methods, high-frequency coefficients play an important role in preserving spatial details of images and low-frequency ones are of vital importance to retain spectral information. In addition, the performance of the wavelet-based methods is quite sensitive to the order of the decomposition. The lower the order is, the better the spectral information retained. The higher the order is, the more the spatial details injected. In our paper, we carry out a two-layer wavelet decomposition. Then high-pass coefficients of the *I* component are replaced by high-pass coefficients of the PAN image. After the reconstruction, we can get the new intensity component in non-salient regions.

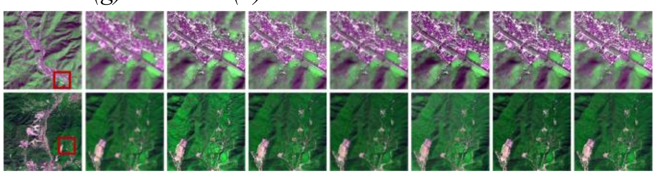
3. EXPERIMENTAL RESULTS AND DISCUSSION

To verify the performance of the proposed method, registered multispectral and panchromatic SPOT5 data are used for experiments. Due to the inexistence of absolute references, we spatially degrade down the original images to obtain the MS images with lower resolution and take original ones as genuine references. We compared our method with three traditional algorithms, IHS [5], PCA [3], WT [6] and two recent ones, RB-CWT [8] and MMGD [10].



(a) (b) (c) (d) (e) (f) (g) (h)

Fig.5. *Comparison of fusion results of six methods. (a) PAN images (b) Degraded MS images; (c) IHS (d) WT (e) PCA (f) MMGD (g) RB-CWT (h) Our method.*



(a) (b) (c) (d) (e) (f) (g) (h)

Fig.6. *Subscenes of fusion results of six methods. (a) Two MS images in Fig.4 (b) Subscenes in MS images; (c) IHS (d) WT (e) PCA (f) MMGD (g) RB-CWT (h) Our method.*

3.1. Visual Analysis

In this section, we select typical results to show the performances of the fusion methods aforementioned. As the results displayed above (see Fig.5 and 6), we can observe that IHS-based and MMGD method produce severe spectral distortion as the color of farmland is obviously altered in the fused image. In addition, the MMGD method sharpens the image too much and produces many spatial artifacts. Apparently, the methods based on the multi-resolution analysis (WT, RB-DWT) provide better performance in maintenance of spectral characteristics. Our proposal keeps the spectral information very well when the details of urban areas are significantly improved.

3.2. Quantitative Assessment

For quantitative assessment, we utilized four objective metrics including average gradient (AG), difference coefficient (DC), relative average spectral error (RASE) and relative dimensionless global error in synthesis (ERGAS). AG indicates the spatial resolution and the visual quality. High values of AG reveal the validity of detail infusion. DC is obtained through comparisons with the MS image. RASE and ERGAS signify the global spectral similarity between the fused image and the reference image [21, 22]. Smaller values of DC, RASE and ERGAS indicate a better performance of spectral maintenance. We select the PAN and MS image of Fig.5 as an example to show the quantitative performance of the methods mentioned above. The results are shown in Table 1.

Table 1. *Quantitative evaluation of the whole image*

Methods	AG	DC	RASE	ERGAS
IHS	5.225	0.124	11.132	4.781
PCA	2.644	0.069	2.480	1.211
MMGD	7.998	0.242	6.984	3.153
RB-CWT	4.491	0.086	1.737	0.902
WT	4.686	0.047	4.054	1.886
Our	**4.847**	**0.041**	**4.012**	**1.876**

Additionally, in order to evaluate the effectiveness of the sub-region strategy, we compared fusion results in non-salient and salient regions respectively (see Table 2 and 3). For non-salient regions, our proposal can not only achieve the lowest spectral distrotion but also preserve enough image details and edge information. For salient regions, our proposal represents a remarkable improvement in spatial resolution.

Table 2. *Quantitative evaluation of non-salient regions*

Methods	AG	DC	RASE	ERGAS
IHS	3.921	0.124	14.797	6.139
WT	3.457	0.037	5.534	2.596
PCA	2.041	0.069	9.878	4.233
MMGD	6.105	0.241	25.613	11.549
RB-CWT	3.265	0.082	9.8880	4.190
Our	**3.441**	**0.029**	**5.932**	**2.798**

Table 3. *Quantitative evaluation of salient regions*

Methods	AG	DC	RASE	ERGAS
IHS	13.375	0.125	7.654	3.457
WT	12.368	0.106	8.877	4.307
PCA	6.408	0.068	13.860	6.525
MMGD	19.818	0.247	23.825	10.733
RB-CWT	12.157	0.114	10.057	4.805
Our	**13.639**	**0.114**	**7.947**	**3.950**

4. CONCLUSION

In this paper, a new fusion method based on salient region extraction is proposed for remote sensing images. The regions in the images are automatically partitioned into salient and non-salient regions by saliency analysis and the two regions are fused separately through a sub-region fusion strategy. It reveals that our proposal significantly subsides the spectral distortion while ensuring an appropriate spatial injection. Compared with the well-known algorithms, our proposal also has an outstanding performance in managing the complex needs of different regions in fused images.

5. REFERENCES

[1] C. Thomas, T. Ranchin, and L. Wald, et al., "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1301–1312, 2008.

[2] C. Pohl, J.L. Van Genderen, "Review article multi-sensor image fusion in remote sensing: Concepts, methods and applications," *International Journal of Remote Sensing*, vol. 19, no. 5, pp. 823-854. 1998.

[3] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogrammetric Engineering & Remote Sensing*, vol. 58, no. 5, pp. 561–567, 1992.

[4] R. Haydn, G.W. Dalke and J. Henkel, et al., "Application of IHS color transform to the processing of multi-sensor data and image enhancement," *Proceedings of the International Symposium on Remote Sensing of Environment*, pp. 599–616, 1982.

[5] M. Choi, "A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1672–1682, 2006.

[6] J. Zhou, D.L. Civco and J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *International Journal of Remote Sensing*, vol. 19, no. 4, pp. 743-757, 1998.

[7] X. Otazu, M. Gonzalez-Ausicana, "Introduction of sensor spectral response into image fusion methods: Application to wavelet-based methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 10, pp. 2376–2385, 2005.

[8] J. Lewis, J. O'Callaghan and G. Nikolov, et al., "Pixel- and region-based image fusion with complex wavelets", *Information Fusion*, vol. 8, pp. 119-130, 2007.

[9] B. Aiazzi, L. Alparone and S. Baronti, et al., "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2300–2312, 2002.

[10] S. Paul, S. Sevcenco and P. Agathoklis, "Multi-Exposure and Multi-Focus Image Fusion in Gradient Domain", *Journal of Circuits Systems and Computers*, vol. 25, no. 10, pp. 1650123 (1-18), 2016.

[11] S. Li, X. Kang and J. Hu, "Image Fusion With Guided Filtering," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2864-2875, 2013.

[12] L. Zhang, A. Li, Z. Zhang, and K. Yang, "Global and local saliency analysis for the extraction of residential areas in high-spatial-resolution remote sensing image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 3750-3763, Jul. 2016.

[13] L. Zhang, J. Chen, and B. Qiu, " Region-of-interest coding based on saliency detection and directional wavelet for remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 1, pp. 23-27, 2017.

[14] X. Hou and L. Zhang, "Saliency Detection: A Spectral Residual Approach," *Computer Vision and Pattern Recognition (CVPR)*, pp. 1-8, 2007 IEEE Conference, 2007.

[15] L. Zhang, K. Yang, and H. Li, "Regions of interest detection in panchromatic remote sensing images based on multiscale feature fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 12, pp. 4704-4716, 2014.

[16] L. Zhang, J. Zhang, S. Wang and J. Chen, "Residential area extraction based on saliency analysis for high spatial resolution remote sensing images," *Journal of Visual Communication and Image Representation*, vol.33, pp. 273-285, 2015.

[17] A. Jain and G. Healey, "A multiscale representation including opponent color features for texture recognition," *IEEE Transactions on Image Processing*, vol.7, no.1, pp. 124-128, 1998.

[18] S.G. Solomon, P. Lennie, "The machinery of colour vision," *Nature Reviews Neuroscience*, vol.8, no.4, pp. 276-286, 2007.

[19] J.P. Gleyzes, A. Meygret and C. Fratter, et al., "SPOT5: system overview and image ground segment," *Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 300-302, 2003 IEEE International, 2003.

[20] T.M. Tu, P.S. Huang, C.L. Hung, et al., "A fast intensity hue-saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 4, pp. 309–312, 2004.

[21] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation," *Photogrammetric Engineering & Remote Sensing*, vol. 66, no. 1, pp. 49–61, 2000.

[22] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Transactions Geoscience and Remote Sensing*, vol. 46, no. 1, pp. 228–236, 2008.