HYPERSPECTRAL IMAGE DENOISING BASED ON GLOBAL AND NON-LOCAL LOW-RANK FACTORIZATIONS

Lina Zhuang and José M. Bioucas-Dias

Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal.

ABSTRACT

The ever increasing spectral resolution of the hyperspectral images (HSIs) is often obtained at the cost of a decrease in the signal-to-noise of the measurements, thus calling for effective denoising techniques. HSIs from the real world live in low dimensional subspaces and are self-similar. The low dimensionality stems from the high correlation existing among the reflectance vectors and the self-similarity is common to images of the real world. In this paper, we exploit the above two properties. The low dimensionality is a global property, which enables the denoising to be formulated just with respect to the subspace representation coefficients, thus greatly improving the denoising performance and reducing the processing computational complexity. The self-similarity is exploited via low-rank tensor factorization of non-local similar 3D-patches. The proposed factorization hinges on optimal shrinkage/thresholding of SVD singular value of low-rank tensor unfoldings. As a result, the proposed method has no parameters, apart from the noise variance. Its effectiveness is illustrated in a comparison with state-of-the-art competitors.

Index Terms— Hyperspectral image denoising, self-similarity, 3D-patches, low-rank tensor factorization.

1. INTRODUCTION

The hyperspectral cameras measure the electromagnetic energy scattered in their instantaneous field view in hundreds or thousands of spectral channels with remarkably high spectral resolution, which enables material identification with precision via spectroscopic analysis [1]. Hyperspectral remote sensing images have been widely used in countless applications, (e.g., earth observation, environmental protection and natural disaster monitoring). However, the measurement noise often precludes the widespread use of HSIs in precise material identification (e.g., precision farming) applications.

Natural images are self-similar. This means that they contain many similar patches at different locations or scales. This characteristic has been recently exploited by the patch-based image restoration methods and holds the state-of-the-art in image denoising. Representative examples of this methodology in single-band images include the non-local means filter [2], the Gaussian mixture model (GMM) learned from the noisy image [3], and the collaborative filtering of groups of similar patches BM3D [4]. Identical ideas have been pursued in multi-band image denoising: BM4D [5], VBM4D [6], and MSPCA-BM3D [7] use collaborative filtering in groups of 3D patches extracted from volumetric data, videos, multispectral data, respectively. DHOSVD [8] applies hard threshold filter in coefficients of higher order SVD of similar patches. In HSI denoising, we refer NAILRMA [9] that uses tensor factorization and FastHyDe [10] that exploits the self-similarity of the HSIs representation coefficients in suitable subspaces.

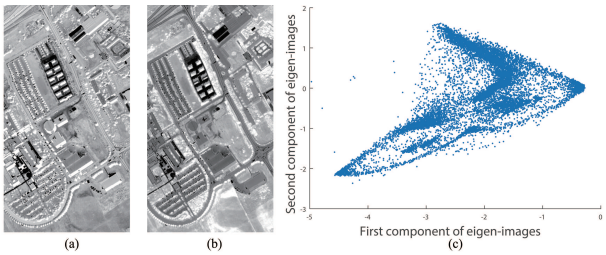

Fig. 1. Pavia HSI. (a) 1st eigen-image (image of the first representation coefficient); (b) 2nd eigen-image; (c) scattergram of (a)-(b).

The computational complexity of most HSI denoising methods is very high, owing to the large number of bands, usually on the order of hundreds. It happens, however, that hyperspectral vectors leave, with very good approximation, in subspaces of low dimension, compared with the number of bands [1], and this subspace may be accurately inferred from the observed HSIs [11]. Therefore, instead of denoising the original HSI, we may formulate the denoising with respect to the HSI subspace representation coefficients. This is the research line followed in FastHyDe [10], which is, to our knowledge, state-of-the-art and faster than the competitors.

The representation coefficients image (RCI) is formed by a few image bands, each linked with a subspace basis vector, hereafter termed eigen-images. FastHyDe denoise the eigen-images independently from each other, thus implicitly assuming that they are statistically independent. An informal justification for the independent processing is that the eigen-images are uncorrelated, owing to the use eigenvectors of the signal correlation matrix as subspace basis. However, uncorrelation

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7-PEOPLE-2013-ITN) under grant agreement n° 607290 SpaRTaN. This work was partially supported by the Fundação para a Ciência e Tecnologia, Portuguese Ministry of Science and Higher Education, project UID/EEA/50008/2013.

does not imply independence as illustrated in Fig. 1, where parts (a) and (b) show two eigen-images of the Pavia HSI and part (c) shows the scattergram of 2D vectors formed by the two gray levels of the same pixel in (a) and (b). The statistical dependence is clearly observed both in the couple (a)-(b) and in the (c). There is still, therefore, room to improve the HSI denoising performance, by taking the statistical dependence between the eigen-images into consideration.

1.1. Contribution

The paper exploits two main characteristics of hyperspectral data: 1) HSIs are well approximated by low-dimensional subspaces [1]; 2) HSIs are self-similar and thus RCIs are also self-similar. The latter property is exploited via low-rank tensor factorization of non-local similar 3D-patches. The proposed factorization hinges on optimal shrinkage/thresholding of SVD singular values of the low-rank tensor unfoldings. As a result, the proposed method has no parameters, apart from the noise variance.

The paper is organized as follows. Section 2 formulates the problem of HSI denoising. Section 3 introduces our new denoising algorithm step by step. Section 4 presents experimental results including comparisons with the state-of-the-art. Section 5 concludes the paper.

2. PROBLEM FORMULATION

2.1. Observation model

Let $\mathbf{X} := [\mathbf{x}_1, \dots \mathbf{x}_n] \in \mathbb{R}^{n_b \times n}$ denote a HSI with n spectral vectors (the columns of $\mathbf{X}$) of size n_b (the number of bands of the sensor). Under the additive noise assumption, the observation model may be written as

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}, \tag{1}$$

where $\mathbf{Y}, \mathbf{N} \in \mathbb{R}^{n_b \times n}$ represent the observed HSI data and noise, respectively.

An usual assumption in HSIs is that the columns (spectral vectors) of matrix $\mathbf{X}$ live in a low-dimensional subspace that may be estimated from the observed data $\mathbf{Y}$ with good approximation [1, 10]. Thus, we write $\mathbf{X} = \mathbf{E}\mathbf{Z}$, with $\mathbf{E} \in \mathbb{R}^{n_b \times k}$ and $k \ll n_b$, and $\mathbf{E}$ holding an orthogonal basis for the signal subspace. Hence, the observation model (1) may be written as

$$\mathbf{Y} = \mathbf{E}\mathbf{Z} + \mathbf{N}. \tag{2}$$

Hereafter, $\mathbf{Z}$ is termed the RCI and its rows the eigen-images.

2.2. RCI Denoising

Assuming that the noise is zero-mean Gaussian independent and identically distributed (other covariance matrices are easily dealt with [10]) and that $\mathbf{E}$ was learned from observated data $\mathbf{Y}$ [11], the RCI denoising problem is formulated as

$$\widehat{\mathbf{Z}} = \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{Y} - \mathbf{E}\mathbf{Z}\|_F^2 + \lambda\phi(\mathbf{Z}) \tag{3}$$

$$= \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{E}^T\mathbf{Y} - \mathbf{Z}\|_F^2 + \lambda\phi(\mathbf{Z}), \tag{4}$$

where $\|\mathbf{X}\|_F^2 = \text{trace}(\mathbf{X}\mathbf{X}^T)$ is the Frobenius norm of $\mathbf{X}$. The first term on the right-hand side represents the data fidelity and accounts for zero-mean Gaussian i.i.d noise. The second term is a regularizer expressing prior information tailored to self-similar RCIs. We remark that (4) is equivalent to a denoising problem where the additive noise is $\mathbf{E}^T\mathbf{N}$, thus with covariance $\sigma^2\mathbf{I}_k$, where $\mathbf{I}_k$ is the identity matrix of size k. Therefore, the noise attenuation resulting from the subspace representation is $\mathbb{E}[\|\mathbf{E}^T\mathbf{N}\|_F^2]/\mathbb{E}[\|\mathbf{N}\|_F^2] = k/n_b$, which is often of the order of tens.

The solution of the optimization (3) is

$$\widehat{\mathbf{Z}} = \Psi_{\lambda\phi}(\mathbf{E}^T\mathbf{Y}), \tag{5}$$

where $\Psi_{\lambda\phi}(\mathbf{U}) = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{U} - \mathbf{X}\|_F^2 + \lambda\phi(\mathbf{X})$ is the so-called denoising operator, or PO of ϕ [12]. Instead of tailoring ϕ to promote self-similar RCIs and then solve (4), we adopt the powerful patch-based framework. Accordingly, $\mathbf{Z}$ is decomposed into 3D-patches, which are denoised and recombined, yielding the estimates $\widehat{\mathbf{Z}}$ and $\widehat{\mathbf{X}} = \mathbf{E}\widehat{\mathbf{Z}}$. The global structure of the proposed RCI denoiser is partially inspired in BM3D [4] and BM4D [5]. The differences concern the formulation with respect to the RCI and the filtering, which is adaptive and based on tensor factorization.

3. GLF - THE PROPOSED METHOD

We propose a HSI denoising algorithm with four steps: 1) Subspace identification; 2) RCI grouping of 3D-patches; 3) low-rank factorization of the grouped 3D-patches; and 4) 3D-patch aggregation. These four steps are shown in Fig. 2. As stated before, the general structure of the steps 2, 3, and 4 is inspired on BM4D. The differences reside in step 1), which is not implemented in BM4D, and in the filtering step, which in our case is based on SVD singular value thresholding/shrinkage, and thus data adaptive, whereas in BM4D is fixed and based on 3D DCT and 1D Haar wavelets. Below, we detail each step. Given that the steps 1 and 3 correspond to, respectively, a global matrix factorization and to local tensor factorizations, we term our method Global Local Factorization (GLF).

The subspace identification step is carried out, for i.i.d. noise, via SVD of the correlation matrix $\mathbf{Y}^T\mathbf{Y}/n$. For the non i.i.d. noise, the observation noise is firstly whitened (see [10]). Below we focus on the GLF steps 2, 3, and 4.

3.1. Grouping RCI 3D-patches

Let $\mathbf{Y}'_i = \mathbf{Z}_i + \mathbf{N}'_i \in \mathbb{R}^{k \times \sqrt{s} \times \sqrt{s}}$ denote 3D-patches of the projected observed HSI $\mathbf{E}^T\mathbf{Y}$, where $\mathbf{Z}_i$ and $\mathbf{N}'_i$ are the corresponding 3D-patches of the clean RCI and of the projected noise $\mathbf{E}^T\mathbf{N}$, $i \in \{1, \dots, n\}$ is the index of the top-left pixel, k is the subspace dimension and $\sqrt{s}$ is the spatial size of the 3D-patch. Given $\mathbf{Y}'_i$, we find, in a neighborhood of pixel i of size 79×79, a set of similar $(b-1)$ 3D-patches, with indexes $i_1, i_2, \dots, i_{b-1}$, according to the Frobenius distance. With $\mathbf{Y}'_i$ and the $(b-1)$ similar patches, we build a

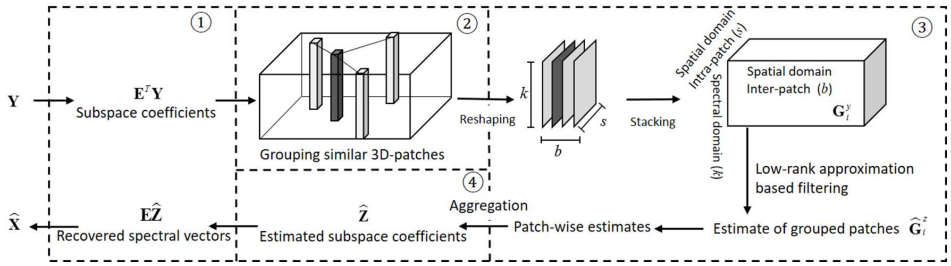

Fig. 2. Flowchart of the proposed denoising algorithm GLF.

3D-tensor $\mathbf{G}_i^y \in \mathbb{R}^{k\times b\times s}$ (see Fig. 2), such that $\mathbf{G}_i^y(:,j,:) = (\mathbf{Y}'_j)_{(1)} \in \mathbb{R}^{k\times s}$, for $j \in \mathcal{I}_i = (i, i_1, i_2, \ldots, i_{b-1})$, and $(\mathbf{Y}'_j)_{(1)}$ is the mode-1 matrix unfolding of tensor $\mathbf{Y}'_j$. By construction, $\mathbf{G}_i^y = \mathbf{G}_i^z + \mathbf{G}_i^n$, where $\mathbf{G}_i^z$ and $\mathbf{G}_i^n$ are the tensors corresponding to the clean RCI and projected noisy 3D-patches, respectively.

3.2. Low-rank factorization of grouped 3D-patches

Due to the way $\mathbf{G}_i^y$, for $i \in \{1,\ldots,n\}$, is constructed, its signal component, $\mathbf{G}_i^z$, is expected to have low-rank, whereas its noise component, $\mathbf{G}_i^n$, is expected to have higher rank, which might be as higher as $\min\{ks, kb, sb\}$ [13]. On the other hand, the lower is the rank of the approximation the higher is noise attenuation. Therefore, by finding the lowest rank approximations for the 3D-tensors $\mathbf{G}_i^y$, we are maximally attenuating the noise while preserving the signal.

Computing a low-rank approximations $\widehat{\mathbf{G}}_i^z$ of $\mathbf{G}_i^z$, from $\mathbf{G}_i^y$, able to minimize the mean square error $\|\widehat{\mathbf{G}}_i^z - \mathbf{G}_i^z\|_F$, is a hard problem [13]. On the other, in the case of matrices and using results from random matrix theory [14], the low-rank approximations for matrices in additive i.i.d. Gaussian noise may be optimally computed, in an asymptotic sense. This computation is carried out via SVD, by properly thresholding and shrinking its singular values [14, 15]. In summary, and taking the observation model (1) for the sake of explanation, the estimated $\mathbf{X}$ is given by $\widehat{\mathbf{X}} = \mathbf{U}T(\mathbf{\Sigma})\mathbf{V}^T$, where $\{\mathbf{U},\mathbf{\Sigma},\mathbf{V}\} = \text{SVD}(\mathbf{Y})$ and $T(\mathbf{\Sigma}) = \text{Diag}[t(\sigma_1), t(\sigma_2), \ldots, t(\sigma_{n_b})]$, with $t : \mathbb{R}_+ \to \mathbb{R}_+$ being the shrinkage/thresholding function shown in [15], expression (10). In this work, we exploit these results by successively applying them to the mode-1, mode-2, and mode-3 matrix unfoldings of tensor $\mathbf{G}^y \in \mathbb{R}^{k\times b\times s}$, as shown in Algorithm 1. The spatial indexes are omitted for simplicity.

Line 3 of Algorithm 1 computes the projection coefficients of the columns $\mathbf{G}^y_{(1)}$ onto the range of $\mathbf{U}_1(:,1:k')$, where $(\sigma_1,\ldots,\sigma_k) = \text{diag}(\mathbf{\Sigma}_1)$ and $k' = \arg\min_l t(\sigma_l)$ subject to $t(\sigma_l) > 0$. Since matrix $\mathbf{U}(:,1:k')$ is orthogonal, the additive noise present in $\widehat{\mathbf{G}}_{(1)}$ has density $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{k'})$. This is important as it enables the subsequent application of the same filtering thresholding results based on random matrix theory [14, 15]. Line 6 is similar to line 3 applied to dimension 2 of the tensor $\widehat{\mathbf{G}}$. In Line 9, we apply not only thresholding as before, but also shrinkage as defined in expressions (7)-(10) in [15]. In lines 11 and 12 the tensor is projected back to the original dimensions.

Algorithm 1 Successive SVD low-rank factorization[1]

1: Filtering in the spectral domain
2: $\{\mathbf{U}_1, \mathbf{\Sigma}_1\} := \text{SVD}\big(\mathbf{G}^y_{(1)}\big)$, $\quad \mathbf{G}^y_{(1)} \in \mathbb{R}^{k\times(bs)}$
3: $\widehat{\mathbf{G}}_{(1)} := \mathbf{U}_1(:,1:k')^T\mathbf{G}^y_{(1)}$, $\quad \widehat{\mathbf{G}} \in \mathbb{R}^{k'\times b\times s}$, $\quad k' \le k$
4: Filtering in spatial domain (intra-patch)
5: $\{\mathbf{U}_2, \mathbf{\Sigma}_2\} := \text{SVD}\big(\widehat{\mathbf{G}}_{(2)}\big)$, $\quad \widehat{\mathbf{G}}_{(2)} \in \mathbb{R}^{b\times(k's)}$
6: $\widehat{\mathbf{G}}_{(2)} := \mathbf{U}_2(:,1:b')^T\widehat{\mathbf{G}}_{(2)}$, $\quad \widehat{\mathbf{G}} \in \mathbb{R}^{k'\times b'\times s}$, $\quad b' \le b$
7: Filtering in spatial domain (inter-patch)
8: $\{\mathbf{U}_3, \mathbf{\Sigma}_3, \mathbf{V}_3\} := \text{SVD}\big(\widehat{\mathbf{G}}_{(3)}\big)$, $\quad \widehat{\mathbf{G}}_{(3)} \in \mathbb{R}^{s\times(k'b')}$
9: $\widehat{\mathbf{G}}_{(3)} := \mathbf{U}_3 T(\mathbf{\Sigma}_3)\mathbf{V}_3^T$, $\quad \widehat{\mathbf{G}} \in \mathbb{R}^{k'\times b'\times s}$
10: Inverse transform to compute $\widehat{\mathbf{G}}^z$
11: $\widehat{\mathbf{G}}_{(2)} := \mathbf{U}_2(:,1:b')\widehat{\mathbf{G}}_{(2)}$, $\quad \widehat{\mathbf{G}} \in \mathbb{R}^{k'\times b\times s}$
12: $\widehat{\mathbf{G}}_{(1)} := \mathbf{U}_1(:,1:k')\widehat{\mathbf{G}}_{(1)}$, $\quad \widehat{\mathbf{G}} \in \mathbb{R}^{k\times b\times s}$
13: **return** $\widehat{\mathbf{G}}^z := \widehat{\mathbf{G}}$

3.3. 3D patch aggregation

The estimated 3D-patches contained in the tensors $\widehat{\mathbf{G}}_i^z$, for $i = 1,\ldots,n$, are added back to their original location in the RCI and then averaged taking into account that a given RCI pixel belongs various 3D-patches. Finally, we compute $\widehat{\mathbf{X}} = \mathbf{E}\widehat{\mathbf{Z}}$.

4. EXPERIMENTS

A noisy hyperspectral dataset (Fig. 3) was simulated as follows: sixteen very low signal-to-noise bands in original Pavia University data[2] were removed as the signal in these wavelength regions is largely attenuated due to water vapor in the atmosphere leading mostly noise. Furthermore, the remaining spectral vectors were then projected onto the signal subspace of dimension 8 learned via performing SVD on correlation matrix. The obtained HSI is

[1]The demos of GLF can be downloaded from `http://www.lx.it.pt/~bioucas/code/GLF_demos.rar`.

[2]Pavia scenes were provided by Prof. Paolo Gamba from the Telecommunications and Remote Sensing Laboratory, Pavia university (Italy) and can be downloaded from `http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes`.

considered the clean HSI to which Gaussian i.i.d. noise with $\sigma \in \{0.02, 0.04, 0.06, 0.08, 0.10\}$ was added to simulate noisy HSIs. Following the same procedure, noisy HSIs of Washington DC Mall data[3] (Fig. 4) subscene were also generated.

The denoising performance of proposed GLF is compared with those of BM4D [5], NAILRMA [9], two versions of FastHyDe [10] (namely each component of subspace representation coefficients is denoised by BM3D [4] or LRCF [15]), "Subspace+BM4D [5]" (meaning BM4D applied in subspace representation coefficients). Parameters related to non-local patch based framework, including patch-size ($N_1 = 10$), size of search region $N_S \times N_S$ ($N_S = 79$), sliding step to process every next reference patch ($N_{\text{step}} = 3$) , and (maximum) cardinality of similar patch groups ($N_2 = 16$), are set to the same in BM3D, BM4D, LRCF and GLF. For quantitative assessment, the peak signal-to-noise (PSNR) index and the structural similarity (SSIM) index [10] of each band are calculated. The mean PSNR (MPSNR) and mean SSIM (MSSIM) results of compared methods in two datasets are reported in Tab. 1 and Tab. 2.

Table 1. Quantitative assessment of different denoising algorithms applied to Pavia University dataset.

Index	Noisy image	BM4D	NAILRMA	FastHyDe (BM3D)	FastHyDe (LRCF)	Subspace +BM4D	GLF
MPSNR	33.98	45.16	46.53	47.71	47.74	47.67	**48.66**
MSSIM	0.8731	0.9903	0.9921	0.9946	0.9944	0.9946	**0.9956**
Time	-	11258	447	**43**	1268	211	470
MPSNR	27.96	40.98	41.80	43.40	43.35	43.35	**44.32**
MSSIM	0.6688	0.9767	0.9787	0.9870	0.9863	0.9869	**0.9893**
Time	-	11224	298	**44**	1212	211	421
MPSNR	24.44	38.60	38.90	41.02	40.94	40.96	**41.86**
MSSIM	0.5084	0.9622	0.9600	0.9791	0.9776	0.9786	**0.9824**
Time	-	11028	272	**46**	1277	215	400
MPSNR	21.94	36.91	36.96	39.39	39.28	39.12	**40.14**
MSSIM	0.3951	0.9472	0.9405	0.9714	0.9691	0.9689	**0.9753**
Time	-	11050	225	**46**	1280	42	400
MPSNR	20.00	35.58	35.37	38.15	38.03	38.03	**38.84**
MSSIM	0.3148	0.9319	0.9170	0.9638	0.9608	0.9617	**0.9682**
Time	-	11187	233	**49**	1302	215	400

Table 2. Quantitative assessment of different denoising algorithms applied to Washington DC Mall dataset.

Index	Noisy image	BM4D	NAILRMA	FastHyDe (BM3D)	FastHyDe (LRCF)	Subspace +BM4D	GLF
MPSNR	33.98	45.03	50.90	54.72	54.71	53.86	**55.22**
MSSIM	0.6382	0.8790	0.9678	0.9939	0.9938	0.9877	**0.9936**
Time	-	15553	505	**31**	810	129	342
MPSNR	27.96	41.11	45.73	49.92	49.93	49.23	**50.43**
MSSIM	0.5367	0.8119	0.9362	0.9840	0.9839	0.9766	**0.9826**
Time	-	15471	335	**31**	729	129	311
MPSNR	24.44	38.90	43.18	47.29	47.28	46.82	**47.84**
MSSIM	0.4634	0.7658	0.9189	0.9776	0.9775	0.9724	**0.9779**
Time	-	15275	262	**31**	733	128	287
MPSNR	21.94	37.37	41.14	45.23	45.25	44.95	**45.76**
MSSIM	0.4032	0.7307	0.8906	0.9618	0.9637	0.9589	**0.9627**
Time	-	15318	215	**30**	748	126	258
MPSNR	20.00	36.17	39.50	43.75	43.75	43.46	**44.13**
MSSIM	0.3530	0.7045	0.8705	0.9570	0.9578	0.9526	**0.9569**
Time	-	15325	236	**31**	708	128	262

The methods working in subspace representation coefficients yield better results than BM4D and NAILRMA, which work in the original spectral space. The subspace dimension input to the former methods was 10, instead of the 8, the true value, which indicates robustness with respect to subspace overestimation. Results in Tab. 1 and Tab. 2 show that GLF yields uniformly the best performance. The quality of reconstructed bands and spectra may also be inferred from Fig. 3 and Fig. 4.

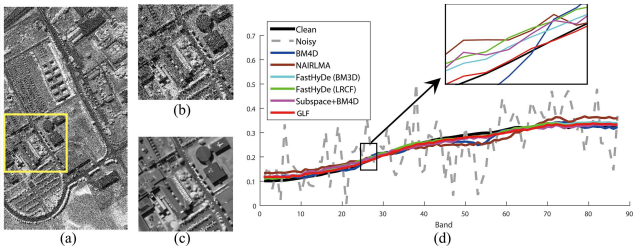

Fig. 3. Pavia University dataset of size $610 \times 339 \times 87$ (MPSNR = 20.00 dB): (a) Noisy image (band 76), (b) A subset of noisy image, (c) A subset of estimated image (MPSNR = 38.84 dB) by GLF, and (d) A denoised spectral signature.

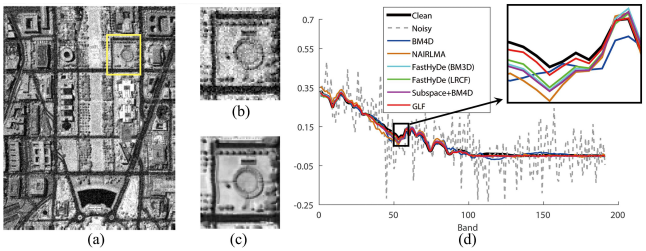

Fig. 4. A subset of Washington DC Mall dataset of size $400 \times 307 \times 191$ (MPSNR = 20.00 dB): (a) Noisy image (band 61), (b) A subset of noisy image, (c) A subset of estimated image (MPSNR = 44.13 dB) by GLF, and (d) A denoised spectral signature.

The running time of denoisers are reported in Tab. 1 and Tab. 2. The time comparison between BM4D and 'Subspace+BM4D' shows that denoising in subspace coefficients highly reduces the computational complexity. FastHyDe takes shortest time. But note that available codes of BM3D, BM4D and LRCF are programmed in C language, and codes of NAILRMA and proposed GLF are implemented in MATLAB. GLF is supposed to be further speed up by programming in C language.

5. CONCLUSIONS

We have proposed GLF, a new HSI denoiser, which represents HSIs in a low-dimensional subspaces. The denoising is formulated in the subspace representation coefficients. Considering components of subspace representation coefficients are self-similar and statistical dependent, they are jointly denoised in a non-local patch-based framework. Furthermore, a new filter based on low-rank approximation is devised for filtering noisy 3D-patches. In a comparison with the state-of-the-art GLF uniformly outperformed the competitors.

[3]This data set is available from the Purdue University Research Repository (`https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html`)

References

[1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Jun. 2005, vol. 2, pp. 60–65 vol. 2.

[3] A. M. Teodoro, M. S. C. Almeida, and M. A. T. Figueiredo, "Single-frame image denoising and inpainting using gaussian mixtures.," in *ICPRAM (2)*, 2015, pp. 283–288.

[4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[5] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 119–133, Jan. 2013.

[6] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 3952–3966, Sep. 2012.

[7] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian, "Denoising of multispectral images via nonlocal groupwise spectrum-PCA," in *Conference on Colour in Graphics, Imaging, and Vision*. Society for Imaging Science and Technology, 2010, vol. 2010, pp. 261–266.

[8] A. Rajwade, A. Rangarajan, and A. Banerjee, "Image denoising using the higher order singular value decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 849–862, 2013.

[9] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.

[10] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising based on low rank and sparse representations," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2016)*, 2016.

[11] J. M. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[12] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[13] N. D. Sidiropoulos, L. D. Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *arXiv preprint arXiv:1607.01668*, 2016.

[14] A. A. Shabalin and A. B. Nobel, "Reconstruction of a low-rank matrix in the presence of Gaussian noise," *Journal of Multivariate Analysis*, vol. 118, pp. 67–76, Jul. 2013.

[15] M. Nejati, S. Samavi, S. Soroushmehr, and K. Najarian, "Low-rank regularized collaborative filtering for image denoising," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 730–734.