



A Task-Driven Invertible Projection Matrix Learning Algorithm for Hyperspectral Compressed Sensing

Shaofei Dai ^{1,2,*}, Wenbo Liu ^{1,2}, Zhengyi Wang ^{1,2} and Kaiyu Li ^{1,2}

- ¹ College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211006, China; wenboliu@nuaa.edu.cn (W.L.); 13685111061@nuaa.edu.cn (Z.W.); LKY_401@nuaa.edu.cn (K.L.)
² Non-Destructive Testing and Monitoring Technology for High-Speed Transport Facilities Key Laboratory of Ministry of Industry and Information Technology, Nanjing 211006, China
* Correspondence: daishaofei@nuaa.edu.cn

Abstract: The high complexity of the reconstruction algorithm is the main bottleneck of the hyperspectral image (HSI) compression technology based on compressed sensing. Compressed sensing technology is an important tool for retrieving the maximum number of HSI scenes on the ground. However, the complexity of the compressed sensing algorithm is limited by the energy and hardware of spaceborne equipment. Aiming at the high complexity of compressed sensing reconstruction algorithm and low reconstruction accuracy, an equivalent model of the invertible transformation is theoretically derived by us in the paper, which can convert the complex invertible projection training model into the coupled dictionary training model. Besides, aiming at the invertible projection training model, the most competitive task-driven invertible projection matrix learning algorithm (TIPML) is proposed. In TIPML, we don't need to directly train the complex invertible projection model, but indirectly train the invertible projection model through the training of the coupled dictionary. In order to improve the accuracy of reconstructed data, in the paper, the singular value transformation is proposed. It has been verified that the concentration of the dictionary is increased and that the expressive ability of the dictionary has not been reduced by the transformation. Besides, two-loop iterative training is established to improve the accuracy of data reconstruction. Experiments show that, compared with the traditional compressed sensing algorithm, the compressed sensing algorithm based on TIPML has higher reconstruction accuracy, and the reconstruction time is shortened by more than a hundred times. It is foreseeable that the TIPML algorithm will have a huge application prospect in the field of HSI compression.

Keywords: compressed sensing; hyperspectral image; invertible projection; coupled dictionary; singular value; task-driven learning



Citation: Dai, S.; Liu, W.; Wang, Z.; Li, K. A Task-Driven Invertible Projection Matrix Learning Algorithm for Hyperspectral Compressed Sensing. *Remote Sens.* **2021**, *13*, 295. <https://doi.org/10.3390/rs13020295>

Received: 8 December 2020

Accepted: 14 January 2021

Published: 15 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Spectral images with a spectral resolution in the 10-2 order of magnitude are called hyperspectral images (HSI). HSI is one of the main components of modern remote sensing [1]. In recent decades, the popularity of hyperspectral technology has continued to rise. One of the main reasons for the higher visibility of hyperspectral imaging is the richness of the spectral information collected by this sensor. This function has positioned the hyperspectral analysis technology as the mainstream solution for land area analysis and the identification and differentiation of visually similar surface materials. Therefore, hyperspectral technology has become more and more important and is widely used in various applications, such as precision agriculture, environmental monitoring, geology, urban surveillance and homeland security, food quality inspection, etc. However, hyperspectral image processing is accompanied by a large amount of data management, which affects real-time performance on the one hand, and, on the other hand, the demand for on-board storage resources. In addition, the latest technological advances are introducing hyperspectral cameras with a higher spectrum and spatial resolution to the market. From

the perspective of onboard processing, communication, and storage, all of these make efficient data processing more challenging [2–5].

In order to solve the storage and transmission problems that need to be faced in processing multi-dimensional hyperspectral image data, data compression technology is usually selected. The original digital signal is refined and expressed, and its storage space and transmission bandwidth requirements are reduced by data compression technology. However, in traditional data compression technology, a large amount of redundant data is collected, and then compression technology is used to remove these redundant data. Although the redundant information of data is reduced, huge resources are wasted in this process [6].

Compressed Sensing (CS) theory was proposed in 2006 [7], and was widely used in wireless networks, imaging technology, target positioning, the direction of arrival estimation, and other fields [8,9].

As shown in Figure 1, in the compressed sensing technology, data compression is set up at the signal acquisition end, the information characteristics of the data are directly collected, and the original signal is reconstructed with high precision at the reconstruction end.

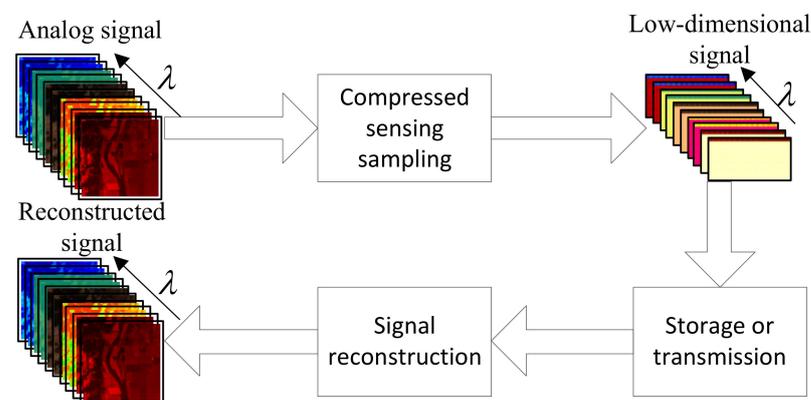


Figure 1. Block diagram of compressed sensing.

As shown in Equation (1), compressed sensing theory shows [7] that an over-complete dictionary $D \in R^{m \times s}$ is used, so that the original signal x is sparse enough under dictionary D , and its sparse coefficient is expressed as $\alpha \in R^s$. Equation (2) is the mathematical expression of compressed sensing. The original signal is observed by the observation matrix $\Phi \in R^{d \times m}$ to obtain the observation signal $y \in R^d$ ($d \ll m$). In this process, signal compression is realized. When the dimension d of the observation matrix is higher than a certain lower limit, the original signal x can be uniquely reconstructed from the observation signal y [10].

$$x = D\alpha \quad (1)$$

$$y = \Phi x \quad (2)$$

In CS theory, the compressibility, sparsity, and incoherence of the signal are fused, and in addition, compression and sampling are combined. The measurement signal is projected into the observation space containing all the effective information of the signal so that the sparse signal of limited dimensions can be sampled at a sampling frequency far less than the Nyquist theorem requires, making it possible to sample less than twice the original signal bandwidth [11].

The compression rate and reconstruction accuracy of the CS system mainly depend on the following three aspects:

(1) Sparse expression

To express the signal in a refined manner, the signal is usually transformed into a new basis or framework, which is also called a dictionary. The more sparsely the data is represented by the dictionary, the larger the compression ratio can be obtained under the condition of ensuring the image reconstruction error. Therefore, dictionary learning plays an important role in data compression based on sparse decomposition. At present, the commonly used dictionaries in compressed sensing include the discrete cosine transform (DCT) dictionary [12,13], the discrete wavelet transform (DWT) dictionary [9,14], and the K-SVD dictionary [15–17], Gabor dictionary [15], CDL dictionary [18], etc.

(2) Establishment of measurement matrix

When the original signal is compressed by the observation matrix, it must obey the Restricted Isometry Property (RIP). Under RIP conditions, the dimension d of the observation matrix of the CS system must be higher than a certain limit value. At the same time, the measurement matrix Φ and the sparse basis D meet incoherence, and the observation signal y can be uniquely reconstructed [7]. At present, in compressed sensing, the commonly used measurement matrices include random Gaussian matrix, random Bernoulli matrix, random Fourier matrix, and random Hadamard matrix [19,20], etc.

(3) Sparse reconstruction

The signal reconstruction problem of compressed sensing is to solve the underdetermined equations $y = \Phi x$ to obtain the original signal x based on the known measurement value y and measurement matrix Φ . In order to ensure that the original signal is efficiently and stably reconstructed, many excellent sparse signals have been proposed. Commonly used sparse reconstruction algorithms include Orthogonal Matching Pursuit (OMP), Stage-wise Orthogonal Matching Pursuit (StOMP), Compressive sampling matching pursuit (CoSaMP), Regularized Orthogonal Matching Pursuit (ROMP), Generalized Regularized Orthogonal Matching Pursuit (GROMP), Smoothed Projected Landweber (SPL) [21–25], etc.

Although compressed sensing technology has been proven to have great research significance in data compression, however, the features of HSI are complex. In order to improve the sparsity of the signal, the redundant dictionary with high redundancy is used as its sparse domain. When HSI is reconstructed, a large number of iterative calculations and inversion operations are required, and the traversal process of finding the optimal atom in each iteration is very time-consuming. A lot of time will be wasted in the sparse reconstruction process, which greatly limits the application of compressed sensing technology in the field of HSI compression.

Aiming at the time-consuming and low reconstruction accuracy of the reconstruction algorithm of compressed sensing technology, in the paper, we proposed a task-driven invertible projection matrix learning algorithm. The problems of long reconstruction time and insufficient reconstruction accuracy of HSI compression algorithm based on solving compressed sensing are solved by our proposed algorithm.

In the algorithm proposed in this paper, prior knowledge of data is used to train an invertible projection matrix $U \in \mathbb{R}^{d \times m}$. The projection matrix can project the original signal x into the low-dimensional observation signal y , and the projection formula can be expressed as $y = Ux$. Different from the traditional observation matrix Φ , the process in which the original signal x is projected by the projection matrix U to the low-dimensional observation signal y is invertible, and the inverse process can be expressed as $x' = U^T y$, where U^T represents the transposition of U , and x' is the reconstruction signal. Compared with the sparse reconstruction of traditional compressed sensing, when the algorithm proposed in this paper is used to train the projection matrix U as the observation matrix of the compressed sensing algorithm. In the signal reconstruction process, there is no need to perform tedious iteration and inversion operations, which results in a lot of time being saved and increases the real-time performance of the compressed sensing algorithm.

On the basis of this algorithm, we study a hyperspectral compressed sensing algorithm based on a task-driven invertible projection matrix learning algorithm. In order to prove the effectiveness of the algorithm proposed in this paper, this algorithm was compared with the most competitive compressed sensing algorithm based on DCT dictionary, CDL dictionary, and K-SVD dictionary. Besides, in order to verify the real-time performance of the signal reconstruction process of the compressed sensing algorithm based on the algorithm proposed in this paper, in the experiment, this algorithm was compared with the current best compressed sensing reconstruction algorithm, such as OMP, StOMP, CoSaMP, GOMP, GROMP, and SPL, etc. [21–25].

2. Principles and Methods

2.1. Constraints on Invertible Projection Transformation

Assuming that the projection matrix $\mathbf{U} \in \mathbf{R}^{d \times m}$, the original signal $\mathbf{x} \in \mathbf{R}^{m \times 1}$ can be accurately reconstructed from the low-dimensional signal \mathbf{y} by the projection transposed matrix \mathbf{U}^T , and the solution to the projection matrix \mathbf{U} can be described as

$$\mathbf{U} = \underset{\mathbf{U}}{\operatorname{argmin}} \left\| \mathbf{x} - \mathbf{U}^T \mathbf{U} \mathbf{x} \right\|_F^2 \quad (3)$$

Model (3) can be equivalent to

$$\begin{aligned} \mathbf{U} &= \underset{\mathbf{U}}{\operatorname{argmin}} \left\| \mathbf{x} - \mathbf{U}^T \mathbf{U} \mathbf{x} \right\|_F^2 \\ &= \underset{\mathbf{U}}{\operatorname{argmin}} \left\| \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{U}^T \mathbf{U} \mathbf{x} \right\|_F^2 \\ &= \underset{\mathbf{U}}{\operatorname{argmin}} \left\| \mathbf{x}^T \mathbf{x} - \mathbf{y}^T \mathbf{y} \right\|_F^2 \end{aligned} \quad (4)$$

where \mathbf{x} is the high-dimensional original signal, and $\mathbf{y} = \mathbf{U} \mathbf{x}$, \mathbf{y} is the low-dimensional signal of the high-dimensional original signal \mathbf{x} .

The high-dimensional signal \mathbf{x} and its low-dimensional signal \mathbf{y} have the same or similar sparse representation coefficients $\boldsymbol{\alpha} \in \mathbf{R}^{s \times 1}$ [26]. It can be described as

$$\begin{cases} \mathbf{x} = \mathbf{D} \boldsymbol{\alpha} \\ \mathbf{y} = \mathbf{P} \boldsymbol{\alpha} \end{cases} \quad (5)$$

where $\mathbf{D} \in \mathbf{R}^{m \times s}$ is the high-dimensional dictionary of high-dimensional signal \mathbf{x} , $\mathbf{P} \in \mathbf{R}^{d \times s}$ ($d \ll m$) is the low-dimensional dictionary of low-dimensional signal \mathbf{y} , and dictionary \mathbf{P} is constructed by \mathbf{D} through linear projection mapping \mathbf{U} . The projection formula is $\mathbf{P} = \mathbf{U} \mathbf{D}$ [26].

Therefore, model (3) can be described as

$$\begin{aligned} \mathbf{U} &= \underset{\mathbf{U}}{\operatorname{argmin}} \left\| \mathbf{x} - \mathbf{U}^T \mathbf{U} \mathbf{x} \right\|_F^2 \\ &= \underset{\mathbf{U}}{\operatorname{argmin}} \left\| \boldsymbol{\alpha}^T \mathbf{D}^T \mathbf{D} \boldsymbol{\alpha} - \boldsymbol{\alpha}^T \mathbf{P}^T \mathbf{P} \boldsymbol{\alpha} \right\|_F^2 \\ &= \underset{\mathbf{U}}{\operatorname{argmin}} \left\| \mathbf{D}^T \mathbf{D} - \mathbf{P}^T \mathbf{P} \right\|_F^2 \\ &= \underset{\mathbf{U}}{\operatorname{argmin}} \left\| \mathbf{D}^T \mathbf{D} - \mathbf{U} \mathbf{D} \right\|_F^2 \end{aligned} \quad (6)$$

From the derivation process of model (6) we can know that the optimal projection matrix \mathbf{U} under model (3) is equivalent to the optimal value under model (6). Since the high-dimensional dictionary \mathbf{D} can be regarded as fixed when solving model (6), the optimal projection matrix \mathbf{U} in the solution model (6) can be regarded as the optimal $\mathbf{P} = \mathbf{U} \mathbf{D}$, so

that model (6) is obtained the minimum Frobenius norm. The problem of calculating the projection matrix \mathbf{U} can be converted to calculating the low-dimensional matrix \mathbf{P} indirectly, which can be described as

$$\mathbf{P} = \underset{\mathbf{P}}{\operatorname{argmin}} \left\| \mathbf{D}^T \mathbf{D} - \mathbf{P}^T \mathbf{P} \right\|_F^2 \quad (7)$$

In model (7), \mathbf{P} can be considered as a principal components analysis (PCA) dimensionality reduction of dictionary \mathbf{D} . Although PCA has the best performance in the sense of mean square error, the PCA algorithm requires the covariance matrix of the signal to be calculated in advance, and many covariance calculations are required. Different from the idea of PCA, the low-dimensional dictionary $\mathbf{P} (\mathbf{P} = \mathbf{U}\mathbf{D})$ is trained by model (7) to make model (7) obtain the optimal value. From the derivation process of model (6), we can know that this projection matrix \mathbf{U} is also the optimal value of model (3). Therefore, Equations (3) and (7) have the same extreme points. In the process of training model (7), in order to ensure sufficient reconstruction accuracy of the signal, we only need to ensure that model (7) obtains the optimal value, then model (3) will also be optimal. Model (3), in order to solve the invertible transformation problem, is converted to the problem of solving model (7), that is, extracting principal components from dictionary \mathbf{D} . In order to ensure that more principal components of high-dimensional dictionary \mathbf{D} are retained in low-dimensional dictionary \mathbf{P} , dictionary \mathbf{D} is required to concentrate energy. We will give a solution in Section 2.2.

2.2. Singularity Transformation

For a general dictionary after dimensionality reduction, more principal components cannot be retained [16], so the singular value decomposition of dictionary \mathbf{D} can be described as

$$\mathbf{D} = \mathbf{M}\mathbf{\Lambda}\mathbf{V}^T = \begin{bmatrix} \mathbf{M}_h & \mathbf{M}_l \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_h & 0 \\ 0 & \mathbf{\Lambda}_l \end{bmatrix} \begin{bmatrix} \mathbf{V}_h & \mathbf{V}_l \end{bmatrix}^T \quad (8)$$

where \mathbf{M} is the left singular matrix of matrix \mathbf{D} , $\mathbf{\Lambda}$ is the singular value of matrix \mathbf{D} , \mathbf{V} is the right singular matrix of matrix \mathbf{D} , and $\mathbf{\Lambda}_h$ is the first h rows and h columns of $\mathbf{\Lambda}$.

Take $\mathbf{U} = \mathbf{M}_h^T$, then

$$\mathbf{D}^T \mathbf{D} = \mathbf{V}\mathbf{\Lambda}^T \mathbf{\Lambda}\mathbf{V}^T = \mathbf{V}(\mathbf{\Lambda}_h^T \mathbf{\Lambda}_h + \mathbf{\Lambda}_l^T \mathbf{\Lambda}_l)\mathbf{V}^T \quad (9)$$

$$\mathbf{P}^T \mathbf{P} = (\mathbf{U}\mathbf{D})^T \mathbf{U}\mathbf{D} = \mathbf{V}_h \mathbf{\Lambda}_h^T \mathbf{\Lambda}_h \mathbf{V}_h^T \quad (10)$$

Therefore, the first h singular values $\mathbf{\Lambda}_h$ of the matrix \mathbf{D} are larger (the last l singular values $\mathbf{\Lambda}_l$ are smaller and not 0), and more principal components of the high-dimensional dictionary \mathbf{D} are retained in the low-dimensional dictionary \mathbf{P} , in other words, $\| \mathbf{D}^T \mathbf{D} - \mathbf{P}^T \mathbf{P} \|_F^2$ is smaller.

In order to make the first h dimension of dictionary \mathbf{D} gather more energy, it should satisfy:

- (1) The expression coefficient of signal \mathbf{x} under dictionary \mathbf{D} is sparse enough;
- (2) the first h singular values of dictionary \mathbf{D} are large enough, and the last l singular values are small enough, and the value is not 0 [16].

The singular values often correspond to the important information implicit in the matrix, and the importance is positively related to the magnitude of the singular values. The more "singular" the matrix is, the less singular value contains more matrix information, and the smaller the information entropy of the matrix, and the more relevant its row (or column) vectors are [27]. Aiming at the characteristics of matrix singular values, in the paper, matrix singular transformation is proposed to increase the redundancy of the dictionary without reducing the expressive ability of the dictionary and the rank of the matrix.

The singular value transformation matrix Θ is defined as

$$\Theta(t, r) = Q(1, 0) \dots \underbrace{Q(n-t+1, r) \dots Q(n, r)}_t = \begin{bmatrix} 1 & 0 & \dots & r & \dots & r \\ 0 & 1 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 1-r & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & 1-r \end{bmatrix}_{n \times n} \quad (11)$$

where, the variables t and r are transformation adjustment parameters, and $Q(i, r)$ represents the values of the first row and the i -th row of the i -th column vector of the unit square $E_{n \times n}$ are r and $1 - r$, respectively.

Let the dictionary $D = [d_1 \ d_2 \ d_3 \ \dots \ d_n]$ and the original data set be $X = [x_1, x_2, \dots, x_n]$ ($x_i \in \mathbf{R}^m$). Assuming data x_i under dictionary D indicates that the sparse coefficient has four non-zero terms, expressed as $\alpha_i = [\dots \ a_0 \ \dots \ a_m \ \dots \ a_p \ \dots \ a_q \ \dots]^T$ ($\alpha_i \in \mathbf{R}^s$). The regulating parameter of singular value transformation $t = n - 1$.

The result of dictionary D undergoing singular value Θ transformation is:

$$\hat{D} = D\Theta(n-1, r) = [d_1 \ \text{d}w_2 \ \dots \ \text{d}w_n] \quad (12)$$

and

$$\text{d}w_k = rd_1 + (1-r)d_k \quad (13)$$

Assuming that the data x_i under the dictionary \hat{D} represents the sparsity coefficient $\beta = [b_1 \ b_2 \ b_3 \ \dots \ b_n]^T$, then:

$$D\alpha_i = a_0d_0 + a_md_m + a_pd_p + a_qd_q \quad (14)$$

$$\hat{D}\beta = b_1d_1 + b_2\text{d}w_2 + \dots + b_n\text{d}w_n \quad (15)$$

$$x_i = D\alpha_i = \hat{D}\beta \quad (16)$$

According to Equations (14)–(16), it can be determined that the coefficient of data x_i is still sparse under dictionary \hat{D} , and the sparse coefficient is:

$$\begin{cases} b_1 = \frac{r(a_0+a_m+a_p+a_q)}{1-r} \\ b_0 = \frac{a_0}{(1-r)} \\ b_m = \frac{a_m}{(1-r)} \\ b_p = \frac{a_p}{(1-r)} \\ b_q = \frac{a_q}{(1-r)} \\ b_i = 0 \\ i \neq 1, 0, m, p, q \end{cases} \quad (17)$$

Therefore, not only the redundancy of dictionary D can be increased by the singular value Θ transformation, which makes the energy of dictionary D more concentrated, but it also hardly affects the expression ability of dictionary D .

2.3. Task-Driven Invertible Projection Matrix Learning

In order to obtain a reversible projection transformation, we have to use prior knowledge to solve the invertible projection matrix. However, as we all know, it is very difficult to solve the low-dimensional invertible projection matrix U directly. Fortunately, it has been proven in Section 2.1 that the low-dimensional invertible projection matrix U can be solved indirectly by solving model (7). Model (7) is the problem of extracting principal components from dictionary D . Based on the theory in Section 2.2, a task-driven invertible projection matrix learning algorithm (TIPML) is proposed. In this algorithm, dictionary learning and singular value Θ transformation are used as the core of the algorithm, and a dual-loop iterative training mechanism is established based on the core.

As shown in Figure 2, in order to ensure that more principal components of high-dimensional dictionary D are retained in the low-dimensional dictionary P . In TIPML, the singular value Θ transformation is first used to make the energy of the dictionary D more concentrated. Next, the sparse representation coefficient A of the training data under the low-dimensional dictionary P is used to train the high-dimensional dictionary D . In this step, the low-dimensional dictionary P is used to train the high-dimensional dictionary D , which further increases the coupling between the two dictionaries and lets the principal components of the high-dimensional dictionary D be more retained in the low-dimensional dictionary P .

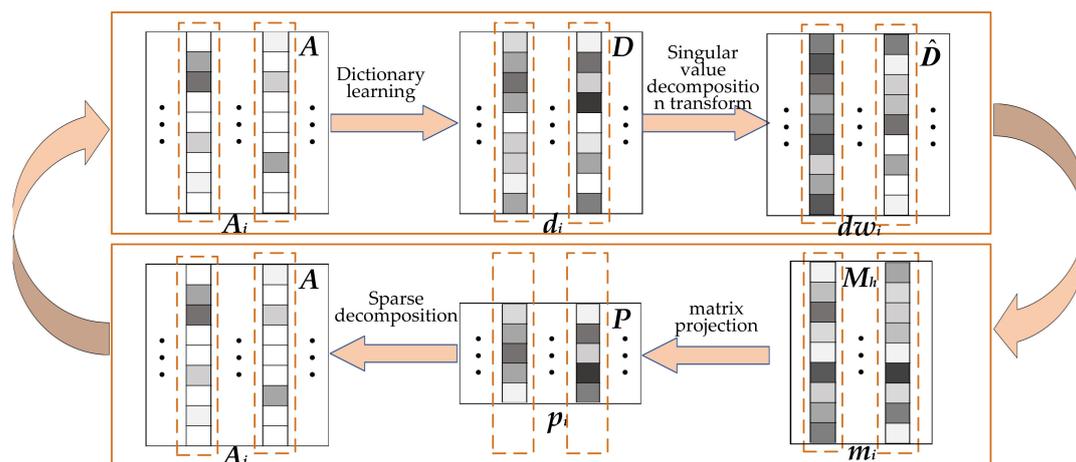


Figure 2. Block diagram of TIPML (task-driven invertible projection matrix learning algorithm).

The specific implementation steps of the TIPML algorithm are shown in Algorithm 1.

Algorithm 1: TIPML

Input: Training data set X , number of iterations T , the singular value Θ transformation parameter t, r , low-dimensional dictionary P dimension h , Signal sparsity K , and dimension of the data column N .

Output: Projection matrix U .

1: Initialization: Split the data set X into data columns x_i ($N \times 1$), $i = 1, 2, \dots$, i is the index of the data column x_i . Assuming $X = [x_1, x_2, \dots, x_n]$ ($x_i \in \mathbb{R}^N$). The initial dictionary D is set as a DCT dictionary, $D = [d_1, d_2, \dots, d_n]$, ($d_i \in \mathbb{R}^N$).

2: Repeat

3: Do singular value Θ transformation to dictionary D : $D = D\Theta(t, r)$.

4: Singular value decomposition: $D = M\Lambda V^T$.

5: Calculate low-dimensional dictionary: $P = M_h^T D$.

6: Based on the low-dimensional dictionary P , the OMP algorithm is used to sparse the data set X to obtain the sparse coefficient $A = [a_1, a_2, \dots, a_n]$, update the index $j = 1$ of the dictionary atom.

7: Repeat

8: The error is calculated: $E_j = X - \sum_{l \neq j} d_l a_l^T$.

9: The error is decomposed by SVD (rank-1 decomposition) into: $E_j \approx u\lambda v^T$.

10: Update dictionary: $d_j = u$, Update sparsity coefficient: $a_j = \lambda v$.

11: $j = j + 1$

12: Until $j > n$

13: Until $\frac{\sum_i \Lambda_i}{\sum_i \Lambda}$ is big enough, or reach the maximum number of iterations T .

14: Calculate the projection matrix: $U = M_h^T$.

3. Simulation Experiment and Results

In order to verify the effectiveness of the algorithm proposed in this paper, the AVIRIS hyperspectral data image Indian Pine dataset [28] was selected as the experimental dataset. These hyperspectral data include 220 bands, and each pixel is stored in a 16-bit integer format. In the experiment, the most competitive sparse expression dictionary and reconstruction algorithm in compressed sensing were selected and compared with the algorithm proposed in this paper. Sparse expression dictionaries included DCT dictionary, KSVD dictionary, and CDL dictionary, and reconstruction algorithms included StOMP, OMP, GOMP, GROMP, CoSaMP, and SPL. Besides, the data compression algorithm PCA [29] has also been selected for comparison with the algorithm proposed in this paper. The average peak signal-to-noise ratio (PSNR), Spectral Angle Mapper (SAM), and reconstruction time T were selected as the evaluation indicators of the experiment, and the PSNR and reconstruction time T of the reconstructed image was obtained by setting different sampling rates (SR). The software and hardware environment of the test experiment are CPU: Intel(R) Core (TM) i7-9750H 2.6GHz, 8GB memory, and Windows 10 and MatLab 2019a.

The PSNR calculation formula is

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}^2}{\frac{1}{m \cdot n} \sum_{i=1, j=1}^{m, n} [x(i, j) - x'(i, j)]^2} \right) \quad (18)$$

where x is the original image, x' is the reconstructed image, m and n respectively represent the length and width of the image, and MAX is the maximum value of image pixels.

The sampling rate calculation formula is

$$\text{SR} = \frac{M_\lambda}{N_\lambda} \quad (19)$$

where N_λ is the dimension of the original data, and M_λ is the dimension of compressed sensing sampling.

The SAM calculation formula is

$$\text{SAM} = \cos^{-1} \frac{x^T x'}{(x^T x)^{\frac{1}{2}} (x'^T x')^{\frac{1}{2}}} \quad (20)$$

where x is the original image, x' is the reconstructed image, x^T and x'^T are the transpose of matrix x and x' , respectively, and \cos^{-1} is the arc cosine function.

As shown in Figure 3, the simulation experiment mainly included three stages, namely offline learning, data sampling (encoding), and data reconstruction (decoding). In the experiment, the data of the Indian Pine data set was divided into two parts: The training data set and the test data set. The transformation adjustment parameters of the TIPML algorithm were set to 't = 255' and 'r = 0.3'. In addition, the data set was processed in blocks with a block size of 16×16 , and each small block was arranged into a column vector with a size of 256×1 . First, the training set Z was used as the TIPML training sample to obtain the projection matrix U . Then, the projection matrix U was used as a reduced-dimensional sampling matrix of the test data X to obtain the observation data Y . Finally, the transposed U^T of the projection matrix was used to back-project the low-dimensional sampling data Y to obtain the reconstructed data.

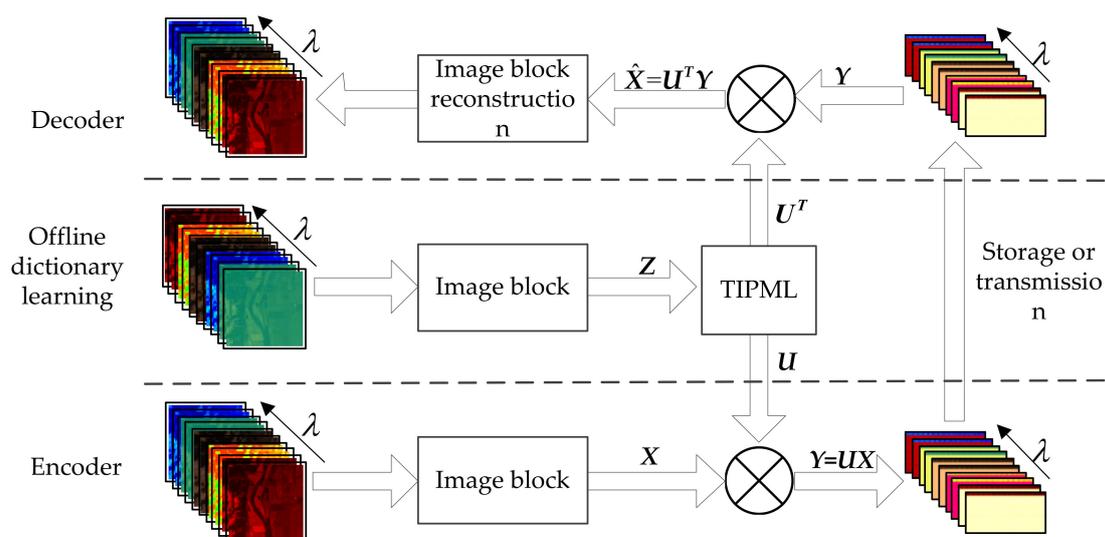


Figure 3. Flow chart of simulation experiment.

As shown in Figures 4–6, the PSNR of the reconstructed image of the algorithm proposed in this paper, CDL-OMP algorithm, and KSVD-StOMP algorithm under different sampling rates (SR) are respectively given. By comparing the reconstructed images, it can be known that the KSVD-StOMP algorithm has the lowest image reconstruction accuracy at the same sampling rate. The reconstruction accuracy of the algorithm proposed in this paper is visually similar to that of the CDL-OMP algorithm. At low sampling rates (SR = 0.1), there is still a good reconstruction effect.

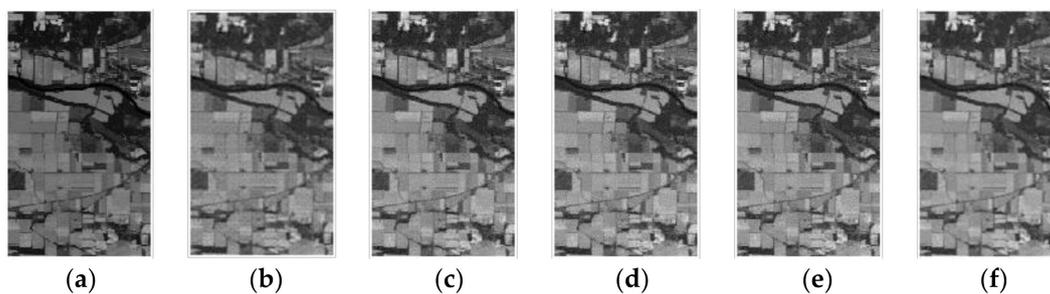


Figure 4. The original image and reconstructed image with band number 180, the compressed sensing algorithm is based on the TIPML algorithm, (a) The original images; (b) The sampling rate of the reconstructed image is SR = 0.1; (c) The sampling rate of the reconstructed image is SR = 0.2; (d) The sampling rate of the reconstructed image is SR = 0.3; (e) The sampling rate of the reconstructed image is SR = 0.4; (f) The sampling rate of the reconstructed image is SR = 0.5.

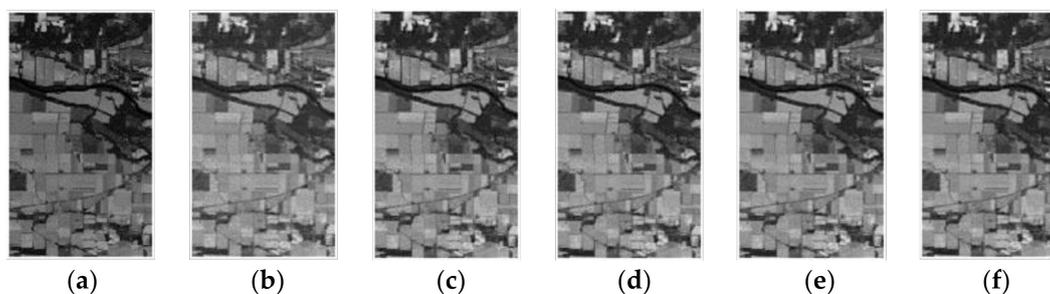


Figure 5. The Original image and reconstructed image with band number 180, the compressed sensing algorithm is based on the CDL-OMP algorithm, (a) The original images; (b) The sampling rate of the reconstructed image is SR = 0.1; (c) The sampling rate of the reconstructed image is SR = 0.2; (d) The sampling rate of the reconstructed image is SR = 0.3; (e) The sampling rate of the reconstructed image is SR = 0.4; (f) The sampling rate of the reconstructed image is SR = 0.5.

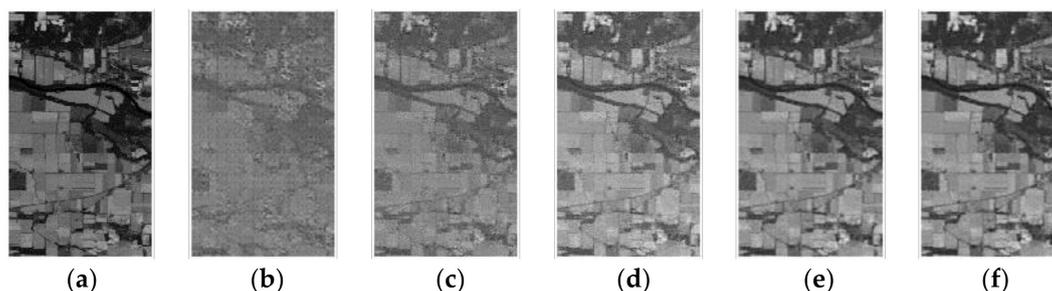


Figure 6. The Original image and reconstructed image with band number 180, the compressed sensing algorithm is based on KSVD-StOMP algorithm, (a) The original image; (b) The sampling rate of the reconstructed image is SR = 0.1; (c) The sampling rate of the reconstructed image is SR = 0.2; (d) The sampling rate of the reconstructed image is SR = 0.3; (e) The sampling rate of the reconstructed image is SR = 0.4; (f) The sampling rate of the reconstructed image is SR = 0.5.

As shown in Figure 7, at different sampling rates, the original spectral line and the reconstructed spectral line at the (256, 1) pixel point, the experimental comparison methods are the algorithm proposed in this paper, CDL-OMP algorithm, and KSVD-StOMP algorithm. Comparing the experimental results, it can be seen that when the sampling rate SR = 0.2 in Figure 7b, the reconstructed spectrum line of the algorithm proposed in this paper almost coincides with the original spectrum line. When the sampling rate is low, the reconstructed spectrum of the proposed algorithm and the original spectrum have some errors, but the reconstruction effect is still slightly better than the CDL-OMP algorithm, and far better than the KSVD-StOMP algorithm. This is because when SR = 0.1 in Figure 7a, the dimension of the projection matrix $M_\lambda = 26$, which is much smaller than the original signal dimension $N_\lambda = 256$. When the original signal is projected into a low-dimensional space, too much information is lost, resulting in reduced reconstruction accuracy.

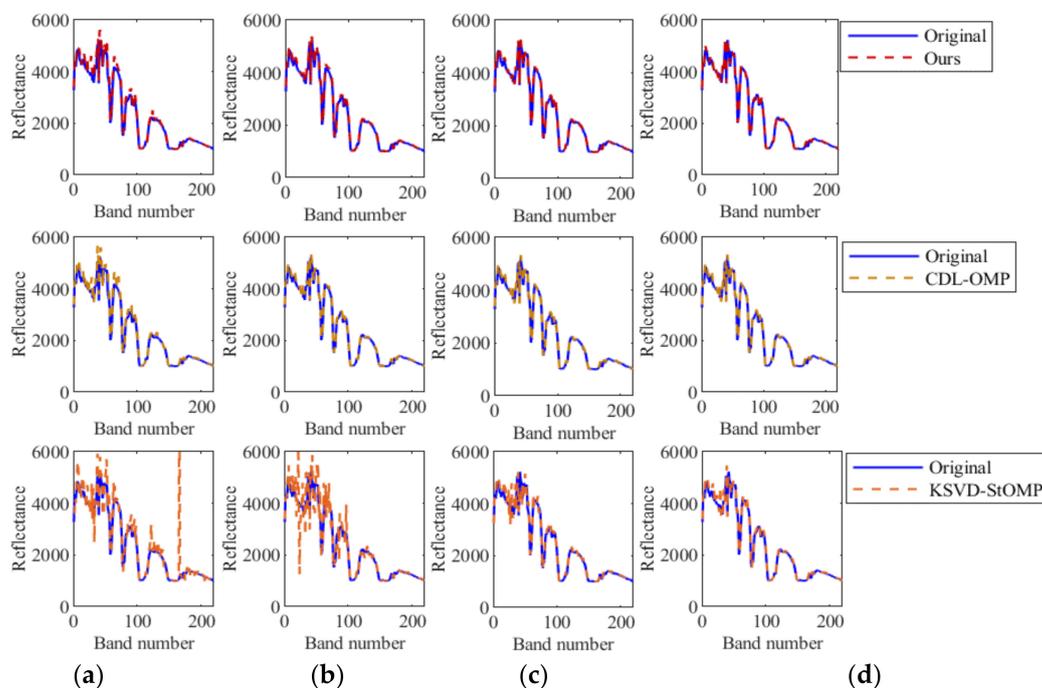


Figure 7. The original and reconstructed spectral lines of the pixel at coordinates (256, 1). From top to bottom, they are TIPML algorithm, CDL-OMP, and KSVD-StOMP. From the first column to the fourth column (a) The sampling rate SR = 0.1; (b) The sampling rate SR = 0.2; (c) The sampling rate SR = 0.3; (d) The sampling rate SR = 0.4.

In order to further verify the superiority of the algorithm proposed in this paper, more compressed sensing algorithms and PCA algorithm were selected. As shown in Table 1, the PSNR under the same sampling rate in the table is the average of all PSNRs of 220-band

images. The results in the comparison table show that the reconstructed image accuracy of the compressed sensing method based on the algorithm proposed in this paper is better than other algorithms, and it still has a higher reconstruction accuracy at low sampling rates.

Table 1. The reconstructed average peak signal-to-noise ratio (PSNR) (dB) of hyperspectral images in all bands under different sampling rates.

Methods	Sampling Rate				
	0.1	0.2	0.3	0.4	0.5
Ours	61.20	63.39	64.17	64.86	65.63
PCA	54.09	56.15	57.84	59.46	61.10
KSVD-StOMP	51.06	54.10	59.21	60.87	61.51
KSVD-OMP	31.37	28.30	26.93	27.08	27.20
KSVD-GOMP	33.85	30.31	29.02	28.61	28.74
KSVD-GROMP	34.39	33.59	33.57	35.04	35.72
KSVD-CoSaMP	48.70	49.61	50.30	50.18	50.26
DCT-SPL	38.27	46.83	54.10	58.81	60.70
DCT-StOMP	30.44	30.97	31.52	32.17	32.96
CDL-StOMP	60.67	63.06	63.33	63.33	63.33
CDL-OMP	60.67	63.06	63.33	63.33	63.33
CDL-CoSaMP	48.79	49.93	49.94	49.94	49.94
CDL-GROMP	14.83	22.90	28.39	33.27	37.21
CDL-GOMP	9.19	13.73	17.85	21.60	24.74

As shown in Table 2, the SAM under the same sampling rate in the table is the average of all SAMs of 220-band images. For the convenience of comparison, the SAM data in the table is displayed in the form of scientific notation. The results in the comparison table show that the reconstructed image SAM still has a lower SAM at low sampling rates. In other words, the reconstructed spectrum is more similar to the original spectrum.

Table 2. The reconstructed average Spectral Angle Mapper (SAM) (Scientific notation) of hyperspectral images in all bands under different sampling rates.

Methods	Sampling Rate				
	0.1	0.2	0.3	0.4	0.5
Ours	1.17×10^{-8}	1.16×10^{-8}	1.18×10^{-8}	1.18×10^{-8}	1.14×10^{-8}
PCA	1.20×10^{-8}	1.19×10^{-8}	1.22×10^{-8}	1.21×10^{-8}	1.18×10^{-8}
KSVD-StOMP	8.39×10^{-4}	1.16×10^{-8}	1.19×10^{-8}	1.18×10^{-8}	1.15×10^{-8}
KSVD-OMP	1.45×10^{-1}	6.06×10^{-3}	5.03×10^{-4}	1.18×10^{-8}	1.32×10^{-3}
KSVD-GOMP	5.34×10^{-2}	1.15×10^{-8}	4.82×10^{-3}	1.18×10^{-8}	1.97×10^{-3}
KSVD-GROMP	1.22×10^{-3}	1.16×10^{-8}	3.19×10^{-2}	1.17×10^{-8}	1.07×10^{-2}
KSVD-CoSaMP	1.16×10^{-8}	1.16×10^{-8}	1.19×10^{-8}	1.16×10^{-8}	1.15×10^{-8}
DCT-SPL	1.14×10^{-8}	1.18×10^{-8}	1.15×10^{-8}	1.17×10^{-8}	1.15×10^{-8}
CDL-StOMP	1.17×10^{-8}	1.16×10^{-8}	1.19×10^{-8}	1.17×10^{-8}	1.15×10^{-8}
CDL-OMP	1.17×10^{-8}	1.16×10^{-8}	1.19×10^{-8}	1.17×10^{-8}	1.15×10^{-8}
CDL-CoSaMP	1.65×10^{-3}	7.16×10^{-1}	1.18×10^{-8}	1.75×10^{-2}	8.31×10^{-2}
CDL-GROMP	4.55×10^{-4}	1.20×10^{-4}	1.32×10^{-3}	9.83×10^{-4}	2.88×10^{-4}
CDL-GOMP	4.55×10^{-4}	1.20×10^{-4}	1.32×10^{-3}	9.83×10^{-4}	2.88×10^{-4}

For compressed sensing algorithms, the running time T of the reconstruction algorithm is a very important evaluation index. As shown in Table 3, the running time T under the same sampling rate in the table is the mean value of the running time of 220-band image reconstruction, and the time unit recorded in the table is milliseconds. By comparing the experimental results, it can be seen that the running time T of the algorithm proposed in this paper is much lower than that of the traditional compressed sensing algorithm, and the running time is shortened by at least a hundred times. This is because the projection matrix

U trained by the algorithm proposed in this paper is approximately invertible. In the reconstruction process, only the low-dimensional sampling matrix Y is left multiplied by U^T . However, traditional compression sparse reconstruction algorithms require multiple matrix inversion and iterative operations. These operations are very time-consuming. Therefore, the real-time performance of the algorithm proposed in this paper is much higher than other algorithms. Besides, compared to data compression algorithm PCA, when we know the covariance matrix of the source in advance and find the eigenvalues, the time consumption of the TIPML algorithm is similar to that of the PCA algorithm. However, in practical applications, the covariance and eigenvalue of the source are unknown. We need to calculate the covariance and eigenvalue of the source, which will cause a lot of time to be consumed. Therefore, compared to traditional compressed sensing algorithms and traditional PCA data compression algorithms, the TIPML algorithm is more competitive.

Table 3. The reconstructed average time (millisecond) of hyperspectral images in all bands under different sampling rates.

Methods	Sampling Rate				
	0.1	0.2	0.3	0.4	0.5
Ours	0.3	0.3	0.4	0.4	0.4
PCA	0.3	0.4	0.4	0.4	0.4
KSVD-StOMP	210.8	443.7	667.1	934.6	1235.0
KSVD-OMP	264.9	401.8	545.3	594.0	613.5
KSVD-GOMP	101.7	170.4	206.2	246.8	252.3
KSVD-GROMP	84.6	131.7	181.0	186.7	189.3
KSVD-CoSaMP	302.7	527.7	597.3	829.7	877.4
DCT-SPL	354.8	355.1	354.8	357.0	359.1
DCT-StOMP	199.4	437.6	732.2	1085.2	1517.4
CDL-StOMP	228.2	503.3	763.0	1054.0	1384.2
CDL-OMP	209.5	462.7	696.1	955.6	1261.4
CDL-CoSaMP	341.3	553.0	635.0	895.7	943.9
CDL-GROMP	75.5	111.8	150.4	153.0	157.3
CDL-GOMP	167.7	253.3	317.4	365.8	375.5

4. Conclusions

HSI is the main tool for remote sensing and earth observation. The amount of valuable industrial and scientific data retrieved on the ground can be greatly increased by data compression technology. The resources of spaceborne equipment are very precious, and compression is placed on the sampling end by compressed sensing technology, which can save a lot of time and resources for hyperspectral imaging technology. Therefore, compressed sensing technology has huge application prospects in the field of hyperspectral compression. However, compressed sensing technology needs to solve underdetermined equations. In traditional algorithms, a lot of time and storage resources are spent in the sparse reconstruction process. Therefore, the high complexity of data reconstruction is also the biggest drawback of compressed sensing technology.

Aiming at the high computational complexity of compressed sensing technology and insufficient reconstruction accuracy, a task-driven invertible projection matrix learning algorithm was proposed by us. Our main contribution:

- (1) Derived the equivalent model of the invertible projection model theoretically, which converts the complex invertible projection training model into a coupled dictionary training model;
- (2) proposed a task-driven invertible projection matrix learning algorithm for invertible projection model training;
- (3) based on a task-driven reversible projection matrix learning algorithm, established a compressed sensing algorithm with strong real-time performance and high reconstruction accuracy.

Experimental verification has shown that the compressed sensing technology based on the algorithm proposed in this paper not only has higher reconstruction accuracy than traditional compressed sensing technology, but also has improved real-time performance by more than a hundred times. It is foreseeable that the algorithm proposed in this paper will have great application prospects in the field of hyperspectral image compression. In addition, the algorithm proposed in this paper can't only be used in the fields of one-dimensional signal compression, two-dimensional image compression, and data denoising, but it will also have greater research value on hardware platforms such as FPGA and embedded devices due to its extremely low the complexity.

Author Contributions: Conceptualization, S.D.; methodology, S.D.; software, S.D.; validation, S.D.; formal analysis, S.D.; investigation, S.D.; resources, Z.W.; data curation, S.D.; writing—original draft preparation, S.D.; writing—review and editing, S.D.; visualization, S.D.; supervision, W.L and K.L.; project administration, W.L. and K.L.; funding acquisition, W.L. and K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program of China grant number 2018YFB2003304, National Natural Science Foundation of China grant number 61871218, Fundamental Research Funds for the Central Universities grant number NJ2019007 and NJ2020014, National Key Research and Development Program of China grant number 2017YFF0107304, 2017YFF0209700, 2016YFB1100205, and 2016YFF0103702.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data included in this study are available upon request by contact with the corresponding author.

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

References

- Parente, M.; Kerekes, J.; Heylen, R. A Special Issue on Hyperspectral Imaging [From the Guest Editors]. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 6–7. [[CrossRef](#)]
- Vohland, M.; Jung, A. Hyperspectral Imaging for Fine to Medium Scale Applications in Environmental Sciences. *Remote Sens.* **2020**, *12*, 2962. [[CrossRef](#)]
- Saari, H.; Aallos, V.-V.; Aukujärvi, A.; Antila, T.; Holmlund, C.; Kantojärvi, U.; Mäkynen, J.; Ollila, J. Novel Miniaturized Hyperspectral Sensor for UAV and Space Applications. In Proceedings of the SPIE 7474, Sensors, Systems, and Next-Generation Satellites XIII, 74741M, SPIE Remote Sensing, Berlin, Germany, 31 August–3 September 2009. [[CrossRef](#)]
- Renhorn, I.G.E.; Axelsson, L. High spatial resolution hyperspectral camera based on exponentially variable filter. *Opt. Eng.* **2019**, *58*, 103106. [[CrossRef](#)]
- Pu, H.; Lin, L.; Sun, D.-W. Principles of Hyperspectral Microscope Imaging Techniques and Their Applications in Food Quality and Safety Detection: A Review. *Compr. Rev. Food Sci. Food Saf.* **2019**, *18*, 853–866. [[CrossRef](#)] [[PubMed](#)]
- Wang, L.; Lu, K.; Liu, P. Compressed Sensing of a Remote Sensing Image Based on the Priors of the Reference Image. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 736–740. [[CrossRef](#)]
- Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]
- Hsu, C.; Lin, C.; Kao, C.; Lin, Y. DCSN: Deep Compressed Sensing Network for Efficient Hyperspectral Data Transmission of Miniaturized Satellite. *IEEE Trans. Geosci. Remote Sens.* **2020**. [[CrossRef](#)]
- Biondi, F. In *Compressed Sensing Radar—New Concepts of Incoherent Continuous Wave Transmissions*; IEEE: Piscataway, NJ, USA, 2015; pp. 204–208.
- Candes, E.J.; Romberg, J.; Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **2004**, *52*, 489–509. [[CrossRef](#)]
- Candes, E.J.; Tao, T. Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies. *Inf. Theory IEEE Trans.* **2006**, *52*, 5406–5425. [[CrossRef](#)]
- Xue, H.; Sun, L.; Ou, G. Speech reconstruction based on compressed sensing theory using smoothed L0 algorithm. In Proceedings of the 2016 8th International Conference on Wireless Communications & Signal Processing (WCSP), Yangzhou, China, 13–15 October 2016; pp. 1–4. [[CrossRef](#)]
- Luo, H.; Zhang, N.; Wang, Y. Modified Smoothed Projected Landweber Algorithm for Adaptive Block Compressed Sensing Image Reconstruction. In Proceedings of the 2018 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, 16–17 July 2018; pp. 430–434. [[CrossRef](#)]

14. Matin, A.; Dai, B.; Huang, Y.; Wang, X. Ultrafast Imaging with Optical Encoding and Compressive Sensing. *J. Light. Technol.* **2019**, *37*, 761–768. [[CrossRef](#)]
15. Dao, P.; Li, X.; Griffin, A. Quantitative Comparison of EEG Compressed Sensing using Gabor and K-SVD Dictionaries. In Proceedings of the 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), Shanghai, China, 19–21 November 2018; pp. 1–5. [[CrossRef](#)]
16. Sana, F.; Katterbauer, K.; Al-Naffouri, T. Enhanced recovery of subsurface geological structures using compressed sensing and the Ensemble Kalman filter. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 3107–3110. [[CrossRef](#)]
17. Dehkordi, R.; Khosravi, H.; Ahmadyfard, A. Single image super-resolution based on sparse representation using dictionaries trained with input image patches. *IET Image Process.* **2020**, *14*, 1587–1593. [[CrossRef](#)]
18. Silong, Z.; Yuanxiang, L.; Xian, W. Nonlinear dimensionality reduction method based on dictionary learning. *Acta Autom. Sin.* **2016**, *42*, 1065–1076.
19. Vehkaperä, M.; Kabashima, Y.; Chatterjee, S. Analysis of Regularized LS Reconstruction and Random Matrix Ensembles in Compressed Sensing. *IEEE Trans. Inf. Theory* **2016**, *62*, 2100–2124. [[CrossRef](#)]
20. Ziran, W.; Huachuang, W.; Jianlin, Z. Structural optimization of measurement matrix in image reconstruction based on compressed sensing. In Proceedings of the 2017 7th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC), Macau, China, 21–23 July 2017; pp. 223–227. [[CrossRef](#)]
21. Zhang, W.; Tan, C.; Xu, Y. Electrical Resistance Tomography Image Reconstruction Based on Modified OMP Algorithm. *IEEE Sens. J.* **2019**, *19*, 5723–5731. [[CrossRef](#)]
22. Qian, Y.; Zhu, D.; Yu, X. SAR high-resolution imaging from missing raw data using StOMP. *J. Eng.* **2019**, *2019*, 7347–7351. [[CrossRef](#)]
23. Davenport, M.; Needell, D.; Wakin, M. Signal Space CoSaMP for Sparse Recovery with Redundant Dictionaries. *IEEE Trans. Inf. Theory* **2013**, *59*, 6820–6829. [[CrossRef](#)]
24. Chen, S.; Feng, C.; Xu, X. Micro-motion feature extraction of narrow-band radar target based on ROMP. *J. Eng.* **2019**, *2019*, 7860–7863. [[CrossRef](#)]
25. Trevisi, M.; Akbari, A.; Trocan, M. Compressive Imaging Using RIP-Compliant CMOS Imager Architecture and Landweber Reconstruction. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 387–399. [[CrossRef](#)]
26. Wei, X. Reconstructible Nonlinear Dimensionality Reduction via Joint Dictionary Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 175–189. [[CrossRef](#)] [[PubMed](#)]
27. Shufang, X. *Theory and Method of Matrix Calculation*; Peking University Press: Beijing, China, 1995; p. 7.
28. Baumgardner, M.F.; Biehl, L.L.; Landgrebe, D.A. 220 Band AVIRIS Hyperspectral Image Data Set: 12 June 1992 Indian Pine Test Site 3. In *Purdue University Research Repository*; Purdue University: West Lafayette, IN, USA, 2015. [[CrossRef](#)]
29. Lee, C.; Youn, S.; Jeong, T.; Lee, E.; Serra-Sagristà, J. Hybrid Compression of Hyperspectral Images Based on PCA with Pre-Encoding Discriminant Information. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1491–1495. [[CrossRef](#)]