Fast and Stable Hyperspectral Multispectral Image Fusion Technique Using Moore–Penrose Inverse Solver

Jian Long, Yuanxi Peng *, Tong Zhou, Liyuan Zhao and Jun Li

State Key Laboratory of High Performance Computing, College of Computer Science, National University of Defense Technology, Changsha 410073, China; jian_long@nudt.edu.cn (J.L.);
zhoutongnudt@163.com (T.Z.); zhaoliyuan17@nudt.edu.cn (L.Z.); lijun14@nudt.edu.cn (J.L.)
* Correspondence: pyx@nudt.edu.cn

Abstract: Fusion low-resolution hyperspectral images (LR-HSI) and high-resolution multispectral images (HR-MSI) are important methods for obtaining high-resolution hyperspectral images (HR-HSI). Some hyperspectral image fusion application areas have strong real-time requirements for image fusion, and a fast fusion method is urgently needed. This paper proposes a fast and stable fusion method (FSF) based on matrix factorization, which can largely reduce the computational workloads of image fusion to achieve fast and efficient image fusion. FSF introduces the Moore–Penrose inverse in the fusion model to simplify the estimation of the coefficient matrix and uses singular value decomposition (SVD) to simplify the estimation of the spectral basis, thus significantly reducing the computational effort of model solving. Meanwhile, FSF introduces two multiplicative iterative processes to optimize the spectral basis and coefficient matrix to achieve stable and high-quality fusion. We have tested the fusion method on remote sensing and ground-based datasets. The experiments show that our proposed method can achieve the performance of several state-of-the-art algorithms while reducing execution time to less than 1% of such algorithms.

Keywords: hyperspectral imaging super-resolution; image fusion; matrix factorization

1. Introduction

Hyperspectral images with rich spectral information play an important role in agriculture [1], medicine [2], geological surveying [3], remote sensing [4], computer vision [5,6], and other fields. Generally, the resolution of hyperspectral images directly affects the application effect of hyperspectral images. However, the spatial resolution of hyperspectral images directly obtained with hyperspectral cameras is relatively low. On the contrary, the multispectral image directly obtained by the multispectral camera has a higher spatial resolution. Obtaining high-resolution hyperspectral images by improving hardware processes is usually costly, so fusing high-resolution multispectral images with low-resolution hyperspectral images is one of the most economical ways to obtain high-resolution hyperspectral images.

In the practical application of the fusion method, two images of the same area obtained by the hyperspectral camera and the multispectral camera can be pre-processed by image alignment [7] and enhancement. Then, the feature points in the two images can be extracted and matched separately, and then the two images can be geometrically aligned. After alignment, the two images can be used as input images for the fusion algorithm. The application flow chart is shown in Figure 1a. However, in this paper, in order to compare the fusion ability of various algorithms more effectively, we use the original image in the dataset as the target HR-HSI in the experimental part, simulate the LR-HSI obtained by the hyperspectral camera by spatially blurred downsampling of the target HR-HSI, and simulate the HR-MSI obtained by the multispectral camera by spectral downsampling of the target HR-HSI. The LR-HSI and HR-MSI involved in the fusion are already aligned images, so that we can easily check the fusion performance of the fusion algorithm by

Citation: Long, J.; Peng, Y.; Zhou, T.; Zhao, L.; Li, J. Fast and Stable Hyperspectral Multispectral Image Fusion Technique Using Moore–Penrose Inverse Solver. Appl. Sci. 2021, 11, 7365. https://doi.org/10.3390/app11167365
comparing the similarity between the fused images and the target image (the better the similarity, the better the fusion effect). The application flow chart is shown in Figure 1b.

The earliest methods of spatial-spectral image fusion can be traced back to pan-sharpening techniques, which fuse low spatial resolution RGB images with high resolution panchromatic images to obtain high-resolution RGB images. Representative pan-sharpening techniques include component substitution [8], methods based on principal component analysis [9], and methods based on compressed sensing [10]. Since panchromatic images contain only one piece of spectral information, pan-sharpening techniques usually lead to spectral distortion. To enhance the fusion quality, HR-MSI is usually used instead of panchromatic images to fuse with LR-HSI. Typical fusion methods include Bayesian-based fusion methods, spectral unmixing-based fusion methods, matrix decomposition-based fusion methods, and tensor decomposition-based fusion methods.

Eismannet et al. first used the Bayesian method based on maximum posterior probability (MAP) estimation to realize fusion. This method uses a stochastic mixed model (SMM) to estimate potential spectral scene features and formulates a cost function to optimize the corresponding hyperspectral image data estimated by LR-HSI and HR-MSI. The optimization of MAP-SMM is carried out in the principal component subspace. The idea of using two input images to merge spectral information in a subspace is the primary source of inspiration for many later fusion methods [11–13]. Compared with pan-sharpening technology, this fusion method has made a significant breakthrough in improving the spatial resolution of hyperspectral images [14,15].

With the development of fusion technology, fusion methods based on spectral unmixing [16–18] have been proposed, based on the fact that hyperspectral images often have a large number of so-called “mixed pixels”, i.e., overlapping spectral responses of several different surface materials, due to the much broader pixel coverage. The spectral unmixing approach tries to separate these responses, i.e., to recover the underlying pure material spectra (called endmembers) and their relative proportions (called abundances). Since LR-HSI and HR-MSI capture the same scene, the underlying pure matter members are identical. Therefore, it can extract a spectral dictionary from one image to interpret the other, thus achieving high spectral image super-resolution. Based on linear spectral unmixing, the literature [19] has proposed a coupled spectral unmixing method, which solves the super-resolution problem by alternating the spectral unmixing of HR-MSI and LR-HSI simultaneously while updating spectral and abundance. However, the fusion quality of this method is closely related to the estimation of the number of underlying endmembers. The requirement for estimating the underlying endmembers is high, and it is not easy to achieve an accurate estimation. Therefore, in the literature [20], a sparse constraint is added to the fusion. As a result, a hyperspectral, multispectral image fusion algorithm based on sparse representation is proposed (NSSR). Through updating the number of the underlying endmembers by an effective non-negative dictionary learning algorithm, NSSR effectively improves the fusion quality. However, this method demands more parameters to be determined, and the fusion quality fluctuates wildly.

In recent years, matrix factorization-based fusion methods have started to emerge. The matrix factorization-based fusion methods assume HR-HSI as the product of the spectral basis matrix and a coefficient matrix. They estimate the spectral basis from LR-HSI and coefficient matrix from HR-MSI, respectively, to achieve hyperspectral super-resolution. The authors of [21] propose the coupled non-negative matrix factorization (CNMF) technology, which extracts the spectral basis and coefficient matrix by performing non-negative matrix factorization (NMF) on LR-HSI and HR-MSI, respectively. However, since the NMF is usually not unique, the results produced in [21] are not always good. In [13], a standard linear inverse problem model for LR-HSI and HR-MSI fusion was developed to solve the optimized spectral basis and coefficient matrices by introducing full variational regularization terms in the pairwise matrix equations. However, this approach usually requires splitting of variables requiring relatively more computational time as well. Further studies began to add a priori information to the fusion to improve the quality of
the fusion. For example, in [22], by adding spatial similarity to the fusion, and in [23], by using the nonlocal coefficient representation method and the self-similarity constraint to impose a priori constraints on the fusion, the quality of image fusion can be improved to some extent. However, the fusion process becomes more complicated due to the added constraints, which leads to a further increase in the fusion time and makes the efficiency for fusion decrease. Thus, for some application scenarios that require real-time response, this fusion method is not suitable.

Unlike matrix factorization-based methods, tensor factorization-based methods can preserve the three-dimensional characteristics of image data. This has led to the continuous application of tensor factorization-based methods to image processing work, such as multi-frame data denoising [24], completion [25], compression perception [26], and classification [27]. Tensor factorization-based methods for solving hyperspectral image fusion have also been proposed. Based on the Tucker decomposition, [28] views HR-HSI as the product of a kernel tensor and three-factor matrices and estimates the kernel tensor and factor matrices from HR-MSI and LR-HSI, respectively. The autocorrelation of hyperspectral images is cited in [29], and the method of nonlocal tensor factorization is proposed to effectively use the nonlocal similarity information to realize hyperspectral image fusion. However, this method based on Tucker decomposition requires more intermediate variables to be estimated, which requires more computational effort and usually results in a slower fusion rate. A fusion method based on canonical polyadic decomposition (CPD) is proposed in [30], which decomposes the target HR-HSI into the sum of vector products of three dimensions. It effectively reduces the computational effort of tensor decomposition-based methods. However, this approach assumes that the ranks of the three dimensions are the same (which is not valid in some cases), so this usually leads to undesirable fusion results. In [31], the advantages of Tucker decomposition and CP decomposition are combined, and a fusion method based on tensor ring factorization is proposed. This method views the target HR-HSI as a matrix product of three dimensions and achieves hyperspectral image super-resolution by estimating the spectral dimension matrix from LR-HSI and the two matrices of spatial dimension from HR-MSI. This fusion method usually requires the estimation of three matrices, which also has a significant computation and can lead to poor timeliness of fusion.

Although the recently proposed tensor factorization-based method can effectively preserve the three-dimensional characteristics of hyperspectral images, the tensor factorization-based method also has to be operated by converting to a two-dimensional matrix, and the fusion quality is not substantially improved. Compared with the tensor factorization-based method, the matrix factorization-based method has better timeliness. This paper proposes a fast and stable fusion method based on matrix factorization to ensure the fusion quality and shorten the fusion time. This method first solves the fuzzy estimation of the spectral basis through a simple singular value decomposition method. Then, through the Moore–Penrose inversion operation, it calculates the fuzzy estimate of the spectral coefficient matrix and then obtains the optimized spectral base and coefficient matrix estimate through a multiplicative iterative optimization process. Finally, the exact HR-HSI estimate is obtained by doing another multiplicative iterative optimization operation on the fuzzy version of the HR-HSI obtained in the previous step.

Overall, this paper presents a fast and stable hyperspectral multispectral image fusion technique. The significant contributions of this work are as follows.

- It is the first time that the Moore–Penrose inverse has been used to determine the solution of the spectral coefficient matrix, which simplifies the solution process of the fusion model, dramatically reduces the computational workload, and shortens the fusion time.
- A multiplicative iterative optimization process is used to optimize the spectral basis and coefficient matrices, enabling further enhancement of the fused results and improving the quality of the fusion.
- Previous fusion methods stop optimizing the fusion results after obtaining the spectral basis and coefficient matrices. In contrast, the method proposed in this paper continues to optimize the rough HR-HSI estimates after fusion to further improve the fusion quality after obtaining the spectral basis and coefficient matrices.

The rest of this paper is organized as follows. Section 2 details the proposed method. Section 3 gives the experimental settings. The results and analysis are presented in Section 4, and the conclusion is given in Section 5.

![Flowchart for achieving super-resolution using a fusion approach.](image)

**Figure 1.** Flowchart for achieving super-resolution using a fusion approach.

### 2. Proposed Method

Our proposed method is a matrix factorization-based fusion method, which requires first transforming the 3D hyperspectral image data into 2D matrix data and then performing the corresponding calculations. Usually, we denote the target HR-HSI as $Z \in \mathbb{R}^{W \times H \times L}$, with $W \times H$ pixels and $L$ bands; the LR-HSI involved in the fusion as $\mathcal{H} \in \mathbb{R}^{w \times h \times L}$, with $w \times h$ pixels and $L$ bands; and the HR-MSI as $\mathcal{M} \in \mathbb{R}^{W \times H \times l}$, with $W \times H$ pixels and $l$ bands, and their mode-3 matricization matrices are $Z \in \mathbb{R}^{L \times WH}$, $H \in \mathbb{R}^{L \times wh}$, and
\( \mathbf{M} \in \mathbb{R}^{l \times WH} \), respectively. We denote a three-dimensional tensor as \( \mathbf{X} \in \mathbb{R}^{l_1 \times l_2 \times h} \), its \((i, j, k)\) elements as \( x_{ijk} \) or \( X_{ijk} \), and its Frobenius norm as \( \| \mathbf{X} \|_F = \sqrt{\sum_{i,j,k} |x_{ijk}|^2} \).

2.1. Fusion Model

In the fusion-based approach, LR-HSI and HR-MSI are usually assumed to be the spatial blur downsampling and spectral downsampling of HR-HSI, respectively. Furthermore, the spectral response function can be estimated from the hyperspectral camera and multispectral camera parameters, and the estimation of spatial blur downsampling can be approximated by several spatial blur estimation functions. Thus, the following equation can be obtained:

\[
\mathbf{H} = \mathbf{ZBS} \tag{1}
\]

where \( \mathbf{B} \in \mathbb{R}^{WH \times WH} \) and \( \mathbf{S} \in \mathbb{R}^{WH \times wh} \) represent the convolution blur and the downsampling matrix, respectively. The HR-MSI can be seen as the spectrally downsampled version of HR-HSI, i.e.,

\[
\mathbf{M} = \mathbf{RZ} \tag{2}
\]

where \( \mathbf{R} \in \mathbb{R}^{l \times L} \) denotes the spectral response matrix. As mentioned above, both the spectral response function and spatial fuzzy downsampling can be obtained. Therefore, the matrices \( \mathbf{B}, \mathbf{S}, \) and \( \mathbf{R} \) in Equations (1) and (2) are all determined, so solving HR-HSI means solving the unknown matrix \( \mathbf{Z} \) in the above matrix equation. In matrix factorization-based methods, it is usually assumed that the spectral basis exists in a lower subspace \([13]\), i.e.,

\[
\mathbf{Z} = \mathbf{DC} \tag{3}
\]

where \( \mathbf{D} \in \mathbb{R}^{L \times L_s} \) and \( \mathbf{C} \in \mathbb{R}^{L_s \times WH} \) are the spectral bases and spectral basis coefficient matrix, respectively. Therefore, the following equations can be obtained:

\[
\mathbf{H} = \mathbf{DCBS} \tag{4}
\]
\[
\mathbf{M} = \mathbf{RDC} \tag{5}
\]

2.2. Moore–Penrose Inverse

From Equations (4) and (5), only the matrices \( \mathbf{D} \) and \( \mathbf{C} \) are unknown, and combining Equation (3), \( \mathbf{Z} \) can be solved by solving the matrices \( \mathbf{D} \) and \( \mathbf{C} \). Since the spectral information in HR-HSI is mainly from LR-HSI, the idea that LR-HSI and HR-HSI contain the same spectral basis was proposed and verified in \([22]\). Therefore, we can perform SVD on LR-HSI to obtain the preliminary spectral basis.

\[
\mathbf{H} = \mathbf{U} \Sigma \mathbf{V}^T \tag{6}
\]

Since the first few singular values of a matrix can contain more than 90% of the matrix information, we save the first \( q \) singular values in \( \Sigma \) to obtain the following spectral basis, i.e.,

\[
\mathbf{D} = \mathbf{U}(:, 1 : q) \tag{7}
\]

In Equation (5), only the matrix \( \mathbf{C} \) is unknown, and the matrix \( \mathbf{C} \) can be obtained by inverting the matrix \( \mathbf{RD} \).

\[
\mathbf{C} = (\mathbf{RD})^{-1} \mathbf{M} \tag{8}
\]

Since the matrix \( \mathbf{RD} \) is not necessarily a square matrix, the inverse of the matrix \( \mathbf{RD} \) in Equation (8) is not unique, which will cause errors in the solution results. However, the Moore–Penrose inverse (“+” inverse) is unique among all inverse matrices of the matrix,
so the fuzzy estimate of the matrix $C$ can be obtained by introducing the Moore–Penrose inverse as follows.

$$C = (RD)^+ M$$  \hspace{1cm} (9)

According to the definition of the matrix Moore–Penrose inverse, we can solve the Moore–Penrose inverse of the matrix $RD$ through the Algorithm 1.

\begin{algorithm}
\caption{Moore–Penrose inverse}
\begin{algorithmic}
\Require $RD$
\State $[U_1, \Sigma_1, V_1] = \text{SVD}(RD)$
\State $[m, n] = \text{size}(\Sigma_1)$
\State $K = \min(m, n)$
\For{$k = 1:K$}
\If{$\Sigma_1(k, k) = 0$}
\State $\Sigma_1(k, k) = 1/\Sigma_1(k, k)$
\EndIf
\EndFor
\State $\Sigma_2 = \text{zeros}(n, m)$
\For{$k = 1:K$}
\State $\Sigma_2(k, k) = \Sigma_1(k, k)$
\EndFor
\State $(RD)^+ = V_1 \Sigma_2 U_1^T$
\Return $(RD)^+$
\end{algorithmic}
\end{algorithm}

2.3. Optimized Fusion Solution

From above, we have obtained the fuzzy estimates of spectral basis $D$ and coefficients $C$. However, the fused images generated from the fuzzy estimates $D$, $C$ and the fusion quality are unsatisfactory. Therefore, based on Equation (4), we can optimize the spectral basis $D$ by solving the optimization procedure of the following equation.

$$\min_D \|H - DCBS\|_F^2$$  \hspace{1cm} (10)

In Equation (10), we assume that $X = CBS$; then, Equation (10) can be resolved as

$$\min_D \|H - DX\|_F^2$$  \hspace{1cm} (11)

We extend the multiplicative update rule in [21] to obtain an approximate solution for solving Equation (11), i.e.,

$$D = D \times (HX^T)/(DXX^T)$$  \hspace{1cm} (12)

After obtaining the optimized spectral base $D$, we can obtain the HR-HSI fuzzy estimation matrix $Z_1$ as

$$Z_1 = DC$$  \hspace{1cm} (13)

Obtaining the HR-HSI estimation result $Z_1$, fusion quality still has room for further improvement. Combining Equation (2), we can obtain

$$M = RZ_1$$  \hspace{1cm} (14)

Combining the known matrix $M$ and $R$ to further optimize the fusion result $Z_1$,

$$\min_{Z_1} \|M - RZ_1\|_F^2$$  \hspace{1cm} (15)

Similarly, we use the extended multiplication iteration rule to solve Equation (15), and we can obtain

$$Z_1 = Z_1 \times (R^T M)/(R^T RZ_1)$$  \hspace{1cm} (16)
After iterative optimization, we can obtain a more accurate fused image $Z = Z_1$. Finally, our proposed method can be described by Algorithm 2, where $q$ represents the parameter to be adjusted.

**Algorithm 2**  
**FSF Algorithm**

Require: $H$ $M$ $R$ $B$ $S$ $q$

1. Get $D$ by Equations (6) and (7)
2. Get $C$ by Algorithm 2
3. for $k = 1:K$
   1. Update $D$ by Equation (12)
4. end for
5. $Z_1 = DC$
6. for $k = 1:K$
   1. Update $Z_1$ by Equation (16)
7. end for
8. $Z = Z_1$
9. return $Z$

3. Experiments

To validate the superior performance of our proposed method, we compare it with six state-of-the-art methods on two commonly used datasets, Pavia University and CAVE. Six commonly used quality metrics are used to measure the effectiveness of the fusion.

3.1. Date Set

We use two datasets, among which Pavia University [32] belongs to the remote sensing dataset. It is one of two scenes acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy. It contains 103 spectral bands (0.43–0.86 µm), and each band includes 610 × 610 pixels, but some of the samples in the images contain no information and have to be discarded before the analysis. The geometric resolution is 1.3 m. After removing the absorbing bands and some undesirable pixel points, we finally keep 93 bands and 256 × 256 pixel points as the reference HR-HSI image. We generate the 4-band HR-MSI (4 × 256 × 256) using the IKONOS class reflection spectral response filter [12] as a spectral response function. We used a 7 × 7 Gaussian blur (standard deviation 2) for the reference image and then sampled down every 32 pixels in both spatial dimensions to simulate LR-HSI (93 × 8 × 8).

The CAVE dataset [33] is a ground-based hyperspectral dataset containing 32 high-quality indoor HSIs, each containing 512 × 512 pixels in 31 spectral bands, of which the first two bands are blurred. Its spectral range is 400 nm–700 nm, and the wavelength interval is 10 nm. We remove the first two bands and finally keep 29 bands and 512 × 512 pixel points as the reference HR-HSI image (29 × 512 × 512). Also, to simulate LR-HSI, we use 7 × 7 Gaussian blur (standard deviation 2) and then sample down every 32 pixels in both spatial dimensions (29 × 16 × 16). To simulate HR-MSI, we use Gf-1-16m multispectral camera response to generate the three-band HR-MSI (RGB image) (3 × 512 × 512).

3.2. Compared Methods

We use six state-of-the-art algorithms for comparison, including three methods based on Tucker decomposition: low tensor-train rank (LTTR)-based method [34], coupled sparse tensor factorization (CSTF) [28], nonlocal sparse tensor factorization (NLSTF) [29], and a method based on CP decomposition Super-resolution TEnsor-REcOnstruction (STEREO) [30] (we use the non-blind version). In addition, we use the matrix decomposition-based method Hysure [13] and the spectral unmixing-based method non-negative structured sparse representation (NSSR) [20]. To demonstrate the reliability of the experimental results, all source codes were obtained through open access, and the codes for the various experiments were run in the same environment. This paper’s experiments were implemented in MATLAB R2014a on localhost with 2.40 GHz Intel i7-7700 CPU and 8.0 GB DDR3.
3.3. Quantitative Metrics

To comprehensively evaluate the superior performance of our proposed method, we used six quantitative metrics in our experiments. These include the peak signal-to-noise ratio (PSNR) [22] based on an error-sensitive quality metric; an index to evaluate the mean square error between the fused image and the reference image (ERGAS) [35]; the spectral angle mapper (SAM) [12] to evaluate the spectral angle shift of the fused image; and the universal quality index (UIQI) [36] and the structural similarity (SSIM) [37], the last of which is an important evaluation metric for evaluating the time required for fusion.

The first five metrics evaluate the similarity between the fused image and the target image from several aspects, and the better the result of the metrics, the better the effect of the fusion. A higher similarity between the fused image and the target image indicates that the LR-HSI obtained after spatial blurring downsampling achieves better resolution enhancement after fusion.

3.4. Parameters Discussion

For our proposed method FSF, only one parameter is to be adjusted, the dimensionality of spectral bases q. When extracting spectral bases by SVD on LR-HSI, the dimensionality of spectral bases will affect the fusion results. Usually, the spectral bases exist in a small subspace, and we vary the dimension of the spectral bases from 1 to 10 to find the most suitable number of spectral bases. Figure 2a shows the PSNR values of the fused images when the number of spectral bases is varied from 1 to 10 in the Pavia University dataset, and it can be seen that the best fusion results are obtained when the dimension of spectral bases is 4. In Figure 2b, corresponding to the results on the CAVE dataset, it is clear that the best results are obtained when the spectral basis dimension is 3. The other methods used for comparison, CSTF https://github.com/renweidian/CSTF (accessed date: 9 August 2021), Hysure https://github.com/alfaiate/HySure (accessed date: 9 August 2021), LTTR https://github.com/renweidian/LTTR (accessed date: 9 August 2021), NLSTF https://github.com/renweidian/NLSTF (accessed date: 9 August 2021), STEREO https://sites.google.com/site/harikanats/ (accessed date: 9 August 2021), and NSSR http://see.xidian.edu.cn/faculty/wsdong (accessed date: 9 August 2021), adjusted the parameters that affect their fusion results to achieve the best fusion results. The best parameters after tuning corresponding to these methods are shown in Tables 1 and 2.

![Figure 2](image-url)  
**Figure 2.** FSF fusion quality on two datasets with different number of subspaces.
To fully demonstrate the timeliness and stability of our proposed method, we conduct comparative experiments in the following order. First, we conduct comparative experiments on the CAVE and Pavia University datasets to test the fusion effect and the excellent timeliness of our proposed method. Then, we test the fusion effects of the various fusion methods at downsampling factors of 4, 8, 16, 32, and 64 to verify the stability of our proposed method. Note that the downsampling factor is the number of pixel points per pixel point interval taken when simulating LR-HSI. For example, when the downsampling factor is 4, the size of the LR-HSI simulated from Pavia University data is $93 \times 64 \times 64$, and when the downsampling factor is 16, the size of the corresponding LR-HSI is $93 \times 16 \times 16$. The larger the downsampling factor is, the stronger the super-resolution capability of the fusion is. Here, we take the maximum downsampling factor to 64, i.e., to achieve a super-resolution of 64 times the spatial dimension, and the size of the LR-HSI involved in the fusion is only $93 \times 4 \times 4$ at this time.

4. Results and Analysis

We first conducted experiments on the Pavia University and CAVE datasets at a downsampling factor of 32. Then, to describe the experimental results objectively and fairly, we show the subjective results of the fused images and the objective results of the six fusion metrics, respectively. Figure 3 shows false-color images from LR-HSI, fusion images, and ground truth images. It can be seen that the fused image is clearer than LR-HSI.

4.1. Timeliness of The FSF

Figure 4 shows the fusion results of the compared methods on the Pavia University dataset. When there are more black pixels in the error image, the fusion result is better. It is evident from the figure that NLSTF has the worst fusion result, and the error image contains a large number of white spots. The STEREO fusion result is also unsatisfactory, and the
error image contains many white pixel points. The main reason for the unsatisfactory fusion results of NLSTF and STEREO is that the downsampling factor influences the fusion effect of NLSTF and STEREO, and the larger the downsampling factor is, the worse the fusion effect is. The fusion quality of NSSR is usually closely related to the selection of endmembers, and the fusion quality of pixel points at the edges of endmembers is generally poor. Therefore, we can see that the error images of the fusion results obtained by the NSSR method also contain many noticeable white pixels. Among all the methods, only our proposed method, FSF, gives the best fusion results with the least number of white pixel points in the error image.

Table 3 corresponds to the quantitative metrics of the fusion results of the competing approaches on the Pavia University dataset. Our proposed method achieves the best results in all metrics, and only the CSTF and Hysure methods achieve fusion results similar to ours when the time required for fusion is not taken into account. When considering the running time, CSTF and Hysure consume 289.643 s and 60.61 s, respectively, while our proposed method FSF takes only 0.26 s, consuming less than 0.1% of the time for CSTF and less than 1% that for Hysure. Aside from our FSF, the fastest algorithm is STEREO; this is because STEREO reduces the tensor calculation in three dimensions to a one-dimensional vector to calculate, which significantly improves the speed of fusion, but even STEREO takes 8.088 s.

Figure 5 shows the fusion results of the compared methods on the CAVE dataset. Similarly, the NLSTF and STEREO methods obtained the worst fusion results due to the poor fusion of these two methods at 32× super-resolution. Our FSF method still achieves better fusion results.

Table 3. The quantitative results of the experiment on the Pavia University dataset. The best results are shown in bold among all methods.

| Method   | PSNR  | ERGAS | SAM   | UIQI  | SSIM  | Time   |
|----------|-------|-------|-------|-------|-------|--------|
| Best Value | +∞   | 0     | 0     | 1     | 1     | 0      |
| CSTF     | 43.072 | 0.144 | 2.123 | 0.993 | 0.989 | 289.643 |
| Hysure   | 42.313 | 0.158 | 2.255 | 0.992 | 0.988 | 60.61  |
| LTTR     | 32.932 | 0.623 | 7.263 | 0.918 | 0.937 | 485.533|
| NLSTF    | 29.551 | 0.672 | 7.388 | 0.924 | 0.936 | 85.319 |
| STEREO   | 27.051 | 0.812 | 9.108 | 0.767 | 0.689 | 8.088  |
| NSSR     | 38.27  | 0.229 | 2.466 | 0.985 | 0.981 | 60.676 |
| FSF(Ours)| 43.077 | 0.141 | 2.058 | 0.993 | 0.989 | 0.26   |

Table 4. The quantitative results of the experiment on the CAVE dataset. The best results are shown in bold among all methods.

| Method   | PSNR  | ERGAS | SAM   | UIQI  | SSIM  | Time   |
|----------|-------|-------|-------|-------|-------|--------|
| Best Value | +∞   | 0     | 0     | 1     | 1     | 0      |
| CSTF     | 40.691 | 0.463 | 7.953 | 0.914 | 0.957 | 386.388|
| Hysure   | 40.263 | 0.421 | 15.23 | 0.934 | 0.957 | 239.961|
| LTTR     | 34.482 | 1.138 | 12.268| 0.81  | 0.914 | 768.284|
| NLSTF    | 24.63  | 1.521 | 14.475| 0.717 | 0.813 | 246.843|
| STEREO   | 31.275 | 0.695 | 17.171| 0.738 | 0.835 | 25.213 |
| NSSR     | 38.379 | 0.552 | 14.337| 0.922 | 0.952 | 125.605|
| FSF(Ours)| 41.067 | 0.388 | 13.926| 0.937 | 0.959 | 0.561  |

Table 4 corresponds to the quantitative metrics of the fusion results of the competing approaches on the CAVE dataset. In terms of the quantitative evaluation metrics of fusion,
our method achieves similar or even slightly better fusion results than Hysure and CSTF methods. However, in terms of the time required for fusion, Hysure and CSTF required 239.961 s and 386.388 s, respectively, while our proposed FSF method required only 0.561 s. Thus, the time required was less than 0.2% of that for CSTF and less than 0.3% of that for Hysure. In addition to FSF, the shortest fusion time, STEREO, also required 25.213 s. Thus, FSF needed only 1/20 of STEREO, while the fusion effect was much better than STEREO.

Figure 3. First row: LR-HSI, fusion images, and ground truth images’ false-color images (formed by bands 30th, 60th, and 90th) in Pavia university. Second row: LR-HSI, fusion images, and ground truth images’ false-color images (formed by bands 10th, 19th, and 28th) of oil_painting in the CAVE database.

Figure 4. First row: the competing approaches to reconstructed images of Pavia University at 30th band. Second row: the corresponding error images of the 30th band. Third row: the competing approaches to reconstructed images’ false-color images formed by bands 30th, 60th, and 90th. Fourth row: the corresponding error images of the false-color images.
Figure 5. First row: the competing approaches to reconstructed images of oil_painting in the CAVE database at 10th band. Second row: the corresponding error images of the 10th band. Third row: the competing approaches to reconstructed images’ false-color images formed by bands 10th, 19th, and 28th. Fourth row: the corresponding error images of the false-color images.

4.2. Stability of The FSF

Figure 6 shows the PSNR value of the fusion image when the downsampling factors are 4, 8, 16, 32, and 64 for the comparison method. Our proposed method FSF can maintain a high fusion effect under different downsampling factors, while other methods will decrease the fusion quality as the downsampling factor increases. The stability of our proposed method FSF is based on the fact that in matrix singular value decomposition, only the first few singular values need to be retained to recover the information of the original matrix. Furthermore, CSTF and Hysure are also stable for the downsampling factors since both methods also use SVD to extract part of the fusion information in the fusion process. On the other hand, the NLSTF method only achieves better fusion results when the downsampling factor is 4, while the fusion results are the worst in all other cases since NLSTF uses a nonlocal clustering operation before fusion, limiting the super-resolution capability.
5. Conclusions

In this paper, we propose a new fast and stable fusion algorithm based on matrix factorization. Unlike the traditional matrix factorization approach, our proposed factorization method takes the lead in using the Moore–Penrose inverse to simplify the estimation process of the spectral basis coefficient matrix, which significantly reduces the computation time of the fusion model. In addition, to further stabilize the quality of the fusion, we use two multiplicative iterative processes to optimize the quality of the fused image. In addition, we are the first to propose a different optimization method for the first generation of spectral basis and coefficient matrices to make the fusion results more accurate. Finally, our method is tested on two commonly used public datasets, verifying that our proposed method has better timeliness and stability than other state-of-the-art algorithms. In future research, more a priori information can be introduced into the fusion model proposed in this paper to improve the quality of fusion further.

Author Contributions: Conceptualization, Y.P. and T.Z.; methodology, J.L. (Jian Long); software, J.L. (Jian Long); validation, J.L. (Jian Long), J.L. (Jun Li) and L.Z.; data curation, J.L.; writing—original draft preparation, J.L. (Jian Long); writing—review and editing, Y.P., T.Z., J.L. (Jun Li) and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the National Key Research and Development Program of China (Nos. 2017YFB1301104 and 2017YFB1001900) and the National Natural Science Foundation of China (Nos. 91648204 and 61803375).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Sample Availability: The code in the article is available from the authors.
28. Li, S.; Dian, R.; Fang, L.; Bioucas-Dias, J.M. Fusing Hyperspectral and Multispectral Images via Coupled Sparse Tensor Factorization. *IEEE Trans. Image Process.* 2018, 27, 4118–4130. [CrossRef] [PubMed]
29. Dian, R.; Fang, L.; Li, S. Hyperspectral Image Super-Resolution via Non-local Sparse Tensor Factorization. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3862–3871. [CrossRef]
30. Kanatsoulis, C.I.; Fu, X.; Sidiropoulos, N.D.; Ma, W. Hyperspectral Super-Resolution: A Coupled Tensor Factorization Approach. *IEEE Trans. Signal Process.* 2018, 66, 6503–6517. [CrossRef]
31. He, W.; Chen, Y.; Yokoya, N.; Li, C.; Zhao, Q. Hyperspectral Super-Resolution via Coupled Tensor Ring Factorization. *arXiv* 2020, arXiv:2001.01547 [CrossRef]
32. Dell’Acqua, F.; Gamba, P.; Ferrari, A.; Palmason, J.A.; Benediktsson, J.A.; Arnason, K. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *IEEE Geosci. Remote Sens. Lett.* 2004, 1, 322–326. [CrossRef]
33. Yasuma, F.; Mitsunaga, T.; Iso, D.; Nayar, S.K. Generalized Assorted Pixel Camera: Postcapture Control of Resolution, Dynamic Range, and Spectrum. *IEEE Trans. Image Process.* 2010, 19, 2241–2253. [CrossRef] [PubMed]
34. Dian, R.; Li, S.; Fang, L. Learning a Low Tensor-Train Rank Representation for Hyperspectral Image Super-Resolution. *IEEE Trans. Neural Netw. Learn. Syst.* 2019, 30, 2672–2683. [CrossRef]
35. Wald, L. Quality of high resolution synthesised images: Is there a simple criterion? In Third Conference Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images; SEE/URISCA: Nice, France, 2000; pp. 99–103.
36. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* 2002, 9, 81–84. [CrossRef]
37. 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]