A Novel Pansharpening Method with Multi-scale Mutual-structure Perception

Zhenzhen Fan, Xu Li(✉), and Xugang Wu

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

Abstract. Network geographic information system (WebGIS), a branch of Internet of Things (IoT), is the extension and development of traditional GIS on the network, which integrates and upgrades many remote sensing information processing tasks such as pansharpening, mapping, classification, etc. Pansharpening technology allows a detailed redisplay of the earth surface by fusing a panchromatic (PAN) and a multispectral (MS) image acquired over the same area. The extraction of the details from the PAN image and their subsequent injection into the MS image are still the key points for the pansharpening design. In this paper, we propose a novel pansharpening method which considers the multi-scale mutual-structures of the PAN and MS image pair as a new reference for details extraction and applies a weighted details injection strategy to generate the fused MS image. The experimental results show that our proposed method can obtain high-quality sharpened results and has superior performance with respect to some recent approaches.

Keywords: Pansharpening · Remote sensing · Mutual-structure

1 Introduction

WebGIS is the development of traditional GIS technology in the IoT, in which the task of remote sensing information visualization is one of the key technologies. Nowadays with the rapid development of Earth Observation technology, many high-performance commercial satellites such as QuickBird, GeoEye-1, WorldView-2/3/4 when capturing the same scene, can provide the panchromatic (PAN) image and multispectral (MS) image at almost the same time. Due to the physical constraints of the acquisition devices, PAN and MS images usually have different resolutions, although both of them contain the same scene [1]. PAN band without spectral diversity often covers a wide spectral range (usually from visual to near infrared) and has very high spatial resolution. In contrast, MS image has lower spatial resolution but higher spectral resolution. Therefore, PAN and MS images contain similar or common structures but inconsistent details such as fine edges or textures. Pansharpening has been an important tool in remote sensing field whose aim is to enhance the spatial resolution of the MS image with the aid of the PAN spatial information and the pansharpened products are dedicated to many applications [2, 3] such as Google Earth, classification, change detection, etc.

B. Li et al. (Eds.): IoTaaS 2020, LNICST 346, pp. 783–792, 2021.
https://doi.org/10.1007/978-3-030-67514-1_62

To date, a large number of pansharpening methods have been proposed which can be roughly classified into three major categories, i.e., the component substitution based methods, the multi-resolution analysis based methods, and the variational optimization based methods [4]. Essentially, the effective extraction of the details from the PAN image and the accurate details injection into the MS image always play the key roles in the pansharpening approaches. In recent years, the popularity of the filtering-based pansharpening methods is attributed to the powerful performance of feature-aware filters. Some pioneer attempts with edge-preserving filters are representative [5–7]. However, classical edge-preserving filters process only a single image to either preserve edges or remove textures, which can hardly separate structure from details since edge strength and object scale are completely different concepts. Joint filters such as guided filter (GF) [8], rolling guidance filter (RGF) [9] are more suitable for structural features perception. Li et al. proposed a fusion strategy based on GF, which decomposes the input image into a base layer having large scale structures and a detail layer with small scale details [10]. A method adopting RGF has been developed, aiming to sharpen the large-scale agricultural fragmented landscapes [11]. Mutual-structure refers to the common structural information shared in both images. Shen et al. explicitly defined the concept of mutual-structure and proposed a joint filtering process, a.k.a. mutual-structure for joint filtering (MSF), to extract the common structure from depth and RGB image pair [12]. Then, Liu et al. first introduced MSF to their pansharpening method, which enforces a direct multi-scale MSF to decompose PAN and MS images into high frequency and low frequency sub-bands and implements different fusion rules to obtain the integrated sub-bands [13]. Since PAN and MS images have different spatial resolutions, during the details extraction process, it is important to recognize the objects with different sizes or the structures of various scales. How to accurately extract the details from PAN image compared with MS image is still a core question regarding pansharpening performance. Motivated by this, we propose a novel pansharpening method which takes common multi-scale structures between the PAN and MS images into account and tries to filter out their mutual-structures as a new reference for details extraction. A weighted details injection is designed to guarantee the proper injection and faithful spectral preservation. The experiment is carried out on WorldView-3 data set. The subjective and objective evaluations show that the proposed method can produce high-quality pansharpened results and outperform some recent methods.

2 Methodology

The original MS images are upsampled to the size of the PAN image using bicubic interpolation. The upsampled N-band low-resolution MS images are denoted as $\boldsymbol{LRM}_i$, where $i = 1 \ldots N$. The flowchart of the proposed method is shown in Fig. 1. It consists of three main parts: multi-scale structure filtering, mutual-structure perception, and weighted details injection.

2.1 Multi-scale Structure Filtering

Although MSF has the capability of capturing the mutual-structures from an image pair, it sometimes introduces halo artifacts to the filtering outputs due to its local filtering formulation. It is believed that conducting direct multi-scale MSF to the PAN and MS image pair will accumulate more artifacts and may result in poor pansharpened result. Different from the method [13], we first carry out a multi-scale structure filtering to the PAN image before the mutual-structure perception. In scale-space theory [14], the variance (σ^2) of Gaussian filter is referred to as the scale parameter. When applying Gaussian with varying σ to an image, the structures are suppressed differently according to their size. Inspired by this, we perform Gaussian filtering to the PAN image $\boldsymbol{P}$ at small, medium, and large scales, individually. Then the Gaussian filtering outputs are denoted as $\boldsymbol{P}_{\mathrm{S}}$, $\boldsymbol{P}_{\mathrm{M}}$ and $\boldsymbol{P}_{\mathrm{L}}$ as shown in Fig. 1 which are expected to have the similar structural features to the MS image at different scales. The corresponding deviations are denoted as σ_{S}, σ_{M} and σ_{L}, respectively. The value of deviation becomes higher, the larger structures will be suppressed. Therefore, we set σ_{S} = 3, σ_{M} = 15 and σ_{L} = 50 based on a large number of experiments. Gaussian filtering is calculated as

$$\boldsymbol{P}_{\sigma^2} = g * \boldsymbol{P} \tag{1}$$

in which Gaussian kernel $g(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$ and $*$ denotes convolution.(x, y) is the coordinate.

2.2 Mutual-Structure Perception

In this stage, we employ MSF to capture the mutual-structures between {$\boldsymbol{P}_{\mathrm{S}}$, $\boldsymbol{P}_{\mathrm{M}}$, $\boldsymbol{P}_{\mathrm{L}}$} and MS image. MSF is a joint filtering process to extract the common structures from an image pair and has two inputs: target image and reference image [12]. The function of MSF is defined as

$$(\boldsymbol{T}, \boldsymbol{R}) \ = \ \mathrm{MSF}(\boldsymbol{T}_0, \boldsymbol{R}_0, \lambda, \beta, \varepsilon_1, \varepsilon_2) \tag{2}$$

in which $\boldsymbol{T}$ and $\boldsymbol{R}$ are the mutual-structures corresponding to the target image $\boldsymbol{T}_0$ and the reference image $\boldsymbol{R}_0$, respectively. λ and β control the deviation to $\boldsymbol{R_0}$ and $\boldsymbol{T}_0$, respectively. We set λ and β in range 30−300, which satisfy $\lambda_1 > \lambda_2 > \lambda_3$ and $\beta_1 > \beta_2 > \beta_3$. ε_1 and ε_2 control the smoothness of $\boldsymbol{R}$ and $\boldsymbol{T}$ respectively. We set them around 1E-5. Since the MS image has more than one band, the average of N-band MS image is used to represent the intensity component (***INT***). We take $\boldsymbol{P}_{\mathrm{S}}$, $\boldsymbol{P}_{\mathrm{M}}$, $\boldsymbol{P}_{\mathrm{L}}$ as the target image respectively and ***INT*** as the reference for MSF under the assumption that the reference image should contain correct structure information at different scales. Specifically, the multi-scale mutual-structure perception process can be implemented as the following:

$$\boldsymbol{INT} = \frac{1}{N}\sum\nolimits_{i}^{N} (\boldsymbol{LRM}_1 + \boldsymbol{LRM}_2 + \ldots \boldsymbol{LRM}_{\mathrm{N}}) \tag{3}$$

$$(\boldsymbol{T}_1, \boldsymbol{R}_1) = \mathrm{MSF}(\boldsymbol{P}_\mathrm{S}, \boldsymbol{INT}, \lambda_1, \beta_1, \varepsilon_1, \varepsilon_2) \tag{4}$$

$$(\boldsymbol{T}_2, \boldsymbol{R}_2) = \mathrm{MSF}(\boldsymbol{P}_\mathrm{M}, \boldsymbol{INT}, \lambda_2, \beta_2, \varepsilon_1, \varepsilon_2) \tag{5}$$

$$(\boldsymbol{T}_3, \boldsymbol{R}_3) = \mathrm{MSF}(\boldsymbol{P}_\mathrm{L}, \boldsymbol{INT}, \lambda_3, \beta_3, \varepsilon_1, \varepsilon_2) \tag{6}$$

Through multiple iterations and setting constraints, we can obtain the desired mutual structures $\boldsymbol{T}_1$, $\boldsymbol{T}_2$, and $\boldsymbol{T}_3$. As shown in Fig. 1, the filtering out images, i.e., $\boldsymbol{T}_1$, $\boldsymbol{T}_2$, $\boldsymbol{T}_3$ correspond to the mutual structures at small, medium, and large scales, respectively. The scale of the small structures in $\boldsymbol{T}_1$ is as small as the minimum spatial size of the objects in $\boldsymbol{INT}$, e.g. cars.

2.3 Weighted Details Injection

In order to reduce the spatial and spectral information distortions, we design a weighted details injection to avoid improper "over" or "under" injection. First, we match the PAN image $\boldsymbol{P}$ with the intensity component $\boldsymbol{INT}$ to obtain the histogram-matched PAN image named as $\boldsymbol{P\_M}$:

$$\boldsymbol{P\_M} = \boldsymbol{P} \times \frac{\sigma_\mathrm{INT}^2}{\sigma_\mathrm{P}^2} + \mu_\mathrm{INT} - \frac{\sigma_\mathrm{INT}^2}{\sigma_\mathrm{P}^2} \times \mu_P \tag{7}$$

where σ^2 and μ denote the variance and mean of the corresponding subscripted images, respectively.

Then, the PAN details at different scales can be calculated from the difference between $\boldsymbol{P\_M}$ and the mutual-structures:

$$\boldsymbol{D}_1 = \boldsymbol{P\_M} - \boldsymbol{T}_1 \tag{8}$$

$$\boldsymbol{D}_2 = \boldsymbol{P\_M} - \boldsymbol{T}_2 \tag{9}$$

$$\boldsymbol{D}_3 = \boldsymbol{P\_M} - \boldsymbol{T}_3 \tag{10}$$

From Fig. 1, it is clear that the extracted multi-scale details $\boldsymbol{D}_1, \boldsymbol{D}_2, \boldsymbol{D}_3$ represent the spatial details of the PAN image from large to medium, and then to small scales, progressively. To eliminate the redundancy among $\{\boldsymbol{D}_1, \boldsymbol{D}_2, \boldsymbol{D}_3\}$, we use the weighted sum of $\boldsymbol{D}_1$, $\boldsymbol{D}_2$, and $\boldsymbol{D}_3$ to estimate a synthesized details $\boldsymbol{D}$ by (11). The weight coefficients k_1, k_2, and k_3 can be obtained through (12)–(14), in which $()^\mathrm{T}$ represents the transpose operation.

$$\boldsymbol{D} = k_1 \times \boldsymbol{D}_1 + k_2 \times \boldsymbol{D}_2 + k_3 \times \boldsymbol{D}_3 \tag{11}$$

$$k_1 = \frac{(\boldsymbol{D}_1 - \boldsymbol{D}_2 - \boldsymbol{D}_3)^\mathrm{T} \times \boldsymbol{D}_1}{\left[(\boldsymbol{D}_1 - \boldsymbol{D}_2 - \boldsymbol{D}_3)^\mathrm{T} \times (\boldsymbol{D}_1 - \boldsymbol{D}_2 - \boldsymbol{D}_3)\right]} \tag{12}$$

$$k_2 = \frac{(\boldsymbol{D}_2 - \boldsymbol{D}_1 - \boldsymbol{D}_3)^{\mathrm{T}} \times \boldsymbol{D}_1}{\left[(\boldsymbol{D}_2 - \boldsymbol{D}_1 - \boldsymbol{D}_3)^{\mathrm{T}} \times (\boldsymbol{D}_2 - \boldsymbol{D}_1 - \boldsymbol{D}_3)\right]} \tag{13}$$

$$k_3 = 1 - k_1 - k_2 \tag{14}$$

The details injection into each MS band is modulated by a gains matrix which keeps the proportion among the MS bands unvaried. The injection gains matrix is calculated as

$$\boldsymbol{O}_i = \frac{\boldsymbol{LRM}_i}{\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{LRM}_i} \tag{15}$$

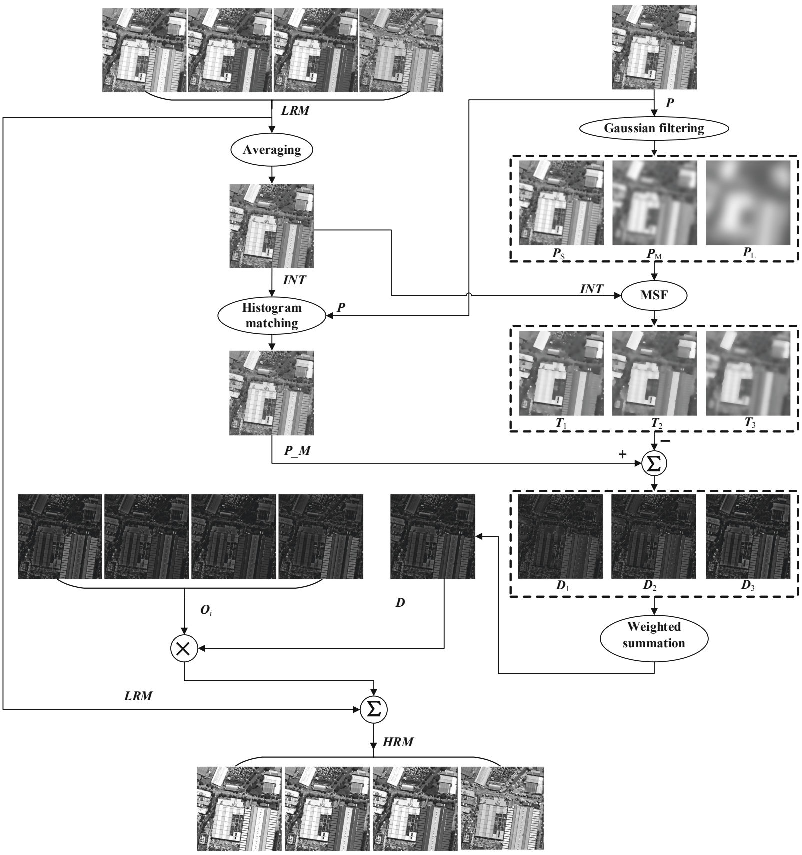

Fig. 1. Flowchart of the proposed method.

Finally, the pansharpened MS image $\boldsymbol{HRM}_i$ is generated by injecting the spatial information $\boldsymbol{D}$ into each band of the upsampled MS image $\boldsymbol{LRM}_i$ by a gains matrix $\boldsymbol{O}_i$. The sharpened MS image is obtained by

$$\boldsymbol{HRM}_i = \boldsymbol{O}_i \times \boldsymbol{D} + \boldsymbol{LRM}_i \tag{16}$$

3 Experimental Results

3.1 Dataset and Evaluation Indexes

To validate the proposed method, we use WorldView-3 satellite imagery as the test data which consists of 8-band MS image with 1.24 m spatial resolution and a 0.31 m PAN band. The data set taken in October of 2014 covers the harbor area of Sydney, Australia. The original size of the MS and PAN images are 400 $\times$ 400 pixels and 1600 $\times$ 1600 pixels, respectively. Both images have been registered. Figure 2(a) shows a PAN sub-scene of 800 $\times$ 800 pixels and Fig. 2(b) reports the same area of the resampled MS image.

Three evaluation indexes, i.e., spectral angle mapper (SAM) [15], relative dimensionless global error in synthesis (ERGAS) [16], and Q4 [17], are selected to assess the quality of the results, which are the most widely used in the pansharpening studies because of their robustness to the difference between experimental datasets [4]. For comparison purpose, we also implemented and tested Kaplan's method [6], Dong's method [18], and Liu's [13] method.

3.2 Subjective Analysis

Figure 2(c0-c3)-(f0-f3) show the RGB composition of the pansharpened MS bands corresponding to Dong's, Kaplan's, Liu's and our method, respectively. In order to facilitate observation and comparison, three parts framed by three colored-squares (red, yellow, green) are zoomed up and shown in rows 3–4, rows 5–6 and rows 7–8, respectively. The order of the row arrangement is roughly according to the scale of the main features from large to small. In rows 3–4 of Fig. 2, the prominent object is the big white roof. It is easy to find that Fig. 2(d0-d3) and (e0-e3) present more blurring compared with Fig. 2(c0-c3) and (f0-f3) especially around the contour of the roof. In addition, the light blue part on the roof disappears in Fig. 2(c0-c3) which mean Dong's method produces the worst spectral distortions. In rows 5–6, most ground objects are presented on medium scale such as the red roofs and crowns of trees. Compared with Fig. 2(b0-b3), Fig. 2(c0-c3)-(e0-e3) show the spectral distortions more or less in which the yellow roof in Fig. 2(c0-c3) appear too bright, and the color of the vegetation part in Fig. 2(d0-d3) are distorted from dark to bright. Furthermore, Liu's method (Fig. 2 (e0-e3)) generates the worst visual result especially on the red roofs. Our method in rows 5–6 present the high-quality in terms of the spectral fidelity. Although Fig. 2(d0-d3) have sufficient spatial details, their color distortions are obvious. Rows 7–8 of Fig. 2 display the ground scene full of small objects including the unfinished building

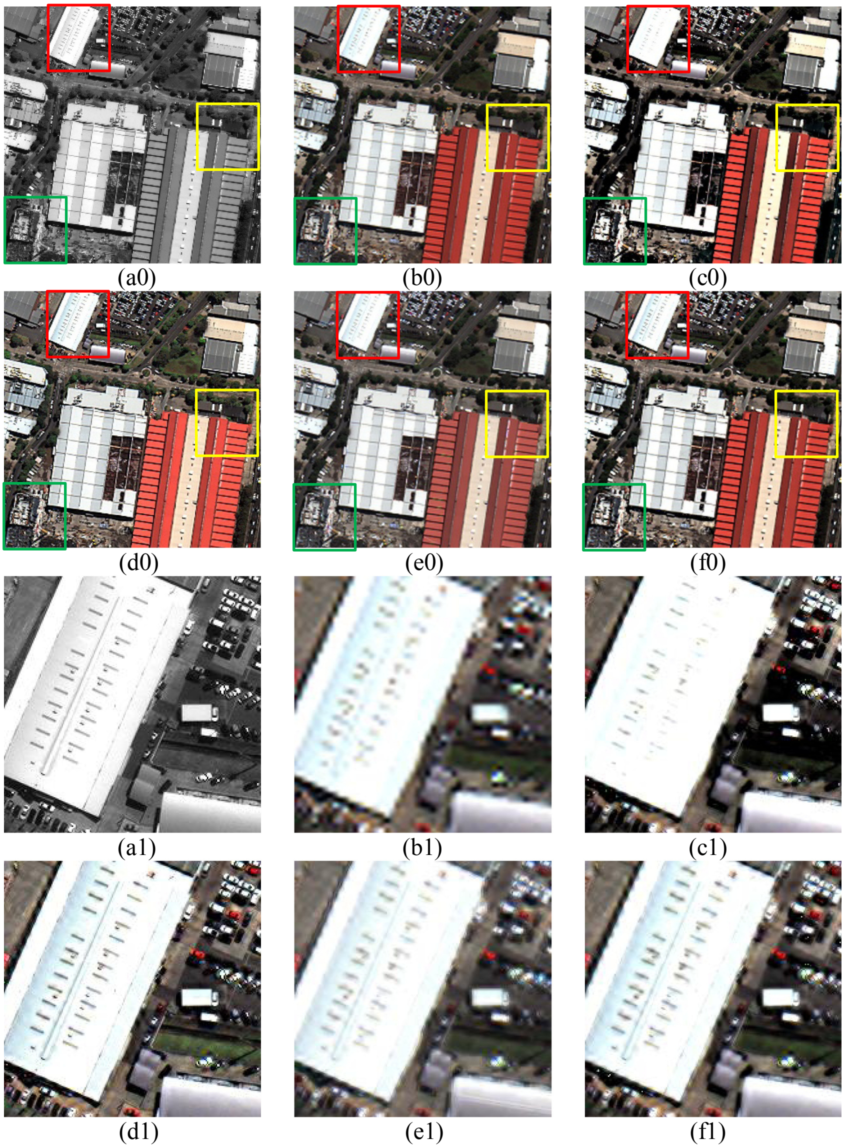

Fig. 2. A sub-scene of the test WorldView-3 dataset with 800 × 800 pixels: (a0-a3) PAN image, (b0-b3) MS image, (c0-c3) Dong's method [18], (d0-d3) Kaplan's method [6], (e0-e3) Liu's method [13], (f0-f3) our method. In order to observe more clearly, three color (red, yellow, green) boxed parts of the results are zoomed up and shown in rows 3–4, rows 5–6 and rows 7–8, respectively. (Color figure online)

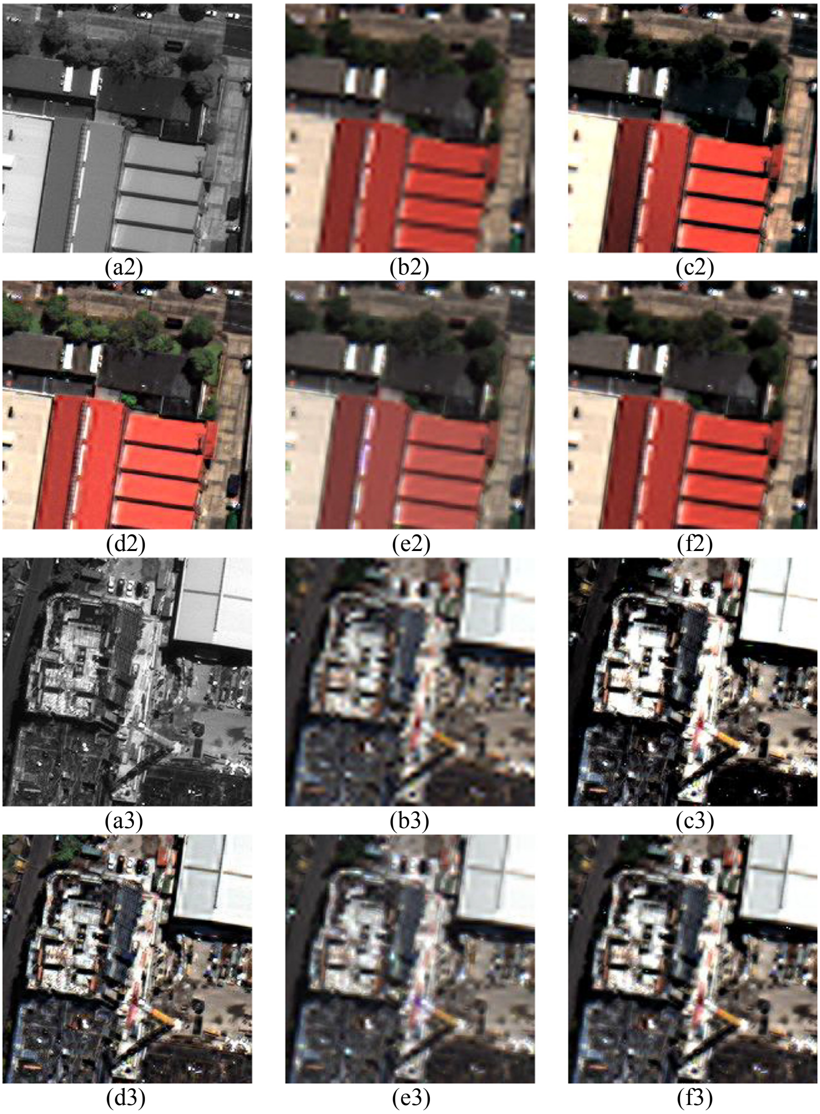

Fig. 2. (*continued*)

site and cars. It is clear that Liu's method (Fig. 2(e0-e3)) is the most blurred partly because of the accumulated artefacts from the direct multi-scale MSF process. Figure 2 (d0-d3) have blurred parts around the fine details which indicate the insufficient details injection of Kaplan's method. Dong's method (Fig. 2(c0-c3)) show better visual effect than Fig. 2(d0-d3) and 2(e0-e3), but it still presents the evident spectral distortions concentrated on the high reflection area. Our method (Fig. 2(f0-f3)) in the rows 7–8 obtains the best visual result compared with the other three methods. From the subjective analysis and comparison of Fig. 2, we can see that our method produces high-quality result at different scales and outperforms the compared methods.

3.3 Objective Analysis

Table 1 lists the quantitative evaluation of the pansharpened results. The values in parentheses are the ideal values for the evaluation indexes. Kaplan's method has the worst performance in terms of all indexes which is consistent with the visual analysis. The advantage of edge-perception seems not obvious. According to SAM and ERGAS scores, Liu's method obtains good results in the second place. The Q4 index shows that Dong's and our method are better, which indicates multi-scale GF and MSF are helpful to improve the spatial quality. Our method achieves the best scores which demonstrates that its multi-scale mutual-structure perception and weighted details injection are more effective for the spatial improvement and spectral preservation than the compared methods.

Table 1. Quality indexes of the pansharpened results.

Index	Dong's method [18]	Kaplan's method [6]	Liu's method [13]	Ours
SAM (0)	6.019	7.045	5.741	**3.514**
ERGAS (0)	15.711	31.009	13.593	**9.620**
Q4 (1)	0.815	0.809	0.814	**0.823**

4 Conclusion

We propose a novel multi-scale mutual-structure perception based pansharpening method focusing on the accurate details extraction and the following injection. The characteristic of MS and PAN image pair having similar structures and inconsistent details motivates us to combine Gaussian filtering and MSF to achieve the purpose of the multi-scale mutual-structure perception. A weighted details injection is also designed to guarantee that the PAN details can be properly injected into the MS image. We conduct the experiment on WorldView-3 satellite imagery and our method performs well both visually and objectively. The experimental results demonstrate that the proposed method can produce high-quality pansharpened results and outperform some recently published methods.

Acknowledgements. This research is supported by the Seed Foundation of Innovation and Creation for Graduate Students in NPU (No. CX2020016), Key R&D Plan of Shaanxi Province (No. 2020GY-034), and Shanghai Key Laboratory of Multidimensional Information Processing in East China Normal University (No. 2019KEY001).

References

1. Alparone, L., Aiazzi, B., Baronti, S., Garzelli, A.: Remote Sensing Image Fusion. CRC Press, New York (2015)
2. Li, S., Kang, X., Fang, L., Hu, J., Yin, H.: Pixel-Level image fusion: a survey of the state of the art. Inf. Fusion **33**, 100–112 (2017)
3. Wang, T., Fang, F., Li, F., Zhang, G.: High-quality Bayesian pansharpening. IEEE Trans. Image Process. **28**(1), 227–239 (2019)
4. Meng, X., Sheng, H., Li, H., Zhang, L., Fu, R.: Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: practical discussion and challenges. Inf. Fusion **46**, 102–113 (2019)
5. Yin, H., Li, S.: Pansharpening with multi-scale normalized non-local mean filter. IEEE Trans. Geosci. Remote Sens. **53**(10), 1–12 (2015)
6. Kaplan, N.H., Erer, I.: Bilateral filtering-based enhanced pansharpening of multispectral satellite images. IEEE Geosci. Remote Sens. Lett. **11**(11), 1941–1945 (2014)
7. Song, Y., Wu, W., Liu, Z.: An adaptive pansharpening method by using weighted least squares filter. IEEE Geosci. Remote Sens. Lett. **13**(1), 18–22 (2016)
8. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. **35** (6), 1397–1409 (2013)
9. 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
10. Li, S., Kang, X., Hu, J.: Image fusion with guided filtering. IEEE Trans. Image Process. **22**(7), 2864–2875 (2013)
11. 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)
12. Shen, X., Zhou, C., Xu, L., Jia, J.: Mutual-structure for joint filtering. Int. J. Comput. Vision **125**(1), 19–33 (2017)
13. Liu, S., Li, X., Zhang, X.: Remote sensing image fusion algorithm based on mutual-structure for joint filtering using saliency detection. J. Electron. Imaging **28**(3), 033007 (2019)
14. Lindeberg, T.: Scale-space theory: a basic tool for analyzing structures at different scales. J. Appl. Stat. **21**, 225–270 (1994)
15. Palsson, F., Sveinsson, R., Ulfarsson, M.O., Benediktsson, J.A.: Quantitative quality evaluation of pansharpened imagery: consistency versus synthesis. IEEE Trans. Geosci. Remote Sens. **54**(3), 1247–1259 (2016)
16. Ranchin, T., Aiazzi, B., Alparone, L., Baronti, S., Wald, L.: Image fusion – the ARSIS concept and some successful implementation schemes. ISPRS J. Photogram. Remote Sens. **58**, 4–18 (2003)
17. Alparone, L., Baronti, S., Garzelli, A., Nencini, F.: A global quality measurement of pan-sharpened multispectral imagery. IEEE Geosci. Remote Sens. Lett. **1**(4), 313–317 (2004)
18. 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