

# Optical and SAR image fusion method with coupling gain injection and guided filtering

Yukai Fu<sup>1</sup>,<sup>a</sup> Shuwen Yang,<sup>a,b,c,\*</sup> Heng Yan,<sup>a</sup> Qing Xue,<sup>a</sup> Zhuang Shi,<sup>a</sup> and Xiaoqiang Hu<sup>a</sup>

<sup>a</sup>Lanzhou Jiaotong University, Faculty of Geomatics, Lanzhou, China

<sup>b</sup>National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou, China

<sup>c</sup>Gansu Province Engineering Laboratory for National Geographic State Monitoring, Lanzhou, China

**Abstract.** Significant radiometric differences and weak grayscale correlations exist between optical and SAR images. As a result, there are severe spectral and spatial distortions in the fused images. We propose a fusion method of optical and SAR remote sensing images that couples the gain injection method and the guided filter. The proposed method is based on the fusion framework of generalized intensity-hue-saturation non-subsampled contourlet transform, and the gain injection is used for the low-frequency coefficient fusion to reduce the spectral distortion. Then, the divergence is used as the activity measure operator to calculate the initial weight template for the high-frequency coefficients. The guided filter is used to optimize the edge details of the initial weight template. The fused high-frequency coefficients are obtained by weighted average. Through comparison experiments with existing fusion methods, the results show that the proposed method has the best quality of fusion and the proposed method has the best performance. © *The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: [10.1117/1.JRS.16.046505](https://doi.org/10.1117/1.JRS.16.046505)]

**Keywords:** images fusion; SAR and optical images; gain injection; guided filter.

Paper 220270G received May 3, 2022; accepted for publication Oct. 5, 2022; published online Nov. 7, 2022.

## 1 Introduction

SAR has strong penetrating power and can generate remote sensing images without being restricted by weather and time, and the structural features of the images are apparent. However, the lack of spectral information and severe noise make SAR image interpretation difficult. Optical images are rich in texture and spectral information, but the imaging conditions are volatile and easily obscured by clouds *az*.<sup>1</sup> The pixel-level fusion of optical and SAR images can integrate both advantages and obtain complementary information, which is of great significance to overcoming the limitations of single-source remote sensing images and improving the interpretation capability of images. The fused images of SAR and optical images have been widely used in many fields to improve the interpretation of remote sensing images, such as land cover classification,<sup>2,3</sup> sea ice identification,<sup>4,5</sup> biomass estimation,<sup>6</sup> change detection,<sup>7,8</sup> flood monitoring,<sup>9,10</sup> and urban feature extraction and classification.<sup>11,12</sup>

Pixel-level fusion methods of optical and SAR images can be divided into four categories: component substitution (CS) fusion methods, multi-scale decomposition (MSD) fusion methods, hybrid methods based on CS and MSD, and model-based methods. The most used method is the hybrid method, which integrates the advantages of both CS and MSD methods. Compared to single methods, hybrid methods can reduce spatial structure and spectral distortion in the fused image and are more suitable for optical and SAR image fusion.<sup>13</sup> Hong et al. proposed a method based on the intensity-hue-saturation (IHS) and wavelet transform. This method uses global statistical features as the active measure and then achieves fusion by directly replacing sub-bands. However, this method ignores the specificity of individual image elements and may introduce

\*Address all correspondence to Shuwen Yang, [yangshuwen@mail.lzjtu.cn](mailto:yangshuwen@mail.lzjtu.cn)

a large amount of noise, leading to significant spectral distortions in the fusion results.<sup>14</sup> Subsequently, Han et al. performs IHS transform and à trous wavelet transforms on the images to be fused and then uses local statistical parameters as activity measures to calculate the pixel-wise fusion weight. The fusion weight estimated by this method fully considers the unique characteristic of a single pixel and the influence of neighboring pixels on the central pixel. The resulting fused image retains more spatial structure and spectral information.<sup>15</sup> To further reduce the spatial distortion in the fused images, Anandhi and Valli<sup>16</sup> calculated the fusion weights based on non-subsampled contourlet transform (NSCT) with minimum likelihood ratio, local gradient, and maximum edge intensity as the active measure operators, which can retain more edge and contour features in the fused image. Kulkarni et al. used a hybrid method of the principal component analysis (PCA) and discrete wavelet transform (DWT) transform as the base fusion framework, calculated the fusion weights using the image element local energy as the active measure operator, and performed a weighted average fusion of the components further to reduce the spectral distortion in the fused images.<sup>17</sup> Zhou et al. used an adaptive IHS fusion method based on phase coherence feature preservation to fuse SAR and optical remote sensing images, and more spectral and spatial structure information was retained in the results.<sup>18</sup>

Although many scholars have used the hybrid method as the basic framework and continuously introduced better-performing activity measure operators and improved the fusion weight calculation method for multi-scale components, there are still two problems:

1. Because the fusion method of multi-scale component weighted averaging cannot overcome the nonlinear radiometric differences between SAR and optical images, significant spectral distortions are inevitable in the fusion results.
2. The noise in SAR images is serious. However, the existing multi-scale feature activity measurement operator has poor noise immunity. The fusion weight template calculated this way cannot effectively reduce the spatial and spectral distortions caused by noise.

Given the problems of existing fusion methods, this paper proposed a coupled gain injection and guided filtering method for optical and SAR image fusion. The proposed method uses generalized intensity-hue-saturation non-subsampled contourlet transform (GIHS-NSCT) as the basic fusion framework. First, GIHS extracts the luminance component  $I$  of the optical image. NSCT then decomposes the  $I$  and SAR images into multi-scale and multi-directional. Next, the low-frequency coefficients of  $I$  and SAR are fused using the gain injection method. The gain injection method is used by solving the unique features of the low-frequency coefficients of the SAR image and injecting the unique features into the low-frequency coefficients of  $I$  as gain. Fusing only the unique features of the low-frequency coefficients of SAR images can effectively reduce the spectral distortion.

## 2 Fundamental Theories and Methods

This section introduces some fundamental theories involved in the proposed fusion method, including the GIHS ensemble method, the NSCT method, and the guided filter.

### 2.1 GIHS Fusion Method

GIHS extends the classical IHS method of the CS class fusion method. Compared with IHS, it can acquire the luminance components of images with more than three channels. It does not require forward and inverse transformation of the image color space, which is computationally tiny and improves the fusion efficiency.<sup>19</sup> Therefore, GIHS is widely used in image fusion,<sup>20,21</sup> and we extend it to optical and SAR image fusion. The fused image calculation process of the GIHS method is as follows:

$$F_i = M_i + \lambda(\text{SAR} - I), \quad (1)$$

$$I = \sum_{i=1}^B \omega_i M_i, \quad (2)$$

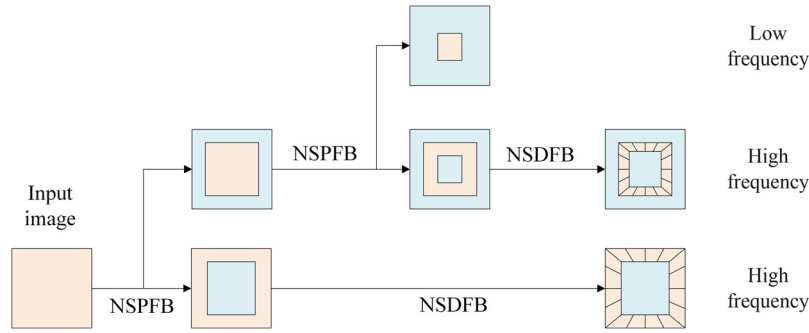


Fig. 1 Decomposition framework of NSCT.

where  $F$ ,  $M$ , and SAR represent the fused image, optical image, and SAR image, respectively.  $I$  is the brightness component of the optical image and  $B$  is the number of bands.  $\lambda$  and  $\omega$  represent the corresponding weights of each band of the optical image, respectively.

### 2.2 NSCT Method

NSCT is an image MSD method proposed by Da Cunha et al.<sup>22</sup> It consists of the non-subsampling pyramid filter bank (NSPFB) and the non-subsampling directional filter bank (NSDFB). NSPFB can perform MSD of images. Its non-down sampling decomposition can reduce the distortion of image elements caused by up-sampling and down-sampling processes and has translation invariance. NSDFB is a multi-directional filter bank that decomposes the image into multiple directions and preserves multi-directional detail features. The NSCT method that coupled NSPFB and NSDFB has the advantages of multi-scale, multi-directional, and non-down sampling.<sup>23</sup> So NSCT is widely used in image fusion.<sup>24,25</sup> The schematic diagram of NSCT MSD is shown in Fig. 1.

### 2.3 Guided Filter

The guided filter is an edge-preserving filter based on a local linear model. It works by correcting the noisy image with reference to the guiding image and has the properties of noise reduction and edge retention. Therefore, guided filters are widely used to optimize fusion weight maps in image fusion.<sup>26,27</sup> The guiding image is the key to determining the filtering effect, which can be the same or different from the input image but must be given in advance. The guided filter is implemented by a sliding calculation of the local window. For a square sliding window  $w_k$  of size  $r \times r$ , the linear relationship between the guiding image  $G$  and the output image  $O$  can be expressed as

$$O_i = a_k G_i + b_k, \quad \forall i \in w_k, \tag{3}$$

where  $(a_k, b_k)$  is the linearity factor of the sliding window  $w_k$ . The linear coefficients are significant, and solving for them is a least-squares optimization process. Optimization aims to solve a set of  $(a_k, b_k)$  such that the difference between the input image window  $T$  and the output image window  $O$  is minimized. Based on the above, the optimization objective function can be defined as follows:

$$E(a_k, b_k) = \sum_{i \in w_k} [(a_k G_i + b_k - T_i)^2 + \epsilon a_k^2], \tag{4}$$

where  $\epsilon$  is the regularization parameter and  $\epsilon > 0$ ,  $a_k$  and  $b_k$  are calculated as

$$a_k = \frac{\sum_{i \in w_k} G_i T_i - \mu_k \bar{T}_k}{r^2 (\sigma_k^2 + \epsilon)}, \tag{5}$$

$$b_k = \overline{T_k} - a_k \mu_k, \quad (6)$$

where  $\mu_k$  and  $\sigma_k^2$  are the average values and variances values of  $G$  and the window  $w_k$ .  $\overline{T_k}$  is the mean value of the window  $w_k$  in  $T$ .

### 3 Proposed Fusion Method

In this section, we elaborate on the implementation process of the proposed fusion method, including the basic framework, the rules for low-frequency coefficients, and high-frequency coefficients fusion.

#### 3.1 Basic Framework

We use the hybrid method of GIHS-NSCT to fuse optical and SAR images. The overall methodological framework of the algorithm is shown in Fig. 2. In Fig. 2, the input optical and SAR images have been registered using the method proposed by Yan et al.<sup>28</sup> and can be directly used for pixel-level fusion. GIHS-NSCT first acquires the luminance image of the optical image with GIHS. Then the luminance and SAR images are multi-scale fused based on NSCT to obtain the fused luminance image. Finally, the original optical and the new luminance image are fused using GIHS. The key to determining the quality of fusion is the feature maps and fusion weight maps corresponding to the low-frequency and high-frequency coefficients. The quality of the feature map depends on the feature extraction method and the feature measurement used. The key to the fusion weight map lies in the fusion weight calculation method and the activity measurement operator. The main steps of the GIHS-NSCT fusion method are as follows:

1. Perform basic pre-processing of optical and SAR images, respectively, and upsample the optical image to the exact resolution as the SAR image. Then register it with the SAR image and crop out the joint region to get the input image Optical and SAR.
2. Use the GIHS method to obtain the luminance component  $I$  of the optical image and adjust the grayscale range of SAR to the same as  $I$  to get  $SAR^*$ .
3. Get the approximate images  $\{L^I, L^{SAR^*}\}$  of low-frequency sub-bands and the detailed images  $\{H_{j,k}^I, H_{l,k}^{SAR^*}\}$  of high-frequency sub-bands of  $I$  and  $SAR^*$  with NSCT.
4. Calculate the feature maps  $\{FL^I, FL^{SAR^*}\}$  of the approximate image and the fusion weight maps  $\{PL^I, PL^{SAR^*}\}$  for the gain injection and feature maps  $\{FH_{j,k}^I, FH_{j,k}^{SAR^*}\}$  and fusion weight maps  $\{PH_{j,k}^I, PH_{j,k}^{SAR^*}\}$  of the detailed images.
5. Perform gain injection and take the weighted average on  $\{L^I, L^{SAR^*}\}$  and  $\{H_{j,k}^I, H_{l,k}^{SAR^*}\}$  according to the fusion weight to obtain  $L^{new}$  and  $H_{j,k}^{new}$ .
6. Perform inverse NSCT on  $L^{new}$  and  $H_{j,k}^{new}$  to get the fused luminance component  $I^{new}$ .
7. Fuse the optical image and  $I^{new}$  use GIHS. When fusing,  $\lambda$  is set to 1, and  $\omega_i$  is set to  $1/B$ .

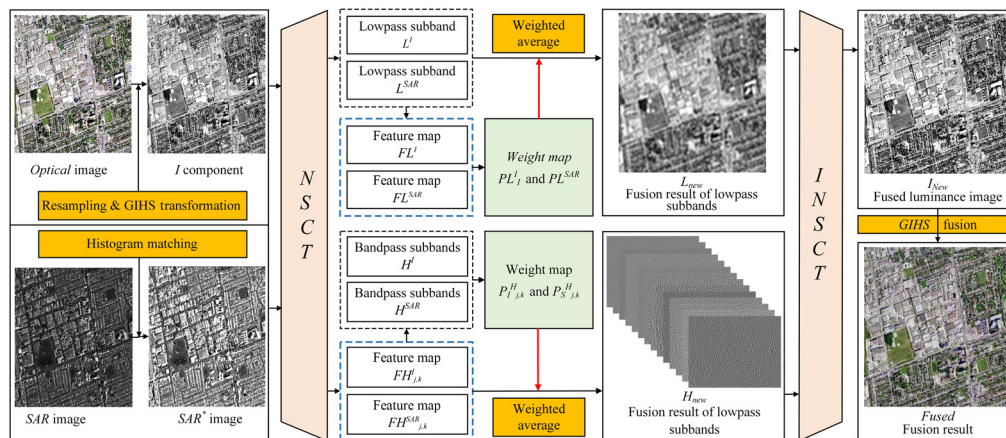


Fig. 2 Framework of the proposed method.

### 3.2 Rule for Low-frequency Coefficients

The low-frequency sub-band approximates the image, which contains the main contour features. The low-frequency sub-bands are also crucial for determining the fused image’s spectral distortion. Therefore, considering the significant nonlinear radiometric differences between optical and SAR images, we use the feature gain method for fusion when calculating the fused low-frequency coefficients. The fusion is weighted only at specific features of the low-frequency sub-band of the SAR image. The weights of the SAR image elements at non-specific features are all 0, while the weights of the optical image elements are set to 1. This fusion method can effectively reduce the spectral distortion in the fused image caused by the nonlinear radiation difference.<sup>29</sup> The fusion process of the low-frequency sub-band is shown in Fig. 3. The images used in Fig. 3 are rendered for easy observation.

In the fusion process, the common features of the low-frequency sub-bands of SAR and I are firstly calculated according to the Eq. (7):

$$FL^{common} = \min\{FL^I, FL^{SAR}\}. \tag{7}$$

Since the low-frequency sub-bands are the approximation of the image features, we take the low-frequency sub-bands of I and SAR directly as the feature maps, which is to let  $FL^I = L^I$ ,  $FL^{SAR} = L^{SAR}$ . Thus, the common feature  $FL^{common}$  of the low-frequency sub-bands of I and SAR images is calculated as follows:

$$FL^{common} = \min\{FL^I, FL^{SAR}\} = \min\{L^I, L^{SAR}\}. \tag{8}$$

Based on Eq. (8), the peculiar features of the LF sub-band of the SAR image are given as follows:

$$FLP^{SAR} = FL^{SAR} - FL^{common} = L^{SAR} - FL^{common}. \tag{9}$$

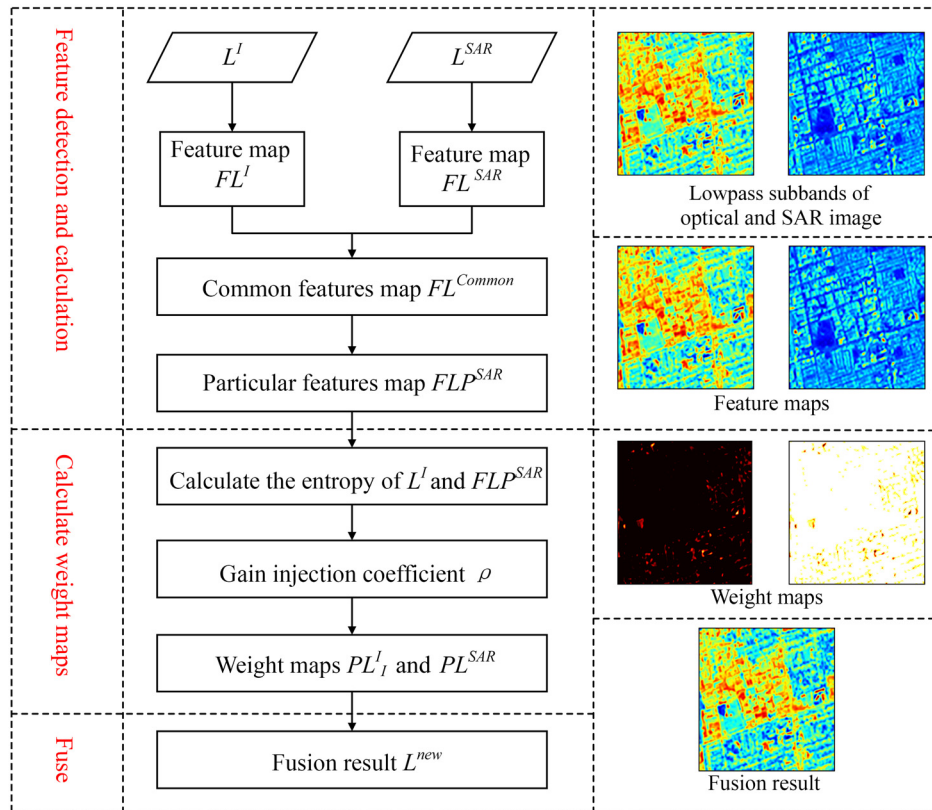


Fig. 3 Fusion process of low-frequency sub-bands.

The method of the fused low-frequency sub-band calculated based on the feature gain injection method is given as

$$L_{\text{new}} = L^I + \rho \times \text{FLP}^{\text{SAR}}, \quad (10)$$

where  $L_{\text{new}}$  is the sub-band of fused low-frequency,  $\rho$  is the injection coefficient of the unique features of the low-frequency sub-band of SAR, calculated as

$$\rho = \frac{\text{entropy}(\text{FLP}^{\text{SAR}})}{\text{entropy}(\text{FLP}^{\text{SAR}}) + \text{entropy}(\text{FLP}^I)}, \quad (11)$$

where  $\text{entropy}(\cdot)$  denotes the entropy of the corresponding image.

### 3.3 Rule for High-frequency Coefficients

The high-frequency sub-bands are multi-directional detailed images of the original image, rich in details and textures. Meanwhile, the high-frequency coefficients are also crucial in determining the degree of spatial distortion of the fused image. Therefore, when fusing high-frequency, we introduce the image divergence, which is sensitive to the points near the texture edge, as the feature activity metric to accurately extract and describe the point features of the high-frequency sub-bands. The divergence of a point in the image precisely describes its degree of clustering in the gradient field. The larger divergence value indicates a greater divergence of the point in the gradient field and a higher probability of the point being a feature point on the edge of the texture.<sup>30</sup> Therefore, using divergence as the active measure in high-frequency fusion can accurately describe the feature saliency of all image elements and thus acquire complete feature maps.

The NSCT method allows us to obtain detailed images of the source in multiple directions at multiple scales. Each detailed image can be considered as a single-channel image in a two-dimensional (2D) cartesian coordinate space, and the divergence of the image is calculated in the gradient field. For a 2D field  $U(x, y)$ , the gradient at  $(x, y)$  is calculated as

$$\text{grad}U(x, y) = \nabla U(x, y) = \left[ \frac{\partial U}{\partial x}, \frac{\partial U}{\partial y} \right]. \quad (12)$$

For a 2D vector field, the divergence of  $V(x, y)$  at  $(x, y)$  is formulated as

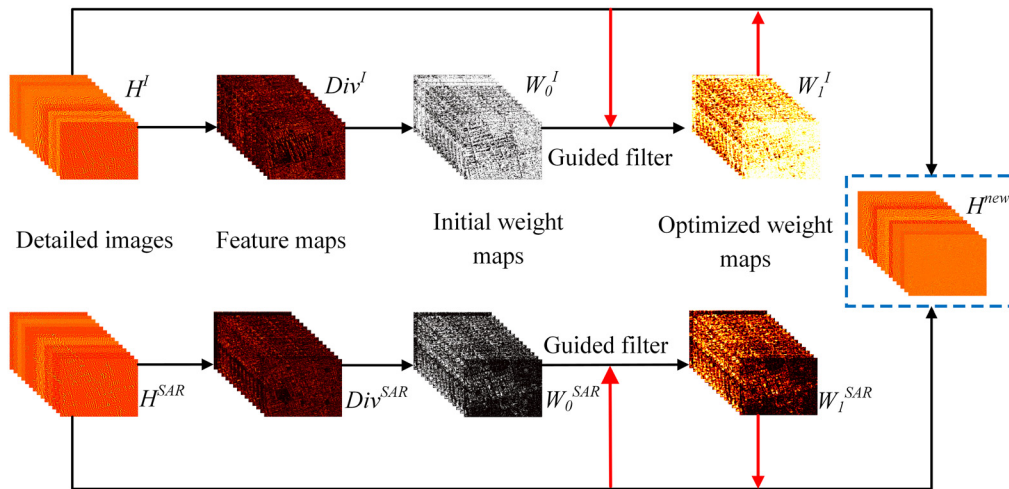
$$\text{div}V(x, y) = \nabla V(x, y) = \frac{\partial V}{\partial x} + \frac{\partial V}{\partial y}. \quad (13)$$

Based on the Eqs. (12) and (13), the divergences of an image are given as

$$\text{div}(U(x, y)) = \nabla(\nabla U(x, y)) = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2}. \quad (14)$$

Since SAR images are seriously polluted by noise, there is still some noise in the speckle-filtered SAR images, which may reduce the fusion quality. Unfortunately, according to the calculation principle of divergence, the image divergence is the second-order image gradient, and the gradient operator is not robust to the noise in the image. Therefore, the divergence is used as the activity measure for the fusion of high-frequency sub-bands to calculate the fusion weights, which is challenging to overcome the influence of noise on fusion quality. To address the above problems, we utilized the guided filter in the fusion process of high-frequency sub-bands, optimized the weight maps obtained from the divergence calculation, and used the strong correlation between pixels to improve the fusion quality of detail images.<sup>31</sup>

The fusion process of high-frequency sub-bands is shown in Fig. 4. First, acquire the high-frequency sub-bands  $\{H^I, H^{\text{SAR}}\}$  of SAR and I separately, and then calculate feature maps  $\{\text{Div}^I, \text{Div}^{\text{SAR}}\}$  based on divergence. Second, initial weight maps  $\{W_0^I, W_0^{\text{SAR}}\}$  are determined with the maximum divergence rule. Third, we use  $\{H^I, H^{\text{SAR}}\}$  as guiding images to optimize



**Fig. 4** Fusion process of high-frequency sub-bands.

initial weight maps  $\{W_0^I, W_0^{SAR}\}$  and acquire the optimized weight maps  $\{W_1^I, W_1^{SAR}\}$ . Finally,  $\{W_1^I, W_1^{SAR}\}$  are used to fuse the detailed images of SAR and I by a weighted average method.

## 4 Experiment

This section introduces the datasets used in the experiments, the indicators for objective evaluation of the fusion results, and the comparative analysis of the experimental results. All experiments were performed using MATLAB2020a on a computer with NVIDIA Quadro P4000 GPU and Intel Xeon W-2102 CPU.

### 4.1 Datasets

We arranged three sets of experiments, and the datasets used in the experiments consisted of three groups of optical and SAR images. The datasets used in the experiments contain three groups of optical and SAR images. In experiment 1, the main scene of the data is farmland, which includes a scene of airborne SAR images and a scene of Google Optical images with sub-meter resolution. In experiment 2, the main scene of the data is the city, which includes a scene of the GaoFen-3 SAR image and a scene of the GaoFen-1 multispectral image with meter-level resolution. In experiment 3, the main scenes of the data are mountains and lakes, including one scene of Sentinel-1 SAR image and one scene of Landsat8 image, with 10-m level resolution. Through three sets of experiments, the proposed algorithm's effectiveness is verified from multi-source, multi-scale, and multi-scene perspectives. It is worth stating that the test images used in the experiment were completely pre-processed. The SAR images are processed in the SARscape toolbox in ENVI, which includes import, multilooking, speckle filtering, geocoding, and radiometric calibration. The specific parameters of the experimental data set are shown in Table 1.

### 4.2 Evaluation Metrics

To evaluate the performance of the fusion methods, root mean square error (RMSE),<sup>32</sup> Erreur relative globale adimensionnelle de synthèse (ERGAS),<sup>33</sup> universal image quality index (UIQI),<sup>34</sup> spectral angle mapper (SAM),<sup>35</sup> and quality with no reference (QNR)<sup>36</sup> are used to evaluate the quality of fusion results quantitatively. Among them, SAM measures the degree of spectral distortion of the fusion result, and the smaller the value, the smaller the spectral distortion. UIQI also called the  $Q$  index, measures the fused image's correlation, luminance, and contrast distortion. Its value ranges from  $[-1, 1]$ , and a higher  $Q$  value indicates higher image quality. RMSE measures the global spectral distortion, and the smaller the value, the

**Table 1** Data information of the experiment.

Data no.	Scene	Source and type	Size (pixel)	GSD (m)
1	Farmland	Airborne (Ka-band with VV polarization)	378 × 404	0.25
		Google (RGB)	192 × 205	0.5
2	City	GF-3 (C-band with HV polarization)	1159 × 1211	3
		GF-1 (R G B NIR)	435 × 454	8
3	Mountain + lake	Sentinel-1 (C-band with HH polarization)	875 × 1576	20
		Landsat8 (SWIR1 NIR R)	583 × 1051	30

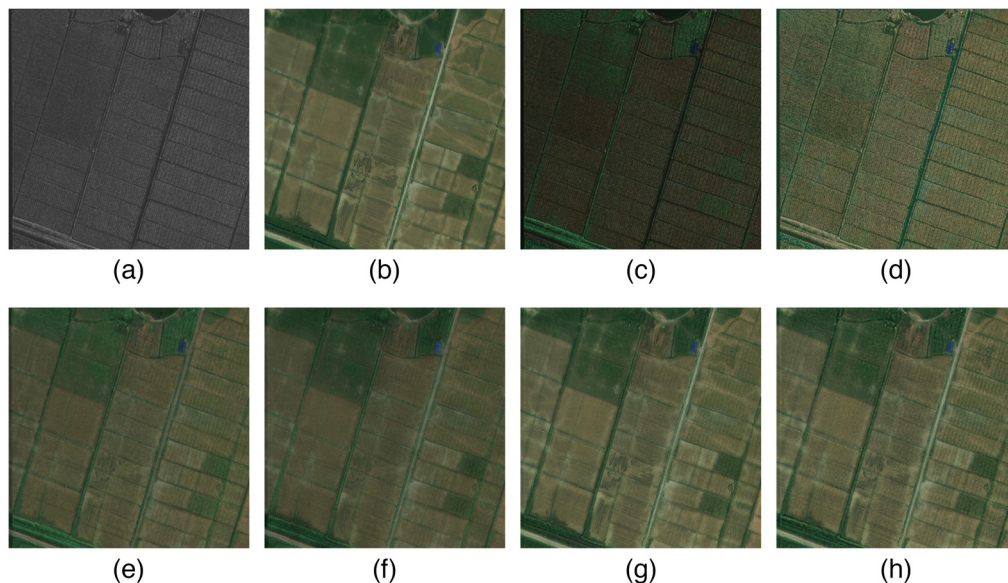
smaller the global spectral distortion. ERGAS can reflect the overall image quality, and the smaller the ERGAS value of the fused image, the higher the fusion quality. QNR is a comprehensive evaluation index that contains two parts: spectral distortion  $D_\lambda$  and spatial distortion  $D_\beta$ . The smaller the value of  $D_\lambda$  and  $D_\beta$ , the smaller the spectral and spatial distortion, while the larger the value of QNR, the higher the quality.

### 4.3 Experimental Results and Comparison

The comparison methods used in the experiments include the IHS<sup>37</sup> and PCA<sup>38</sup> methods that belong to the CS class, the NSCT-PC<sup>39</sup> method that belongs to the MSD class, and the IHS-wavelet<sup>14</sup> and the NSCT-mean<sup>40</sup> method in the hybrid method that couples CS and MSD.

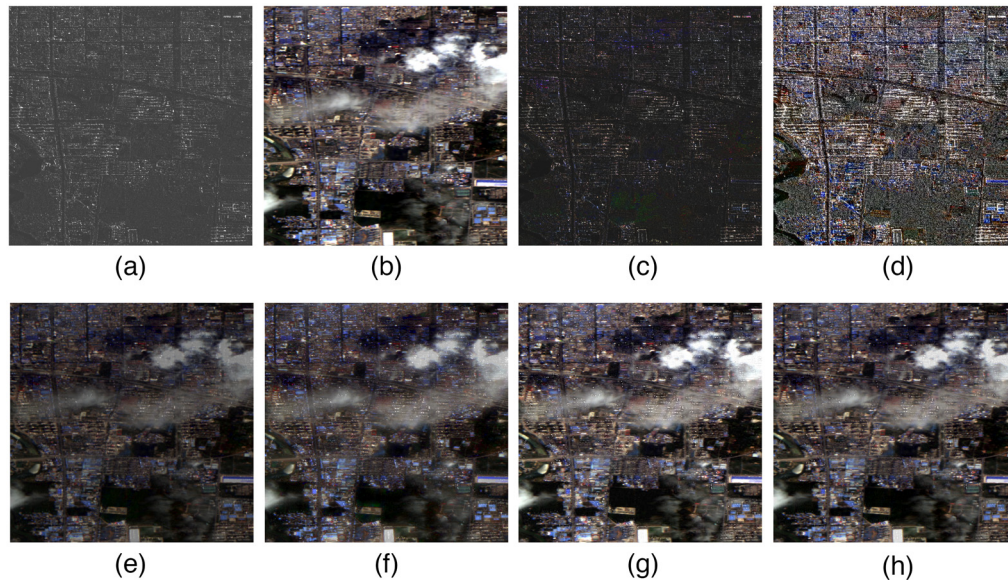
#### 4.3.1 Subjective evaluation

The fusion results of experiment 1 are shown in Fig. 5. The IHS and PCA methods of the CS class can inject the spatial structure information of the SAR image into the optical image more completely. Still, simultaneously, they also cause severe spectral distortion. In contrast, the hybrid methods can effectively reduce the spectral distortion while retaining more spatial information. The IHS-wavelet and the NSCT-mean methods show different degrees of global brightness reduction than the original optical image. NSCT-PC and the proposed method have similar



**Fig. 5** Results of experiment 1. (a) S1A SAR image; (b) Landsat8 optical image; (c) IHS; (d) PCA; (e) IHS-Wavelet; (f) NSCT-mean; (g) NSCT-PC; and (h) ours.





**Fig. 6** Results of experiment 2. (a) S1A SAR image; (b) Landsat8 optical image; (c) IHS; (d) PCA; (e) IHS-wavelet; (f) NSCT-mean; (g) NSCT-PC; and (h) ours.

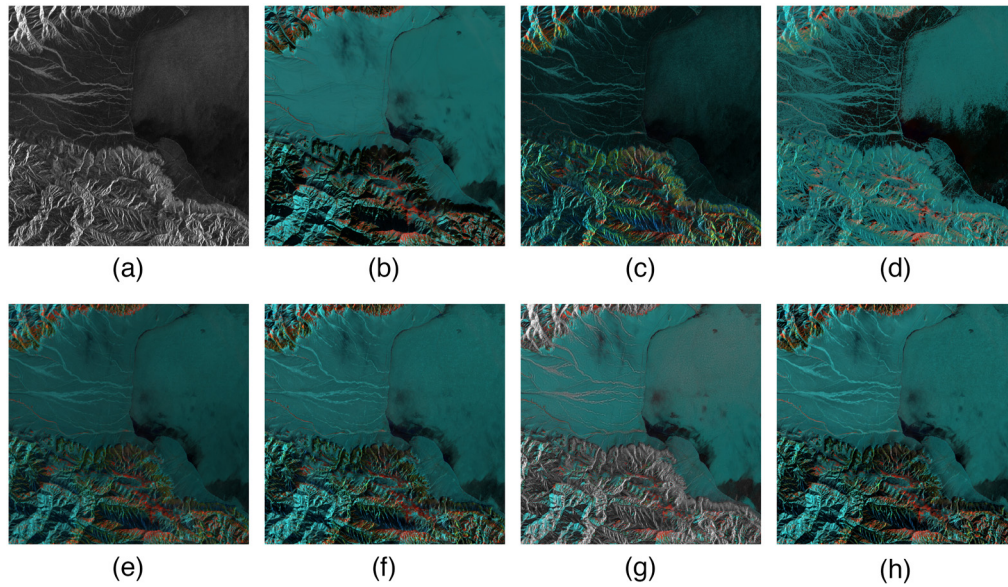
results in spectral retention, while the fused image obtained by the proposed method has more distinct features; therefore, the proposed method has the best fusion performance in experiment 1.

The fusion results of experiment 2 are shown in Fig. 6. The IHS and PCA methods of the CS class can thoroughly remove the clouds when fusing SAR and optical remote sensing images affected by cloud occlusion. Although the direct component replacement method does not need to take into account the information of the optical images and can thoroughly remove the occluded clouds, it severely distorts the spectral information in the fusion results. The principle of the CS method determines this. The direct component replacement method does not need to consider the information of the optical image. It uses the SAR image replacement directly, which can remove the occluded clouds, but it also severely distorts the spectral information in the fusion result. The fusion results of the hybrid methods inject the spatial information of the SAR image into the part occluded by the cloud in the optical image. Such methods cannot thoroughly remove the obscured clouds but retain more spectral information and effectively inject spatial information. Among them, the spectral preservation of the IHS-Wavelet and NSCT-mean methods are relatively low. The spectral protection of the NSCT-PC method and the proposed method achieve similar results. Nonetheless, since the features of NSCT-PC injection are less evident than those of the proposed method, the fused image of the proposed method is of the highest quality in experiment 2 in a comprehensive view.

The fusion results of experiment 3 are shown in Fig. 7. The main scenes of the experimental data are mountains and lakes. By fusing Landsat8 and Sentinel-1 SAR images, the distinct features in the SAR images are injected into the optical images, making fusion images rich with structural and spectral information. In terms of structural feature integrity, the PCA and NSCT-PC methods of injection do not achieve the expected results of the experiment. IHS, IHS-Wavelet, NSCT-mean, and the proposed method are all capable of injecting intact, stereoscopic structural features from SAR images into optical images. Among them, the fusion results of NSCT-mean and the proposed method have similar results, but the mountainous features in the fused image of the NSCT-mean are not as evident as those of the proposed method. Thus, the proposed fusion method performs the best in experiment 3.

#### 4.3.2 Objective evaluation

Tables 2–4 show the evaluation results of the fusion results for the three groups of experiments. As seen from the three tables, compared with the IHS and PCA methods of the CS class, the



**Fig. 7** Results of experiment 3. (a) S1A SAR image; (b) Landsat8 optical image; (c) IHS; (d) PCA; (e) IHS-wavelet; (f) NSCT-mean; and (g) NSCT-PC; and (h) ours.

**Table 2** Results of the quantitative evaluation of the methods in experiment 1.

Methods	SAM	RMSE	ERGAS	$Q$	$D_\lambda$	$D_\beta$	QNR
IHS	0.7045	42.9644	31.5469	0.4285	0.0466	0.9364	0.2462
PCA	0.0065	24.0366	28.3508	0.5028	0.0304	0.6629	0.5717
IHS-wavelet	0.2938	14.6728	10.8125	0.8911	0.0492	0.3992	0.7558
NSCTMean	0.1751	15.7895	11.6339	0.8725	0.0453	0.4543	0.7218
NSCT-PC	0.1141	4.0874	7.1397	0.9676	0.0234	0.2658	0.7170
Ours	0.0527	5.5012	4.1620	0.9380	0.0120	0.3193	0.8201

**Table 3** Results of the quantitative evaluation of the methods in experiment 2.

Methods	SAM	RMSE	ERGAS	$Q$	$D_\lambda$	$D_\beta$	QNR
IHS	0.5301	87.7858	86.239	0.0256	0.0772	0.9535	0.0429
PCA	0.0478	71.5813	119.5852	0.0547	0.1224	0.1404	0.7544
IHS-wavelet	0.1229	44.1552	45.9721	0.6163	0.0522	0.1241	0.8302
NSCT-mean	0.5134	27.1348	39.0338	0.6896	0.0406	0.2014	0.7662
NSCT-PC	0.0308	11.4651	61.9446	0.8518	0.0351	0.1364	0.8332
Ours	0.0252	7.331	28.7343	0.9237	0.0127	0.0867	0.9017

hybrid method can retain more spectral information of the optical images while injecting the spatial information of SAR images into the optical images completely and clearly. Therefore, the hybrid method is more suitable for fusing SAR and optical remote sensing images. SAM, RMSE, ERGAS,  $Q$ , and  $D_\lambda$  can measure the spectral distortion of the fused images. From the index results, the spectral distortion of IHS and PCA methods of the CS class is the most severe, while the spectral distortion of the proposed method is the smallest. The evaluation results of

**Table 4** Results of the quantitative evaluation of the methods in experiment 3.

Methods	SAM	RMSE	ERGAS	$Q$	$D_\lambda$	$D_\beta$	QNR
IHS	3.8838	47.9851	44.2295	0.3351	0.0814	0.7806	0.2016
PCA	0.2925	41.4908	45.3523	0.5117	0.0182	0.4974	0.4935
IHS-wavelet	1.4272	23.9023	22.0357	0.7901	0.0217	0.3854	0.6013
NSCT-mean	0.3672	17.8382	16.7255	0.8501	0.0289	0.418	0.5652
NSCT-PC	8.7935	33.8339	49.6500	0.7344	0.2220	0.5764	0.3296
Ours	0.2519	14.897	15.0998	0.8905	0.0597	0.3599	0.6019

$D_\beta$  show that among the hybrid methods, IHS-wavelet and NSCT-mean have similar spatial information retention. In contrast, the proposed method has the highest spatial information retention. In terms of the comprehensive quality of the fusion results, the evaluation of the QNR showed that the hybrid methods have higher comprehensive qualities of the fused images than the methods of the CS and MSD classes.

## 5 Conclusions

We have made improvements in two aspects to solve the problems that fused images of SAR and optical remote sensing images often have significant spectral and spatial distortion.

1. We use gain injection to achieve fusion in the low-frequency sub-band fusion process. Since the gain injection is performed only at the unique features of the SAR image for feature gain injection, this effectively reduces the spectral distortion of the fused image.
2. We introduced the guide filter when fusing high-frequency sub-bands. The noise reduction features and edge preservation of the guide filter are used to optimize the fusion weight template, effectively reducing the fused image's spatial distortion.

A limitation of the proposed method in this paper is that it is time-consuming. Therefore, reducing the time consumption of the fusion process will be one direction of our next research in the future. Our experiments found that the time consumption is mainly concentrated in NSCT MSD and reconstruction. Therefore, we will try some fast MSD methods, such as fast NSCT,<sup>41</sup> fast finite Shearlet transform,<sup>42</sup> etc. In addition, we plan to improve the fusion quality by using some active metric operators that are robust to noise and nonlinear radiometric differences, such as phase congruency features.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 41761082 and 41861055), the National Key Research and Development Program of China (Grant No. 2017YFB0504201). No potential conflict of interest was reported by the authors.

## References

1. C. Pohl and J. L. Van Genderen, "Review article multisensor image fusion in remote sensing: concepts, methods and applications," *Int. J. Remote Sens.* **19**(5), 823–854 (1998).
2. D. Amarsaikhan et al., "Fusing high-resolution SAR and optical imagery for improved urban land cover study and classification," *Int. J. Image Data Fusion* **1**(1), 83–97 (2010).
3. D. Luo et al., "Fusion of high spatial resolution optical and polarimetric SAR images for urban land cover classification," in *Third Int. Workshop on Earth Observ. and Remote Sens. Appl. (EORSA)*, pp. 362–365 (2014).

4. M. Liu et al., "PCA-based sea-ice image fusion of optical data by HIS transform and SAR data by wavelet transform," *Acta Oceanol. Sin.* **34**(3), 59–67 (2015).
5. S. Sandven, "Sea Ice monitoring in the European Arctic Seas using a multi-sensor approach," in *Remote Sensing of the European Seas*, V. Barale and M. Gade, eds., pp. 487–498, Springer (2008).
6. M. E. J. Cutler et al., "Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: an assessment of predictions between regions," *ISPRS J. Photogramm. Remote Sens.* **70**, 66–77 (2012).
7. Y. Zeng et al., "Image fusion for land cover change detection," *Int. J. Image Data Fusion* **1**(2), 193–215 (2010).
8. M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *IEEE Trans. Image Process.* **21**(4), 2141–2151 (2012).
9. J. Avendano et al., "Flood monitoring and change detection based on unsupervised image segmentation and fusion in multitemporal SAR imagery," in *12th Int. Conf. Electr. Eng., Comput. Sci. and Autom. Control*, pp. 1–6 (2015).
10. A. D'Addabbo et al., "SAR/optical data fusion for flood detection," in *IEEE Int. Geosci. and Remote Sens. Symp.*, pp. 7631–7634 (2016).
11. T. Riedel, C. Thiel, and C. Schmullius, "Fusion of optical and SAR satellite data for improved land cover mapping in agricultural areas," in *Proc. Envisat Symp.* (2007).
12. A. Salentinig and P. Gamba, "Combining SAR-based and multispectral-based extractions to map urban areas at multiple spatial resolutions," *IEEE Geosci. Remote Sens. Mag.* **3**(3), 100–112 (2015).
13. S. C. Kulkarni and P. P. Rege, "Pixel level fusion techniques for SAR and optical images: a review," *Inf. Fusion* **59**, 13–29 (2020).
14. G. Hong, Y. Zhang, and B. Mercer, "A wavelet and IHS integration method to fuse high resolution SAR with moderate resolution multispectral images," *Photogramm. Eng. Remote Sens.* **75**(10), 1213–1223 (2009).
15. N. Han, J. Hu, and W. Zhang, "Multi-spectral and SAR images fusion via Mallat and À trous wavelet transform," in *18th Int. Conf. Geoinf.*, pp. 1–4 (2010).
16. D. Anandhi and S. Valli, "An algorithm for multi-sensor image fusion using maximum a posteriori and nonsubsampling contourlet transform," *Comput. Electr. Eng.* **65**, 139–152 (2018).
17. S. C. Kulkarni, P. P. Rege, and O. Parishwad, "Hybrid fusion approach for synthetic aperture radar and multispectral imagery for improvement in land use land cover classification," *J. Appl. Remote Sens.* **13**(3), 034516 (2019).
18. Z. Shunjie et al., "Fusion algorithm of SAR and visible images for feature recognition," *J. Hefei Univ. Technol.* **41**(7), 900–907 (2018).
19. J. Zhang et al., "Cloud detection in high-resolution remote sensing images using multi-features of ground objects," *J. Geovisual. Sp. Anal.* **3**(2), 1–9 (2019).
20. X. Zhou et al., "A GIHS-based spectral preservation fusion method for remote sensing images using edge restored spectral modulation," *ISPRS J. Photogramm. Remote Sens.* **88**, 16–27 (2014).
21. M. Chikr El-Mezouar et al., "An IHS-based fusion for color distortion reduction and vegetation enhancement in IKONOS imagery," *IEEE Trans. Geosci. Remote Sens.* **49**(5), 1590–1602 (2011).
22. A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampling contourlet transform: theory, design, and applications," *IEEE Trans. Image Process.* **15**(10), 3089–3101 (2006).
23. Z. Wang et al., "Infrared and visible image fusion via hybrid decomposition of NSCT and morphological sequential toggle operator," *Optik* **201**, 163497 (2020).
24. C. Zhao, Y. Guo, and Y. Wang, "A fast fusion scheme for infrared and visible light images in NSCT domain," *Infrared Phys. Technol.* **72**, 266–275 (2015).
25. Z. Zhu et al., "A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain," *IEEE Access* **7**, 20811–20824 (2019).
26. Y. Yang et al., "Remote sensing image fusion based on adaptive IHS and multiscale guided filter," *IEEE Access* **4**, 4573–4582 (2016).

27. Q. Jiahui, L. Yunsong, and D. Wenqian, “Guided filter and principal component analysis hybrid method for hyperspectral pansharpening,” *J. Appl. Remote Sens.* **12**(1), 1–18 (2018).
28. H. Yan et al., “HR optical and SAR image registration using uniform optimized feature and extend phase congruency,” *Int. J. Remote Sens.* **43**(1), 52–74 (2022).
29. X. J. Chong and C. Xuejiao, “Comparative analysis of different fusion rules for SAR and multispectral image fusion based on NSCT and IHS transform,” in *Int. Conf. Comput. and Computational Sci.*, pp. 271–274 (2015).
30. Z. Sheng et al., “Divergence-based multifocuses image fusion,” *J.-Huazhong Univ. Sci. Technol. Nat. Sci. Ed.* **35**(4), 7 (2007).
31. Q. Li et al., “Pansharpening multispectral remote-sensing images with guided filter for monitoring impact of human behavior on environment,” *Concurr. Comput. Pract. Exp.* **33**(4), e5074 (2018).
32. L. He et al., “HyperPNN: hyperspectral pansharpening via spectrally predictive convolutional neural networks,” *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **12**(8), 3092–3100 (2019).
33. L. He et al., “Pansharpening via detail injection based convolutional neural networks,” *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **12**(4), 1188–1204 (2019).
34. D. Li et al., “A universal hypercomplex color image quality index,” in *IEEE Int. Instrum. and Meas. Technol. Conf. Proc.*, pp. 985–990 (2012).
35. L. Sui et al., “Fusion of hyperspectral and multispectral images based on a Bayesian nonparametric approach,” *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **12**(4), 1205–1218 (2019).
36. C. Han et al., “A remote sensing image fusion method based on the analysis sparse model,” *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **9**(1), 439–453 (2016).
37. J. R. Harris, R. Murray, and T. Hirose, “Ihs transform for the integration of radar imagery with geophysical data,” *12th Can. Symp. Remote Sens. Geosci. Remote Sens. Symp.*, pp. 923–926 (1989).
38. İ. Kösesoy et al., “A comparative analysis of image fusion methods,” in *20th Signal Process. and Commun. Appl. Conf.*, pp. 1–4 (2012).
39. G. Bhatnagar, Q. J. Wu, and Z. Liu, “Directive contrast based multimodal medical image fusion in NSCT domain,” *IEEE Trans. Multimedia* **15**(5), 1014–1024 (2013).
40. Y. Wei, Z. Yong, and Y. Zheng, “Fusion of GF-3 SAR and optical images based on the nonsubsampling contourlet transform,” *Acta Opt. Sin.* **38**(11), 1110002 (2018).
41. D. Wang et al., “Optimization of the oil drilling monitoring system based on the multisensor image fusion algorithm,” *J. Sens.* **2021**, 5229073 (2021).
42. L. Tan and X. Yu, “Medical image fusion based on fast finite Shearlet transform and sparse representation,” *Comput. Math. Methods Med.* **2019**, 1–14 (2019).

**Yukai Fu** is currently working toward his MS degree in surveying and mapping from Lanzhou Jiaotong University, Lanzhou, China. His research focuses on remote sensing image processing and analysis.

**Shuwen Yang** received his BS degree from the China University of Geosciences, Wuhan, China, in 1999, and his MS and PhD degrees from the School of Earth Sciences, China University of Geosciences, China, in 2004 and 2011, respectively. Since 2004, he has been with Lanzhou Jiaotong University, where he is currently a professor in the Faculty of Geomatics. His research focuses on image processing and pattern recognition.

**Heng Yan** is currently working toward his MS degree in surveying and mapping from Lanzhou Jiaotong University, Lanzhou, China. His research focuses on image processing and pattern recognition.

**Qing Xue** is currently working toward his MS degree in surveying and mapping from Lanzhou Jiaotong University, Lanzhou, China. His research focuses on remote sensing image processing and analysis.

**Zhuang Shi** is currently working toward his PhD degree in surveying and mapping from Lanzhou Jiaotong University, Lanzhou, China. His research focuses on remote sensing image processing and analysis.

**Xiaoqiang Hu** is currently working toward his MS degree in surveying and mapping from Lanzhou Jiaotong University, Lanzhou, China. His research focuses on remote sensing image processing and analysis.