Hyperspectral image super-resolution through clustering-based sparse representation

Fenghua Guo[1] · Caiming Zhang[1] · Mingli Zhang[2]

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
Promoting the spatial resolution of hyperspectral sensors is expected to improve computer vision tasks. However, due to the physical limitations of imaging sensors, the hyperspectral image is often of low spatial resolution. In this paper, we propose a new hyperspectral image super-resolution method from a low-resolution (LR) hyperspectral image and a high resolution (HR) multispectral image of the same scene. The reconstruction of HR hyperspectral image is formulated as a joint estimation of the hyperspectral dictionary and the sparse codes based on the spatial-spectral sparsity of the hyperspectral image. The hyperspectral dictionary is learned from the LR hyperspectral image. The sparse codes with respect to the learned dictionary are estimated from LR hyperspectral image and the corresponding HR multispectral image. To improve the accuracy, both spectral dictionary learning and sparse coefficients estimation exploit the spatial correlation of the HR hyperspectral image. Experiments show that the proposed method outperforms several state-of-art hyperspectral image super-resolution methods in objective quality metrics and visual performance.

Keywords Hyperspectral imaging · Sparse representation · Structural prior

1 Introduction

Hyperspectral imaging sensors provide the ability to sample a scene's spectral properties more densely. Whereas a normal RGB imaging sensors roughly divides the observation into red, green, and blue, a hyperspectral imaging sensors can easily obtain thirty or more distinct bands across the visible spectrum. Having detailed spectral information can be very useful

✉ Caiming Zhang
czhang@sdu.edu.cn

1 School of Software, Shandong University, Jinan 250101, People's Republic of China

2 McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada

to remote sensing [44] and computer vision tasks ranging from object recognition and tracking [9], as RGB alone is often insufficient to identify the materials within a scene. However, due to the physical limitations of imaging sensors, the hyperspectral image (HSI) is often of low spatial resolution. This is due to the fact that hyperspectral imaging systems need a large number of exposures to simultaneously acquire many bands within a narrow spectral window. To ensure sufficient signal-to-noise ratio, long exposures are often necessary, resulting in the sacrifice of spatial resolution. Simply increasing the spatial resolution of image sensors would not be effective for hyperspectral imaging because the average amount of photons reaching the sensors would be further reduced leading to even lower signal-to-noise ratio. Consequently, the low spatial resolution hyperspectral images (LR-HSIs) are often fused with high spatial resolution multispectral images (HR-MSIs) to reconstruct high spatial resolution hyperspectral images (HR-HSIs). This procedure is referred to as HSI super-resolution or HSI-MSI fusion.

The goal of this paper is to recover a HR-HSI $Z \in \mathrm{R}^{L \times N}$ from a LR-HSI $X \in \mathrm{R}^{L \times n}$ and a HR-MSI image $Y \in \mathrm{R}^{l \times N}$ of the same scene, where L is the number of spectral bands of Z ($L \gg l$), N and n ($N = W \times H$, $n = w \times h$, $w \ll W$ and $h \ll H$) denote pixel number in Z and X respectively. The reconstruction of Z from X and Y is an ill-posed inverse problem. For such ill-posed inverse problems, regularization is a popular tool for exploiting the prior knowledge about the unknown. Sparsity prior has been shown effective for solving hyperspectral image reconstruction [1, 2, 22, 24, 39], the target image Z can be written as [17, 23],

$$Z = DA + E \tag{1}$$

where $D = [d_1, \cdots, d_K] \in \mathrm{R}_+^{L \times K}$ is the spectral dictionary, $K(K \geq L)$ is the number of atoms, $A = [a_1, \cdots, a_N] \in \mathrm{R}_+^{K \times N}$ is the sparse coefficient matrix, and E is the approximation error. Both X and Y result from linear degradations of Z,

$$X = ZH + E_1, Y = PZ + E_2 \tag{2}$$

where $H \in \mathrm{R}^{N \times n}$ denotes the blurring and down-sampling degradation operator, P is a spectral transformation matrix, E_1 and E_2 denote the approximation error matrix. Many LR-HSI and HR-MSI fusion methods rely on a similar linear model of (2) [8, 24, 28, 29, 31, 35, 36, 38, 42, 46]. Combining (1) and (2) leads to

$$X = DAH + E_1 = DB + E_1, Y = PDA + E_2 = \overline{D}A + E_2 \tag{3}$$

where $B = AH \in \mathrm{R}^{K \times n}$ denotes the transformed sparse coefficient matrix, $\overline{D} = PD$ denotes the transformed spectral dictionary. The spectral dictionary D and coefficient matrix A in (3) are unknown. For scenes satisfying the sparsity assumption, coefficient matrix A have to be sparse. It follows that D and A can be solved by sparse matrix decomposition.

Besides the spectral information, some methods [8, 30–32, 36] also used the spatial structure to regularize the fusion problem. Dong et al. [8] proposed a non-negative structured sparse representation (NSSR) method which exploited the clustering-based sparsity of hyperspectral images. The limitation of [8] is that the structured sparse representation is only used to estimate the coefficient matrix, not for the spectral dictionary learning.

In this paper, we present a novel hyperspectral image super-resolution method. The main contributions of our work are as follows.

(1) A non-negative clustering-based sparse representation (NNCSR) model is proposed. The reconstruction of HR hyperspectral image is formulated as a joint estima-tion of the

hyperspectral dictionary and the sparse coefficients based on the spatial-spectral sparsity of the hyperspectral image. To improve the accuracy, both spectral dictionary and sparse coefficients exploit the clustering-based sparsity of hyperspectral images.

(2) We present an experimental valuation, where the proposed method is compared against several state-of-the-art approaches. Results of these experiments show our method outperforms these state-of-the-art approaches in objective quality metrics and visual performance.

2 Related work

Reconstructing a HR-HSI from a LR-HSI and HR-MSI, although challenging, is a crucial inverse problem that has been addressed in various scenarios [18–21]. As this inverse problem is generally ill-posed, introducing prior distributions to regularize the target image has been widely explored and some methods use Bayesian inference to regularize the problem [2, 12, 36, 37]. Based on prior knowledge and on the observation model, these Bayesian fusion methods build the posterior distribution, which is the Bayesian inference engine. The work [12, 36] computed the maximum a posteriori (MAP) estimator in an optimization framework to solve the fusion problem. The work [37] formed the likelihoods of the observations and developed a fast Sylvester based solver. Akhtar et al. [2] proposed a method using non-parametric Bayesian dictionary learning.

Many fusion methods are based on matrix factorization (MF) [1, 8, 22, 24, 26, 30–32, 36, 38, 39, 42] or nonnegative matrix under- approximation (NMU) [4]. Assuming that the HR-HSI only contains a small number of pure spectral signatures, these approaches first unfold the HR-HSI as a matrix and then factor it into a basis matrix and a coefficient matrix. The work [22] firstly introduced MF into fusion by factorizing the HSI into a dictionary of basis vectors and a set of sparse coefficients. The work [39] proposed a non-negative sparse framework to integrate LR-HSI and RGB data into a HR hyperspectral set of data. Akhtar et al. [1] exploited signal sparsity, non-negativity and the spatial structure in the scene. A coupled non-negative matrix factorization (CNMF) method [42] alternately applied NMF unmixing to LR hyperspectral and HR multispectral data. Since non-negative matrix factorization is not always unique [25, 26], the results of [42] are often not satisfactory. A similar fusion and unmixing framework was introduced in [24]. The common point of [24, 42] is to learn the spectral bases from the LR-HSI and sparse coefficients from the HR-MSI alternately instead of using both LR-HSI and HR-MSI jointly. To fully exploit the data information, Wei et al. [38] used an optimization formulation similar to [24, 42], but both LR-HSI and HR-MSI images contributed to the estimation of spectral bases and sparse coefficients.

Recently, the work [6, 7, 27, 40] used tensor factorization. Different from MF based methods, the hyperspectral image was approximated by a core tensor multiplied by dictionaries of the width, height, and spectral modes. Dian et al. [7] first proposed a nonlocal sparse tensor factorization for HSI-MSI fusion, where they approximate the HR-HSI by dictionaries of three modes and a sparse core tensor. Li et al. [27] solved the fusion problem by simultaneously conducting sparse Tucker decomposition on the HR-MSI and LR-HSI, where the core tensor and three dictionaries are alternatively updated until convergence. Xu et al. [40] propose a non-local tensor sparse representation model, and the main difference between [7, 40] is that they use different tensor sparse representation model.

Besides the spectral information, some methods [6, 8, 30–32, 36] also used the spatial structure to regularize the fusion problem. Veganzones et al. [32] exploited the low rank of the

hyperspectral images, while the work [30, 36] exploited the low intrinsic dimensionality of hyperspectral images. Dong et al. [8] proposed a non-negative structured sparse representation (NSSR) method. They exploited the clustering-based sparsity of hyperspectral images. A similar work was proposed in [31]. The limitation of [8, 31] is that the structured sparse representation is only used to estimate the coefficient matrix, not for the spectral dictionary learning. Dian et al. [6] proposed a subspace based low tensor multi-rank (LTMR) regularization method for HSI super-resolution. The LTMR mainly exploited two prior information of HR-HSI: high correlations among spectral bands and non-local self-similarities. Their work achieved state-of-art performance.

Recently, considerable improvements in performance have been achieved by exploiting learning methods [15, 16]. Spectral information and spatial structure can be learned from the images. To fully exploit the data information, our work exploits the clustering-based sparsity of the high spatial resolution hyperspectral images for both the coefficient matrix estimation and the spectral dictionary learning.

3 Hyperspectral image super-resolution

In this section, we explain the proposed hyperspectral image super-resolution method. The flowchart of the proposed method is illustrated in Fig. 1. The reconstruction of HR-HSI Z is formulated as a joint estimation of the hyperspectral dictionary D and the sparse coefficient matrix A based on the spatial-spectral sparsity of the hyperspectral image. First, we learn the hyperspectral dictionary D from the LR-HSI X. Once the spectral dictionary D is estimated, we estimate the sparse coefficient matrix A from LR-HSI X and the corresponding HR-MSI Y. To improve the accuracy, both spectral dictionary learning and sparse coefficients estimation are imposed a structural sparsity constraint, which exploits the clustering-based sparsity of hyperspectral images - namely reconstructed spectral pixels should be similar to those learned centroids. At last, we reconstruct the HR-HSI Z with the hyperspectral dictionary D and the sparse coefficient matrix A.

3.1 NNCSR model

For spectral dictionary D, it is reasonable to assume that the LR-HSI X contains the same spectral dictionary as the HR-HSI Z, but due to the spectral degradation P, only l spectral channels cannot contain sufficient information to reconstruct L ($L \gg l$) spectral channels [24], thus spectral dictionary D is estimated by solving the following sparse nonnegative matrix decomposition problem
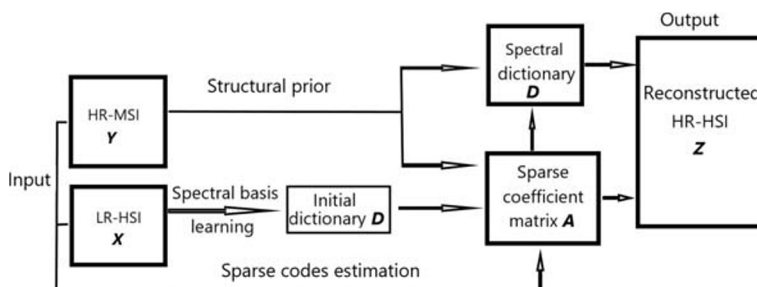

Fig. 1 Flowchart of the proposed hyperspectral image super-resolution method

$$(\boldsymbol{D}, \boldsymbol{B}) = \underset{\boldsymbol{D},\boldsymbol{B}}{argmin} \frac{1}{2} \|\boldsymbol{X} - \boldsymbol{DB}\|_2^2 + \lambda_1 \|\boldsymbol{B}\|_1, s.t. \mathbf{b}_i \geq 0, d_k \geq 0 \qquad (4)$$

where $\boldsymbol{D} = [\boldsymbol{d}_1, \cdots, \boldsymbol{d}_K] \in \mathrm{R}_+^{L \times K}$ is the spectral dictionary, $K(K \geq L)$ is the number of atoms, $\boldsymbol{B} = [\boldsymbol{b}_1, \cdots, \boldsymbol{b}_N] \in \mathrm{R}_+^{K \times N}$ is the sparse coefficient matrix. In this formulation, $\frac{1}{2} \|\boldsymbol{X} - \boldsymbol{DB}\|_2^2$ is a term modeling data fidelity, $\|\boldsymbol{B}\|_1$ is the regularization prior on the image to recover, and λ_1 is a parameter controlling the trade-off between these two terms.

Once the spectral dictionary $\boldsymbol{D}$ is estimated, the coefficient matrix A can be estimated from both the HR-MSI image $\boldsymbol{Y}$ and the LR-HSI $\boldsymbol{X}$ by solving the following sparse nonnegative matrix decomposition problem

$$A = \underset{A}{argmin} \left\| \boldsymbol{Y} - \overline{\boldsymbol{D}} A \right\|_2^2 + \tau \|\boldsymbol{X} - \boldsymbol{DAH}\|_2^2 + \eta_2 \|A\|_1, \text{s.t.} \boldsymbol{\alpha}_i \geq 0 \qquad (5)$$

where $A = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_N] \in \mathrm{R}_+^{K \times N}$ is the sparse coefficient matrix. In this formulation, $\left\| \boldsymbol{Y} - \overline{\boldsymbol{D}} A \right\|_2^2$ and $\|\boldsymbol{X} - \boldsymbol{DAH}\|_2^2$ are two terms of modeling data fidelity, $\|A\|_1$ is the regularization prior on the image to recover, and τ, η_2 are two parameters controlling the trade-off between these terms. A typical natural scene usually contains a collection of similar patches from all over the image. These non-local similar patches can be exploited to enhance the performance of image restoration tasks [10, 11]. But the $l1$-norm nonnegative sparse model of (4) and (5) cannot exploit the spatial correlations among local and nonlocal similar neighbors. To address this issue, we propose the following non-negative clustering-based sparse representation (NNCSR) model,

$$(\boldsymbol{D}, \boldsymbol{B}, \boldsymbol{U}) = \underset{\boldsymbol{D},\boldsymbol{B},\boldsymbol{U}}{argmin} \frac{1}{2} \|\boldsymbol{X} - \boldsymbol{DB}\|_2^2 + \lambda_1 \|\boldsymbol{B}\|_1 + \lambda_2 \sum_{q=1}^{Q} \sum_{i \in S_q} \left\| \boldsymbol{D}\boldsymbol{\alpha}_i - \boldsymbol{\mu}_q \right\|_2^2, s.t. \mathbf{b}_i \geq 0, \mathbf{d}_k \geq 0 \, (6)$$

$$(\boldsymbol{A}, \boldsymbol{U}) = \underset{\boldsymbol{A},\boldsymbol{U}}{argmin} \left\| \boldsymbol{Y} - \overline{\boldsymbol{D}} A \right\|_2^2 + \tau \|\boldsymbol{X} - \boldsymbol{DAH}\|_2^2 + \eta_1 \sum_{q=1}^{Q} \sum_{i \in S_q} \left\| \boldsymbol{D}\boldsymbol{\alpha}_i - \boldsymbol{\mu}_q \right\|_2^2 + \eta_2 \|A\|_1, \text{s.t.} \boldsymbol{\alpha}_i \geq 0 \quad (7)$$

where μ_q denotes the centroid of the q-th cluster $S_q = \left\{ i \middle| \left\| \overline{\boldsymbol{y}_i} - \overline{\boldsymbol{y}_q} \right\|_2^2 < t \right\}$, $\overline{\boldsymbol{y}_i}$ and $\overline{\boldsymbol{y}_q}$ denote the image patches of $\boldsymbol{Y}$ centered at positions i and q respectively. Recently, clustering learning has been widely used [13, 14, 43]. The efficient k-Nearest Neighbour (k-NN) clustering method is used to group similar spectral pixels for each spectral pixel [18]. Due to the structural similarity between $\boldsymbol{Z}$ and $\boldsymbol{Y}$, we perform the k-NN clustering on the HR-MSI image patches to search for similar neighbors of $\overline{\boldsymbol{y}_q}$. The vector μ_q is computed as

$$\boldsymbol{\mu}_q = \sum_{i \in \boldsymbol{S}_q} w_i (\boldsymbol{D}\boldsymbol{a}_i) \qquad (8)$$

where $w_i = \frac{1}{c} exp\left(-\left\| \overline{\boldsymbol{y}_i} - \overline{\boldsymbol{y}_q} \right\|_2^2 / h \right)$ is the weighting coefficients. The proposed NNCSR model exploits the structural prior that the reconstructed spectral pixels should be similar to those learned centroids.

In practice, both $\boldsymbol{D}$ and $\boldsymbol{a}_i$ for each pixel $\boldsymbol{z}_i$ of $\boldsymbol{Z}$ are unknown, μ_q cannot be directly computed via (8). This difficulty can be overcome by estimating μ_q from the current estimates of $\boldsymbol{D}$ and $\boldsymbol{a}_i$. We estimate $\boldsymbol{D}$ and $\boldsymbol{a}_i$ alternatively. For notation convenience, Eq. (6) and (7) can be rewritten as

$$(\boldsymbol{D}, \boldsymbol{B}, \boldsymbol{U}) = \underset{\boldsymbol{D}, \boldsymbol{B}, \boldsymbol{U}}{argmin} \frac{1}{2} \|\boldsymbol{X} - \boldsymbol{DB}\|_2^2 + \lambda_1 \|\boldsymbol{B}\|_1 + \lambda_2 \|\boldsymbol{DA} - \boldsymbol{U}\|_2^2, s.t. \mathbf{b}_i \geq 0, \mathbf{d}_k \geq 0 \qquad (9)$$

$$(\boldsymbol{A}, \boldsymbol{U}) = \underset{\boldsymbol{A}, \boldsymbol{U}}{argmin} \left\| \boldsymbol{Y} - \overline{\boldsymbol{D}} \boldsymbol{A} \right\|_2^2 + \tau \|\boldsymbol{X} - \boldsymbol{DAH}\|_2^2 + \eta_1 \|\boldsymbol{DA} - \boldsymbol{U}\|_2^2 + \eta_2 \|\boldsymbol{A}\|_1, s.t. \boldsymbol{\alpha}_i \geq 0 \quad (10)$$

where $\boldsymbol{U} = [\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_q, ..., \boldsymbol{\mu}_Q]$.

3.2 Spectral dictionary learning

First, the spectral dictionary $\boldsymbol{D}$ can be learned from $\boldsymbol{X}$ by solving (4),i.e., the first two terms of (9). As both $\boldsymbol{D}$ and $\boldsymbol{B}$ are constrained to be nonnegative, existing dictionary learning algorithms and online dictionary learning cannot be used. In [39], the alternating direction method of multipliers (ADMM) is adopted to convert a constrained dictionary learning problem into an unconstrained version and solved the unconstrained dictionary learning problem via alternative optimization. For a fixed $\boldsymbol{D}$, the subproblem with respect to $\boldsymbol{B}$ becomes

$$\boldsymbol{B} = \underset{\boldsymbol{B}}{argmin} \frac{1}{2} \|\boldsymbol{X} - \boldsymbol{DB}\|_2^2 + \lambda_1 \|\boldsymbol{B}\|_1, s.t. \mathbf{b}_i \geq 0 \qquad (11)$$

which is convex and can be efficiently solved by any convex optimization solver. For fast convergence rate, we use ADMM technique to solve (11). We reformulate (11) into

$$\boldsymbol{B} = \underset{\boldsymbol{B}}{argmin} \frac{1}{2} \|\boldsymbol{X} - \boldsymbol{DS}\|_2^2 + \lambda_1 \|\boldsymbol{B}\|_1, s.t. \mathbf{B} = \mathbf{S}, \mathbf{b}_i \geq 0 \qquad (12)$$

Applying ADMM [3], there is the following augmented Lagrangian function.

Algorithm 1 Spectral dictionary learning

Algorithm 1 Spectral dictionary learning

Input $\boldsymbol{X}$, K, λ_1, λ_2;
Initialize $\boldsymbol{D}$ by randomly selecting the columns of $\boldsymbol{X}$ with normalization;
For $i=0,1,\ldots\ldots,num_1$ **do**
   Solving (11) via ADMM:
   For $t=1,2,\ldots,T_1$ **do**
      Update $\boldsymbol{S}^{(t+1)}, \boldsymbol{B}^{(t+1)}$ via (16);
      Update $\boldsymbol{R}^{(t+1)}$ via (15);
      $\mu := \gamma \mu (\gamma > 1)$;
      Update $\boldsymbol{B} = \boldsymbol{B}^{(t+1)}$ if $t=T1$;
   End for
   Update $\boldsymbol{D}$ using current $\boldsymbol{B}$ via (18);
   Output $\boldsymbol{D}$ if $i=num_1$;
End for

$$L_\mu(\boldsymbol{B},\boldsymbol{S},\boldsymbol{R}) = \frac{1}{2}\|\boldsymbol{X}-\boldsymbol{DS}\|_2^2 + \lambda_1\|\boldsymbol{B}\|_1, +\mu\left\|\boldsymbol{S}-\boldsymbol{B}+\frac{\boldsymbol{R}}{2\mu}\right\|_2^2, s.t.\boldsymbol{B}=\boldsymbol{S},\mathbf{b}_i\geq 0 \qquad (13)$$

where $\boldsymbol{R}$ is the Lagrangian multiplier ($\mu > 0$). Then, solving (12) consists of the following alternative iterations,

$$\boldsymbol{S}^{(t+1)} = \underset{S}{argmin}\, L_\mu\left(\boldsymbol{B}^{(t)},\boldsymbol{S},\boldsymbol{R}^{(t)}\right)\boldsymbol{B}^{(t+1)} = \underset{B}{argmin}\, L_\mu\left(\boldsymbol{B},\boldsymbol{S}^{(t+1)},\boldsymbol{R}^{(t)}\right), s.t.\mathbf{b}_i\geq 0 \quad (14)$$

where t is the iteration number and Lagrangian multiplier $\boldsymbol{R}$ is updated by

$$\boldsymbol{R}^{(t+1)} = \boldsymbol{R}^{(t)} + \mu\left(\boldsymbol{S}^{(t+1)}-\boldsymbol{B}^{(t+1)}\right) \qquad (15)$$

The sub-problems in (14) have closed-form solutions,

$$\begin{aligned} \boldsymbol{S}^{(t+1)} &= \left(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}+2\mu\boldsymbol{I}\right)^{-1}\left(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{X}+2\mu\left(\boldsymbol{B}^{(t)}-\frac{\boldsymbol{R}^{(t)}}{2\mu}\right)\right) \\ \boldsymbol{B}^{(t+1)} &= \left[\boldsymbol{Soft}\left(\boldsymbol{S}^{(t+1)}+\frac{\boldsymbol{R}^{(t)}}{2\mu},\frac{\lambda_1}{2\mu}\right)\right]_+ \end{aligned} \qquad (16)$$

where Soft(.) denotes a soft-shrinkage operator and $[x]_+ = \max\{x, 0\}$. For a fixed $\boldsymbol{B}$, spectral dictionary $\boldsymbol{D}$ can be updated by solving

$$\boldsymbol{D}= \underset{D}{argmin}\,\|\boldsymbol{X}-\boldsymbol{DB}\|_2^2, s.t.\mathbf{d}_k\geq 0 \qquad (17)$$

Eq. (17) can be solved using block coordinate descent [8] or be solved analytically

$$\boldsymbol{D}=\boldsymbol{XB}^T\left(\boldsymbol{BB}^T\right)^{-1} \qquad (18)$$

The algorithm for the initial spectral dictionary learning is summarized in Algorithm 1. Once the initial spectral dictionary $\boldsymbol{D}$ is estimated, sparse coefficient matrix $\boldsymbol{A}$ can be estimated by solving (10). We will elaborate on the sparse coefficient estimation in the next section. When the sparse coefficient matrix $\boldsymbol{A}$ is estimated, the last term of (9) can be solved analytically,

$$\boldsymbol{D}=\boldsymbol{UA}^T\left(\boldsymbol{AA}^T\right)^{-1} \qquad (19)$$

3.3 Sparse coefficient estimation

With respect to the learned spectral dictionary $\boldsymbol{D}$, we estimate the sparse coefficient matrix $\boldsymbol{A}$ by solving (10). Eq. (10) is convex and can be efficiently solved by any convex optimization solver. For fast convergence, we use ADMM technique to solve (10), and obtain the following Lagrangian function:

$$L_\mu(\boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{S}, \boldsymbol{R}_1, \boldsymbol{R}_2) = \left\| \begin{bmatrix} \boldsymbol{Y} \\ \tau \boldsymbol{X} \\ \eta_1 \boldsymbol{DS} \end{bmatrix} - \begin{bmatrix} \overline{\boldsymbol{D}}\boldsymbol{S} \\ \tau \boldsymbol{ZH} \\ \eta_1 \boldsymbol{U} \end{bmatrix} \right\|_2^2 + \eta_2 \|\boldsymbol{A}\|_1 + \mu \left\| \boldsymbol{DS} - \boldsymbol{Z} + \frac{\boldsymbol{R}_1}{2\mu} \right\|_2^2$$

$$+ \mu \left\| \boldsymbol{S} - \boldsymbol{A} + \frac{\boldsymbol{R}_2}{2\mu} \right\|_2^2, \text{s.t.} \boldsymbol{a}_i \ge 0 \tag{20}$$

where $\boldsymbol{R}_1$, $\boldsymbol{R}_2$ are Lagrangian multipliers ($\mu > 0$). Minimizing the augmented Lagrangian function leads to the following iterations:

$$A^{(t+1)} = \underset{A}{\arg\min}\, L_\mu\left(A, Z^{(t)}, S^{(t)}, R_1^{(t)}, R_2^{(t)}\right), Z^{(t+1)}$$

$$= \underset{Z}{\arg\min}\, L_\mu\left(A^{(t+1)}, Z, S^{(t)}, R_1^{(t)}, R_2^{(t)}\right), S^{(t+1)}$$

$$= \underset{S}{\arg\min}\, L_\mu\left(A^{(t+1)}, Z^{(t+1)}, S, R_1^{(t)}, R_2^{(t)}\right), \tag{21}$$

where the Lagrangian multipliers are updated by

$$\boldsymbol{R}_1^{(t+1)} = \boldsymbol{R}_1^{(t)} + \mu\left(\boldsymbol{DS}^{(t+1)} - \boldsymbol{Z}^{(t+1)}\right), \boldsymbol{R}_2^{(t+1)} = \boldsymbol{R}_2^{(t)} + \mu\left(\boldsymbol{S}^{(t+1)} - \boldsymbol{A}^{(t+1)}\right) \tag{22}$$

All sub-problems in (20) can have closed-form solutions,

$$S = \left[\overline{D}^T\overline{D} + (\eta_1 + \mu)D^T D + \mu I\right]^{-1}\left[\overline{D}^T Y + \eta_1 D^T U + \mu D^T\left(Z^{(t)} - \frac{R_1^{(t)}}{2\mu}\right) + \mu\left(A^{(t)} - \frac{R_1^{(t)}}{2\mu}\right)\right],$$

$$Z = \left[\tau X H^T + \mu\left(DS^{(t)} + \frac{R_1}{2\mu}\right)\right]\left(\tau HH^T + \mu I\right)^{-1}, \; A = \left[Soft\left(S^{(t)} - \frac{R_2^{(t)}}{2\mu}, \frac{\eta_2}{2\mu}\right)\right]_+ \tag{23}$$

As the matrix to be inverted in the equation of updating $\boldsymbol{Z}$ is large, we use conjugate gradient algorithm to compute the matrix inverse. The overall algorithm for reconstruction $\boldsymbol{Z}$ is summarized in Algorithm 2.

4 Experiments and discussion

4.1 Experimental datasets

To verify the performance of our proposed method, we have test the algorithm on two different public datasets of hyperspectral images. The first dataset is Columbia computer vision laboratory (CAVE) [41]. The CAVE dataset consists of 32 indoor HSIs. The images have 31 spectral bands, and each band has a size of 512×512. We take the hyperspectral images from the dataset as ground-truth images and use these images to simulate LR-HSIs and HR-MSI images. As in [8, 22], the original HR-HSIs $\boldsymbol{Z}$ are downsampled by averaging over disjoint 32×32 blocks to simulate the LR-HSIs $\boldsymbol{X}$. Similar to [8, 42], HR RGB images $\boldsymbol{Y}$ are generated by downsampling the hyperspectral images $\boldsymbol{Z}$ along the spectral dimension using the spectral transform matrix derived from the response of a Nikon D700 camera.

Algorithm 2 Hyperspectral image super-resolution with NNCSR model.

Algorithm 2   Hyperspectral image super-resolution with NNCSR model

----
Learn the initial spectral dictionary D from X using Algorithm 1;
Obtain the sample sets S_q through k-NN search on Y;
Set parameters k, η_1 and η_2;
Initialize U=0, A=0, S=0, Z=0, R_1=0, R_2=0;
Solving (10) via ADMM：
For t=0,1,……,$num2$ **do**
  Compute $S^{(t+1)}$,$Z^{(t+1)}$ and $A^{(t+1)}$ using current D, U via (23);
  Update $R_1^{(t+1)}$ and $R_2^{(t+1)}$ via (22);
  $\mu:=\rho\mu(\rho>1)$;
  Update U using current D, A via (8);
End for
Update D with current U, S via (19);
Update U with current D, A via (8);
Reset A=0, S=0, Z=0, R_1=0, R_2=0;
Update S, Z and A using current D, U via (23);
Output the reconstructed Z.

The second dataset is Pavia University [5], acquired by the reflective optics system imaging spectrometer optical sensor over the area of Pavia University. The Pavia University have the 115 spectral bands and 610×340 spectral pixels. As in [6, 7], by removing the bands of low SNR, we reduce the HSI as 93 bands and select the up-left $256 \times 256 \times 93$-pixel-size image as the ground truth. To simulate the LR-HSI, we firstly filter each band of HR-HSI by a 7×7 Gaussian blur (standard deviation 2) and then downsample every four pixels in two spatial modes. The HR-MSI of four bands is simulated by using IKONOS-like reflectance spectral response filter [36].

4.2 Compared approaches

We compare the proposed NNCSR method with Bayesian sparse representation (BSR) [36], MF [22], CNMF [42], NSSR [8], CSTF [27] and low tensor multi-rank (LTMR) [6], which represent the state-of-the-arts of modern hyperspectral image super-resolution method. Among the compared approaches, BSR is based on Bayesian inference, CSTF and LTMR are based on tensor factorization, MF, CNMF and NSSR are based on matrix factorization.

4.3 Quantitative metrics

Five indexes are utilized to measure the quality of the fusion results.
The first index is the peak signal-to-noise ratio (PSNR) for HSI defined as,

$$PSNR\left(Z, \widetilde{Z}\right) = \frac{1}{L}\sum_{i=1}^{L} PSNR\left(Z_i, \widetilde{Z}_i\right) \tag{24}$$

where $\boldsymbol{Z} \in R^{L \times N}$ and $\widetilde{\boldsymbol{Z}} \in R^{L \times N}$ are ground-truth and reconstructed HSIs respectively, $\boldsymbol{Z}_i$ and $\widetilde{\boldsymbol{Z}}_i$ represent i-th band of $\boldsymbol{Z} \in R^{L \times N}$ and $\widetilde{\boldsymbol{Z}} \in R^{L \times N}$. The higher the PSNR, the better the quality of the reconstructed image.

The second index is the root mean square error (RMSE) defined as

$$\mathrm{RMSE}\left(\boldsymbol{Z}, \widetilde{\boldsymbol{Z}}\right) = \sqrt{\frac{\left\|\boldsymbol{Z} - \widetilde{\boldsymbol{Z}}\right\|_F^2}{LN}} \tag{25}$$

where $\boldsymbol{Z} \in R^{L \times N}$ and $\widetilde{\boldsymbol{Z}} \in R^{L \times N}$ are scaled to the range [0, 255]. RMSE is a measure of the estimation error. The smaller the RMSE, the better the quality of the reconstructed image.

The third index is the spectral angle mapper (SAM) [36, 45], which is defined as the averaged angle between the estimated pixel $\widetilde{z}_j$ and the ground truth pixel z_j, i.e.

$$\mathrm{SAM}\left(\boldsymbol{Z}, \widetilde{\boldsymbol{Z}}\right) = \frac{1}{LN} \Sigma_{j=1}^{LN} arcos \frac{\widetilde{z}_j^T z_j}{\left\|\widetilde{z}_j\right\|_2 \|z_j\|_2} \tag{26}$$

SAM is given in degrees. The smaller SAM, the less spectral distortions.

The fourth index is the relative dimensionless global error in synthesis (ERGAS) [33] defined as

$$\mathrm{ERGAS}\left(\boldsymbol{Z}, \widetilde{\boldsymbol{Z}}\right) = \frac{100}{d} \sqrt{\frac{1}{L} \Sigma_{i=1}^{L} \frac{MSE\left(\boldsymbol{Z}_i, \widetilde{\boldsymbol{Z}}_i\right)}{\mu_{\widetilde{\boldsymbol{Z}}_i}^2}} \tag{27}$$

where d is the spatial downsampling factor, and $\mu_{\boldsymbol{Z}^{\sim i}}$ is the mean value of $\widetilde{\boldsymbol{Z}}_i$. ERGAS reflects the overall quality of the reconstruction image. The smaller ERGAS, the better quality of the reconstructed results.

The fifth index is the universal image quality index (UIQI) [34], which is calculated on a sliding window of size 32×32 and averaged on all windows. The UIQI for two windows $\mathbf{a}$ and $\mathbf{b}$ is given by

$$\mathrm{UIQI}(\mathbf{a}, \mathbf{b}) = \frac{4\sigma_{\mathbf{ab}}^2}{\sigma_{\mathbf{a}}^2 + \sigma_{\mathbf{b}}^2} \cdot \frac{\mu_{\mathbf{a}} \mu_{\mathbf{b}}}{\mu_{\mathbf{a}}^2 + \mu_{\mathbf{b}}^2} \tag{28}$$

where $\sigma_{\mathbf{ab}}$ is the sample covariance between $\mathbf{a}$ and $\mathbf{b}$, and $\sigma_{\mathbf{a}}$ and $\mu_{\mathbf{a}}$ denote the standard deviation and mean value of $\mathbf{a}$, respectively.

4.4 Parameters discussion

In our method, four key parameters: the number of atoms K, the parameter k of k-NN, the regularization parameters η_1 and η_2 need to be discussed.

To discuss the effects of these parameters, we plot the PSNR curves of the fused results of *Balloons* (an HSI in the CAVE dataset) varying from these parameters in Fig. 2. As we can see from Fig. 2 (a), the proposed method performs best when the number of atoms is in the range of 75 ∼ 125 and is insensitive to the variation in the values of K in this range. Therefore, 75 atoms are enough to preserve the spectral information. In our implementation, we set $K = 75$.

From Fig. 2 (b), we can see the PSNR rises as k changed from 1 to 60, and then it keeps relatively stable as k increases further. Therefore we can select $k = 60$. From Fig. 2(c), we can see that the performance of the proposed method rises as η_1 varies from 0.01 to 0.025, and then it declines as η_1 grows further. The PSNR for *Balloons* rises as η_2 varies from $3*10^{-5}$ to $8*10^{-5}$, and then it declines as η_2 is greater than $8*10^{-5}$. So we set $\eta_1 = 0.025$ and $\eta_2 = 8*10^{-5}$. Similar observations can be obtained for other test images.

4.5 Experimental results

Table 1 reports the average quality metrics of competing methods on the CAVE dataset. We highlight the best results in bold. For fair comparison, the results of BSR [36], MF [22], CNMF [42] and NSSR [8] are directly obtained from [8]. From Table 1, we can see that our NNCSR method performs consistently better than other testing methods according to these quality metrics, which indicates that HR-HSIs recovered by our method have better spatial and spectral qualities in CAVE dataset. For example, the PSNR of our method is greater than that of the state-of-the-art CSTF method by 1.12 dB, and the SAM is reduced by 0.22 degrees compared with NSSR. The improvement mainly comes from one aspect. Both NNCSR and NSSR exploit the clustering-based sparsity of hyperspectral images. The main difference is that NNCSR exploits the clustering-based sparsity for both sparse coefficient estimation and dictionary learning, while the NSSR only for sparse coefficient estimation, not for dictionary
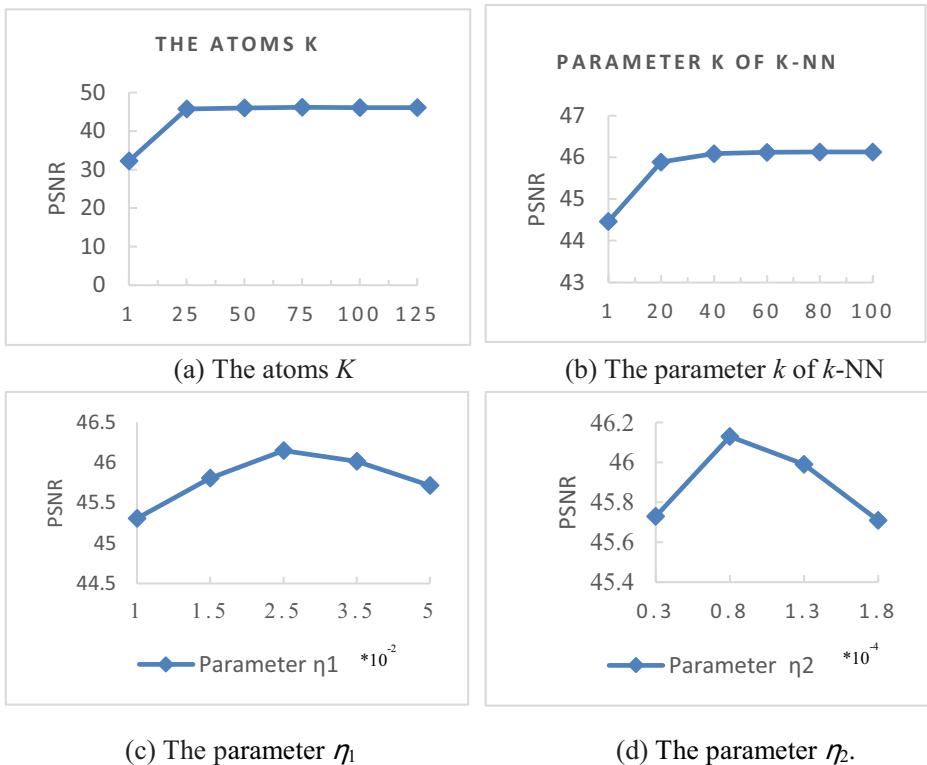

(a) The atoms K

(b) The parameter k of k-NN

(c) The parameter η_1

(d) The parameter η_2.

Fig. 2 The PSNR as functions of parameters K, k, η_1 and η_2, respectively. (a) The atoms K, (b) The parameter k of k-NN, (c) The parameter η_1, (d) The parameter η_2

Table 1 The average results of the competing methods on the CAVE dataset

Method	MF	CNMF	BSR	NSSR	CSTF	NNCSR
PSNR	39.37	39.53	39.16	42.26	41.41	**42.53**
RMSE	3.03	2.93	3.13	2.21	3.06	**2.15**
SAM	6.12	5.48	6.75	4.33	8.75	**4.11**
ERGAS	0.40	0.39	0.37	0.30	0.39	**0.29**
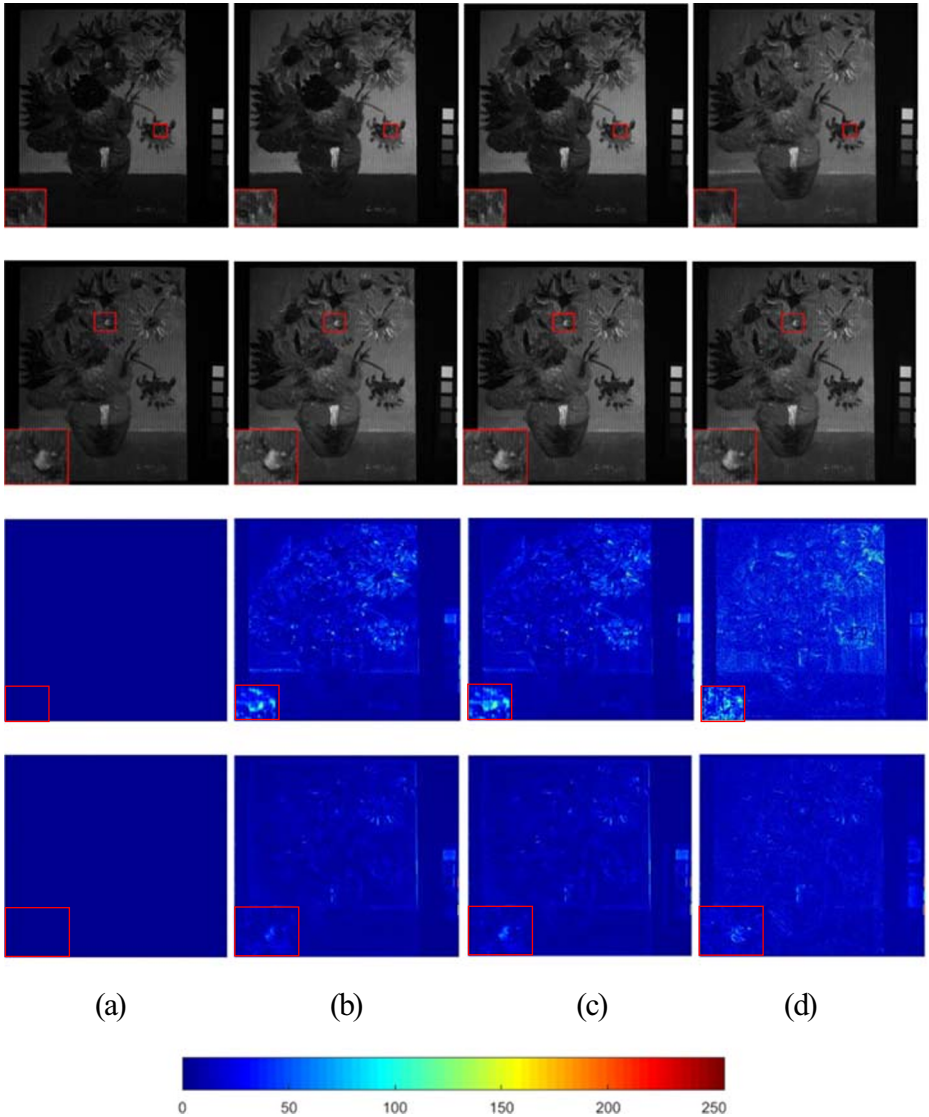
Fig. 3 Reconstructed images and corresponding error images of *Oil-Painting* from CAVE dataset [41] at band 16 and 26. The first two rows show the reconstructed images. The next two rows show the error images. (a) Ground truth, (b) NNCSR, (c) NSSR [8], (d) CSTF [27]

Table 2 The results of the competing methods on the Pavia University dataset

Method	NSSR	CSTF	LTMR	NNCSR
SAM	1.439	1.745	1.493	**1.431**
ERGAS	0.794	0.932	0.805	**0.793**
UIQI	0.996	0.995	0.996	**0.996**

learning. The CSTF does not exploit the clustering-based sparsity. For visual comparison, the recovered HR-HSIs, corresponding error images of the competing approaches at band 16 and 26 for *Oil_Painting* (an HSI in CAVE dataset) are shown in Fig. 3. A representative region of each reconstructed image is magnified. As can be seen from the magnified images and error images, the HR-HSI reconstructed by NNCSR has fewer artifacts than that produced by NSSR and CSTF.

The quality metrics of the competing approaches on the Pavia University are shown in Table 2. For fair comparison, the results of LTMR [6] are directly obtained from its original paper. Our NNCSR method has obvious advantages over CSTF in terms of all quality indexes. Meanwhile, the LTMR and NSSR delivers comparable performance to the NNCSR. Figure 4 shows the reconstructed images and error images at band 30 of the competing approaches for the Pavia University. We can see that all testing approaches can reconstruct the spatial structures of the HR-HSIs. As we can see from the error images, the HR-HSIs produced by NNCSR and NSSR have fewer errors than that produced by CSTF.
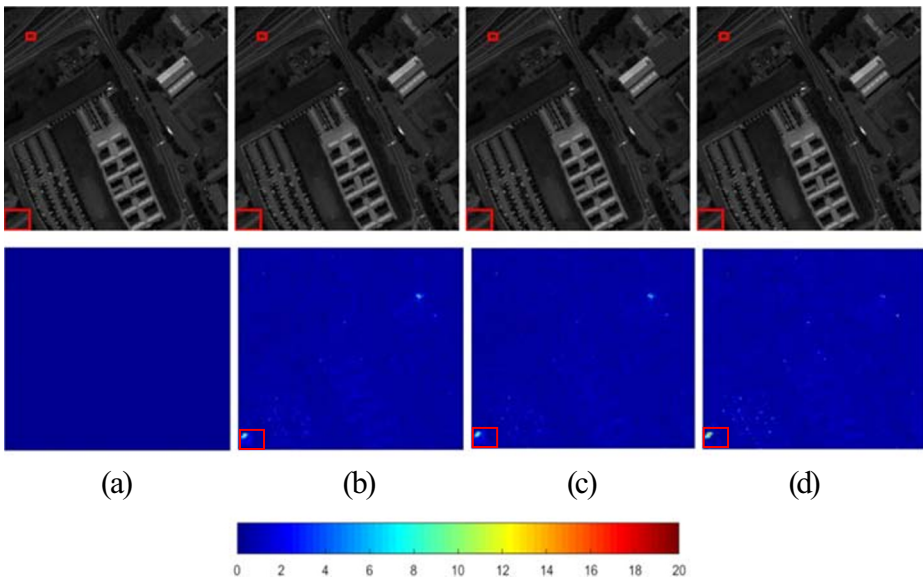

(a)   (b)   (c)   (d)

Fig. 4 Reconstructed images and corresponding error images of the competing approaches for the Pavia University at 30-th band. The first row shows the reconstructed images. The second row shows the error images. (a) Ground truth; (b) NNCSR; (c)NSSR [8]; (d)CSTF [27]

Table 3 Average run time (seconds) of different methods on a test image of size $512 \times 512 \times 31$

CSTF	NSSR	NNCSR
173	129	135

4.6 Computation efficiency

In algorithm 2, the computational complexity mainly consists of three parts: 1) the initial spectral dictionary learning, 2) k-NN nearest neighbor searching each patch within a window of size $h \times h$, 3) the sparse coding. The complexity of solving the initial dictionary is $O(num_1 T_1(K^2 n + K^3 + KLn))$ in algorithm 1. The complexity of computing the k-NN per patch is $O(Nk_1 h^2)$. The complexity of updating S is the same as that of D, i.e. $O(K^2 L + KN^2 + K^3 + KLN)$. The complexity of updating Z and A is $O(LN\sqrt{N}logN)$ and $O(N)$ respectively. The propose algorithm is implemented with Matlab R2016a on an Intel(R) Core(TM) i7-6700T CPU @2.80GHz 2.81GHz. A running time comparison between the proposed method and other competing methods on a $512 \times 512 \times 31$ test image is shown in Table 3. From Table 3, it can be seen that the proposed NNCSR method is quicker than CSTF [27]. And it is slightly slower than NSSR [8], this is because NNCSR exploits the clustering-based sparsity prior for both dictionary learning and sparse codes estimation, while NSSR [8] exploits the clustering-based sparsity prior only for sparse codes estimation.

5 Conclusions

A non-negative sparse representation model for hyperspectral image super-resolution is proposed. The hyperspectral dictionary is learned from the LR hyperspectral image. The sparse codes with respect to the learned dictionary are estimated from LR hyperspectral image and the corresponding HR multispectral image. To improve the accuracy, both spectral dictionary learning and sparse coefficients estimation exploit the clustering-based sparsity of the hyperspectral image. Experiments show that the proposed method outperforms several state-of-the-art methods in objective quality metrics and visual performance.

References

1. Akhtar N, Shafait F, Mian A (2014) Sparse spatio-spectral representation for hyperspectral image super-resolution. Proc. Eur. Conf. Comput. Vis. (ECCV), pp 63–78
2. Akhtar N, Shafait F, Mian A (2015) Bayesian sparse representation for hyperspectral image super resolution. Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp 3631–3640
3. Boyd N, Parikh E, Chu BP, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn 3(1):1–122
4. Casalino G, Gillis N (2017) Sequential dimensionality reduction for extracting localized features. Pattern Recogn 63:15–29

5. Dell'Acqua F, Gamba P, Ferrari A, Palmason JA, Benediktsson JA, Arnason K (2004) Exploiting spectral and spatial information in hyperspectral urban data with high resolution. IEEE Geosci Remote Sens Lett 1(4):322–326
6. Dian R, Li S (2019) Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization. IEEE Trans Image Process 28(10):5135–5146
7. Dian R, Fang L, Li S (2017) Hyperspectral image super-resolution via non-local sparse tensor factorization. Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp 3862–3871
8. Dong W, Fu F, Shi G, Cao X, Wu J, Li G, Li X (2016) Hyperspectral image super-resolution via non-negative structured sparse representation. IEEE Trans Image Process 25(5):2337–2352
9. Du B, Zhang Y, Zhang L, Tao D (2016) Beyond the sparsity-based target detector: a hybrid sparsity and statistics-based detector for hyperspectral images. IEEE Trans Image Process 25(11):5345–5357
10. Guo F, Zhang C (2019) Edge preserving mixed noise removal. Multimed Tools Appl 78(12):16601–16613
11. Guo F, Zhang C, Zhang M (2018) Edge-preserving image denoising. IET Image Process 12(8):1394–1401
12. Hardie RC, Eismann MT, Wilson GL (2004) MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor. IEEE Trans Image Process 13(9):1174–1184
13. Hong C, Yu J, Wan J, Tao D, Wang M (2015) Multimodal deep autoencoder for human pose recovery. IEEE Trans Image Process 24(12):5659–5670
14. Hong C, Yu J, Zhang J, Jin X, Lee K (2019) Multimodal face-pose estimation with multitask manifold deep learning. IEEE Trans Ind Inform 15(7):3952–3961
15. Huang F, Zhang X, Li C, Li Z, He Y, Zhao Z (2018) Deep multi-view representation learning for social images. Appl Soft Comput 73:106–118
16. Huang F, Zhang X, Li Z (2018) Learning joint multimodal representation with adversarial attention networks. ACM Multimedia Conference, Seoul, Korea, pp 1874–1882
17. Iordache M-D, Bioucas-Dias J, Plaza A (2011) Sparse unmixing of hyperspectral data. IEEE Trans Geosci Remote Sens 49(6):2014–2039
18. Jian M, Lam K (2013) A novel face-hallucination scheme based on singular value decomposition. Pattern Recogn 46(11):3091–3102
19. Jian M, Lam K (2014) Face-image retrieval based on singular values and potential-field representation. Signal Process 100:9–15
20. Jian M, Lam K (2015) Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition. IEEE Trans CSVT 25(11):1761–1772
21. Jian M, Cui C, Nie X, Zhang H, Nie L, Yin Y (2019) Multi-view face hallucination using SVD and a mapping model. Inf Sci 488:181–189
22. Kawakami R, Wright J, Tai Y-W, Matsushita Y, Ben-Ezra M, Ikeuchi K (2011) High-resolution hyperspectral imaging via matrix factorization. IEEE CVPR 2329–2336
23. Keshava N, Mustard JF (2002) Spectral unmixing. IEEE signal process. Mag 19(1):44–57
24. Lanaras C, Baltsavias E, Schindler K (2015) Hyperspectral superresolution by coupled spectral unmixing. Proc. IEEE Int. Conf. Comput. Vis., pp 3586–3594
25. Lee DD, Seung HS (1999) Learning the parts of objects by nonnegative matrix factorization. Nature 401(6755):788–791
26. Lee DD, Seung SH (2001) Algorithms for non-negative matrix factorization. Proc. NIPS, pp 556–562.
27. Li S, Dian R, Fang L, Bioucas-Dias JM (2018) Fusing Hyperspectral and multispectral images via coupled sparse tensor factorization. IEEE Trans Image Process 27(8):4118–4130
28. Molina R, Katsaggelos AK, Mateos J (1999) Bayesian and regularization methods for hyperparameter estimation in image restoration. IEEE Trans Image Process 8(2):231–246
29. Molina R, Vega M, Mateos J, Katsaggelos AK (2008) Variational posterior distribution approximation in Bayesian super resolution reconstruction of multispectral images. Appl Comput Harmon Anal 24(2):251–267
30. Simoes M, Bioucas-Dias J, Almeida L, Chanussot J (2015) A convex formulation for hyperspectral image superresolution via subspace-based regularization. IEEE Trans Geosci Remote Sens 53(6):3373–3388
31. Sudheer Babu R, Sreenivasa Murthy KE (2018) Enhanced joint estimation-based Hyperspectral image super resolution. In: Dash S, Naidu P, Bayindir R, Das S (eds) Artificial intelligence and evolutionary computations in engineering systems, Advances in intelligent systems and computing, vol 668. Springer, Singapore, pp 503–516
32. Veganzones MA, Simoes M, Licciardi G, Yokoya N, Bioucas-Dias JM, Chanussot J (2016) Hyperspectral super resolution of locally low rank images from complementary multisource data. IEEE Trans Image Process 25(1):274–288
33. Wald L (2000) Quality of high resolution synthesised images: is there a simple criterion. Proc. Int. Conf. Fusion Earth Data, Nice, France, pp 99–103
34. Wang Z, Bovik AC (2002) A universal image quality index. IEEE signal process. Lett 9(3):81–84

35. Wei Q, Dobigeon N, Tourneret J-Y (2014) Bayesian fusion of hyperspectral and multispectral images. Proc. IEEE ICASSP, Florence, Italy, pp 3176–3180
36. Wei Q, Bioucas-Dias J, Dobigeon N, Tourneret J-Y (2015) Hyperspectral and multispectral image fusion based on a sparse representation. IEEE Trans Geosci Remote Sens 53(7):3658–3668
37. Wei Q, Dobigeon N, Tourneret J (2015) Fast fusion of multi-band images based on solving a Sylvester equation. IEEE Trans Image Process 24(11):4109–4121
38. Wei Q, Bioucas-Dias J, Dobigeon N, Tourneret J-Y, Chen M, Godsill S (2016) Multiband image fusion based on spectral unmixing. IEEE Trans Geosci Remote Sens 54(12):7236–7249
39. Wycoff E, Chan TH, Jia K, Ma WK, Ma Y (2013) A nonnegative sparse promoting algorithm for high resolution hyperspectral imaging. IEEE ICASSP 1409–1413
40. Xu Y, Wu Z, Chanussot J, Wei Z (2019) Nonlocal patch tensor sparse representation for hyperspectral image super-resolution. IEEE Trans Image Process 28(6):3034–3047
41. Yasuma F, Mitsunaga T, Iso D, Nayar SK (2010) Generalized assorted pixel camera: post-capture control of resolution, dynamic range and spectrum. IEEE Trans Image Proc 19(9):2241–2253
42. Yokoya N, Yairi T, Iwasaki A (2012) Coupled non-negative matrix factorization unmixing for hyperspectral and multispectral data fusion. IEEE trans Geosci. Remote Sens 50(2):528–537
43. Yu J, Tao D, Wang M, Rui Y (2015) Learning to rank using user clicks and visual features for image retrieval. IEEE Trans Cybern 45(4):767–779
44. Yuan Y, Zheng X, Lu X (2017) Discovering diverse subset for unsupervised hyperspectral band selection. IEEE Trans Image Process 26(1):51–64
45. Zhang Y, De Backer S, Scheunders P (2009) Noise-resistant wavelet based Bayesian fusion of multispectral and hyperspectral images. IEEE Trans Geosci Remote Sens 47(11):3834–3843
46. Zhang Y, Duijster A, Scheunders P (2012) A Bayesian restoration approach for hyperspectral images. IEEE Trans Geosci Remote Sens 50(9):3453–3462