A Spectral-spatial Jointed Spectral Super-resolution and Its Application to HJ-1A Satellite Images

Xiaolin Han, Huan Zhang, Jing-Hao Xue, and Weidong Sun, Member, IEEE

Abstract—To generate a high-spatial-resolution hyperspectral (HHS) image from a high-spatial-resolution multispectral (HMS) image, both the spatial and spectral information should be considered simultaneously, if we want to build a more accurate mapping from HMS to HHS. To this end, a spectral and spatial jointed spectral super-resolution method is proposed in this paper, using an end-to-end learning strategy for each subspace with the cluster-based multi-branch BPNN. More specifically, in addition to the spectral similarity, a modified superpixel segmentation is introduced to jointly take spatial contextual information into account, and a new framework with it is given. Comparisons on the CAVE dataset shown that, our proposed method outperforms other relative state-of-thet-art methods more than 0.3 in RMSE and more than 1.0 in SAM index. Especially, an exemplary application is demonstrated using the synchronized observation data collected by the multispectral and hyperspectral sensors mounted on the HJ-1A satellite at the same time.

Index Terms—spectral super-resolution, spectral and spatial jointed, subspace-based learning, HJ-1A satellite image.

I. INTRODUCTION

HYPERSONTICAL(HS) remotely sensed satellite images contain hundreds of bands, can be generally applied in environmental monitoring, landcover classification and object recognition using its rich spectral information. Nevertheless, with the increasing of spectral bands, the spatial resolution of HS images will decrease [1]. In contrast, multispectral (MS) images contain fewer bands with limited spectral information but can usually provide more detailed structural information with higher spatial resolution. Luckily, if we have enough corresponding samples between MS and HS images, the lost spectral information can be reconstructed through a spectral mapping established by the end-to-end learning from MS to HS image. It means that, through the end-to-end deep learning with a batch of known image-pairs of high-spatial-resolution multi-spectral (HMS) and high-spatial-resolution hyperspectral (HHS) image of the same scene, a desired spectral mapping from MS to HS can be established, and then a desired corresponding HHS image can be achieved from a given HMS image through spectral super-resolution.

Dictionary learning and neural network learning are two effective ways to achieve spectral super-resolution. Based on dictionary learning, a spectral super-resolution method is proposed by Han et al. [2] utilizing HMS image and the spectral library at the same time. Based on sparse representation, a spectral super-resolution method is proposed by Arad et al. [3], where HHS images can be obtained using a spectral dictionary pre-learned with a batch of RGB-HHS image-pairs and the sparse coefficients estimated from the given high-spatial-resolution RGB image. Besides, an improved version of this method [3] is proposed by Aeschbacher et al. [4], for better performance of the spectral super-resolution. In addition, they also provided a shallow learning method based on A+ [5].

Neural network is another effective way to build the spectral mapping when we have large amount of known image-pairs, such as HMS and HHS images. For natural color images, a nonlinear mapping based method is proposed by Nguyen et al. [6], where HHS images are reconstructed from RGB images using a well-trained radial basis function network. A 56-layer convolutional neural network (CNN) based method is proposed by Galliani et al. [7], where an end-to-end mapping is established by the known RGB and HHS image-pairs. Besides, a CNN based spectral reconstruction method is proposed by Can et al. [8], achieved by shallow learning and residual blocks. And data augmentation [9] is also utilized to increase the number of training samples for improving the mapping performance. In addition, Shi et al. [10] also propose a CNN based method by using dense blocks for a more accurate solution. Martínez et al. [11] proposed a pre-trained CNN method with U-Net architecture, to avoid a large number of training images. In the above CNN-based methods, image patches are used as the learning basis, while the size of patch is $64 \times 64$ [7] or $36 \times 36$ [8]. But, the possible combinations of the above large-sized patches may lead to a large indexing space, which further results in a huge requirement of training samples.

Since the use of the spectral domain as an index space is more effective, we have proposed a spectral super-resolution method [12] utilizing back propagation neural network (BPNN) [13], to establish the nonlinear spectral mapping between spectrum-pairs. Although our previous method proposed the idea of multi-branch BPNN based spectral super-resolution, there are two problems still remaining in this work. Firstly, this previous method is only focused on the natural color images; and secondly, as only the spectral information is considered in this previous method, it suffers from the problem of “different objects with same spectrum” during the spectral mapping.

On the other hand, superpixel segmentation has been widely used in landcover classification [14], spectral unmixing [15], due to its powerful ability to deal with local spatial contextual relationships. Meanwhile, it is also expected that the use of different variants could bring better effects, such as a variant version of simple linear iterative clustering (SLIC) proposed
by Xu et al. for spectral unmixing [16]. Considering that spectral information and spatial contextual information are both significant in spectral super-resolution for the remotely sensed satellite images, and superpixel segmentation may give a solution to the problem of “different objects with the same spectrum” by taking spatial contextual information into account, a modified multi-branch BPNN and superpixel based spectral super-resolution method (termed as MS$^3$) considering spectral and spatial similarity jointly is proposed in this paper. The proposed method effectively reduces the size of the index space, and expands to multispectral domain. The main contributions are shown as follows.

1) A spectral and spatial jointed scheme with superpixel segmentation and spectrum clustering is proposed, to exploit the spectral information and spatial contextual information simultaneously.

2) An exemplary application scenario of our proposed MS$^3$ method and its corresponding data flow are supplied, using synchronized observation data of multispectral and hyperspectral sensors mounted both on the HJ-1A satellite at the same time.

The rest of this paper is organized as follows. The modified MS$^3$ method is presented in Section II. Comparison results on the mechanism level of various related methods are shown in Section III. Section IV demonstrates its actual applications on the HJ-1A satellite, followed by a conclusion in Section V.

II. PROPOSED METHOD

As a spectral degraded version of HHS image $X \in \mathbb{R}^{\lambda \times X \times N}$, an HMS image $Y \in \mathbb{R}^{\lambda \times Y \times N}$ can be expressed as:

$$Y = LX + N_Y,$$

(1)

where $L \in \mathbb{R}^{\lambda \times \lambda_X}$ denotes the spectral response function, $N$ denotes the number of pixels in each band, $\lambda_X$ and $\lambda_Y$ ( $\lambda_X \gg \lambda_Y$ ) represent the number of spectral bands in $X$ and $Y$, respectively. As a common assumption in the multispectral imaging [17], a zero-mean Gaussian noise is added in the above model, represented as $N_Y$.

Solving $X$ from $Y$ in Eq.(1) is a typical ill-posed inverse problem with infinite solutions, since a large amount of spectral information is lost during the MS imaging process. However, the solution of Eq.(1) can become unique, if we put some prior constraints on it using a batch of known sample pairs of $Y$ and its corresponding $X$. Clearly, the inverse of $L$, i.e., the estimation of spectral mapping from $Y$ to $X$, will become easier when the spectrum-pairs of $Y$ and its corresponding $X$ are similar, and furthermore, the smaller the estimation space is, the more accurate the spectral mapping will be. Based on the above analyses, a subspace of the whole estimation space will be formed firstly by grouping the pixels of HMS image into several blocks, according to the similarity of their local textures, then Eq.(1) can be rewritten as:

$$Y^{(j)} = LX^{(j)} + N^{(j)}_Y,$$

(2)

where $Y^{(i)}$ and $X^{(i)}$ denote spectrum-pairs in the $i^{th}$ block, $K_b$ is the total number of blocks. Meanwhile, we should also exploit the information from spectral similarity, and further divide the above blocks into several subspaces according to their spectra similarity, and then rewrite Eq.(1) as:

$$Y^{(j)} = LX^{(j)} + N^{(j)}_Y,$$

(3)

where $Y^{(i)}$ and $X^{(i)}$ denote spectrum-pairs in the $j^{th}$ subspace, and $K_c$ is the total number of subspaces. That is, each subspace $c(j)$ may contain several blocks $b(i)$ according to the spectral correlation of their typical spectra, i.e. $Y^{(j)} = \{Y^{(j)b(i)}\}_{b(i)\in c(j)}$. As can be seen in Eq.(3), the spectral mapping from $Y$ to $X$ can be established by the spectrum-pairs consisting of spectra $Y^{(j)i}$ and the corresponding spectra $X^{(j)i}$ in each subspace. According to Eq.(3), the HMS image $X$ can be reconstructed from a given HMS image $Y$ through the established spectral mapping.

The framework of our MS$^3$ method is depicted in Fig.1, consisting of two stages: training and spectral super-resolution. In the training stage, a modified superpixel segmentation is firstly implemented for HMS image $Y$, grouping its pixels into several blocks $Y^{(i)}$, then the typical spectra $Y^{(i)b(i)}$ in each block is unsupervised clustered into several clusters, and the cluster centers $\mu^c$ are then obtained. The typical spectra in each cluster and the spectra in its blocks form an independent subspace. For subspace $c(j)$, the spectra from HMS images $Y^{(j)i}$ and those from their corresponding HMS images $X^{(j)i}$ are used as a learning basis to construct one sub-branch BPNN, to establish the spectral mapping for this subspace, and all the $K_c$ subspaces together formed a $K_c$-branch BPNN. In the spectral super-resolution stage, the superpixel segmentation is first applied on the given HMS image, and the typical spectra $Y^{(i)b(i)}$ in each block are supervised clustered into $K_c$ subspaces using the cluster centers $\mu^c$ obtained during the training stage. Finally, the spectrum of a given HMS image within subspace $c(j)$ will be putted into the $j^{th}$-branch of the pre-trained multi-branch BPNN model, and be mapped to a reconstructed spectrum $x^{(j)i} \in \mathbb{R}^{\lambda_X}$ in the final desired HHS image. The overall algorithm is summarized in Algorithm 1, where $Z^c = \{Z^{(j)c(i)}\}_{j=1...K_p}$ denotes a set of spectra in $K_p$ blocks/clusters. The details of each part are described as follows.

A. Modified Superpixel Segmentation

Spatial contextual information refers to combining pixels with similar texture characteristics into one block. It can
be regarded as a superpixel segmentation problem that can be solved by SLIC algorithm [18], which is widely used in various applications [14][15]. However, the original SLIC algorithm is proposed for 3-band RGB images, not fit for the multispectral images. Here, a modified SLIC algorithm for the HMS image is given, to utilize the spatial contextual information over the whole bands of the HMS image. In this modified SLIC algorithm, the similarity of spatial contextual information is also employed as a measure of distance, along with the spatial distance between two pixels $p_i$ and $p_j$:

$$D = \sqrt{d_i^2 + \frac{d_j^2}{S} \cdot m^2}, \quad d_c = \sqrt{\sum_{l=1}^{3v} (y_{il} - y_{jl})^2},$$

\hspace{1cm} \text{(4)}$$

where $S = \sqrt{(N/K_i)}$, $y_{il}$ denotes the $l$th band of spectrum $y_i \in \mathbb{R}^{3v}$ at pixel $p_i = (p_{ix}, p_{iy})$, $m \in \{1, 40\}$ [18] balances the distance and contextual (dis)similarity, $d_s$ denotes the spatial distance between pixels $p_i$ and $p_j$, and $d_c$ denotes the spatial contextual (dis)similarity between these two pixels.

B. Subspace-based Multi-branch BPNN

After applying the modified SLIC algorithm, the spectra of pixels in one block of the HMS image are similar. To gather similar spectra in different blocks of the entire image, typical spectra in each block are further divided into $K_c$ subspaces. Here, using the spectral angle as a measure of the similarity, the above dividing process can be described as:

$$\arg \min_{c} \sum_{j=1}^{K_c} \sum_{c} \left( 1 - \frac{y^{(i)}_{c} \cdot \mu^{(j)} }{||y^{(i)}||_2 ||\mu^{(j)}||_2} \right),$$

\hspace{1cm} \text{(5)}$$

where $y^{(i)} \in \mathbb{R}^{3v}$ denotes the typical spectrum in the $i$th block of $Y$, $\mu^{(j)} \in \mathbb{R}^{3v}$ represents the cluster center in $c(j)$ subspace. The above spectra partition process can be treated as a problem of spectrum clustering that can be achieved by the unsupervised clustering method, such as the k-means++ algorithm [19], for its consistent superiority in speed and clustering accuracy. After the above block-based clustering, the spectral vectors in each cluster form a subspace, thereby forming the spectrum-pairs $Y^{(i)}(j)$ and $X^{(i)}(j)$.

To achieve the spectral mapping established in each subspace as described in Eq.(3), one separated sub-branch of the BPNN is used to represent the above nonlinear mapping problem. Here, because BPNN has a strong capability for nonlinear mapping even if there are only a few hidden layers, a three-layer BPNN is used. After training, $K_c$-branch nonlinear spectral mappings from HMS to HHS images for different subspaces are established.

C. Supervised Clustering and Reconstruction

During the spectral super-resolution process, pixels in the given HMS image are firstly grouped into several blocks according to Eq.(4). To obtain similar spectra in the entire image, typical spectrum $y^{(i)}$ in each block of the given HMS image are supervised clustered into $K_c$ subspaces utilizing the clustering centers $\mu^{c}$ obtained during the training stage.

<table>
<thead>
<tr>
<th>Table I: Comparison on the CAVE dataset with different methods.</th>
</tr>
</thead>
<tbody>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
<tr>
<td>rRMSE</td>
</tr>
<tr>
<td>rRMSE$^{CIE}$</td>
</tr>
<tr>
<td>rRMSE$^{CIE-2}$</td>
</tr>
<tr>
<td>rRMSE$^{CIE-4}$</td>
</tr>
<tr>
<td>SAM</td>
</tr>
</tbody>
</table>

This can be achieved by the following supervised clustering method:

$$\arg \min_{c} \sum_{j=1}^{K_c} \left( 1 - \frac{y^{(i)}_{c} \cdot \mu^{(j)} }{||y^{(i)}||_2 ||\mu^{(j)}||_2} \right),$$

\hspace{1cm} \text{(6)}$$

After the supervised clustering, spectra $Y^{(j)}(j) \in \mathbb{R}^{3v}$ of the given HMS image within subspace $c(j)$ ($j = 1,...,K_r$) can be obtained. Finally, the spectra $Y^{(j)}(j)$ will be putted into the $j$th-branch of the pre-trained multi-branch BPNN model, and be mapped to a reconstructed spectrum $X^{(j)}$ in the final desired HHS image.

III. Comparisons Between Different Relative Methods

To validate the performance of our proposed MS$^3$ method$^1$, relative state-of-the-art methods proposed by Galliani et al. [7], Arad et al. [3], Aeschbacher et al. [4], Can et al. [8] and Han et al. [12] are utilized as comparisons on the CAVE dataset. Following the experimental settings in the above methods, HMS images are generated from the corresponding HHS images using CIE 1964 spectral response functions, with a dimension of $512 \times 512 \times 3$. In this part of experiment, the averaged value of spectra in each block is used as the typical spectrum, the number of blocks is set to 10000, and the number of clusters is set to 160, which is the same as the method in [12]. As for our proposed MS$^3$ method, instead of dividing the images into two sets [4], we divide the spectra into two sets as a 2-fold cross-validation, and the numbers of training and test spectra are the same with those of the method proposed by Aeschbacher et al. [4]. For a fair comparison, in our proposed MS$^3$ method, the selection and experimental settings on the numbers of training and testing spectra are the same as those in methods [12] and [4]. Particularly, a 3-10-31 BPNN is used to achieve the non-linear spectral mapping in each sub-branch, which is the same with Han et al. [12]. In addition, during the training stage, BPNNs is trained with 85% of the training samples, and the remaining 15% samples are used for validation to mitigate overfitting. The mean squared error between the input spectrum and the mapped spectrum is used as the loss function; the activation functions in the hidden and output layers are set to sigmoid and linear functions, respectively; and the maximum number of epochs during training is up to 100. For a like-for-like comparison, spectral angle mapper (SAM) [17]; absolute RMSE, RMSE$_G$ and RMSE$_{G-2}$, relative rRMSE, rRMSE$_G$ and rRMSE$_{G-2}$, defined in [3] [7] and [8] are used as full-reference evaluation indexes to evaluate the spectral super-resolution performance.

$^1$https://github.com/xiaolin-han/MS3-method
Table I quantitatively shows the comparisons of different methods on the CAVES dataset. In Table I, the spectral super-resolution results of the other relative methods are obtained according to their original papers, with short line indicating no relevant information available, bold and underline indicating the best and second best results, respectively. It can be seen that our proposed MS method outperforms the other relative methods in both spatial and spectral domains, especially in terms of the indexes of RMSE and SAM. The RMSE index is decreased more than 0.3 compared with the CBPNN method [12], and more than 1.4 compared with the method of [8], and even more than 4.4 compared with the method of [3]. Besides, our proposed method is significantly improved in the indexes of other RMSE variants, where RMSE$_{2,3}$ is improved even more than 16.6. The above performance indicates that our proposed MS method performs better in the spatial domain. Furthermore, our proposed MS method improves over 6.1 and 1.0 in SAM index compared with the methods of [7] and [12], respectively, which is much significant for the HHS images.

IV. APPLICATIONS ON THE HJ-1A SATELLITE DATASET

A. HJ-1A Satellite and Its Dataset

For the environment and disasters monitoring, a small optical satellite named “HJ-1A” was launched by China on September 6, 2008, and put into the service in March 2009 [21], and a follow-up satellite was also launched on September 27, 2020. One multispectral sensor and one hyperspectral sensor are both mounted on the HJ-1A satellite, and a synchronized observation dataset obtained on March 23, 2019, in Shandong Province of China [22], are used in this part of experiment. Fig.2(a) shows the location and coverage of the original multispectral and hyperspectral images. The original multi-spectral image was obtained by a CCD camera with near infrared and visible spectral bands, at a spatial resolution of 30 meters with 720 km width, in a dimension of 13,804×15,222×4, shown in Fig.2(b); the original hyperspectral image was captured by a hyperspectral camera with 115 bands ranging from 450nm to 950nm, at a spatial resolution of 100 meters with 50 km width. Taking serious noise affects into account, four pieces of the original hyperspectral image with a dimension of 589×583×95 are used as the HHS images, shown in Fig.2(c).

B. Spectral Super-Resolution for the HJ-1A Satellite

From the coverage of two sensors shown in Fig.2(a), we can see that, if we take the multispectral image as the HMS image and the hyperspectral image as the HHS image, it will have a clear practical significance to generate an HHS image covering the whole observation area of the HMS image. Fig.3 shows the data flow of our proposed MS method in the spectral super-resolution of HJ-1A satellite data. As can be seen from the data flow along black arrows in Fig.3, after matching and cropping, the overlapped parts of HMS and HHS images are obtained, and will be used to establish the spectral mapping between them through our proposed MS method. In the spectral super-resolution stage, when the HMS image is used as an input image, the final HHS image covering the whole observation area of the HMS image can be obtained, as shown along the red arrows in Fig.3.

Specifically, the matching of HMS and HHS images is an important part in the spectral super-resolution of HJ-1A satellite data. Considering that only translation, rotation and scaling exist between the HHS and HMS images, a spatial matching method for this kind of nonreflective similarity transformation [23] is used here, to match and crop the overlapped parts of the HMS image and the HHS images. The four pieces of HMS and HHS image-pairs used in this part of experiment, in which the spectra from #38, #40 and #53 image-pairs are randomly selected as the training spectrum-pairs and the last one is used for full-reference evaluation.

C. Experimental Results on the HJ-1A Satellite Dataset

In this part of experiments, our proposed method has 70,000 blocks and 20 clusters. A 4-10-95 BPNN is used for each sub-branch to establish the corresponding spectral mapping. The rest of experimental settings are the same with Section III.

The first two lines of Table II show the comparison results on the overlapping part over the land area. It shows that without any denoising process, our proposed MS method has a better performance than that of CBPNN. Specifically, SAM is reduced over 0.01, and MSE is reduced over 0.37. Moreover, by using our proposed method and the data flow given in Fig.3, the reconstructed HHS image covering the whole observation area of the HMS image is shown in Fig.4. The slight difference in false color image is due to the incomplete correspondence between the bands of the two images. Compared with the false color HMS image in Fig.4(a), Fig.4(b) shows a high degree of similarity in the spectral structure and spectra indicating a good spectral super-resolution performance qualitatively. Most of all, it can give a whole view observation of hyperspectral image with high-spatial-resolution.

D. Discussion on Parameter Selection

To analyze the effects of the proposed MS method on key parameters, Fig.5 plots the MSE and SAM curves as
function of $K_5$ and $K_c$ on the HJ-1A dataset #39, respectively. As shown in Fig.5(a), our proposed method has a better performance when $K_5$ is around 70,000 than using more blocks. As can be seen from Fig.5(b) that, as $K_c$ increases, the SAM curve shows a downward trend. It indicates that the proposed method has a better performance in spectra reconstruction with more clusters. However, the MSE curve shows a slight worsening trend when $K_c$ is bigger than 20. This may be due to the fact that, as the number of clusters increases, the number of training samples in each cluster will decrease, which can lead to a worse performance in each sub-branch BPNN, thus further affecting the quality of the final HHS image.

E. Experimental Results on the other HJ-1A Satellite Dataset

To demonstrate the performance of our proposed method on different landcover types, a new HJ-1A dataset covering the sea area shown in Fig.6(a) is also used, which was obtained on Sep. 30, 2019 in Hainan Province of China.

Under the same experimental settings as Section IV-C, the last two lines of Table II show the comparison results over the sea area. Besides, Fig.6(c) shows the overall reconstruction result of the HMS image. The conclusions of these two HJ-1A datasets are consistent with that of the CAVE dataset, and the accuracy of spectral super-resolution between land and sea areas are basically the same.

V. Conclusion

In this paper, we proposed a spectral-spatial jointed spectral super-resolution model to generate a desired HHS image from one given HMS image. Specifically, a modified superpixel segmentation algorithm and then an unsupervised clustering on typical spectra are applied to the HMS image, to exploit both the spatial and spectral similarity jointly, and a new framework with it is given. One separated sub-branch is trained to establish the corresponding spectral mapping for each subspace, and different sub-branches formed a multi-branch BPNN. Utilizing the pre-trained multi-branch BPNN, the desired HHS image can be obtained from a given HMS image. Experimental results on the CAVE dataset demonstrated that our proposed MS$^3$ method outperformed the other relative methods both in the spatial and spectral domains. As an exemplary application of the spectral super-resolution, MS$^3$ method has also been extensively studied on the two HJ-1A datasets.

VI. REFERENCES

[23] https://www2.mathworks.cn/help/images/ref/cpselect.html