



Xiaoxiao Feng¹, Luxiao He¹, Qimin Cheng^{2,*}, Xiaoyi Long³ and Yuxin Yuan⁴

- ¹ State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China; fengxx2018@whu.edu.cn (X.F.); heluxiao@whu.edu.cn (L.H.)
- ² School of Electronics Information and Communications, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan 430074, China
- ³ School of Computer Science, Wuhan University, Wuhan 430072, China; 2018302100026@whu.edu.cn
- ⁴ Hongyi Honor College, Wuhan University, Wuhan 430072, China; yuanyuxin@whu.edu.cn
- * Correspondence: chengqm@hust.edu.cn

Received: 12 February 2020; Accepted: 15 March 2020; Published: 21 March 2020



Abstract: Hyperspectral (HS) images usually have high spectral resolution and low spatial resolution (LSR). However, multispectral (MS) images have high spatial resolution (HSR) and low spectral resolution. HS–MS image fusion technology can combine both advantages, which is beneficial for accurate feature classification. Nevertheless, heterogeneous sensors always have temporal differences between LSR-HS and HSR-MS images in the real cases, which means that the classical fusion methods cannot get effective results. For this problem, we present a fusion method via spectral unmixing and image mask. Considering the difference between the two images, we firstly extracted the endmembers and their corresponding positions from the invariant regions of LSR-HS images. Then we can get the endmembers of HSR-MS images based on the theory that HSR-MS images and LSR-HS images are the spectral and spatial degradation from HSR-HS images, respectively. The fusion image is obtained by two result matrices. Series experimental results on simulated and real datasets substantiated the effectiveness of our method both quantitatively and visually.

Keywords: hyperspectral image; multispectral image; remote sensing; temporal difference; spectral unmixing; endmember spatial information

1. Introduction

Hyperspectral (HS) images are playing important roles in agriculture, medical science, remote sensing, and other fields because of its high spectral resolution [1]. The spectrum width of hyperspectral sensor is narrow, resulting in insufficient energy. In order to maintain a certain signal-to-noise ratio (SNR), the spatial resolution has to be sacrificed. Therefore, multispectral (MS) images often have higher spatial resolution than HS images. High spatial resolution (HSR) information helps to obtain more accurate locations and shapes of ground objects [2], and spectral information helps to identify types of features from images [3]. Spectral information fusion of HS images and MS images is helpful for image retrieval [4], classification [5,6], image analysis [7,8] and information extraction [2,9]. Fusing low spatial resolution HS (LSR-HS) images and HSR-MS images can obtain HSR-HS images [10], which is very helpful for fine mapping, ground objects identification, and so on.



Pansharpening is a special case of HS–MS images fusion problem [11], which enhances the spatial resolution of LSR-MS image by fusing it with the panchromatic image [12]. With the development of hyperspectral imaging technology, these methods are gradually applied to hyperspectral images [13]. For instance, component substitution (CS)-based methods convert the LSR-HS image to another orthogonal space and replace the decomposed spatial components with one of the multispectral bands [14,15], including Gram–Schmidt adaptive (GSA) [16]. Another representative pansharpening method is multi-resolution analysis (MRA), which utilizes the multiresolution decomposition way to obtain the spatial information of panchromatic image and combine it into the resampled HS bands. The well-known methods include generalized Laplacian pyramid (GLP) [17].

Another kind of HS–MS fusion method is to take HSR-MS image and LSR-HS image as degradation results of HSR-HS image in spectrum and space, respectively, and then inversing two images to reconstruct the target image. This category method mainly includes the Bayesian-based method and spectral unmixing-based method. The Bayesian-based method uses the posterior probability distribution of the target image for image fusion given LSR-HS and HSR-MS images [18–20]. For instance, Eismann et al. proposed a Bayesian-based method under the maximum a posteriori (MAP) estimation, which estimated the underlying spectral characteristics of the ground objects by a stochastic mixing model (SMM) and optimized the estimated HS images relative to the input images by a cost function [21].

The spectral unmixing-based fusion method has been widely studied because of its physically straightforward representation of the fusion process. Each pixel is a mixture of some endmember spectra and corresponding fractional abundances [22]. The approach based on spectral unmixing is obtaining endmember matrix and abundance matrix from LSR-HS image and HSR-MS image using spectral mixture model, respectively. Then, two resulting matrices are reconstructed as the HSR-HS image. For instance, coupled nonnegative matrix factorization (CNMF) [23] alternately using NMF unmixing the LSR-HS image and HSR-MS image to get the endmember and abundance matrices, respectively. A method proposed by Akhtar et al. [24] is to extract the endmember spectra as a spectral basis first, and then with the basis fixed use a pursuit method reconstruct the image. The method proposed by Lanaras et al. [25] is to obtain the endmembers by alternately updating method and use the projection gradient method to get abundance matrix of HSR-MS image.

Most HS–MS fusion methods require a high correlation between LSR-HS images and HSR-MS images. The correlation between the two images is affected by various factors such as sensor characteristics, temporal difference, weather, imaging angle, etc., and the real data is often difficult to meet the requirements [25]. To solve the problem, we propose a new HS–MS fusion method based on linear spectral mixture model (LSMM), which investigates the endmember signatures' spatial relationship between LSR-HS image and HSR-MS image. First, in order to ensure the consistency of the endmembers positions in two images, the mask is generated according to the differences between the two images. Then, the proposed method eliminates the variation area by masking the large difference regions. After that, we extract the endmembers and their corresponding positions from masked LSR-HS image and locate them according to the endmember spatial information in HSR-MS image. Next, based on the obtained endmember matrix, the abundance matrix is calculated by nonnegative linear regression. Finally, the HSR-HS image is reconstructed by endmember matrix of LSR-HS image and abundance matrix of HSR-MS image. Our fusion method is physical clear and has very simple model with less calculation, and it is robust to the image temporal difference. Experimental results based on the simulated images and real images indicate our method can effectively integrate the spatial and spectral information, and outperform than other compared methods.

2. Methodology

2.1. The Proposed Method Overview

LSMM is widely used to represent the observation images because of its effectiveness and straightforward [22]. The LSMM considers the ground to be flat and assumes that incoming sunlight radiation is reflected off the surface only in a single reflection, not multiple scattering [26]. Each pixel in the scene can be modeled as a linear combination of all the spectra of the materials and the proportional contribution of each spectra. Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \Re^{L \times N}$ denotes the observation image, *L* is the number of bands, *N* is the number of all pixels. For the *j*-th pixel, $\mathbf{x}_j = [x_{1j}, \dots, x_{Lj}]^T \in \Re^L$, the LSMM models the pixel as:

$$\mathbf{x}_j = \sum_{i=1}^P s_{i,j} \mathbf{a}_i + \mathbf{n}_j = \mathbf{A} \mathbf{s}_j + \mathbf{n}_j$$
(1)

where $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_P] \in \Re^{L \times P}$ is the mixing matrix that consists of *P* distinct endmember spectral signiatures, *P* is the number of endmember signatures in the scene, $\mathbf{s}_j = [s_{1j}, \dots, s_{Pj}]^T \in \Re^P$ is the abundance vector of \mathbf{x}_j , and $\mathbf{n}_j \in \Re^L$ is the additive noise term. Because of the physical constraints, the abundance is nonnegtive $s_{ij} \ge 0$ and sum-to-one $\sum_{i=1}^{P} s_{ij} = 1$. Correspondingly, the image has *N* pixels can be modeled as:

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N} \tag{2}$$

Based on LSMM, the LSR-HS image X_h and HSR-MS image X_m can be modeled as:

$$\begin{cases} \mathbf{X}_m = \mathbf{A}_m \mathbf{S}_m + \mathbf{N}_m \\ \mathbf{X}_h = \mathbf{A}_h \mathbf{S}_h + \mathbf{N}_h \end{cases}$$
(3)

where $\mathbf{A}_m \in \Re^{L_m \times P}$ is the endmember matrix of HSR-MS image \mathbf{X}_m , L_m is the band number, P is the endmember number. $\mathbf{S}_m \in \Re^{P \times N_m}$ denotes the abundance matrix, N_m denotes the number of pixels. $\mathbf{N}_m \in \Re^{L_m \times N_m}$ denotes the additive noise term. For LSR-HS image \mathbf{X}_h , $\mathbf{A}_h \in \Re^{L_h \times P}$ is the endmember matrix, and L_h is the band number. $\mathbf{S}_h \in \Re^{P \times N_h}$ is the abundance matrix of \mathbf{X}_h , $\mathbf{A}_h \in \Re^{L_h \times P}$ is the endmember of pixels. $\mathbf{N}_h \in \Re^{L_h \times N_h}$ is the noise term of LSR-HS image. If we upsample the LSR-HS image to the same size as the HSR-MS image, the number of pixels will be the same $N_m = N_h = N$. Based on the physical meaning of abundance matrix and endmember matrix, all the matrices in Equation (3) are nonnegative.

Under the assumption that HSR-MS image and LSR-HS image is the spectral and spatial degradation from HSR-HS image, respectively. We assume that A_m and A_h denote the same endmember signatures, and the difference between them is only in the spectral resolution. In this way, HSR-HS image Z can be expressed as:

$$\mathbf{Z} \approx \mathbf{A}_h \mathbf{S}_m \tag{4}$$

where $\mathbf{Z} \in \Re^{L_h \times N}$ has *N* pixels with L_h bands. \mathbf{A}_h and \mathbf{S}_m provides the spectral information and spatial information for \mathbf{Z} , respectively.

From Equation (4), we can see that the key of HS–MS fusion based on spectral unmixing is how to ensure A_m and A_h are the same endmembers. Many methods assume that HSR-MS image and LSR-HS image are obtained from spatial and spectral degradation of HSR-HS image, respectively [27]. For this reason, we propose a HS–MS fusion method based on endmember spatial information and regional mask. Figure 1 shows the overall flowchart of the proposed method. This method consists of three main parts: regional mask, LSR-HS image unmixing and HSR-MS image unmixing.

Since it is difficult to obtain the simultaneous LSR-HS image and HSR-MS image of the same area in the real case, our method proposes a regional mask prepossessing based on the ground objects variation caused by temporal difference. The method masks the areas with large differences between two images and extracts the endmembers from areas with small changes. In this way, endmembers of the fusion part from two images are consistent. This part will be described in detail in Section 2.2.

In this paper, the proposed fusion method is based on the spectral unmixing model, and the most classical and practical endmember extraction method Vertex Component Analysis (VCA) [28] is used to obtain the LSR-HS image endmember signatures. Based on the convex geometry theory, the endmember spectra and corresponding positions of the pure pixels can be obtained by VCA. This part will be described in detail in Section 2.3.

Based on the principle of one-to-one match between LSR-HS image and HSR-MS image endmember positions, we can obtain the pure pixels' positions in HSR-MS image from LSR-HS image endmember positions. Since LSR-HS image can be regarded as the spatial downsampling result of HSR-MS image, the pixels' mean value of the region is selected as the HSR-MS endmember results. We take the endmembers' pixels positions of LSR-HS image as the center of the region and the resolution ratio of the two images as the diameter. Then, according to LSMM model and HSR-MS endmember matrix, the corresponding abundance matrix is obtained by using nonnegative constrained least squares (NCLS) [29]. This part will be described in detail in Section 2.4.

After obtaining \mathbf{A}_h and \mathbf{S}_m from LSR-HS image \mathbf{X}_h and HSR-MS image \mathbf{X}_m , respectively, we reconstruct the HSR-HS image $\hat{\mathbf{Z}}$ according to Equation (4).



Figure 1. Overall flowchart of the proposed hyperspectral–multispectral (HS–MS) fusion method (Image in the flowchart has six endmembers).

2.2. Regional Mask of the LSR-HS Image and HSR-HS Image Variation Area

Under the premise of accurate registration, the image position and actual position of the two images should correspond to each other. Due to the temporal difference, two images must have some variation regions. Therefore, this paper uses the regional mask caused by temporal difference to preprocess the image, which can eliminate the difference between LSR-HS and HSR-MS images.

Firstly, according to the band wavelength of Landsat8 (Blue: 450–515 nm, Green: 525–600 nm, Red: 630–680 nm, Near infrared (NIR): 845–885 nm), select the bands in the corresponding wavelength range

in \mathbf{X}_h , and obtain four matrices as the spectral patches, Blue band patches: $\mathbf{X}_h^B = [\mathbf{X}_h^{B1}, \dots, \mathbf{X}_h^{Bn}]$, Green band patches: $\mathbf{X}_h^G = [\mathbf{X}_h^{G1}, \dots, \mathbf{X}_h^{Gn}]$, Red band patches: $\mathbf{X}_h^R = [\mathbf{X}_h^{R1}, \dots, \mathbf{X}_h^{Rn}]$ and NIR band patches: $\mathbf{X}_h^{NIR} = [\mathbf{X}_h^{NIR1}, \dots, \mathbf{X}_h^{NIRn}]$. We integrate four spectral patches and simulate a new LSR-MS image $\mathbf{X}'_h = [\mathbf{X}_h^B, \mathbf{X}_h^G, \mathbf{X}_h^R, \mathbf{X}_h^{NIR}] \in \Re^{L_m \times N}$ with four bands by averaging $\mathbf{X}_h^B, \mathbf{X}_h^G, \mathbf{X}_h^R, \mathbf{X}_h^{NIR}$, respectively.

Then, HSR-MS image X_m and simulated LSR-MS image X'_h are normalized separately to eliminate the spectral difference. Since the near infrared band is sensitive to vegetation information, we make a subtraction between of the two images on the near infrared band to obtain regions with variation. We assume that if the results larger than *thresh*1, the features in this region vary greatly, so we take this area as the mask region, defined as Equation (5).

$$\mathbf{M}_{(i,j)} = \begin{cases} 1, if \left(\mathbf{X}_{(i,j)} - \mathbf{X'}_{(i,j)} \le thresh1 \right) \\ 0, else \end{cases}$$
(5)

$$\mathbf{X}_{h}^{mask} = \mathbf{X}_{h} \cdot \mathbf{M} \tag{6}$$

$$\mathbf{X}_{m}^{mask} = \mathbf{X}_{m} \cdot \mathbf{M} \tag{7}$$

After determine the mask matrix **M**, we can get masked HSR-HS image X_h^{mask} and LSR-MS image X_m^{mask} images via Equations (6) and (7), respectively. In X_h^{mask} and X_m^{mask} , we assume that the ground objects are consistent in species and positions. In this paper, we focus on the spectrum of vegetation, rather than specific plants. So, we use mask without considering the vegetation types, or seasonal changes.

2.3. LSR-HS Image Endmember Extracting

Hyperspectral images can be modeled as a collection of spectral signatures, called endmembers. We use the commonly used VCA [28] to extract endmembers, becuase its accuracy is higher and it is easy to implement.

Generally, the HSIs unmixing problem based on LSMM is solved by using the square Frobenius norm of residual matrix as the cost function [30]. We use Equation (8) to extract endmembers from the LSR-HS image.

$$\min_{\mathbf{A}_{h}, \mathbf{S}_{h}} \left\| \mathbf{X}_{h}^{mask} - \mathbf{A}_{h} \mathbf{S}_{h} \right\|_{F}^{2} \tag{8}$$
subject to $\mathbf{A}_{h} \ge 0, \mathbf{S}_{h} \ge 0$

where $\|\cdot\|_F^2$ indicates the Frobenius norm. Because of the physical constraints of endmember matrix and abundance matrix, both of them should be non-negative. According to the convex geometry characteristic of HSIs, hyperspectral points cloud will be located in the simplex after affine transformation. The mixed pixels are locate interior the simplex and the endmembers are locate in the vertex of the simplex. VCA algorithm utilizes the geometric characteristic of HSIs and extracts endmembers by finding the pure pixels locate at the vertices of the simplex, which maximize the volume of the simplex and include all pixels. In this paper, we obtain the endmembers \mathbf{A}_h and corresponding positions \mathbf{P}_h from the masked LSR-HS image \mathbf{X}_h^{mask} by VCA. All the endmembers are used to model the pixels in the dataset.

2.4. HSR-MS Image Unmixing

LSR-HS image can be regarded as the spatial downsampling result of LSR-HS image, in this way, the locations of pure pixels in LSR-HS image, corresponding to a region in HSR-MS image [31]. In terms of spectral information, LSR-HS image has a higher spectral resolution and a larger number of bands,

HSR-MS image can be regarded as the spectral downsampling result of LSR-HS image [27]. Therefore, we assume that the pixels of this region in HSR-MS image constitutes the corresponding pure pixels in LSR-HS image.

In this paper, the spectral information of LSR-HS image and spatial information of HSR-MS image are combined for HSR-HS image fusion. According to endmember positions \mathbf{P}_h in LSR-HS image, we can get the pixel position of each endmember in HSR-MS image (as shown in Figure 2). We take this position as the center point, *scale* as the diameter to get the corresponding endmember region \mathbf{P}_m . The size of each endmember region in HSR-MS image is *scale* × *scale*. Then, by averaging the pixels in the corresponding region, the result is taken as the endmember of the HSR-MS image.



Figure 2. Pure pixel in low spatial resolution (LSR)-HS image and high spatial resolution (HSR)-MS image.

However, in the real case, each endmember corresponding region in HSR-MS image may have more than one kind of ground objects. To solve this problem, we use spectral angle distance d to eliminate the different pixels in the region, which defined in Equation (9).

$$d = \cos^{-1}\left(\frac{\mathbf{x}_{m}(i1,j1)\,\mathbf{x}_{m}(i2,j2)}{\|\mathbf{x}_{m}(i1,j1)\|\,\|\mathbf{x}_{m}(i2,j2)\|}\right)$$
(9)

where $\mathbf{x}_m(i1, j1)$ and $\mathbf{x}_m(i2, j2)$ denote the pixels in the region. We calculate the distance between each pixel in the region, select the pixels with closer distance value and take their mean value as the endmember \mathbf{A}_m of HSR-MS image. In addition, in order to avoid the second type of objects in the region, we sort them according to the distance between pixels, and select the first 90% as the candidate pixels.

After obtaining the endmember matrix \mathbf{A}_m of HSR-MS image, the corresponding abundance matrix \mathbf{S}_m can be calculated by Equation (10).

$$\min_{\mathbf{S}_m} \|\mathbf{X}_m - \mathbf{A}_m \mathbf{S}_m\|_F^2
\text{subject to } \mathbf{S}_m \ge 0$$
(10)

According to Equation (10), we obtain the abundance matrix S_m of HSR-MS image by nonnegative constrained least-squares (NCLS) algorithm.

3. Experiments and Results

In order to test the effectiveness of our method, we design the experiments by Wald's protocol [13]. Then, the proposed method and other six fusion methods, including CNMF [23], fast fusion based on Sylvester equation (FUSE) [18], GLP [17], GSA [16], HS Superresolution (HySure) [20] and MAPSMM [21] are used to fuse and compare the results (The comparative methods' codes are download from

http://naotoyokoya.com/Download.html). Here, **Z** denotes the HSR-HS image as the reference image, and $\hat{\mathbf{Z}}$ denotes the fused image as the estimated HSR-HS image. Some popular metrics are used to evaluate the fused results, Peak-SNR, Spectral Angle Mapper (SAM) [32], Erreur Relative Globale Adimensionnelle de synthèse (ERGAS) [33] and Q2n.

3.1. Performance Evaluation Metrics

1. PSNR

PSNR is for measuring the spatial quality of each band. It is the ratio between the maximum pixel value and the mean square error of the reconstructed image in each band. The PSNR of the *i*-th band is defined as:

$$\operatorname{PSNR}\left(\mathbf{z}_{i}, \hat{\mathbf{z}}_{i}\right) = 10 \cdot \log_{10}\left(\frac{\max\left(\mathbf{z}_{i}\right)^{2}}{\|\mathbf{z}_{i} - \hat{\mathbf{z}}_{i}\|_{2}^{2}/N}\right)$$
(11)

where max $(\hat{\mathbf{z}}_i)$ is the maximum pixel value in the *i* th band of the reference image, *N* is the pixel number of $\hat{\mathbf{z}}$.

2. SAM

SAM [32] is for quantifying the spectral similarity between the estimated and reference spectra in each pixel. The smaller of the SAM value indicates the higher spectral quality, the smaller spectral distortion.

$$SAM = \frac{1}{L_h} \sum \arccos \frac{\hat{\mathbf{z}}_i^{\mathrm{T}} \mathbf{z}_i}{\|\mathbf{z}_i\|_2 \|\hat{\mathbf{z}}_i\|_2}$$
(12)

where $\|\cdot\|_2$ denotes the l_2 norm.

3. ERGAS

EGRAS index [33] describes the global statistical quality of the fused data, the smaller the better.

$$\operatorname{ERGAS} = 100S \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\left\|\hat{\mathbf{Z}}_{i} - \mathbf{Z}_{i}\right\|_{2}^{2}}{\frac{1}{L} \sum_{i=1}^{L} \mathbf{Z}_{i}}\right)^{2}}$$
(13)

where *S* is the ground sampling distance (GSD) ratio between the HSR-MS and LSR-HS images. Q2n

4. Q2n

Q2n is a generalization of the universal image quality index (UIQI) [34] to measure the spatial and spectral quality in monochromatic images. The UIQI between reference image Z and fused image \hat{Z} is defined as:

$$Q2^{n} = \frac{4\sigma_{\hat{\mathbf{z}},\hat{\mathbf{z}}}\hat{\mathbf{z}} \cdot \bar{\mathbf{z}}}{\left(\sigma_{\hat{\mathbf{z}}}^{2} + \sigma_{\mathbf{z}}^{2}\right)\left(\tilde{\mathbf{z}}^{2} + \bar{\mathbf{z}}^{2}\right)'}$$
(14)

where $\mathbf{\bar{z}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{\hat{z}}_i$, $\mathbf{\bar{z}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{z}_i$, $\sigma_{\mathbf{z}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mathbf{z}_i - \mathbf{\bar{z}})^2}$, $\sigma_{\mathbf{\hat{z}}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mathbf{\hat{z}}_i - \mathbf{\bar{z}})^2}$, $\sigma_{\mathbf{\hat{z}},\mathbf{z}} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{\hat{z}}_i - \mathbf{\bar{z}})^2$

We considered two sets of simulation data and one set of real data for experiments to evaluate the performance of the proposed method. Simulation datasets are: 1. ROSIS Pavia center, 2. HYDICE Urban; Real dataset: 3. LSR-HS data is from Orbita Hyper Spectral (OHS) and HSR-MS data is from worldview-3 (shown in Figure 3).

1. ROSIS Pavia center dataset

The city center scene in Pavia is in the northern Italy, and the image was obtained by the Reflective Optics System Imaging Spectrometer (ROSIS-3) sensor with high spatial resolution (1.3 m) and the spectrum range is 430–834 nm (http://www.ehu.eus/ccwintco/index.php). There are 102 effective bands with a size of 1096×715 pixels in the image, and in this experiment we use the 480×480 pixels in the bottom right part the original image as the experiment data. The number of endmembers in this dataset is six.

2. HYDICE Urban dataset

The Urban dataset was obtained by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) over the urban in Copperas Cove (http://www.erdc.usace.army.mil/Media/FactSheets/FactSheetArticleView/tabid/9254/Article/610433/hypercube.aspx). The image includes 210 bands with high spatial resuolution (2 m) and the spectrum range is 400 to 2400 nm. After the removal of low-SNR and water absorption bands (1–4, 76, 87, 101–111, 136–153 and 198–210), 162 bands remain. The image size is 307 × 307 pixels, and in order to get a integer scale, we subset the in image as 304×304 pixels. The number of endmembers in this dataset is six.

We use the original datasets as reference images **Z**. To obtain the LSR-HS images X_h , the reference images are first blurred by 9×9 Gaussian kernel in the spatial domain and then downsampled by a ratio of 4. As to obtain HSR-MS X_m , we directly select the bands whose center wavelengths are the same as landsat 8 from the original reference images. We choose the blue, green, red and NIR channel to simulate the HSR-MS images. For Pavia center dataset, we choose 478 nm, 558 nm, 658 nm and 830 nm, for Urban dataset, we choose 481 nm, 555 nm, 650 nm and 859 nm. Because an HS image with lower spatial resolution contains more noise than an HSR-MS image, we add Gaussian white noises with standard deviation 0.1 and 0.04, respectively [35].

3. Worldview-3 and OHS Hengqin dataset

The third dataset is Hengqin island area of Guangdong Province, China. The HSR-MS image was captured by worldview-3 on 15 January 2018. The data includes four bands: blue, green, red and NIR, with a spatial resolution of 1.332 m. The LSR-MS image was captured by OHS sensor, which commercial satellites launched by Zhuhai Orbita Aerospace Science and Technology Company in 28 April 2018 carry. The LSR-HS image has 32 bands with 2.5 nm spectral resolution, the spectrum range is 400–1000 nm, and the spatial resolution is 10 m. We discard the 32-th band for the noise, so there are 31 bands left. In this experiment, we choose the area with small change of surface features, which includes water body, vegetation, residential area, highway, etc. The size of HSR-MS image and LSR-HS image in the same experimental area is 500×500 and 69×69 , respectively. For the convenience of comparative experiment, LSR-HS image was resampled to 100×100 (shown in Figure 3). In this experiment, *thresh*1 is set as 1.3, and the number of endmembers in this dataset is six.

HS–MS fusion technology obtains HSR-HS images by fusing LSR-HS images and HSR-MS images. Experimental dataset 1 (spatial resolution is 1.3 m) and dataset 2 (spatial resolution is 2 m) are the airborne hyperspectral data, the initial image is HSR-HS data. But for dataset 3, the LSR-HS image is from OHS sensor (10 m), and the HSR-MS image is from worldview3 (1.332 m). We do not have the corresponding area HSR-HS image. So, a ground truth image does not exist for this dataset. In this case, we use mean gradient (MG) [36] to evaluate the fusion results. MG is defined as follows:

$$MG = \left| grd_{x}\left(\hat{\mathbf{Z}} \right) \right| + \left| grd_{y}\left(\hat{\mathbf{Z}} \right) \right|$$
(15)

where $|\cdot|$ is the absolute value. $grd_x(\hat{\mathbf{Z}})$ and $grd_y(\hat{\mathbf{Z}})$ are the gradients of image $\hat{\mathbf{Z}}$ on the *x*- and *y*-axes, respectively. MG can describe the change of texture gray level on the image. It reflects the rate of change in the contrast of small details of the image. In general, the larger the MG is, the more detailed the image information is. The fused HSR-HS image has high spatial resolution, we use MG to evaluate the fusion results. Since MG is originally used to evaluate the accuracy of RGB image not HS image, so we select three bands from fused result to form RGB image (B:3; G:7; R:13), and then evaluate the accuracy. The comparison results of real images showed the applicability of the proposed method.



Figure 3. Experiment datasets. ((a)Pavia HSR-MS image (B:13, G:33, R:58); (b) Pavia LSR-HS image (downsampled ratio:4, add noise); (c) Urban HSR-MS image (B:17, G:37, R:50); (d) Urban LSR-HS image (downsampled ratio:4, add scale); (e) Hengqin HSR-MS image (B:2, G:3, R:4); (f) Hengqin LSR-HS image (B:8, G:13, R:27). Scale 4 is the downsampled ratio.)

3.3. Experimental Results

To quantitatively evaluate the performance of our method, we compare it with the following methods:

- CNMF [23] is a well-known HS–MS fusion method based on the spectral unmixing theory.
- FUSE [18] is one of the Bayesian-based HS–MS fusion algorithms with lower computational cost.
- GLP [17] determines the difference between the high-resolution image and its low-pass version, and multiplies the gain factor to obtain the spatial details of each low-resolution band.
- GSA [16] is an adaptive Gram–Schmidt algorithm, which better preserves the spectral information.
- HySure [20] preserves the edges of the fused image and uses total variation regularization to smooth out noise in homogeneous regions.
- MAPSMM [21] is the classic baysein-based HS–MS fusion method.

3.3.1. Simulated Dataset Fusion Results

We use PSNR, SAM, ERGAS and Q2n to quantitative evaluate the performances of compared methods, and the results are displayed in Table 1. Figures 4 and 5 show the performance of Pavia Center by fusion results, SAM map, respectively. Figures 6 and 7 show the performance of Urban dataset by fusion results, SAM map, respectively. Figure 8 shows the examples of reconstructed pixel values obtained by the seven methods and the corresponding ground truth.



Figure 4. HS–MS fusion results among seven compared methods and ground truth on ROSIS Pavia Center dataset with one demarcated areas zoomed in 1.5 times for easy observation. The composite HSIs are shown with bands 58-33-13 as R-G-B. (a) Coupled nonnegative matrix factorization (CNMF). (b) Fast fusion based on Sylvester equation (FUSE). (c) Generalized Laplacian pyramid (GLP). (d) Gram–Schmidt (GS). (e) HS Superresolution (HySure). (f) MAPSMM. (g) Proposed method. (h) Ground truth.

0.9 0.8 0.7 0.6 (a) CNMF (b) FUSE (d) GSA (c) GLP 0.5 Ν 0.4 0.3 0.2 0.1 Λ (g) Proposed (f) MAPSMM (e) HySure

Figure 5. HS–MS fusion Spectral Angle Mapper (SAM) map results among seven compared methods on ROSIS Pavia Center dataset. (a) CNMF. (b) FUSE. (c) GLP. (d) GS. (e) HySure. (f) MAPSMM. (g) Proposed method. The color scale denotes the spectral angle difference of each pixel between the ground truth to the fused image (from 0 to 1).



Figure 6. HS–MS fusion results among seven compared methods and ground truth on HYDICE Urban dataset with one demarcated areas zoomed in 1.5 times for easy observation. The composite HSIs are shown with bands 50-37-17 as R-G-B. (a) CNMF. (b) FUSE. (c) GLP. (d) GS. (e) HySure. (f) MAPSMM. (g) Proposed method. (h) Ground truth.



Figure 7. HS–MS fusion SAM map results among seven compared methods on HYDICE Urban dataset. (a) CNMF. (b) FUSE. (c) GLP. (d) GS. (e) HySure. (f) MAPSMM. (g) Proposed method. The color scale denotes the spectral angle difference of each pixel between the ground truth to the fused image (from 0 to 1).



Figure 8. Spectral value curves of the pixel in (45, 98) pixel of Pavia dataset and (88, 16) pixel of Urban dataset, respectively, using the seven compared methods compared to the ground truth.

Table 1. Experimental evaluation metrics among seven compared methods on two simulated Dataset (Bold numbers indicate the best performance).

Data	Index	Method								
		CNMF	FUSE	GLP	GSA	HySure	MAPSMM	Proposed		
Pavia	PSNR	31.747	26.766	27.838	30.636	28.061	26.255	32.860		
	SAM	6.357	13.646	8.643	11.774	11.876	9.629	5.700		
	ERGAS	4.973	8.667	7.448	5.460	7.334	8.922	4.799		
	Q2n	0.929	0.836	0.852	0.935	0.888	0.792	0.942		
Urban	PSNR	25.890	23.223	25.370	27.235	24.183	24.002	29.083		
	SAM	8.665	15.190	8.689	11.307	13.946	9.546	8.827		
	ERGAS	7.361	10.240	7.715	6.180	9.008	9.028	5.745		
	Q2n	0.863	0.772	0.817	0.886	0.802	0.772	0.933		

3.3.2. Real Dataset Fusion Results

Table 2 shows the performance of seven compared methods on the real hengqin dataset. MG and SAM evaluates the sharpness and the spectral distortion of the fusion results, respectively. Figure 9 shows the worldview-3 and OHS data fusion results.

Table 2. MG and SAM of HS–MS fusion results among seven compared methods on Hengqin dataset (Bold numbers indicate the best performance).

Method	Hengqin									
	CNMF	FUSE	GLP	GSA	HySure	MAPSMM	Proposed			
MG	57.743	6.852	7.791	14.812	57.570	9.996	69.575			
SAM	4.407	80.276	1.546	4.589	8.239	2.435	4.849			



Figure 9. HS–MS fusion results among seven compared methods on Hengqin dataset with one demarcated areas zoomed in 1.5 times for easy observation. The composite HSIs are shown with bands 27-13-8 as R-G-B. (a) CNMF. (b) FUSE. (c) GLP. (d) GS. (e) HySure. (f) MAPSMM. (g) Proposed method. (h) HSR-MS image. (i) LSR-HS image.

4. Discussion

4.1. Results Discussion

From Table 1, we can see that for two dataset, the proposed method achieves largest PSNR and smallest SAM (except Urban dataset), this means the fused results have the best spatial quality and spectral shape preservation capability, respectively. Moreover, the proposed method also gets the smallest ERGAS value and largest Q2n, this means the result has the overall best quality. For Urban dataset, the SAM of our method result is 8.827, and the smallest SAM 8.665 is from CNMF, it is only a little larger than the best result of the SAM. In addition, the results of the proposed method and CNMF are better than others, this indicates the spectral unmixing-based methods are very promising.

From Figures 4–7 the fusion results and SAM map both indicate our method outperforms than other methods. From the fusion results in Figures 4 and 6, we can see there are some spots and mottled textures in the fused images, except our method. Especially from the enlarged part of the fused image (a)–(f), there is some error texture and spectral distortion.

Figures 5 and 7 is the SAM map of Pavia and Urban dataset, respectively. SAM map shows the spectral distortion distribution and degree of the fusion results. We can see from (a)–(f) that the spectral error is mainly distributed in the shadow casted by trees and buildings, which shows that the proposed method can better preserve the spectrum of the shadow part of the LSR-HS images among the comparative methods. In Figure 8, (a) is a pixel value curve in shadow (45, 98), the fusion result of the proposed method better preserves the spectral details of the ground objects in the image. (b) shows the curves are the spectral value of a pixel (88, 16), which is the vegetation in shadow. The result curve of proposed method is the closest one to the ground truth among other compared methods. In addition, fusion results of CNMF are also good, which indicates the methods based on spectral unmixing are very promising.

For the real dataset from worldview3 and OHS Hengqin area, we can see from Table 2 that the proposed method has the largest MG. As we introduced before, the larger the MG value, the finer the fused image quality in detail. Because the HSR-HSI image does not exist in the real case, we use the LSR-HSI image as the reference image for evaluating the spectral accurate of the fusion image. Our method ensures the spectral reliability and improves the spatial resolution of the image. Figure 9 shows the HS–MS fusion results among seven compared methods on the Henqin dataset. From the results, fused images in (a)–(f) cannot achieve good results and produce the spectral distortion. Although (a) and (e) reconstructed the spatial detail, the water body (enlarged part) still have some unsmooth textures and the whole image have spectral distortions. From (c) and (f), the results keep the spectral information of the LSR-HS image, but the reconstructed edges are blurred. The proposed method outperforms the other approaches in that it keeps the spatial and spectral information to a large extent.

4.2. Time Complexity

Figure 10 shows the running time of all algorithms. The running time of HySure and MAPSMM algorithms are much longer than the rest of MS–HS fusion methods. Although CNMF has a better running time performance of Pavia dataset, its PSNR is lower than that of our method. For Urban dataset, GSA has the smallest time consumption but the PSNR is still smaller than the proposed method. In the Hengqin dataset, our method has the highest MG with moderate running time. In conclusion, our method is slightly slower than some methods but with the best performance. We implemented all algorithms in the experiments under the same hardware configuration: Intel Core i7-8565 U CPU @ 2.00 GHz.



Figure 10. Running time (seconds) of three datasets for different MS–HS fusion algorithms.

5. Conclusions

This paper proposes a new fusion method based on spectral unmixing by reconstructing HSR-HS images from LSR-MS images and HSR-HS images with a temporal difference. The method is very simple with small computation cost. Considering the temporal difference of two input images, we mask the ground feature changed areas, and then obtain consistent endmembers and corresponding pixels' positions from the remaining area in LSR-HS images. According to the LSR-HS image endmember spatial information, we can obtain the abundance matrix of an HSR-MS image. Finally, the HS–MS image is generated by the endmember matrix from the LSR-HS image and abundance matrix from the HSR-MS image. The most important thing is this method works well with the images which have a time difference. Furthermore, experimental results of our method in both simulated and real datasets achieve good visual effect, with the best accuracy among the compared methods.

In future work, we plan to apply the endmember spatial information idea to other real datasets from different sensors to improve the universality of this method. From the real image fusion experiment, our method still needs to improve the ability to reduce the spectral distortion while enhancing the spatial resolution of reconstructed images. Therefore, we also plan to further study the correlation between the spatial structure and spectral characteristics of the details by considering the segmentation of the images.

Author Contributions: X.F. and L.H. conceived and designed the experiments; X.F. performed the experiments, analyzed the data, and wrote the paper; L.H. reviewed the paper and helped with the experiments. Q.C. administrated project and supported with the resources. X.L. and Y.Y. helped with the resources. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National key R&D plan on strategic international scientific and technological innovation cooperation special project under Grant 2016YFE0202300, the National Natural Science Foundation of China under Grants 41771452, 41771454 and 41890820, the Natural Science Fund of Hubei Province in China under Grant 2018CFA007.

Acknowledgments: The authors would like to thank the anonymous reviewers for their hard work and the "Innovation and entrepreneurship training program for university Students (S2019619003)" project.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [CrossRef]
- 2. Shao, Z.; Fu, H.; Li, D.; Altan, O.; Cheng, T. Remote sensing monitoring of multi-scale watersheds impermeability for urban hydrological evaluation. *Remote Sens. Environ.* **2019**, 232, 111338. [CrossRef]
- 3. Shao, Z.; Zhang, L.; Zhou, X.; Ding, L. A novel hierarchical semisupervised SVM for classification of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1609–1613. [CrossRef]

- 4. Shao, Z.; Zhou, W.; Cheng, Q.; Diao, C.; Lei, Z. An effective hyperspectral image retrieval method using integrated spectral and textural features. *Sens. Rev.* **2015**, *35*, 274–281. [CrossRef]
- Li, W.; Wu, G.; Zhang, F.; Du, Q. Hyperspectral image classification using deep pixel-pair features. *IEEE Trans. Geosci. Remote Sens.* 2016, 55, 844–853. [CrossRef]
- Ma, Y.; Zhang, Y.; Mei, X.; Dai, X.; Ma, J. Multifeature-Based Discriminative Label Consistent K-SVD for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2019, 12, 4995–5008. [CrossRef]
- Shao, Z.; Zhang, L. Sparse dimensionality reduction of hyperspectral image based on semi-supervised local Fisher discriminant analysis. *Int. J. Appl. Earth Obs. Geoinf.* 2014, *31*, 122–129. [CrossRef]
- Zhang, L.; Weng, Q.; Shao, Z. An evaluation of monthly impervious surface dynamics by fusing Landsat and MODIS time series in the Pearl River Delta, China, from 2000 to 2015. *Remote Sens. Environ.* 2017, 201, 99–114. [CrossRef]
- Jiang, J.; Ma, J.; Chen, C.; Wang, Z.; Cai, Z.; Wang, L. SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* 2018, 56, 4581–4593. [CrossRef]
- 10. Shao, Z.; Cai, J.; Fu, P.; Hu, L.; Liu, T. Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. *Remote Sens. Environ.* **2019**, 235, 111425. [CrossRef]
- 11. Yokoya, N.; Grohnfeldt, C.; Chanussot, J. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 29–56. [CrossRef]
- 12. He, L.; Wang, M.; Zhu, Y.; Chang, X.; Feng, X. Image Fusion for High-Resolution Optical Satellites Based on Panchromatic Spectral Decomposition. *Sensors* **2019**, *19*, 2619. [CrossRef] [PubMed]
- Loncan, L.; De Almeida, L.B.; Bioucas-Dias, J.M.; Briottet, X.; Chanussot, J.; Dobigeon, N.; Fabre, S.; Liao, W.; Licciardi, G.A.; Simoes, M.; et al. Hyperspectral pansharpening: A review. *IEEE Geosci. Remote Sens. Mag.* 2015, 3, 27–46. [CrossRef]
- 14. Carper, W.; Lillesand, T.; Kiefer, R. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogramm. Eng. Remote Sens.* **1990**, *56*, 459–467.
- 15. Laben, C.A.; Brower, B.V. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening. U.S. Patent 6,011,875, 4 January 2000.
- 16. Aiazzi, B.; Baronti, S.; Selva, M. Improving component substitution pansharpening through multivariate regression of MS + Pan data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3230–3239. [CrossRef]
- 17. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A.; Selva, M. MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 591–596. [CrossRef]
- 18. Wei, Q.; Dobigeon, N.; Tourneret, J.Y. Fast fusion of multi-band images based on solving a Sylvester equation. *IEEE Trans. Image Process.* **2015**, *24*, 4109–4121. [CrossRef]
- 19. Zhang, Y.; De Backer, S.; Scheunders, P. Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3834–3843. [CrossRef]
- Simões, M.; Bioucas-Dias, J.; Almeida, L.B.; Chanussot, J. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Trans. Geosci. Remote Sens.* 2014, 53, 3373–3388. [CrossRef]
- 21. Eismann, M.T.; Hardie, R.C. Application of the stochastic mixing model to hyperspectral resolution enhancement. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1924–1933. [CrossRef]
- 22. Keshava, N.; Mustard, J.F. Spectral unmixing. IEEE Signal Process. Mag. 2002, 19, 44–57. [CrossRef]
- 23. Yokoya, N.; Yairi, T.; Iwasaki, A. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Trans. Geosci. Remote Sens.* 2011, *50*, 528–537. [CrossRef]
- 24. Akhtar, N.; Shafait, F.; Mian, A. Sparse spatio-spectral representation for hyperspectral image super-resolution. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 63–78.
- 25. Lanaras, C.; Baltsavias, E.; Schindler, K. Hyperspectral super-resolution by coupled spectral unmixing. In *IEEE International Conference on Computer Vision;* Springer: Cham, Switzerland, 2015; pp. 3586–3594.
- 26. Wang, X.; Zhong, Y.; Zhang, L.; Xu, Y. Spatial group sparsity regularized nonnegative matrix factorization for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6287–6304. [CrossRef]

- 27. Huang, B.; Song, H.; Cui, H.; Peng, J.; Xu, Z. Spatial and spectral image fusion using sparse matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 1693–1704. [CrossRef]
- 28. Nascimento, J.M.; Dias, J.M. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 898–910. [CrossRef]
- 29. Lawson, C.L.; Hanson, R.J. Solving Least Squares Problems; Siam: Ann Arbor, MI, USA, 1995; Volume 15.
- Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2012, *5*, 354–379. [CrossRef]
- 31. Wycoff, E.; Chan, T.H.; Jia, K.; Ma, W.K.; Ma, Y. A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 1409–1413.
- 32. Kruse, F.A.; Lefkoff, A.; Boardman, J.; Heidebrecht, K.; Shapiro, A.; Barloon, P.; Goetz, A. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Environ.* **1993**, *44*, 145–163. [CrossRef]
- Wald, L. Quality of high resolution synthesised images: Is there a simple criterion? In Proceedings of the Third Conference "Fusion of Earth Data: Merging point Measurements, Raster Maps and Remotely Sensed Images", Sophia Antipolis, France, 26–28 January 2000; pp. 99–103.
- 34. Wang, Z.; Bovik, A.C. A universal image quality index. IEEE Signal Process. Lett. 2002, 9, 81-84. [CrossRef]
- Takeyama, S.; Ono, S.; Kumazawa, I. Hyperspectral and Multispectral Data Fusion by a Regularization Considering. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2152–2156.
- 36. Lu, T.; Wang, J.; Zhang, Y.; Wang, Z.; Jiang, J. Satellite Image Super-Resolution via Multi-Scale Residual Deep Neural Network. *Remote Sens.* **2019**, *11*, 1588. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).