A New Fusion Method for Remote Sensing Images

Xugang Wu, Xu Li(✉), and Zhenzhen Fan

School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, People's Republic of China
lixu@nwpu.edu.cn

Abstract. As a branch of Internet of Things (IoT), network geographic information (WebGIS) has the task of enhancing remote sensing images, including correction, fusion, mapping, etc. Remote sensing images captured by different satellite platforms often present earth objects with various scales. Due to the different spatial resolutions of imaging sensors, the same object will appear in different sizes. Usually panchromatic (PAN) image, without spectral diversity, has high spatial resolution and multispectral (MS) image has lower spatial resolution but more spectral information. Fusing MS and PAN images, also known as pansharpening, will increase the spatial resolution of MS image without changing its spectral information. In the fusion process, inaccurate perception and extraction of the spatial details will lead to spectral and spatial distortions in the fused image. A new multi-scale structure perception based fusion method is proposed in this paper. Firstly, an improved structure-preserving filter is designed by using MS imagery as the reference to capture the structures in the PAN image. Then the structures having different scales are progressively perceived through a multi-scale decomposition scheme and the spatial details are accurately extracted from the PAN image. Finally, the details weighted by the band-dependent gains are injected into each band of MS images. Experimental results are compared with some recently proposed edge-preserving filtering based fusion methods. Visual analysis and objective evaluations demonstrate that the proposed method can produce high-quality fused output and achieve better performance than some state-of-the-arts.

Keywords: Fusion · Remote sensing · Filter · Multispectral image

1 Introduction

With the fast development of IoT, WebGIS integrates more remote sensing applications and migrates tasks from traditional GIS. Increasing the spatial resolution by fusion technique is one of the tasks of remote sensing information visualization enhancement. The high-resolution remote sensing imagery collected by spaceborne imaging sensors, such as QuickBird, IKONOS, WorldView2/3/4 are often made up of a high-spatial resolution PAN image and serval low-spatial resolution MS images [1]. The fusion of PAN and MS images aims at increasing the spatial resolution of MS images with the aid of the spatial details from the PAN image. Thanks to the development of computer vision technology, more and more image filters are employed in the fusion design and

B. Li et al. (Eds.): IoTaaS 2020, LNICST 346, pp. 793–804, 2021.
https://doi.org/10.1007/978-3-030-67514-1_63

the filtering-based fusion methods have evolved into an important branch in the pan-sharpening field. Kaplan [2] et al. estimated the missing fine spatial details extracted by the bilateral filtering decomposition of the PAN image and proposed a bilateral filter luminance proportional method. Joshi [3] et al. took advantage of the edge preserving and detail transferring capabilities of the joint bilateral filter to design a multistage pansharpening algorithm. Yin and Li [4] presented a two-step fusion approach based on a multiscale normalized nonlocal mean filter aiming to address the spatial distortion issues in multi-resolution analysis-based methods. Upla et al. [5] took the advantage of the guided filter and extracted the details from PAN and MS images using a multistage guided filter. However, it may result in details over-injection due to the improper parameter selection. Recently, Dong et al. [6] introduced the guided filter into the hyperspectral pansharpening to reduce both spectral and spatial distortion. In 2014, Zhang et al. [7] proposed an effective scale-aware filter called rolling guidance filter (RGF) which can extract small-scale structures while preserving other content, parallel in terms of importance to previous edge-preserving filters. Then, Lillo-saavedra et al. [8] designed a new RGF-based fusion method for agricultural fragmented landscape. However, it depends more on the input images having objects with large-scale structures or edges. Most of the existing filtering-based methods mainly focus on capturing the edge features by using different edge-aware filters and rarely describe spatial details from the geometric structure perspective.

Generally, in an earth observation image, complex earth objects are widely presented in different spatial/spectral resolutions. Therefore, in the pansharpening, it is important to recognize the objects in different size or the structures of various scales during the detail extraction. The structure-preserving filter (SPF) in [9] have shown great potential in image processing applications, such as image abstraction, image composition, and seam carving, etc. This paper firstly modifies the SPF in [9] to describe the structures more accurately by considering the spatial correlation and similarity between the PAN and MS images. Then a multi-scale filtering decomposition is designed to extract the sufficient details from the PAN image. Finally, such details are modulated by gains matrix and injected into each MS band. The benchmark datasets from several satellites are used to verify the effectiveness of the proposed fusion method. Both subjective analysis and objective evaluations are conducted on the fused results.

The rest of this paper is organized as follows. Section 2 describes the structure-preserving filter. Section 3 presents the multi-scale structure perception based fusion method in details. Section 4 explains the experimental results and evaluations, and conclusion are drawn in Sect. 5.

2 Structure-Preserving Filter

Structure-preserving filter is a simple image smoothing operator designed to extract the structure from the texture via region covariance. Using image features as patch descriptors can capture the local structure and texture information implicitly and makes the approach particularly effective for structure extraction from texture. Due to the

outstanding structure-aware capability, it benefits a variety of image processing applications.

First, let $\boldsymbol{F}(x, y)$ denote a 7-dimensional feature vector representing a pixel located at (x, y) in an input image $\mathbf{G}$:

$$\boldsymbol{F}(x,y) = \left[\boldsymbol{G}(x,y), \left|\frac{\partial \boldsymbol{G}}{\partial x}\right|, \left|\frac{\partial \boldsymbol{G}}{\partial y}\right|, \left|\frac{\partial^2 \boldsymbol{G}}{\partial x^2}\right|, \left|\frac{\partial^2 \boldsymbol{G}}{\partial y^2}\right|, x, y\right]^{\mathrm{T}} \tag{1}$$

where $\boldsymbol{G}(x, y)$ denotes the intensity of the pixel, $\left|\frac{\partial \boldsymbol{G}}{\partial x}\right|, \left|\frac{\partial \boldsymbol{G}}{\partial y}\right|, \left|\frac{\partial^2 \boldsymbol{G}}{\partial x^2}\right|, \left|\frac{\partial^2 \boldsymbol{G}}{\partial y^2}\right|$ are the first and second derivatives of the intensity in both x and y directions. In the feature image $\boldsymbol{F}$, a patch R with the size of $k \times k$ can be described with a 7×7 covariance matrix $\boldsymbol{C}_R$ as follows:

$$\mathbf{C}_R = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{Z}_j - \mu)(\mathbf{Z}_j - \mu)^{\mathrm{T}} \tag{2}$$

with $\mathbf{Z}_j$ denoting every 7-dimensional feature vector inside the patch R and $\boldsymbol{\mu}$ being the mean of these feature vectors. For two image pixels p and q, the corresponding distance measure is defined as:

$$d(\mathrm{p},\mathrm{q}) = \sqrt{(\boldsymbol{\mu}_{\mathrm{p}} - \boldsymbol{\mu}_{\mathrm{q}})(\mathbf{C}_{\mathrm{p}} + \mathbf{C}_{\mathrm{q}})^{-1}(\boldsymbol{\mu}_{\mathrm{p}} - \boldsymbol{\mu}_{\mathrm{q}})^{\mathrm{T}}} \tag{3}$$

in which $\boldsymbol{\mu}_{\mathrm{p}}$, $\boldsymbol{\mu}_{\mathrm{q}}$, $\mathbf{C}_{\mathrm{p}}$ and $\mathbf{C}_{\mathrm{q}}$ denote the means and covariance matrixes of features extracted from these patches centered at pixels p and q, respectively. Based on this measure, the adaptive weight of pixels p and q can be defined as:

$$\omega_{\mathrm{pq}} = \exp\left(-\frac{d(\mathrm{p},\mathrm{q})^2}{2\sigma^2}\right) \tag{4}$$

The structure component of a pixel p is defined as:

$$\boldsymbol{S}(\mathrm{p}) = \frac{1}{H_{\mathrm{p}}}\sum_{\mathrm{q}\in N(\mathrm{p},r)} \omega_{\mathrm{pq}}\boldsymbol{G}(\mathrm{q}) \tag{5}$$

where $N(p, r)$ denotes a squared neighborhood with the radius r centered at pixel p, and $H_{\mathrm{p}} = \sum_{\mathrm{q}} \omega_{\mathrm{pq}}$ is a normalization factor.

3 The Proposed Method

3.1 Modified Structure-Preserving Filtering

In the PAN and MS images, even same object often presents in different structures with various scales. However, the SPF only focus on the filtering input itself without

considering the spatial and structural similarities in image pair. It is necessary to consider such similarity measurements in the filtering process of the PAN image. To more accurately perceive the structure information of the PAN image, the original SPF is improved by substituting the spatial correlation coefficient (SCC) [10] and the structural similarity (SSIM) [11] for the second derivatives in (1):

$$\boldsymbol{F}(x,y) = \left[\boldsymbol{G}(x,y), \left|\frac{\partial \boldsymbol{G}}{\partial x}\right|, \left|\frac{\partial \boldsymbol{G}}{\partial y}\right|, \mathrm{SSIM}, \mathrm{SCC}, x, y\right]^{\mathrm{T}} \tag{6}$$

To calculate SSIM and SCC, the original MS images are interpolated to the same size as the PAN image $\boldsymbol{P}$ through bicubic interpolation. Let $\boldsymbol{LRM}_i$ (i = 1…N) represent the upscaled N-band low-resolution MS images and the average of all $\boldsymbol{LRM}$s is used as the intensity component $\boldsymbol{I}$. The modified PAN image [12] matched with $\boldsymbol{I}$ is denoted as $\boldsymbol{MP}$. Then the SCC and SSIM are calculated from the $\boldsymbol{MP}$ and $\boldsymbol{I}$ image pair. The more spatially similar the two images are, the more accurate the structure-perception is. As a result, the modified SPF (MSPF) regards an image pair as the filtering input in which one is the high-spatial resolution PAN image, the other (MS image) provides the structural reference to the PAN.

3.2 Multi-scale Structure Perception

With reference to the MS imagery, the MSPF decomposes the matched PAN image $\boldsymbol{MP}$ to obtain its structure component $\boldsymbol{S}$ and texture component $\boldsymbol{T}$. A multi-scale decomposition scheme can capture the structures with different scales by varying the covariance window size parameter k and structure patch size parameter r. Specifically, the filtering input is smoothed by increasing k and r at each decomposition. In each decomposition, the extracted structure component is used as the filtering input for the subsequent decomposition. For the L^{th} decomposition, the multistage filtering process can be described as:

$$\boldsymbol{T}_t(\mathrm{p}) = \boldsymbol{S}_t(\mathrm{p}) - \boldsymbol{S}_{t+1}(\mathrm{p}) \tag{7}$$

$$\boldsymbol{P}(\mathrm{p}) = \sum_{t=1}^{L-1} \boldsymbol{T}_t(\mathrm{p}) + \boldsymbol{S}_L(\mathrm{p}) \tag{8}$$

in which $\boldsymbol{T_1}$, $\boldsymbol{T_2}$ … $\boldsymbol{T_{L\text{-}1}}$ represent the texture components that scale up gradually, and $\boldsymbol{S_L}$ is the final structure component having the coarsest scale. Figure 1 illustrates a 2-level decomposition of WorldView-3 (WV-3) satellite image filtered by the proposed MSPF. Figure 1(a)–(c) show that the MSPF can effectively separate texture from structure. By multi-scale filtering the input PAN image (Fig. 1 (a)) guided by the reference image $\boldsymbol{I}$ (Fig. 1(d)), the structures from small to large are successfully perceived step by step. Correspondingly, the details from fine to coarse are extracted effectively shown in Fig. 1(a) and (c).

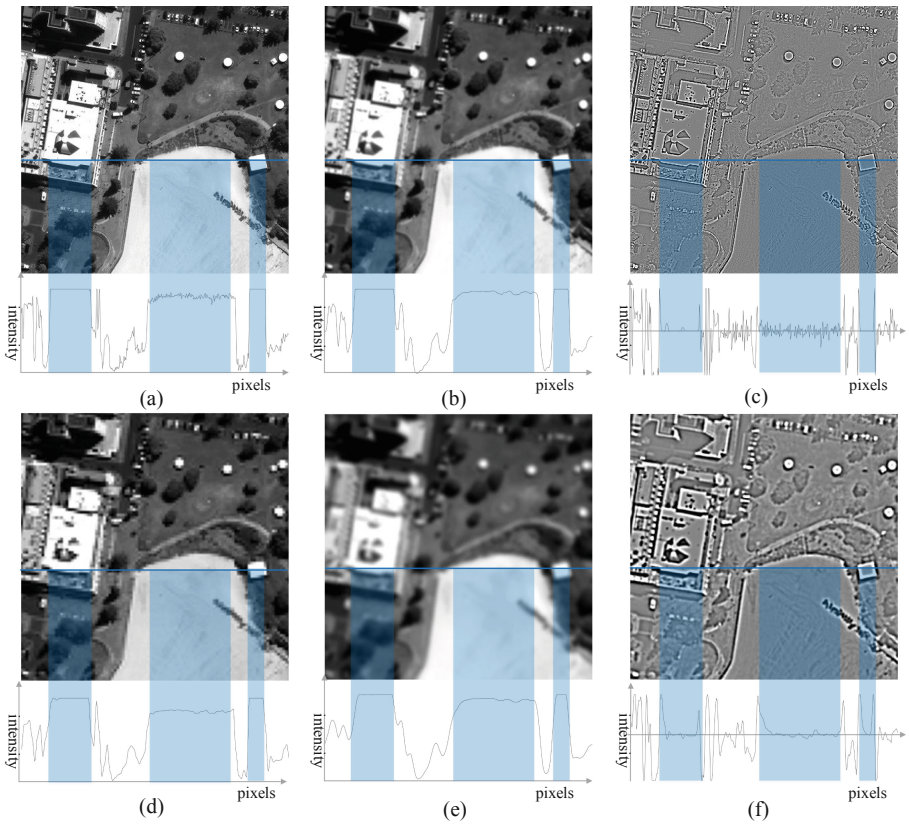

Fig. 1. The 2-level multi-scale structure perception of WV-3 PAN image. (a) The histogram-matched PAN image ***MP***. (b) The 1st level structure component S_1 of (a). (c) The 1st level texture component $\boldsymbol{T}_1$ of (a). (d) The intensity component ***I***. (e) The 2nd level structure component $\mathbf{S}_2$. (f) The 2nd level texture component $\boldsymbol{T}_2$.

3.3 The Proposed Fusion Method

The diagram of the proposed method is displayed in Fig. 2. As the aforementioned denotations, $\boldsymbol{LRM}_i$, ***I***, and ***MP*** represent the i^{th} unsampled MS channel, the average of all ***LRM***s, and the histogram-matched PAN image, respectively. Similarly, the PAN image ***P*** is matched with each $\boldsymbol{LRM}_i$ to obtain $\boldsymbol{MRM}_i$. $MSPF_{k,r,\sigma}(\boldsymbol{A},\boldsymbol{B})$ is used to represent the MSPF process in which ***A*** is the filtering input and ***B*** is the reference.

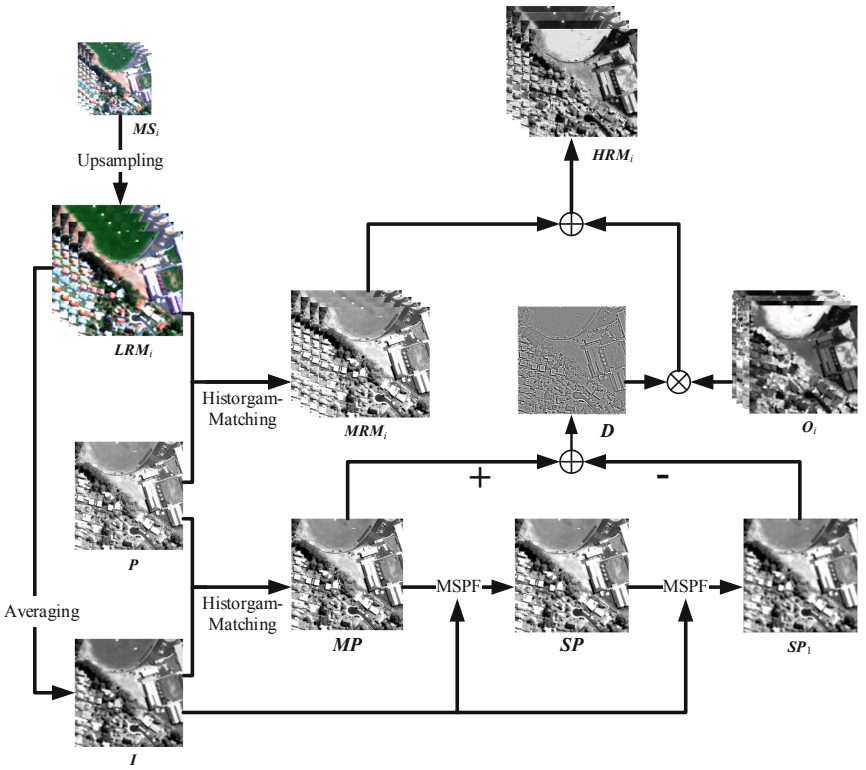

Fig. 2. The schematic diagram of the proposed method.

The fusion procedure can be simply summarized as the follow two steps:

- 1) Multi-scale decomposition
  Generate the 1st and 2nd level structure component images by MSPF with $k_2 = 2k_1 + 1$, $r_2 = 2r_1$, and $\sigma_2 = \sigma_1$:

$$\boldsymbol{SP} = MSPF_{k_1,r_1,\sigma_1}(\boldsymbol{MP},\boldsymbol{I}) \tag{9}$$

$$\boldsymbol{SP}_1 = MSPF_{k_2,r_2,\sigma_2}(\boldsymbol{SP},\boldsymbol{I}) \tag{10}$$

  Then extract the detail component image $\boldsymbol{D}$ as:

$$\boldsymbol{D} = \boldsymbol{MP} - \boldsymbol{SP}_1 \tag{11}$$

- 2) Details injection
  Compute the band-dependent weighting matrix $\boldsymbol{O}_i$ for each $\boldsymbol{LRM}_i$:

$$\boldsymbol{O}_i = N \times \frac{\boldsymbol{LRM}_i}{\sum_{i=1}^{N} \boldsymbol{LRM}_i} \tag{12}$$

The total details $\boldsymbol{D}$ is modulated by the gains matrix and then injected into each MS image $\boldsymbol{MRM}_i$. Then the final fused image is produced and denoted as $\boldsymbol{HRM}_i$.

$$\boldsymbol{HRM}_i = \boldsymbol{O}_i \times \boldsymbol{D} + \boldsymbol{MRM}_i \tag{13}$$

4 Experimental Results

4.1 Experiment Setup

The images acquired by IKONOS, GeoEye-1 (GE-1), and WV-3 are adopted as the test dataset to evaluate the proposed method. IKONOS dataset, which consist of four 4-m MS bands and 1-m PAN band, was built in 2001. It presents part of San Diego urban area, California. GE-1 imagery consists of 2.0-m MS imagery (4 bands) and 0.5-m PAN image, which was taken in February of 2009, shows part of Hobart, Australia. WV-3 dataset contains eight 7.5-m SWIR bands and a 0.3-m PAN band. It was taken in November, 2015 and covers part of the rural area in France. Four indexes are adopted for quantitative evaluation, including spectral angle mapper (SAM), Q4, ERGAS ("relative dimensionless global error in synthesis" in French) [11], and QNR [13]. The ideal value of SAM and ERGAS indexes is 0, the ideal value of Q4 and QNR indexes is 1. For comparison purpose, several state-of-the-art fusion methods, including RGF [8], APS [14], and NND [15], are adopted.

4.2 Selection of the Parameters k and r

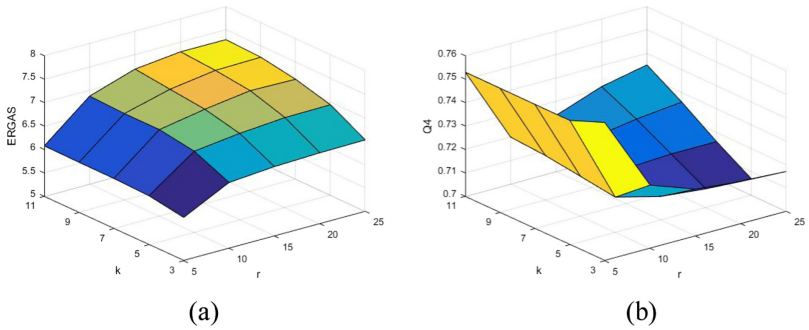

Fig. 3. Change of (a) ERGAS and (b) Q4 with respect to k and r variations.

Since the performance of the MSPF process depends on the choice of the filtering parameters k and r, different parameter selection may affect the quality of the sharpened

results. In this paper, these two parameters are tuned by selection the best ERGAS and Q4 values. Figure 3 shows the ERGAS and Q4 of the fused IKONOS images with different k and r. It is observed that ERGAS index increases and Q4 index decreases continuously with the growth of k and r. When $k = 3$ and $r = 5$, ERGAS index reaches its lowest point corresponding the best value of 5.876 and Q4 obtains its highest value of 0.762 for IKONOS dataset. Additionally, similar trends are also verified for GE-1 and WV-3 datasets.

4.3 Comparison Among Different Methods

4.3.1 Experiment 1—IKONOS Dataset

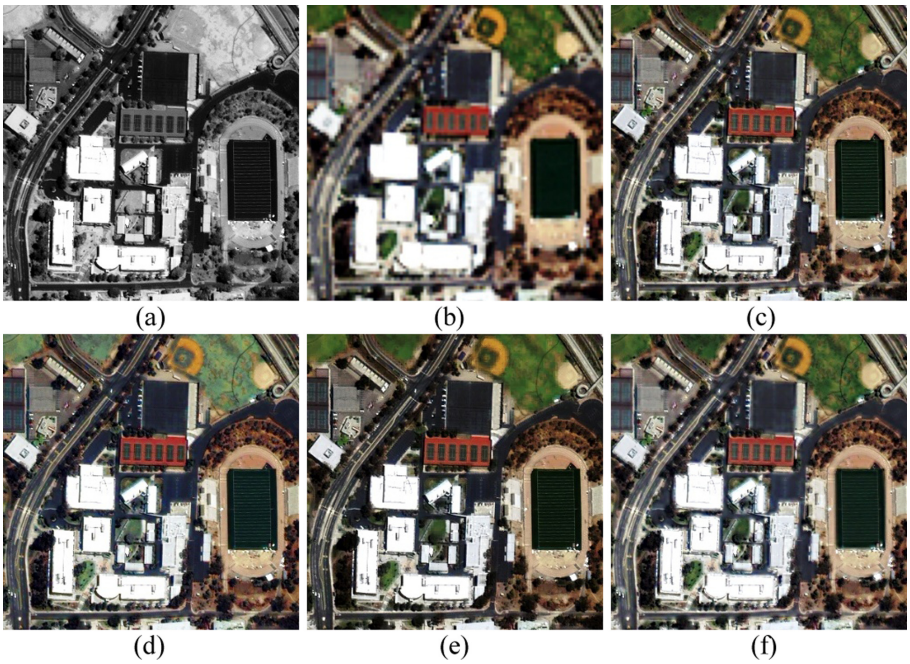

Fig. 4. IKONOS satellite images fusion result. (a) The high-spatial resolution PAN image. (b) The upsampled MS image. (c) The proposed method. (d) APS method. (e) NND method. (f) RGF method.

In this experiment, we set $k_1 = 3$, $r_1 = 5$, and $\sigma_1 = 0.5$. Figure 4(a)-(b) show the IKONOS PAN sub-scene of 400 × 400 pixels and the corresponding MS image (True color composition) upsampled to the same size. The fused results of the proposed, APS, NND, and RGF method are listed in Fig. 4(c)-(f), respectively. Compared with the MS image (Fig. 4(b)), APS method exhibits severe color distortions, especially on the green grass parts. RGF and the proposed methods preserve more spectral information than NND method. Furthermore, the proposed method looks clearer than RGF and shows the best visual effect. The objective evaluations for the four fusion methods are

given in Table 1. It can be seen that NND obtains the 2nd scores in SAM but the worst score in Q4 and ERGAS indexes. Compared with RGF method, APS shows better in SAM and ERGAS. The proposed method achieves the best performance as underlined in the table among all the indexes.

Table 1. Evaluation indexes of IKONOS dataset.

	SAM	Q4	ERGAS
APS	8.050	0.753	6.410
NND	4.344	0.638	11.856
RGF	9.179	0.753	7.124
Proposed	5.759	0.779	5.735

4.3.2 Experiment 2—GeoEye-1 Dataset

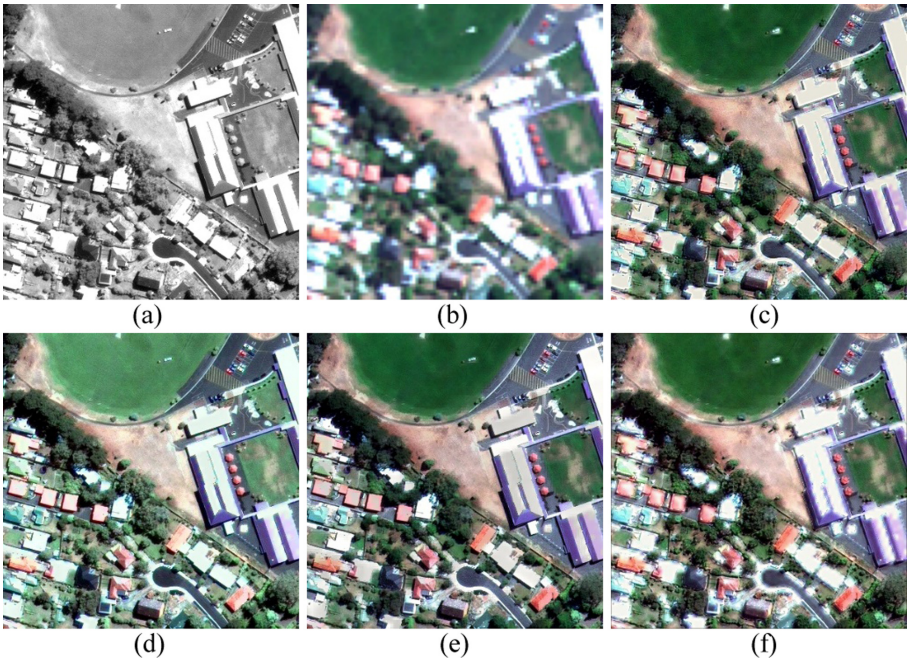

Fig. 5. GE-1 satellite images fusion result. (a) The high-spatial resolution PAN image. (b) The upsampled MS image. (c) The proposed method. (d) APS method. (e) NND method. (f) RGF method.

In Experiment 2, we set $k_1 = 3$, $r_1 = 5$, and $\sigma_1 = 0.5$. Figure 5(a) shows a 400 × 400 pixels PAN sub-scene and Fig. 5(b) is the corresponding upsampled MS scene. Similar to Experiment 1, APS method (Fig. 5(d)) introduces the obvious color distortions in

vegetation area. The highly reflective area at roofs is distorted to gray in NND method (Fig. 5(e)). RGF shows the blurred effect and changes the color around the edges of the red roofs. Although the proposed method produces slight spectral distortions, it looks like the best among the 4 methods. The objective evaluations are listed in Table 2. RGF has the poorest performance in terms of SAM and ERGAS partly due to the higher spatial resolution of GE-1 images. All the evaluation indexes show that the proposed method is optimal, which is consistency with the visual analysis.

Table 2. Evaluation indexes of GE-1 dataset.

	SAM	Q4	ERGAS
APS	9.866	0.712	8.345
NND	8.837	0.698	9.913
RGF	10.606	0.738	10.730
Proposed	3.872	0.792	8.669

4.3.3 Experiment 3—WorldView-3 Dataset

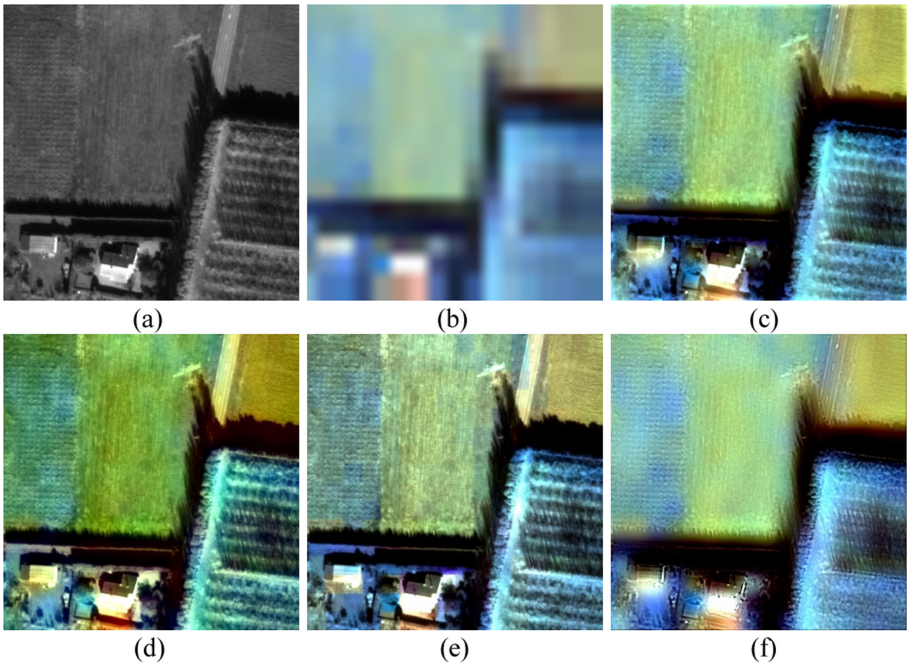

Fig. 6. WV-3 satellite image fusion result. (a) The high-spatial resolution PAN image. (b) The upsampled SWIR image. (c) The proposed method. (d) APS method. (e) NND method. (f) RGF method.

In this experiment, WV-3 PAN and SWIR images are selected and fused together. The original PAN image is 1200 × 1200 pixels while the SWIR is 48 × 48. Figure 6(a)

displays the 400 × 400 pixels sub-scene of the PAN image and Fig. 6(b) is the corresponding SWIR image (Band 8-4-1 composition) resampled to the same size. Since there is no spectral response overlap between the PAN and SWIR bands, the intensity component in APS method is built by averaging all the 8 SWIR bands. It is easy to find that APS method still introduces severe color distortions although the clarity is significant. NND method preserves more spectral information than RGF. The proposed method performs well in both spatial enhancement and spectral preservation. Table 3 reports the assessment result in terms of QNR index. Since the great difference of spatial resolution between PAN and SWIR images, QNR is selected as an appropriate index for quality evaluation. In QNR, D_s and D_λ represent the spatial and spectral distortion, respectively. The ideal values of them are 0. APS method obtains the best values in D_s and RGF has the worst result in D_λ. The proposed method achieves the best values in D_λ and QNR, which means it has the best overall quality in the four methods.

Table 3. QNR evaluation of the fused WV-3 images.

	D_S	D_λ	QNR
APS	0.056	0.398	0.568
NND	0.077	0.400	0.554
RGF	0.069	0.715	0.715
Proposed	0.077	0.114	0.818

5 Conclusion

A new fusion method based on multi-scale structure perception is proposed in this paper. Facing the mutual structure characteristics in the PAN and MS image pair, a modified structure-preserving filter (MSPF) is designed to capture the structure information from the PAN image. With the help of the multi-scale decomposition scheme of MSPF, the structures from small to large are effectively perceived and the corresponding spatial details are extracted accurately. Then the extracted details are injected through a gains matrix. Experiment results over IKONOS, GeoEye-1, and WorldView-3 datasets demonstrate that the proposed method performs well in sharpening high resolution remote sensed images regardless of the relative spectral response between PAN and MS images. The proposed method has better capabilities in preserving spectral information and increasing spatial resolution compared to existing state-of-the-art fusion methods.

Acknowledgements. This work is supported by the Seed Foundation of Innovation and Creation for Graduate Students in NPU (No. CX2020016) and Key R&D Plan of Shaanxi Province (No. 2020GY-034).

References

1. Aiazzi, B., Alparone, L., Baronti, S., Carla, R., Garzelli, A., Santurri, L.: Sensitivity of pansharpening methods to temporal and instrumental changes between multispectral and panchromatic data sets. IEEE Trans. Geosci. Remote Sens. **55**(1), 308–319 (2017)
2. Kaplan, N.H., Erer, L.: Bilateral filtering-based enhanced pansharpening of multispectral satellite images. IEEE Geosci. Remote Sens. Lett. **11**(11), 1941–1945 (2014)
3. Joshi, S., Upla, K.P., Shah, P.K.: Consistent pan-sharpening based on multistage joint and dual bilateral filters. In: 2014 IEEE International Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, pp. 2522–2525. IEEE (2014)
4. Yin, H., Li, S.: Pansharpening with multiscale normalized nonlocal means filter: a two-step approach. IEEE Trans. Geosci. Remote Sens. **53**(10), 5734–5745 (2015)
5. Upla, K.P., Joshi, S., Joshi, M.V., Gajjar, P.P.: Multiresolution image fusion using edge-preserving filters. J. Appl. Remote Sens. **9**(1), 096025 (2015)
6. Dong, W., Xiao, S., Qu, J.: Fusion of hyperspectral and panchromatic images with guided filter. SIViP **12**(7), 1369–1376 (2018). https://doi.org/10.1007/s11760-018-1291-z
7. Zhang, Q., Shen, X., Xu, L., Jia, J.: Rolling guidance filter. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8691, pp. 815–830. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10578-9_53
8. Lillo-Saavedra, M., Gonzalo-Martin, C., Garcia-Pedrero, A., Lagos, O.: Scale-aware pansharpening algorithm for agricultural fragmented landscapes. Remote Sens. **8**(10), 1–19 (2016)
9. Karacan, L., Erdem, E., Erdem, A.: Structure-preserving image smoothing via region covariances. ACM Trans. Graph. **32**(6), 1–11 (2013)
10. Zhou, J., Civco, D.L., Silander, J.A.: A wavelet transform method to merge landsat TM and SPOT panchromatic data. Int. J. Remote Sens. **19**(4), 743–757 (1998)
11. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
12. Laben, C.A., Brower, B.V.: Process for enhancing the spatial resolution of multispectral imagery using pansharpening. U.S. Patents, 6011875 (2000)
13. Alparone, L., Aiazzi, B., Baronte, S., et al.: Multispectral and panchromatic data fusion assessment without reference. Photogram. Eng. Remote Sens. **74**(2), 193–200 (2008)
14. Tu, T., Hsu, C., Tu, P., Lee, C.: An adjustable pan-sharpening approach for IKONOS/QuickBird/GeoEye-1/WorldView-2 imagery. IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. **5**(1), 125–134 (2012)
15. Sun, W., Chen, B., Messinger, D.W.: Nearest-neighbor diffusion-based pansharpening algorithm for spectral images. Opt. Eng. **53**(1), 013107 (2014)