HYPERSPECTRAL AND MULTISPECTRAL IMAGE FUSION USING NON-CONVEX RELAXATION LOW RANK AND TOTAL VARIATION REGULARIZATION

Yue Yuan[1], Qi Wang[1], Xuelong Li[1]*

[1]School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China

ABSTRACT

Hyperspectral (HS) and multispectral (MS) image fusion is an important task to construct an HS image with high spatial and spectral resolutions. In this paper, we present a novel HS and MS fusion method using non-convex low rank tensor approximation and total variation regularization. In specific, the Laplace based low-rank model is formed to exploit spatial-spectral correlation and nonlocal similarity of the HS image, and the second-order total variation is used to describe the local smoothness structure in the spatial domain and adjacent bands. Also, an effective optimization algorithm is designed for the proposed model. In the experiments, we demonstrate the superiority of the proposed method compared to several state-of-the-art approaches.

Index Terms— Hyperspectral (HS) image, multispectral (MS) image, image fusion, low-rank approximation, total variation

1. INTRODUCTION

Hyperspectral (HS) image is famous for its abundant spectral bands, and it has several remote sensing applications [1] [2]. However, as the HS imaging system must maintain a good signal-to-noise ratio (SNR) of images, and the generation of images with high spatial resolution or high spectral resolution requires large amounts of data, resulting in limited spatial resolution of HS images. Fortunately, compared to HS images, multispectral (MS) images have fewer spectral bands and hence achieve higher spatial resolution. Consequently, to increase the spatial resolution of HS images, a popular way is to fuse HS and MS images [3].

In recent years, a series of methods have been developed to settle the fusion of HS and MS images. In particular, this problem can be considered as an extension of pan-sharpening, which fuses the HS image with a panchromatic image. As an implementation, Selva et al. [4] use the linear combination of MS bands to synthesize a high-resolution band for each HS band. Also, HS and MS image fusion is a type of HS super-resolution problem [5].

_____
*Corresponding Author.

Methods specifically for HS and MS image fusion has been developed. In [6], Yokoya et al. first introduce spectral unmixing technology in the multi-source image fusion, and they unmix the two images crosswise using the nonnegative matrix factorization. Considering the smoothing structure of HS images, the total variation is used to regularize the model in [7]. Recently, tensors are utilized to preserve the spatial structure information taking into account the three-dimensional structure of the HS image [8] [9]. A fusion net is proposed in [10] using the observation models and low-rank prior.

Inspired by the above works, we propose a novel non-convex relaxation low rank and total variation (NLRTV) based image fusion method for HS and MS images. Its main contributions are summarized below.

1) An effective fusion model for HS and MS images is proposed in a novel point of view. Considering the nonlocal similarity and the spatial-spectral correlation of HS images, we add the non-convex low-rank constraint term. Meanwhile, the second-order total variation constraint term is adopted to describe the spatial and spectral smoothness.

2) An effective optimization strategy is presented for the proposed model. Using Laplace based non-convex low-rank minimization as an approximation of rank function, the iteratively reweighted nuclear norm algorithm [11] is utilized to settle this problem.

2. PROPOSED METHOD

2.1. Problem Formulation

Let $\mathcal{X} \in \mathbb{R}^{W \times H \times S}$ be the desired high resolution HS (HRHS) image, where W, H and S donate image width, height and the number of spectral bands, respectively. Let $\mathcal{Y} \in \mathbb{R}^{W \times H \times s}$ denote the high resolution MS (HSMS) image and $\mathcal{L} \in \mathbb{R}^{w \times h \times S}$ denote the low resolution HS (LRHS) image, where $s < S$, $w < W$ and $h < H$. The relationship between them can be expressed by the following observation model

$$\mathcal{L}_{(3)} = \mathcal{X}_{(3)}P + N_{\mathcal{L}} \tag{1}$$

$$\mathcal{Y}_{(3)} = D\mathcal{X}_{(3)} + N_{\mathcal{Y}}, \tag{2}$$

where $\mathcal{X}_{(i)}$ means the mode-i unfolding of tensor $\mathcal{X}$. P is the spatial down-sampling operator, D is the spectral response matrix and $N_{\mathcal{L}}$ and $N_{\mathcal{Y}}$ are noise. To estimate $\mathcal{X}$ and minimize the error, the following minimization problem is obtained

$$\min_{\mathcal{X}} ||\mathcal{L}_{(3)} - \mathcal{X}_{(3)}P||_F^2 + \lambda||\mathcal{Y}_{(3)} - D\mathcal{X}_{(3)}||_F^2. \quad (3)$$

Here, $||\cdot||_F$ is the Frobenius norm. The parameter λ is used for trade-offs.

2.2. Non-convex Low Rank Tensor Approximation

To make effective use of global structure information, the low-rank representation model which aims at describing the correlation between matrix entries is utilizes in this part.

For an HS cube, its spatial adjacent pixels are correlated and the adjacent bands are also usually highly correlated. Besides, some scenes in the spatial domain are similar even if they are not adjacent to each other, which means the HS cube is nonlocal self-similar. To describe the correlation and the non-local self-similarity of an HS image, it is divided into several patches with size $b \times b \times S$ at first. After clustering similar patches, K clusters are obtained. We represent kth cluster with N_k $(k = 1, ..., K)$ patches as $\mathcal{C}_k \in \mathbb{R}^{b \times b \times S \times N_k}$. Since the spatial information of $\mathcal{X}$ is mainly obtained from $\mathcal{Y}$, the above division and clustering operations are performed on $\mathcal{Y}$, and then the division and clustering results are matched to $\mathcal{X}$. Thus, taking into account the correlation of the first three dimensions and the similarity of the fourth dimension, $\mathcal{C}_k$ is low-rank in all four dimensions, which can be expressed by the minimization problem below

$$\min \sum_{i=1}^{4} \alpha_i \text{rank}(\mathcal{C}_{k(i)}). \quad (4)$$

Since the optimization of the above problem is NP-hard, we replace it by the general nonconvex low rank minimization problem

$$\min \sum_{j=1}^{4} \sum_{i=1}^{m} \alpha_j g(\sigma_i(\mathcal{C}_{k(j)})), \quad (5)$$

where $\sigma_i(Z)$ represents the ith singular values of matrix Z. And we use Laplace penalty [12] as nonconvex surrogate function, its form is

$$g(\theta) = \lambda(1 - exp(-\frac{\theta}{\gamma})). \quad (6)$$

Define the Laplace norm $||Z||_L = \sum_{i=1}^{m} g(\sigma_i(Z))$ for convenience. By exploring the low-rank property of all K tensor $\{\mathcal{C}_k\}_{k=1}^{K}$, the non-convex relaxation low rank term is expressed as the following problem:

$$\min \sum_{k=1}^{K} \sum_{i=1}^{4} \alpha_i ||\mathcal{C}_{k(i)}||_L. \quad (7)$$

2.3. Second-order Total Variation Regularization

The total variation (TV), defined as the total magnitude of the discrete gradients of an image, is popularly used to describe the piecewise-smoothness structure of images. However, second-order TV works better on image super-resolution. It can be more sensitive to areas with large grayscale variations. For the three-dimension structure of HS images, the second-order TV constraint can be modeled as

$$||T_\omega(\mathcal{X})||_1 = \sum_{i,j,k} \omega_1 ||x_{i-1,j,k} + x_{i+1,j,k} - 2x_{i,j,k}||_1$$
$$+ \omega_2 ||x_{i,j-1,k} + x_{i,j+1,k} - 2x_{i,j,k}||_1 \quad (8)$$
$$+ \omega_3 ||x_{i,j,k-1} + x_{i,j,k+1} - 2x_{i,j,k}||_1,$$

where $x_{i,j,k}$ are elements of $\mathcal{X}$. $\omega_i(i = 1, 2, 3)$ are the weight coefficient.

2.4. Proposed Model

Based on the foregoing, the final fusion model is as follows

$$\min_{\mathcal{X}} ||\mathcal{L}_{(3)} - \mathcal{X}_{(3)}P||_F^2 + \lambda||\mathcal{Y}_{(3)} - D\mathcal{X}_{(3)}||_F^2$$
$$+ \sum_{k=1}^{K} \sum_{i=1}^{4} \alpha_i ||\mathcal{C}_{k(i)}||_L + \lambda_2 ||T_\omega(\mathcal{X})||_1. \quad (9)$$

3. OPTIMIZATION ALGORITHM

We detail the process of solving model (9) using ADMM in this section. First, auxiliary variables $\{\mathcal{M}_i\}_{i=1}^{4}$ and F are introduced to obtain the following optimization problem

$$\min_{\mathcal{X},\{\mathcal{M}_i\}_{i=1}^{4},F} ||\mathcal{L}_{(3)} - \mathcal{X}_{(3)}P||_F^2 + \lambda||\mathcal{Y}_{(3)} - D\mathcal{X}_{(3)}||_F^2$$
$$+ \sum_{k=1}^{K} \sum_{i=1}^{4} \alpha_i ||\mathcal{G}_{\mathcal{M}_i}^k||_L + \lambda_2 ||F||_1 \quad (10)$$
$$s.t. \ \mathcal{X} = \mathcal{M}_i, F = T_\omega(\mathcal{X}), i = 1, ..., 4,$$

where $\mathcal{G}_{\mathcal{X}}^k = \mathcal{C}_k$. After integrating the two penalty terms into the objective function, problem (10) becomes

$$L(\mathcal{X}, \mathcal{M}_i, \mathcal{U}_i, F, V)$$
$$= ||\mathcal{L}_{(3)} - \mathcal{X}_{(3)}P||_F^2 + \lambda||\mathcal{Y}_{(3)} - D\mathcal{X}_{(3)}||_F^2$$
$$+ \sum_{i=1}^{4} (\sum_{k=1}^{K} \alpha_i ||\mathcal{G}_{\mathcal{M}_i}^k||_L + \frac{\mu}{2} ||\mathcal{X} - \mathcal{M}_i + \frac{\mathcal{U}_i}{\mu}||_F^2) \quad (11)$$
$$+ \lambda_2 ||F||_1 + \frac{\nu}{2} ||F - T_\omega(\mathcal{X}) + \frac{V}{\nu}||_F^2.$$

Here, $\mathcal{U}_i$ and V are Lagrange multipliers and μ and ν are penalty parameters. For the update of each variable, we proceed by solving the corresponding sub-problem obtained by fixing other variables.

1) *Sub-problem of $\mathcal{X}$:*

$$\min_{\mathcal{X}}||\mathcal{L}_{(3)} - \mathcal{X}_{(3)}P||_F^2 + \lambda||\mathcal{Y}_{(3)} - D\mathcal{X}_{(3)}||_F^2$$
$$+ \sum_{i=1}^{4} \frac{\mu}{2}||\mathcal{X} - \mathcal{M}_i + \frac{\mathcal{U}_i}{\mu}||_F^2 + \frac{\nu}{2}||F - T_\omega(\mathcal{X}) + \frac{V}{\nu}||_F^2. \tag{12}$$

It is easily settled by conjugate gradient (CG) method.

2) *Sub-problem of $\mathcal{M}_i (i = 1, ..., 4)$:*

$$\min_{\mathcal{M}_i} \sum_{k=1}^{K} \alpha_i ||\mathcal{G}_{\mathcal{M}_i}^k||_L + \frac{\mu}{2}||\mathcal{X} - \mathcal{M}_i + \frac{\mathcal{U}_i}{\mu}||_F^2. \tag{13}$$

It is a non-convex non-smooth low rank minimization problem, whose solution can be obtained by iteratively reweighted nuclear norm algorithm [11], and the solution is:

$$\mathcal{G}_{\mathcal{M}_i}^k = \text{Fold}_i(\text{SoftSh}(\mathcal{G}_{\mathcal{X}_{(i)}}^k + \frac{\mathcal{G}_{\mathcal{U}_{i(i)}}^k}{\mu}, \frac{\lambda_1 \alpha_i}{\mu}, w_i)) \tag{14}$$

$$w_i \in \partial g(\sigma_i(\mathcal{G}_{\mathcal{M}_i}^k)) = \frac{\lambda}{\gamma}exp(-\frac{\sigma_i(\mathcal{G}_{\mathcal{M}_i}^k)}{\gamma}). \tag{15}$$

Here, $\text{SoftSh}(Z, \tau, w) = US_{\tau w}(\Sigma)V^T$ is the singular value shrinkage operation, where $Z = U\Sigma V^T$ is the singular value decomposition of matrix Z and $S_{\tau w}(\Sigma) = \text{Diag}\{(\Sigma_{ii} - \tau w_i)_+\}$. $\text{Fold}_i(\cdot)$ is the mode-i folding of a matrix. $\mathcal{M}_i$ is obtained by rearranging patches in $\{\mathcal{G}_{\mathcal{M}_i}^k\}_{k=1}^K$.

3) *Sub-problem of F:*

$$\min_F \lambda_2 ||F||_1 + \frac{\nu}{2}||F - T_\omega(\mathcal{X}) + \frac{V}{\nu}||_F^2. \tag{16}$$

It is efficiently solved by soft-thresholding operator:

$$F = \text{SoftTh}(T_\omega(\mathcal{X}_{(i)}) - \frac{V_{(i)}}{\nu}, \frac{\lambda_2}{\nu}). \tag{17}$$

Here, $\text{SoftTh}(Z, \tau) = U\text{Sign}(\Sigma)(|\Sigma| - \tau)_+V^T$.

4) *Updating Multipliers:*

$$\mathcal{U}_i = \mathcal{U}_i + \mu(\mathcal{X} - \mathcal{M}_i), i = 1, ..., 4, \tag{18}$$

$$V = V + \nu(F - T_\omega(\mathcal{X})). \tag{19}$$

During each iteration, μ and ν increase by a smaller value. Algorithm 1 summarizes the steps to solve the proposed model.

4. EXPERIMENTS

4.1. Experimental setup

Experiments are conducted on two public HS remote sensing data sets, i.e., University of Pavia and Washington DC Mall. The first one has a resolution of 610×610 pixels, with 224 bands. After disposing of bands with noise and water vapor absorption, there are 93 bands left. We select a sub-image

---

Algorithm 1 NLRTV Fusion via ADMM

Input: $\mathcal{L}, \mathcal{Y}, P, D$, parameters λ, λ_1, and λ_2;
  Acquire the division and clustering results of $\mathcal{X}$.
  for $l = 1$ to MaxIter **do**
    Update $\mathcal{X}$ with CG on (12);
    for $i = 1$ to 4 **do**
      Update $\mathcal{M}_i$ by (14) and update $\mathcal{U}_i$ by (18);
    end for
    Update F by (17) and update V by (19);
  end for
Output: $\mathcal{X}$ (fused image).

---

of 128×128 pixels for experiments. The second one has a spatial resolution of 1280×307, with 210 bands. We only use a sub-image of 200×200 pixels, with 191 bands after getting rid of the water absorption ones. These original images in two data sets serve as ground truth. To obtain LRHS images $\mathcal{L}$, we down-sample original HS images by a factor of 4 and 5, respectively. Spectral response matrix D of a Nikon D700 camera is used to generate HRMS images $\mathcal{Y}$. Moreover, we add Gaussian noise with SNR = 35 dB to both LRHS and HRMS images.

To objectively evaluate the performance, four quality metrics are adopted, including peak signal to noise ratio (PSNR), spectral angle mapper (SAM) for evaluating spectral error, relative dimensionless global error in synthesis (ERGAS) for evaluating spatial quality, and Q2^n. Besides, four state-of-the-art methods are used for comparison, including CNMF [6], HYSURE [7], CSTF [8] and LTTR [9].

The unknown variables of our method are initialized to zero. By careful tuning in the experiment, the parameters λ, λ_1, and λ_2 are set to 1, 0.05, and 10^{-7}, respectively.

4.2. Experimental results

The fusion results at 771.67nm of University of Pavia data set are shown in the first row of Fig. 1. For better visual comparison, we introduce the difference image representing the difference between the fused image and ground truth. It can be observed that the reconstruction error of NLRTV is the smallest. The errors in the difference images of HYSURE and NLRTV are smooth, benefiting from their TV term. The second row of Fig. 1 displays the difference images at 1025.75nm of Washington DC Mall data set. LTTR and CSTF, which suffer from processing the fusion on the segmented spatial blocks, have many block errors in their difference images. Nevertheless, NLRTV has fewest distortions of roads and lands than those of others. Table 1 lists the average objective results of two data sets, with the best results marked in bold. Our method achieves the best quantitative evaluation results. Moreover, Fig. 2 shows the PSNRs of LTTR and NLRTV overall bands on two data sets. It is observed that NLRTV achieves satisfactory results on almost all bands.
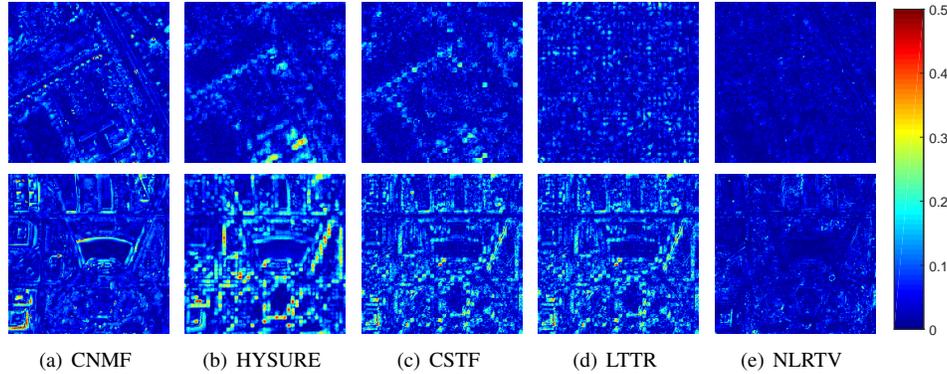
(a) CNMF (b) HYSURE (c) CSTF (d) LTTR (e) NLRTV

Fig. 1. Comparisons of difference images. The first row: University of Pavia. The second row: Washington DC Mall.

Table 1. Experimental evaluation metrics among five compared methods on two data sets.

Method	University of Pavia				Washington DC Mall			
	PSNR	SAM	ERGAS	$Q2^n$	PSNR	SAM	ERGAS	$Q2^n$
CNMF	32.91	2.75	1.85	0.58	28.62	3.64	1.84	0.31
HYSURE	32.21	3.31	2.14	0.72	24.48	5.23	2.58	0.45
CSTF	32.96	2.44	1.84	0.65	26.73	3.81	1.94	0.42
LTTR	35.20	3.30	1.48	0.65	31.06	3.19	1.33	0.68
NLRTV	**38.93**	**1.78**	**0.94**	**0.78**	**32.08**	**2.11**	**1.15**	**0.72**

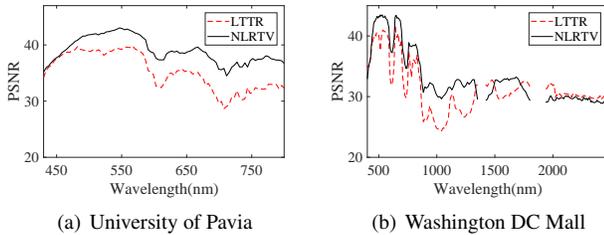
(a) University of Pavia (b) Washington DC Mall

Fig. 2. PSNRs for each band on two data sets.

5. CONCLUSION

In this paper, we propose a novel HS and MS image fusion method, based on non-convex low rank tensor approximation and second-order total variation. It exploits the intrinsic spatial-spectral correlation, nonlocal similarity and the smoothness structure of HS images. Besides, the optimization strategy is proposed for this model. Experiments are performed on two HS data sets, and the experimental results show this method is superior to several advanced methods.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] W. Zhang, X. Lu, and X. Li, "Similarity constrained convex nonnegative matrix factorization for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4810–4822, 2019.

[2] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 911–923, 2019.

[3] R. Dian and S. Li, "Hyperspectral and multispectral image fusion based on spectral low rank and non-local spatial similarities," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 3137–3140.

[4] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hypersharpening: A first approach on SIM-GA data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3008–3024, 2015.

[5] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 1963–1974, 2017.

[6] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2012.

[7] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2015.

[8] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4118–4130, 2018.

[9] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2672–2683, 2019.

[10] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by MS/HS fusion net," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1585–1594.

[11] C. Lu, J. Tang, S. Yan, and Z. Lin, "Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 829–839, 2016.

[12] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0 -minimization," *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 106–121, 2009.