NOISE REDUCTION OF HYPERSPECTRAL IMAGERY USING NONLOCAL SPARSE REPRESENTATION WITH SPECTRAL-SPATIAL STRUCTURE

Yuntao Qian, Minchao Ye, Qi Wang

Institute of Artificial Intelligence, College of Computer Science, Zhejiang University
Hangzhou 310027, China

ABSTRACT

Noise reduction is always an active research area in image processing due to its importance for the sequential tasks such as object classification and detection. In this paper, we develop a sparse representation based noise reduction method for hyperspectral imagery, which is dependent on the assumption that the non-noise component in the signal can be approximated by only a small number of atoms in a dictionary while noise component has not this property. The main contribution of the paper is in introducing nonlocal similarity and spectral-spatial structure of hyperspectral imagery into sparse representation. Non-locality means the self-similarity of image, by which the whole image can be partitioned into some groups containing similar patches. The similar patches in each group is sparsely represented with shared atoms making the signal and noise more easily separated. Sparse representation with spectral-spatial structure can exploit spectral and spatial joint correlations of hyperspectral imagery also making the signal and noise more distinguished, in which 3-D blocks are instead of 2-D patches for sparse coding. The experimental results indicate that the proposed method has a good quality of restoring the true signal from the noisy observation.

Index Terms— Hyperspectral imagery, noise reduction, sparse representation, nonlocal similarity, spectral-spatial structure

1. INTRODUCTION

Hyperspectral imagery is acquired by a new imaging technique, and draws many attentions from various application fields. It can provide much information about spectral and spatial distributions of distinct objects owing to its numerous and continuous spectral bands. The noise of hyperspectral imagery comes from sensor, photon effects, and calibration error. Meanwhile, the increased spatial, spectral, and radiometric resolutions of hyperspectral imaging lead to an increased impact of noise on the results extracted from this kind of imagery. Recently, smoothing filters, anisotropic diffusion, multi-linear algebra, and wavelet shrinkage methods have been exploited for noise reduction of hyperspectral imagery [1, 2]. In this paper, we focus on sparse representation based noise reduction with spectral-spatial structure.

Sparse representation is one of powerful denoising methods, which typically assumes that the true signal can be well approximated by a linear combination of few basis elements [3]. That is, the signal is sparsely represented in the transform domain. Some of existing sparse representation based image denoising methods imply that the whole image has the same properties of signal and noise in everywhere, which can be considered as a global filter ignoring the local details of an image. To overcome this drawback, adaptively sparse representation based on the neighborhood property is proposed, which can be considered as a local filter [2]. However, the information provided by the neighborhood is too limited to preserve the true structure, details and texture of an image.

Since the original nonlocal means algorithm was proposed for image denoising [4], non-locality or self-similarity became to be an very popular strategy for adaptively estimating statistical and geometric structures of signal and noise. Non-locality based denoising algorithms make use of the high degree of self-similarity of any natural image, i.e., every small patch in a natural image has many similar patches in the same image. Inspired by non-locality, a nonlocal sparse representation based noise reduction algorithm is introduced [5], in which sparse representations of the similar patches are recovered via a linear regularized regression model with shared constraint of sparsity, or called multi-task sparse representation. This method is based on the fact that the true signals in these similar patches can be represented by the same subset of basis elements while the noise lacks this consistence of representation.

Most of noise reduction techniques for hyperspectral imagery are based on band-by-band or pixel-by-pixel processing, i.e., they process each band image separately or each pixel's spectral signature separately [6]. But this may lead to loss of correlation between bands or between pixels. In order to further exploit spectral-spatial joint correlations of hyperspectral imagery, we construct the sparse representation of 3-D block instead of 2-D patch or 1-D line segment, which is based on the assumption that the true hyperspectral imagery is generally characterized by a strong spectral and spatial cor-

Thanks to the National Natural Science Foundation of China (No.61171151) for funding.

relation, and conversely, noise sources are commonly to be independent from one another in spectral-spatial domain.

In the section 2, the proposed denoising method called nonlocal sparse representation with spectral-spatial structure is introduced, and its implementation is also discussed. Section 3 shows the experimental results on the real hyperspectral data sets. Conclusions are drawn in section 4.

2. ALGORITHM AND IMPLEMENTATION

2.1. Denoising Model and Sparse Coding

The generic noisy observation model has the form

$$z(x) = y(x) + \eta(x) \tag{1}$$

where x is the position in the signal domain, $z(x)$ is the observed signal, $y(x)$ would be the true signal, and $\eta(x)$ is the noise perturbation.

For each 3-D block $\mathbf{z}^k$, the sparse representation based denoising model can be defined as

$$\mathbf{y}^k = D\hat{W}^k \tag{2}$$

$$\hat{W}^k = \arg\min_{W^k} \frac{1}{2}\|\mathbf{z}^k - DW^k\|_2^2 + \lambda\|W^k\|_1 \tag{3}$$

where D is the dictionary of basis elements, W^k is the vector of coefficients corresponding to the basis elements. The last term in Equation 3 is the sparsity norm of W^k, and λ controls the degree of sparsity. $\mathbf{y}^k$ is the recovered 3-D block from the noisy observation $\mathbf{z}^k$.

This model considers every 3-D blocks independently, regardless of their correlation. Non-locality indicates there is similarity between these overlapped blocks, and exploiting the similarity will benefit the denoising [5]. Therefore, the nonlocal sparse representation is introduced based on the idea that similar blocks should share the same set of basis elements. Suppose we have K noisy signals $\mathbf{z}^k$ to be denoised by sparse representation, and the size of dictionary is P (the number of basis elements). Each signal $\mathbf{y}^k$ has a linear sparse coding

$$\mathbf{y}^k = D\hat{W}^k = \sum_{p=1}^{P} D_p \hat{W}_p^k \tag{4}$$

A group of similar signals has a multi-task sparse coding

$$\hat{W} = \arg\min_{W} \sum_{k=1}^{K} \left\| \mathbf{z}^k - \sum_{p=1}^{P} D_p W_p^k \right\|_2^2 + \lambda \sum_{p=1}^{P} \|W_p\|_2 \tag{5}$$

where W is a $P \times K$ matrix of coefficients, W^k is its kth column, and W_p is its pth row. The constraint of sparsity in Equation 5 is a l_1/l_2 combined norm.

The difference between independent and nonlocal sparse representation is shown in Fig. 1, where $Y = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K]$.

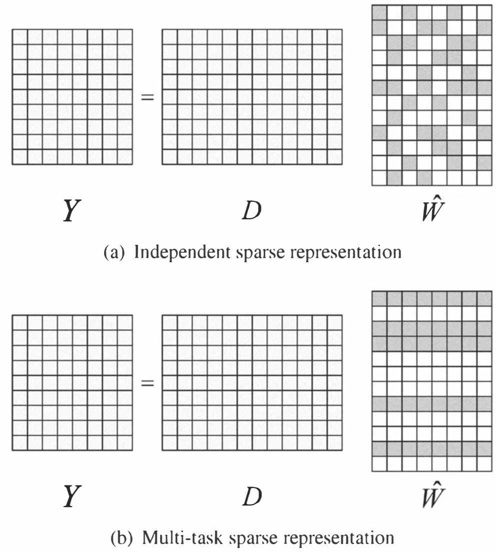


Fig. 1. Independent and nonlocal sparse representation

In matrix $\hat{W}$, the white small squares correspond to zero coefficients and the gray ones correspond to non-zero coefficients. The similar sparse structure can be derived via nonlocal sparse representation.

Furthermore, a single dictionary is always not enough to support a sparse representation for complicated structures. For example, the cosine transform dictionary is not effective in representing impulsive transitions and singularities, whereas wavelet transform dictionary do poorly for textures and smooth transitions. Therefore, it is natural to use several complementary dictionaries for achieving a sparse representation over general signal. Here two dictionaries of 3-D biorthogonal Daubechies wavelet and 3-D cosine are used, i.e., $D = D_{wavelet} \cup D_{cosine}$.

The denoising algorithm of nonlocal sparse representation with spectral-spatial structure can be summarized as:

1. Divide a whole hyperspectral cube into overlapping 3-D blocks with the size of $n \times n \times n$, and each voxel (i, j, k) has a block $\mathcal{N}(i, j, k)$ whose center is this voxel. The overlapping division can avoid blocking effect.

2. Partition the blocks into several groups by affine propagation (AP) clustering algorithm according to their similarity [7]. The similarity is measured by Euclidean distance between a pair of blocks

$$s(i, j, k, i', j', k') = \|\mathcal{N}(i, j, k) - \mathcal{N}(i', j', k')\|_2^2 \tag{6}$$

3. Apply nonlocal sparse representation model for each group of blocks, and estimate the true signals for all blocks in one group by Equations 4 and 5.

4. Calculate the denoised value of each voxel in the hyperspectral cube via weighted average method, as there are several blocks surrounding one voxel.

2.2. Implementation

Two problems should be highlighted in implementation of the algorithm. The first is concerned about step 2. As the number of blocks that is equal to the size of a hyperspectral cube because every voxel has their own subcube, is too large to AP or C-means clustering algorithm. To reduce the computational burden of the clustering algorithm, the grouping is performed within a three-dimensional window of size $S \times S \times S$ centered at the coordinate of the current reference cube. Then we use AP algorithm again for these groups for merging similar groups, in which each group is represented by a data point. This divided-and-conquer scheme is very effective to real applications.

The second is the optimization problem of nonlocal sparse representation in step 3. The non-smoothness of the $\ell_{2,1}$-norm regularization in Equation 5 makes the optimization be a challenging problem. In this paper, we use a fast optimization algorithm proposed by Liu and et al [8]. Consider a general $\ell_{2,1}$ regularized minimization problem

$$\min_{W} f(W) + \lambda \|W\|_{2,1} \tag{7}$$

where $f(\cdot)$ is a differentiable loss function and $\|W\|_{2,1}$ is the $\ell_{2,1}$ norm. The vector of coefficients W^* is obtained via iteratively applying accelerated gradient descent and proximal operator (see Algorithm 1).

Algorithm 1 $\ell_{2,1}$ regularized minimization

Input:
loss function $f(\cdot)$
regularization parameter λ
step size t^0 and affine combination parameter β^0
Output:
vector of coefficients W^*
$k \leftarrow k+1$;
Compute the search point via affine combination:
$S^{(k)} = W^{(k)} + \beta^{(k)}(W^{(k)} - W^{(k-1)})$;
Calculate the gradient descent point $\mathbf{U}^{(k+1)}$ with adaptive step size:
$U^{(k+1)} = S^{(k)} - t^{(k)} \nabla f(S^{(k)})$;
Apply the proximal operator to calculate W^{k+1}:
$W^{(k+1)} = \arg\min_{W} \frac{1}{2}\|W - U^{(k+1)}\|_2^2 + t^{(k)}\lambda\|W\|_{2,1}$;
Update t^{k+1} and β^{k+1} for next iteration.
Repeat the above steps until the difference between $W^{(k+1)}$ and $W^{(k)}$ is smaller than a threshold;
return $W^* = W^{(k+1)}$.

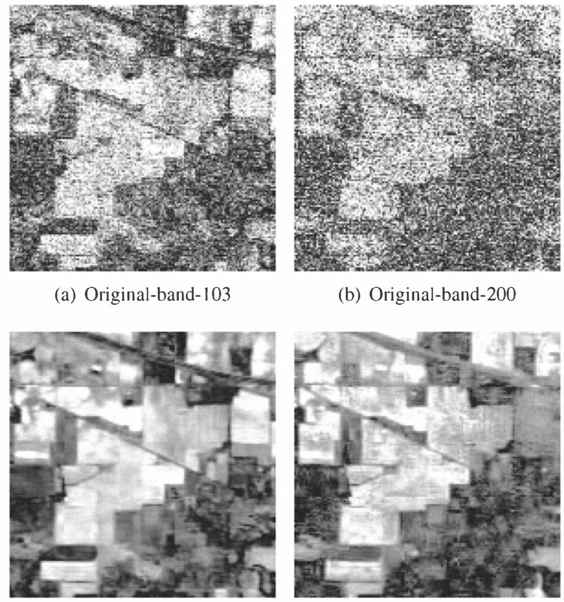

(a) Original-band-103 (b) Original-band-200
(c) Denoised-band-103 (d) Denoised-band-200

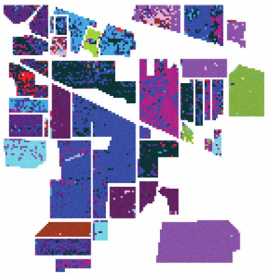
Fig. 2. Denoised results on Indian Pine data set.

3. EXPERIMENTAL RESULTS

We will show some experimental results on two real hyperspectral data sets. The first data set was acquired by NASA AVIRIS instrument over the Indian Pine Test Site in Northwestern Indiana in 1992. The image size is defined as 145×145, and the number of bands is 220. The size of 3-D block is $5 \times 5 \times 5$, and the parameter of sparsity is $\lambda = \frac{1}{2}\sqrt{m}$ where m is the number of blocks in a group.

We evaluate the proposed method by visually comparing the denoised hyperspectral imagery with the original imagery, and quantitatively comparing the classification results based on original and denoised images separately. It can be found from Fig.2 that the noise is reduced and the details are well preserved. Moreover, Fig.3 gives the classification results before and after noise reduction. This hyperspectral imagery contains 16 land-cover classes and 10366 labeled pixels. We randomly selected and 5% and 25% labeled pixels from each class for training, and use the rest for test. It can be clearly seen that the denoised data yield a much better classification performance.

The second data set was acquired by a hyperspectral camera on the ground with a high spatial resolution of 1280×960, and there are 58 spectral bands in total. This data set contain more details than the first one. Due to the high resolution, we use a relatively large block size $9 \times 9 \times 9$. Other parameters remain the same as the first experiment. The denoised results on band 58 are shown in Fig 4. We can find our method works well in both of edge/detail preservation and noise removal.



(a) Classification on original data with 5% training samples (OA=74.84%)

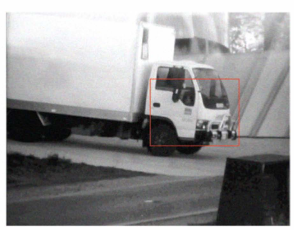
(b) Classification on denoised data with 5% training samples (OA=88.65%)

(c) Classification on original data with 25% training samples (OA=89.34%)

(d) Classification on denoised data with 25% training samples (OA=97.42%)

Fig. 3. Classification results on Indian Pine data set. OA is overall accuracy.

(a) Original-band-58



(b) Denoised-band-58

(c) Part of original-band-58

(d) Part of denoised-band-58

Fig. 4. Denoised results on a hyperspectral data set acquired by a ground-based camera

4. CONCLUSIONS

In this paper, we developed a novel sparse representation based denoising method for hyperspectral imagery, which combines nonlocal sparse coding with spectral-spatial structure to better separate the true signal and noise. Nonlocal sparse coding generates the sparse representations for several similar signal making them share the same basis elements in the dictionary. It considers the consistence of true signals and the randomness of noises so that the obtained sparse representation can more precisely recover the true signals. Instead of 1-d line segments or 2-D patches, 3-D blocks are used for sparse representation, which enables the sparse representation to make the most of the correlations between spectral bands and spatial neighbors of true signal. Experimental results on the real hyperspectral data sets show the effects of the proposed method in noise reduction.

5. REFERENCES

[1] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, 2008.

[2] G. Chen and S. Qian, "Denoising and dimensionality reduction of hyperspectral imagery using wavelet packets, neighbour shrinking and principal component analysis," *International Journal of Remote Sensing*, vol. 30, no. 18, pp. 4889–4895, 2009.

[3] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.

[4] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multisc. Model. Simulat.*, vol. 4, no. 2, pp. 490–530, 2005.

[5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *IEEE 12th International Conference on Computer Vision*, 2009, pp. 2272–2279.

[6] S. Bourguignon, D. Mary, and E. Slezak, "Sparsity-based denoising of hyperspectral astrophysical data with colored noise: Application to the muse instrument," in *2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2010.

[7] J. F. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, pp. 972–976, 2007.

[8] J. Liu, S. Ji, and J. Ye, "Multi-task feature learning via efficient $\ell_{2,1}$-norm minimization," in *Uncertainty in Artificial Intelligence (UAI)*, 2009.