Hyperspectral Image Denoising Based on Nonlocal Low Rank Dictionary Learning

Zeng ZhiHua*, Zhou Bing and Li Cong

City College, Wuhan University of Science and Technology, Hubei Wuhan, 430000, China

Abstract: In allusion to hyperspectral remote sensing image denoising problem, the article proposes an image denoising algorithm based on nonlocal low rank dictionary learning. The basic thought of the algorithm is to make use of the strong correlation among various wavebands of the hyperspectral remote sensing image and meanwhile combine the nonlocal self-similarity and the local sparseness of an image to improve denoising performance. Firstly, combine the strong correlation of waveband images, the nonlocal self-similarity and the local sparseness to establish nonlocal low rank dictionary learning model. Then, adopt iterative method to solve the model to obtain redundant dictionary and sparse representation coefficient. Finally, adopt redundant dictionary and sparse representation coefficient to recover the image. Compared with existing advanced algorithms, due to the adoption of such strong correlation among various wavebands of the hyperspectral image, the algorithm mentioned in the article can well reserve the detailed information of the hyperspectral remote sensing image and improve visual effect. Meanwhile, the test result has verified the effectiveness of the algorithm mentioned in the article.

Keywords: Image Denoising, Hyperspectrum, Remote Sensing Image, Low Rank, Sparse Representation.

1. INTRODUCTION

Able to accurately and comprehensively describe ground object characteristics, hyperspectral remote sensing image is widely applied in agricultural field, geological prospecting, environmental monitoring, military investigation, etc. However, the severe pollution on hyperspectral remote sensing image during the acquisition and transmission processes can significantly influence the subsequent matching, integration, classification, etc [1]. Therefore, it is importantly significant to research hyperspectral remote sensing image denoising problem [2, 3].

Under the assumption that the width of the hyperspectral remote sensing image space is \( W \), the height is \( H \) and the wavebands dimension is \( S \) and the image noise is considered as additive noise [4], so the observation model is:

\[
Y = X + E \tag{1}
\]

Therein, \( Y \in R^{W \times H \times S} \) is the image polluted by noise, \( X \in R^{W \times H \times S} \) is original image, and \( E \in R^{W \times H \times S} \) is additive noise [5].

The existing hyperspectral remote sensing image denoising algorithms are basically divided into two types: independent waveband image denoising algorithm and simultaneous waveband image denoising algorithm. Therein, the independent waveband image denoising algorithm includes NLM algorithm, BM3D algorithm, K-SVD algorithm, NCSR algorithm, etc. [6-9], and the basic thought thereof is to regard the hyperspectral images for different wavebands as independent images and then adopt existing denoising algorithm. Specifically, NLM algorithm and BM3D algorithm adopt the nonlocal self-similarity characteristic of the non-adjacent part in the image for denoising, K-SVD algorithm adopts the local sparseness characteristic of the image for denoising, and NCSR adopts both the nonlocal self-similarity and the local sparseness for denoising [10-12]. But it is usually difficult for these algorithms to have good performance in practical application, because the internal characteristics of the hyperspectral remote sensing image are neglected in these algorithms. Additionally, the simultaneous waveband image denoising algorithm includes NLM3D algorithm and BM4D algorithm [13, 14], and the basic thought thereof is to adopt the internal characteristics of the multi-band images of the hyperspectral remote sensing image for denoising all waveband images simultaneously. NLM3D algorithm and BM4D algorithm are expanded respectively from NLM algorithm and BM3D algorithm, and aim at adopting the three-dimensional nonlocal self-similarity of the hyperspectral remote sensing image for denoising. Compared with independent waveband image denoising algorithms, these algorithms have better effect; but the strong correlation among the waveband images of the hyperspectral remote sensing image is not considered, so these algorithms still can be further improved.

In consideration of adopting the strong correlation among various waveband images of the hyperspectral remote sensing image and meanwhile combining the nonlocal self-similarity and the local sparseness of the image itself, the article proposes an image denoising algorithm based on non-
local low rank dictionary learning: firstly, establish nonlocal low rank dictionary learning model; then, construct corresponding algorithms to solve the model to realize image denoising. Due to the adoption of such strong correlation among various waveband images, the algorithm mentioned in the article can effectively reserve the texture and the detailed information of various waveband images [15-17].

2. DICTIONARY LEARNING

Firstly, simply review the dictionary learning technology. For a given group of samples \( \beta=(\beta_1, \cdots, \beta_N) \in R^{m \times k} \), the purpose of dictionary learning is to find out the redundant dictionary \( D \in R^{m \times k} \) in order to sparsely express each sample as sparsely representation matrix \( T=(T_1, T_2, \cdots, T_K) \in R^{m \times k} \), and this dictionary learning problem is expressed as:

\[
\begin{aligned}
& \underset{D, T}{\text{min}} \quad \| \beta - DT \|_F^2, \\
& \text{s.t.} \quad \| T_k \|_0 \leq T_0, \quad \forall k
\end{aligned}
\]  

(2)

The solving methods for dictionary learning problem (2) usually include MOD algorithm, K-SVD algorithm or online dictionary learning algorithm, etc.

The denoising algorithm in the article is as follows: firstly, establish the nonlocal low rank dictionary learning model for this algorithm; then, construct corresponding algorithms to solve the model to obtain redundant dictionary and sparseness representation coefficient; finally, adopt the results to recover the image [18, 19].

2.1. Nonlocal Low Rank Dictionary Learning Model

The data structure of the hyperspectral remote sensing image is a cube. It is assumed that the image includes \( N \) full-waveband cube data (hereinafter referred to as cube) of which the size is \( \sqrt{N} \times \sqrt{N} \times S \) and which are divided by \( S \) waveband images from the hyperspectral image, namely \( p_n \) (\( n=1, \cdots, N \)), then the dictionary learning can be expressed as:

\[
\begin{aligned}
& \underset{D, p_n}{\text{min}} \quad \sum_{m=1}^{N} \sum_{s=1}^{S} \| p_{n,s} - D \alpha_{n,s} \|_2^2, \\
& \text{s.t.} \quad \| \alpha_{n,s} \|_0 \leq T_0
\end{aligned}
\]  

(3)

Therein, \( p_{n,s} \in R^{m \times S} \) is the image block vector expression form corresponding to the \( s \)-th waveband in the \( n \)-th cube, \( D \) is the redundant dictionary and \( \alpha_{n,s} \) is the sparseness representation coefficient [20].

In consideration of adopting nonlocal self-similarity, K mean value clustering algorithm is used to divide \( N \) cube data into \( K \) types, then the cube data in each type is used to learn to obtain sub-dictionaries, and then the sub-dictionaries are used to express the data in this type. If the cube data included in each type is \( M_k \), then the nonlocal dictionary learning mode of the \( k \)-th type can be expressed as:

\[
\begin{aligned}
& \underset{D_k, \alpha_{m,s}}{\text{min}} \quad \sum_{m=1}^{M_k} \sum_{s=1}^{S} \| p^{(k)}_{m,s} - D_k \alpha^{(k)}_{m,s} \|_2^2, \\
& \text{s.t.} \quad \| \alpha^{(k)}_{m,s} \|_0 \leq T_0
\end{aligned}
\]  

(4)

Therein, \( p^{(k)}_{m,s} \) is the image block vector corresponding to the \( s \)-th waveband of the \( m \)-th cube in the \( k \)-th type, \( D_k \) is the sub-dictionary of the \( k \)-th type, and \( \alpha^{(k)}_{m,s} \) is corresponding sparseness representation coefficient. Since \( p^{(k)}_{m,s} \) can be expressed by \( D_k \), it is indicated that \( \| \alpha^{(k)}_{m,s} \|_0 \leq T_0 \) can be met. Therefore, the nonlocal dictionary learning model (4) of the \( k \)-th type is equivalent to:

\[
\begin{aligned}
& \underset{D_k, \alpha^{(k)}_{m,s}}{\text{min}} \quad \sum_{m=1}^{M_k} \sum_{s=1}^{S} \| p^{(k)}_{m,s} - D_k \alpha^{(k)}_{m,s} \|_2^2, \\
& \text{s.t.} \quad \| \alpha^{(k)}_{m,s} \|_0 \leq T_0
\end{aligned}
\]  

(5)

Since the waveband images of the hyperspectral image have strong correlation, the coefficient matrix \( \alpha^{(k)}_m = [\alpha^{(k)}_{m,1}, \cdots, \alpha^{(k)}_{m,S}] \) corresponding to the full-waveband data of the \( n \)-th cube is a low rank matrix. Add the low rank constraint of matrix \( \alpha^{(k)}_m \) to the nonlocal dictionary learning model (1) to obtain the nonlocal low rank dictionary learning model as follows [21]:

\[
\begin{aligned}
& \underset{D_k, \alpha^{(k)}_{m,s}}{\text{min}} \quad \sum_{m=1}^{M_k} \sum_{s=1}^{S} \frac{\lambda}{2} \| p^{(k)}_{m,s} - D_k \alpha^{(k)}_{m,s} \|_2^2 + \sum_{m=1}^{M_k} \text{rank} \left( \alpha^{(k)}_m \right), \\
& \text{s.t.} \quad \| \alpha^{(k)}_{m,s} \|_0 \leq T_0
\end{aligned}
\]  

(6)

Therein, \( \lambda \) is weighted parameter and \( \text{rank}(\bullet) \) is matrix rank. The matrix rank is usually approximated by nuclear norm \( \| \bullet \|_n \) (sum of matrix eigenvalues), so the nonlocal low rank dictionary learning model is equivalent to:

\[
\begin{aligned}
& \underset{D_k, \alpha^{(k)}_{m,s}}{\text{min}} \quad \sum_{m=1}^{M_k} \frac{\lambda}{2} \| p^{(k)}_{m,s} - D_k \alpha^{(k)}_{m,s} \|_2^2 + \sum_{m=1}^{M_k} \| \alpha^{(k)}_m \|_n, \\
& \text{s.t.} \quad \| \alpha^{(k)}_{m,s} \|_0 \leq T_0
\end{aligned}
\]  

(7)

The model is simplified as:

\[
\begin{aligned}
& \underset{D_k, \alpha^{(k)}_{m,s}}{\text{min}} \quad \sum_{m=1}^{M_k} \frac{\lambda}{2} \| p^{(k)}_{m,s} - D_k \alpha^{(k)}_{m,s} \|_2^2 + \| \alpha^{(k)}_m \|_n \}
\end{aligned}
\]  

(8)

In the matrixes \( p^{(k)}_m = [p^{(k)}_{m,1}, \cdots, p^{(k)}_{m,S}] \) and \( \alpha^{(k)}_m = [\alpha^{(k)}_{m,1}, \cdots, \alpha^{(k)}_{m,S}] \), \( p_n \) is the full-waveband data of the \( m \)-th cube in the \( k \)-th type, and \( \alpha^{(k)}_m \) is the corresponding coefficient matrix.
2.2. Solving of Nonlocal Low Rank Dictionary Learning Model

The nonlocal low rank dictionary learning model is solved basically through two iteration steps: the first step is to fix the sub-dictionary \( D_k \) and update coefficient matrix \( \alpha_m^{(k)} \), the second step is to fix coefficient matrix \( \alpha_m^{(k)} \) and update sub-dictionary \( D_k \).

Fix the sub-dictionary \( D_k \) to convert the full-waveband data of each cube \( p_m^{(k)} (m = 1, 2, \ldots, M_k) \) into minimization problem

\[
\min_{\alpha_m^{(k)}} \frac{\lambda}{2} \left\| p_m^{(k)} - D_k \alpha_m^{(k)} \right\|_F^2 + \left\| \alpha_m^{(k)} \right\|_F^2.
\]

(9)

Then, the optimization problem is actually converted into minimization problem. After auxiliary matrix \( \beta \) is introduced, formula (9) is equivalent to

\[
\min_{\alpha_m^{(k)}} \frac{\lambda}{2} \left\| p_m^{(k)} - D_k \alpha_m^{(k)} \right\|_F^2 + \left\| \alpha_m^{(k)} \right\|_F^2,
\]

s.t. \( \alpha_m^{(k)} - \beta = 0 \)

(10)

The corresponding augmented Lagrange function is

\[
L(\alpha_m^{(k)}, \beta, Z) = \left\| \alpha_m^{(k)} \right\|_F^2 + \frac{\lambda}{2} \left\| p_m^{(k)} - D_k \beta \right\|_F^2 + \left\| Z \right\|_F^2 + \frac{\mu}{2} \left\| \alpha_m^{(k)} - \beta \right\|_F^2.
\]

(11)

Therein, \( Z \) is Lagrange multiplier and \( \mu > 0 \) is fixed constant. Then, the minimization problem of formula (11) is converted into the minimization problem of Lagrange function.

\[
\min_{\alpha_m^{(k)}, \beta, Z} L(\alpha_m^{(k)}, \beta, Z)
\]

(12)

The rapidly converged alternating direction method (ADM) is used to solve the above corresponding optimization problem, with the specific steps as follows. The following iteration format of \( \alpha_m^{(k)} \) is solved through singular value thresholding (SVT):

\[
\alpha_m^{(k)} = \Gamma_{\mu^{-1}}(S) V^T
\]

(13)

Then, \( (U, S, V) = \text{SVD}(\beta - u^T Z) \), \( \Gamma_{\mu}(x) \) is:

\[
\Gamma_{\mu}(x) = \begin{cases} 
  x - 1/\mu & \text{if } x > 1/\mu \\
  x + 1/\mu & \text{if } x < 1/\mu \\
  0 & \text{otherwise}
\end{cases}
\]

Availably, the iteration formats of \( \beta \) and \( Z \) are

\[
\beta = \left(u I + \lambda D_k^T D_k\right)^{-1} \left(u \alpha_m^{(k)} - Z + \lambda D_k^T p_m^{(k)}\right)
\]

(14)

\[
Z = Z - u(\alpha_m^{(k)} - \beta)
\]

(15)

The coefficient matrix \( \alpha_m^{(k)} \) can be obtained through the iteration of formulae (13), (14) and (15).

Through coefficient matrix \( \alpha_m^{(k)} \), formula (15) is into:

\[
\min_{\beta} \left\| p^{(k)} - D_k^{(k)} \beta \right\|_F^2
\]

(16)

Therein, \( p^{(k)} = [p^{(k)}_1, \ldots, p^{(k)}_{M_k}] \) and \( \alpha^{(k)} = [\alpha^{(k)}_1, \ldots, \alpha^{(k)}_{M_k}] \).

The analytical solution expression of above problem is:

\[
D_k = p^{(k)} (\alpha^{(k)})^T (\alpha^{(k)} (\alpha^{(k)})^T)^{-1}
\]

(17)

Through the above two steps of iterations, obtain the sub-dictionary \( D_k \) of the \( k \)-th type and the coefficient matrix \( \alpha_m^{(k)} \) of the full-waveband data \( p_m^{(k)} \) of the \( m \)-th cube. Similarly, obtain the sub-dictionaries of all types and the coefficient matrix corresponding to the full-waveband data belonging to the corresponding type of the cube in order to obtain the sub-dictionary \( D_n \) corresponding to full-waveband data \( p_n \) of each cube and the corresponding coefficient matrix \( \alpha_n^{(k)} \).

2.3. Hyperspectral Remote Sensing Image Denoising

According to section 2.2, obtain the sub-dictionary \( D_n \) corresponding to the full-waveband data \( p_n \) of each cube and the corresponding coefficient matrix \( \alpha_n^{(k)} \) to estimate the full-waveband data \( \hat{p}_n = D_n \alpha_n^{(k)} \) of each recovered cube. Joint the estimated full-waveband data \( \hat{p}_n \) of the cubes according to the positions thereof and average the overlap parts of the cubes to obtain denoised image \( \hat{X} \).

3. DETAILED STEPS AND ANALYSIS OF ALGORITHM

3.1. Detailed Steps of Algorithm

The section will describe the detailed steps of the algorithm in the article, as shown in algorithm 1.

Algorithm 1: denoising algorithm based on nonlocal low rank dictionary learning
Algorithm input: noisy hyperspectral image \( \mathbf{Y} \in \mathbb{R}^{W \times H \times S} \);

Initialization: initial dictionary \( \mathbf{D}^{(0)} \);

Step 1 Divide hyperspectral image into \( N \) mutually overlapped cube full-waveband data \( \{p_n\} \);

Step 2 Divide all \( p_n \) into \( K \) types through K-average value clustering algorithm;

Step 3 Solve the nonlocal low rank dictionary learning models (9) of \( K \) types to obtain the coefficient matrix \( \alpha_n \) of \( p_n \) and the corresponding sub-dictionary \( \mathbf{D}_n \);

Step 4 Estimate the full-waveband data \( \hat{p}_n = \mathbf{D}_n \alpha_n \) of each cube, joint them according to the positions thereof and average the overlapped parts of the cubes to recover the image \( \hat{\mathbf{X}} \).

Output result: denoised image \( \hat{\mathbf{X}} \).

Note of algorithm 1:

(a) Selection of initial dictionary \( \mathbf{D}^{(0)} \): such rapidly realized data dictionary as DCT dictionary and wavelet dictionary shall be selected as initial dictionary \( \mathbf{D}^{(0)} \).

(b) Selection of parameter \( \lambda \): regard the nonlocal low rank dictionary learning problem as multiobjective optimization problem in order to obtain the parameter through \( \alpha \)-method.

3.2. Calculation Complexity Analysis

The calculation quantity of the algorithm in the article is mainly from solving the nonlocal low rank dictionary learning model. The calculation quantity of using AMD algorithm to solve formula (9) is \( O\left( (nSM_k) \right) \) and the calculation quantity of the inversion of formula (16) is also \( O\left( (nSM_k) \right) \). \( N \) times of solving processes are needed, so the total calculation quantity is \( O\left( N \left( nSM_k \right) \right) \).

4. EXPERIMENTAL RESULTS AND ANALYSIS

The section will verify the algorithm performance through experiments and compare K-SVD algorithm and BM4D algorithm. Experiment 1: compare the visual effects of the actual hyperspectral noise images after denoising; Experiment 2: compare the visual effects of the images artificially added with Gaussian noise after denoising; Experiment 3: present the influence of different noise mean square errors on algorithm performance; Experiment 4: present the influence of number of types on algorithm performance; Experiment 5: present the influence of number of overlapped pixels on algorithm performance.

Hyperspectral remote sensing image Indian Pines and Washington DC Mall are selected for the following experiments, wherein the image of Indian Pines includes 220 wavebands and the image of Washington DC Mall includes 191 wavebands, the cube size is \( 8 \times 8 \times 7 \), the overlapped pixels of adjacent cubes are 7, the number of types is \( K = 50 \), and the number of atoms is 512. All experiments are carried out in the computer with CPU of dual-core 3.0GHz and memory of 4Gb, and the algorithm reestablishment performance is measured by peak signal-to-noise ratio (PSNR).

Experiment 1: Compare with the Visual Effect of Actual Image

This experiment aims at comparing the denoising visual effects of three algorithms through actual image denoising. Indian Pines data are adopted and the effect of the denoised images of two representative wavebands ---- waveband 1 and waveband 109 is compared in (Fig. 1). According to (Fig. 1), compared with K-SVD algorithm and BM4D algorithm, the image denoised through the algorithm in the article is more detailed and clearer.

(a) Original Image (b) K-SVD Algorithm (c) BM4D Algorithm (d) Algorithm

Fig. (1). Real hyperspectral remote sensing image to compare the visual effect of noise.
Hyperspectral Image Denoising

The Open Automation and Control Systems Journal, 2015, Volume 7 1817

Experiment 2: Compare with the Visual Effect of the Gaussian Noise Image after Denoising

This experiment aims at denoising the image with Gaussian noise and accordingly comparing the denoising visual effects of the three algorithms. Washington DC Mall data without noise are added with Gaussian noise, the noise mean square error is taken as $\sigma = 30$. The denoising effects of the images of two wavebands (waveband 1 and waveband 100) are compared in (Fig. 1). According to (Fig. 2), compared with other two algorithms, the image denoised through the algorithm in the article is more detailed and clearer, and the visual effect is approximate to the original image.

Experiment 3: Influence of Different Noise Mean Square Errors on Algorithm Performance

The experiment aims at giving PSNR value of the algorithm under different noise mean square errors. The hyperspectral remote sensing image in Experiment 2 is added with Gaussian noise, and the comparison of PSNR values (for waveband 1 and waveband 100) of K-SVD algorithm, BM4D algorithm and the algorithm in the article under different noise mean square errors are as shown in (Table 1). According to (Table 1), due to the comprehensive consideration of the strong correlation among the images of different wavebands, compared with K-SVD algorithm and BM4D algorithm, the algorithm in the article has significantly improved PSNR value and also has strong noise adaptability.

Experiment 4: Influence of Number of Overlapped Pixels of Cube on Algorithm Performance

The experiment aims at showing the influence of the number of overlapped pixels of cube full-waveband data (abbreviated as cube) on algorithm performance. The hyperspectral remote sensing image in Experiment 2 is added with Gaussian noise, the mean square errors are respectively as $\sigma = 20$ and $\sigma = 30$, the number of overlapped pixels of cube is changed from 0 to 7, and other simulation conditions are not changed. PSNR value (mean value of all waveband images) of the algorithm in the article when the number of the overlapped pixels is changed is as shown in (Fig. 3). According to (Fig. 3), along with the increased number of overlapped pixels, the algorithm performance is gradually improved, and when the number of overlapped pixels is more than or equal to 6, the algorithm performance is stable.

Experiment 5: Influence of Number of Types $K$ on Algorithm Performance

The experiment aims at showing the influence of the number of types on algorithm influence. The hyperspectral remote sensing image in Experiment 2 is added with Gaussian noise, the mean square errors are respectively as $\sigma = 20$ and $\sigma = 30$, the number of types are from 10 to 100, and other simulation conditions are not changed. PSNR value (mean value of hyperspectral remote sensing image of all wavebands) of the algorithm in the article when the number of types is changed is as shown in (Fig. 4). According to

Table 1. Comparison of PSNR values of three algorithms.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>Waveband 1</th>
<th>Waveband 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K-SVD Algorithm</td>
<td>BM4D Algorithm</td>
</tr>
<tr>
<td>20</td>
<td>29.19</td>
<td>29.81</td>
</tr>
<tr>
<td>30</td>
<td>27.28</td>
<td>27.31</td>
</tr>
<tr>
<td>50</td>
<td>24.33</td>
<td>24.82</td>
</tr>
<tr>
<td>100</td>
<td>22.17</td>
<td>22.79</td>
</tr>
</tbody>
</table>

Fig. (2). Comparison of visual effects of gaussian noise eliminated images.
(Fig. 4), when the number of types is between 40 and 80, the algorithm in the article can obtain relatively good performance, and the small number of types will cause the insufficient similarity of the cubes of the same type while the large number of types will cause insufficient training samples.

CONCLUSION

In allusion to hyperspectral remote sensing image denoising problem, the article proposes an image denoising algorithm based on nonlocal low rank dictionary learning. The core thought of the algorithm is to make use of the strong correlation among various waveband images of hyperspectral remote sensing image and meanwhile combine the nonlocal self-similarity and the local sparseness of an image. Due to the adoption of such strong correlation of hyperspectral image, the algorithm mentioned in the article has relatively good effect. The experiment results show that PSNR value of the image recovered by the algorithm in the article is more than that of existing advanced algorithms and can well reserve the detailed information of the image and improve visual effect.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES


