Hyper-spectral Impulse Denoising: A row-sparse Blind Compressed Sensing Formulation

[1,2]Angshul Majumdar, [2]Naushad Ansari and [2]Hemant Kumar Aggarwal
{[1]angshul, naushada and hemanta}@iiitd.ac.in

[1]Department of Electrical and Computer Engineering, University of British Columbia
[2]Indraprastha Institute of Information Technology – Delhi

ABSTRACT

This paper addresses the problem of impulse denoising from hyper-spectral images. Impulse noise is sparse; removing impulse noise requires minimizing an l_1-norm data fidelity term. Prior studies have exploited the intra-band spatial correlation (leading to sparsity in transform domain) and inter-band spectral-correlation (joint-sparsity) of hyper-spectral images for Gaussian denoising. In this work, we propose to learn the joint-sparsity promoting dictionary adaptively from the data for impulse denoising problems. Unlike dictionary learning techniques, the sparsifying dictionary is not learnt in an offline training phase. We follow the Blind Compressed Sensing (BCS) framework – dictionary learning and denoising proceeds simultaneously. The optimization problem that arises out of our formulation is solved using the Split Bregman approach. The proposed algorithm, when compared against prior techniques (on real hyper-spectral datasets) shows more than 5dB improvement in PSNR on average.

Index Terms— Hyper-spectral denoising, Impulse Noise, Compressed Sensing, Dictionary Learning.

1. INTRODUCTION

Hyper-spectral images are corrupted by different types of noise – Gaussian noise, impulse noise and streaking artifacts [1]. Strictly speaking, removing streaking artifacts is not a denoising problem – it is an inpainting problem; we will not discuss it in this work. There are a large number of papers on removing Gaussian noise from hyper-spectral images [1-3]. All of them exploit the intra-band spatial redundancy and inter-band spectral correlation for removing Gaussian noise.

Work on impulse denoising for hyper-spectral images is limited. Most of the studies in sparse impulse denoising are limited to single band images [4-8]. The usual technique to remove impulse noise is based on variants of median filtering [4, 5]. More recent techniques exploit the sparsity of the impulse noise and the sparsity of the image in a transform domain to frame an l_1-l_1 minimization problem [6]. The latest techniques [7, 8] learn the sparsifying dictionary for the image in an offline training phase and apply the learnt dictionary for removing impulse noise using l_1-l_1 minimization.

One can apply these techniques [4-8] on each of the spectral bands (of the hyper-sepctral datacube) separately. But such an approach may not yield the best results; since it would not account for the inter-band spectral correlation.

Studies in compressive multi-spectral imaging [9-11] have shown that the images in different spectral bands look very similar to each other and consequently have a common sparse (joint-sparsity) support in certain transform domains. Following these studies, we will exploit the common sparse support of the hyper-spectral images while denoising. But instead of assuming the images to be sparse in a known basis, we will learn the sparsifying basis adaptively from the data following the Blind Compressed Sensing (BCS) framework [12]. In the past the BCS framework had been successful in dynamic MRI reconstruction [13, 14] and EEG signal reconstruction [15].

In the following section we will briefly review some prior studies pertinent to our work. The problem formulation will be described in section 3. The experimental results will be discussed in section 4. Finally the conclusions of the work and future direction of research will be discussed in section 5.

2. LITERATURE REVIEW

Impulse noise is additive in nature; it affects a few pixels but the magnitude of noise is large. The impulse noise model can be expressed as:

$$y = x + n \tag{1}$$

where x is the clean image, n is additive impulse noise (sparse) and y is the corrupted image.

Median filtering techniques [4, 5] have been traditionally used to remove impulse noise; but they only show good results when the number of corrupted pixels is very small. More recent techniques frame denoising as an optimization problem that exploits the sparsity of the image in a transform domain. For example in [6], denoising is framed as:

$$\min_x \|y - x\|_1 + \lambda TV(x) \tag{2}$$

The l_1-norm data fidelity accounts for the sparsity of impulse noise where as the TV penalty assumes the image to be piecewise linear – leading to a sparse representation under finite differencing.

More recent studies [7, 8] believe that the sparsifying dictionary should be learnt adaptively in order to improve denoising results. They employ dictionary learning techniques to learn the sparsifying dictionary (*D*) and employ the learnt dictionary to remove impulse noise:

$$\min_z \|y - Dz\|_1 + \lambda \|z\|_1 \quad (3)$$

Here it is assumed that the image can be represented in a sparse fashion in *D*, i.e. $x=Dz$; once *z* is recovered from (3), obtaining the image is trivial.

To the best of our knowledge there has not been any work on impulse denoising for hyper-spectral images. But, Gaussian noise removal from such images is a well studied topic. In fact there are several studies which study the problem of compressive hyper-spectral imaging in the presence of Gaussian noise [9, 10]. These studies recover the image by solving the following optimization problem:

$$\min_Z \|Y - BW^T Z\|_F^2 + \lambda \|Z\|_{2,1} \quad (4)$$

where *Y* is the noise corrupted image, *B* is the acquisition operator, *W* is the wavelet transform, *Z* are the wavelet coefficients of image stacked as columns of the matrix.

The Frobenius norm arises owing to the Gaussian nature of noise. The $l_{2,1}$-norm is defined as the sum of the l_2-norms of the rows. The $l_{2,1}$-norm penalty assumes that the hyper-spectral images have a common sparse support in the wavelet transform domain.

The $l_{2,1}$-norm with BCS has been used to represent common sparse support in other domains as well. For example in [15] it is used to recover a multi-channel EEG ensemble. BCS has also been used for recovering dynamic MRI sequences from partial K-space samples [13, 14].

It must be noted that the BCS framework can only be utilized when we try to recover sparse multiple measurement vectors (MMV). For single measurement vector recovery problems it is not possible to estimate the sparsifying dictionary during signal recovery. Because estimating the dictionary from a single sample will not be robust.

To summarize the discussion from this section; we learn the following from prior studies:

- Impulse noise is sparse and therefore requires a l_1-norm data fidelity term
- Hyper-spectral images have a common sparse support in transform domain. The $l_{2,1}$-norm penalty is a good choice to exploit the joint-sparsity.
- Since hyper-spectral images have typically many bands, it is possible to learn the sparsifying dictionary on the fly (during signal estimation).

3. PROPOSED DENOISING APPROACH

We assume each band of the hyper-spectral datacube to be corrupted by sparse impulse noise. The noise model is expressed as:

$$y_c = x_c + n \quad (5)$$

where *c* denotes the spectral band.

In a compact form, this can be expressed as:

$$Y = X + N \quad (6)$$

where $Y = [y_1 | ... | y_C]$ and $X = [x_1 | ... | x_C]$ assuming *C* bands in all.

In order to remove noise, we need to exploit the spatio-spectral correlation of the hyper-spectral datacube (*X*). If we assume that the image is sparse in wavelet domain (*W*), the recovery can be posed as follows:

$$\min_Z \|Y - W^T Z\|_1 + \lambda \|Z\|_{2,1} \quad (7)$$

The l_1-norm accounts for sparse noise and the $l_{2,1}$-norm accounts for spatio-spectral correlations. Here we abuse the notations slightly. The l_1-norm is supposed to be defined over a vector, but we do not use the 'vec' notation to keep the expressions uncluttered.

Following the success of the BCS framework [13-15], we propose to learn the sparsifying dictionary from the data. Wavelet transforms yield a sparse representation for different kinds of images; but if we are interested in a particular class of images, learning the sparsifying basis is likely to yield a sparser representation. We assume that the images are row-sparse in a learnt basis *D*, so that

$$X = DZ \quad (8)$$

Therefore, the recovery can be framed as:

$$\min_{Z,D} \|Y - DZ\|_1 + \lambda_1 \|Z\|_{2,1} + \lambda_2 \|D\|_F^2 \quad (9)$$

Here both the dictionary *D* and the row-sparse coefficient matrix *Z*, need to be learnt. The l_2-norm penalty on the dictionary is for regularization.

3.1. Optimization Algorithm

To the best of our knowledge there are no algorithms to solve (9). In this work we follow the Split Bregman technique [16] to derive an algorithm to solve the said problem. We introduce three proxy variables, $P=Y-DZ$, $Q=Z$ and $R=D$. We add terms relaxing the equality constraints of each quantity and its proxy, and in order to enforce equality at convergence, we introduce Bregman variables B_1, B_2, B_3 and B_4. The new objective function turns out to be:

$$\begin{aligned}&\min_{Z,D,P,Q,R} \|P\|_1 + \lambda_1 \|Q\|_{2,1} + \lambda_2 \|R\|_F^2 \\ &+\mu \|P - (Y - DZ) - B_1\|_F^2 + \mu_1 \|Q - Z - B_2\|_F^2 \\ &+\mu_2 \|R - D - B_3\|_F^2\end{aligned} \quad (10)$$

The variable splitting allows us to express (10) as an alternating minimization of the following (easier) sub-problems:

$$\text{P1:}\min_{Z} \mu \|P-(Y-DZ)-B_1\|_F^2 + \mu_1 \|Q-Z-B_2\|_F^2$$

$$\text{P2:}\min_{D} \mu \|P-(Y-DZ)-B_1\|_F^2 + \mu_2 \|R-D-B_3\|_F^2$$

$$\text{P3:}\min_{P} \|P\|_1 + \mu \|P-(Y-DZ)-B_1\|_F^2$$

$$\text{P4:}\min_{Q} \lambda_1 \|Q\|_{2,1} + \mu_1 \|Q-Z-B_2\|_F^2$$

$$\text{P5:}\min_{R} \lambda_2 \|R\|_F^2 + \mu_2 \|R-D-B_3\|_F^2$$

Apart from P3 and P4, the rest are least squares minimization problems which can be solved efficiently using Conjugate Gradient techniques. The sub-problem P3 is an l_1-norm regularized least squares problems which can be solved via iterative soft thresholding [17]. The row-sparse $l_{2,1}$-norm minimization problem can be solved efficiently using the modified iterative Thresholding algorithm [18].

The final step of the Split Bregman technique is to update the relaxation variables:

$$B_1 \leftarrow P - Y + DZ - B_1$$
$$B_2 \leftarrow Q - Z - B_2$$
$$B_3 \leftarrow R - D - B_3$$

There are two stopping criterions for the Split Bregman algorithm. Iterations continue till the objective function converges (to a local minima); The other stopping criterion is a limit on the maximum number of iterations. We have kept it to be 200.

4. EXPERIMENTAL EVALUATION

Two hyperspectral datacubes were used for performing experiments. One is of Reno city, NV, USA available from [19]. This image is from High Resolution Imager (HRI) sensor having 2m spatial resolution and 5 nm band spacing covering spectral range of 395-2450 nm. The second dataset of of Washington DC mall available from [20]. This image is of Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor having 1m spatial resolution and 10-nm band spacing covering spectral range of 400-2500 nm. We used patches of size 64 × 64 × 64 from both the images for experiments. A portion of the pixels were corrupted by salt-and-pepper noise.

As we mentioned before, we are not aware of prior studies in hyper-spectral image denoising. Prior studies could only remove impulse noise from single band images. As a baseline, we have compared our results with two such techniques – adaptive median filtering [4] and l_1-TV [6].

Our proposed approach is called $l_{2,1}$-BCS. It required specifying several parameters. The parameters were tuned on a validation set. For tuning the parameters we employed a sub-optimal yet effective strategy based on the L-curve method [21]. For the first parameter λ_1 we set the other parameter (λ_2) to zero and use the L-curve method to find it. To tune the second parameter λ_2, we fix λ_1 to the obtained value and again use the L-curve method to determine λ_2. Such a technique, although sub-optimal have showed good results in practice before [10, 14, 15]. We did not fine tune the parameters and varied them on the log scale (100, 10, 1, 0.1 etc.). We obtained the values λ_1=10 and λ_2=10^{-1}. We found these values to be robust to changing noise levels.

Our algorithm requires specifying some hyper-parameters. The Bregman relaxation variables were initialized to unity. The internal variables (μ's) were fixed by trial and error to yield the best results on the validation dataset. As before, we only varied the values on a log scale. The following values were used: μ=10, μ_1=10, μ_2=10^3.

In order to check the effect of learning the sparsifying dictionary as opposed to a fixed dictionary we made use of (7) for impulse denoising. We assumed that the images are sparse in wavelet transform. We used our algorithm to solve (7) by putting the parameters and hyper-parameters corresponding to the dictionary learning penalty to zeroes. We call this formulation (7) as the l_1-$l_{2,1}$ technique.

Table 1. PSNR after image denoising

Dataset	Noise	AMF	l_1-TV	$l_{2,1}$-BCS	l_1-$l_{2,1}$
WDC mall	10%	31.62	41.09	**44.62**	38.17
	30%	23.17	39.01	**42.78**	39.78
	50%	16.78	31.38	**35.03**	31.94
Reno	10%	35.98	35.27	**47.74**	35.61
	30%	25.16	32.57	**44.79**	33.87
	50%	16.13	29.27	**38.16**	31.31

* x% noise means x% of the total pixels are corrupted.

It is not surprising that our proposed method always yields the best results. The PSNR from our proposed method is significantly better than the rest. This is because we exploit the spatio-spectral correlation in the hyper-spectral datacube which previous methods [4, 6] do not. It must be remembered that we get these results from coarse tuning of the parameters (λ_1 and λ_2); we do not vary them with changing noise levels. It may be possible to yield even better results by fine tuning them.

Comparing our method with recovery using fixed dictionary (l_1-$l_{2,1}$) shows that, learning the dictionary indeed improves the denoising performance. For the WDC mall image, the improvement is 5 dB (on average) where as for the Reno image the improvement is more significant (almost 10 dB on average).

For small amount of noise (10%) the l_1-TV seems to yield almost as good a result as l_1-$l_{2,1}$ technique. This is because finite differencing (used in the TV norm) has a sparser representation compared to wavelets. The improved sparsity offsets the advantage of spatio-spectral correlation and the l_1-TV yields better denoising. However, when the noise level increases, exploiting the spatio-spectral correlation improves the results and we see slightly better performance by exploiting spatio-spectral correlation in the $l_{2,1}$-l_1 method.

The adaptive median filter can only yield decent denoising results when the number of corrupted pixels is small. As more pixels are corrupted, median filtering fails, since the assumption that the median of the neighbourhood is the true value becomes less plausible.

For visual quality assessment a randomly chosen band from the Reno image is shown in Fig. 1. The denoising results are shown for 30% noisy pixels. One can see that the AMF yields a noisy output. The l_1-TV method yields an overtly smooth image. With fixed basis (l_1-$l_{2,1}$) the image is sharp but consists of denoising artifacts. The best result is obtained with our proposed method. The recovered image is sharp and bereft of any artifact. The denoised image looks as good as the original visually.

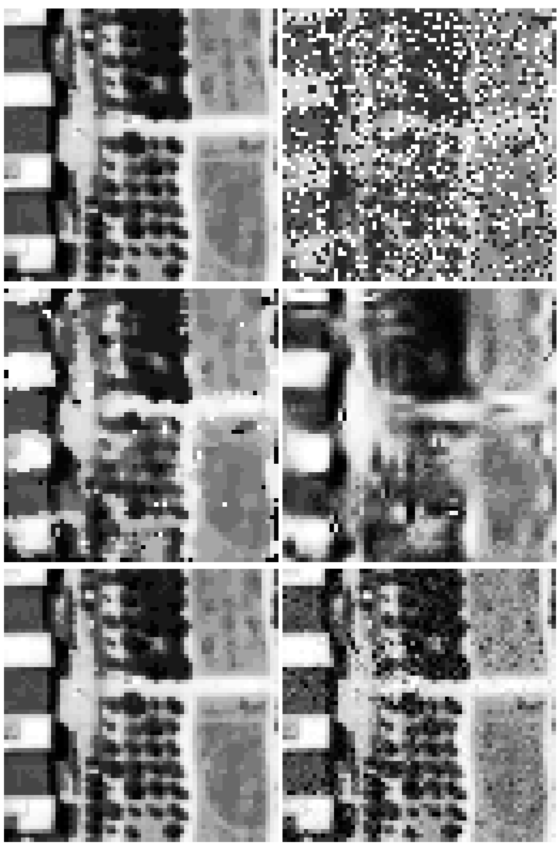

Fig. 1. 1st Row 1st Column – Original; 1st Row 2nd Column – Noisy (30% corrupted pixels); 2nd Row 1st Column – Adaptive Mean Filter; 2nd Row 2nd Column – l_1-TV; 3rd Row 1st Column – Proposed $l_{2,1}$-BCS; 3rd Row 2nd Column – l_1-$l_{2,1}$.

Our proposed formulation is non-convex since the variables *D* and *Z* appear in a product form. There is no guarantee that our algorithm will reach a global minima. But it does reach a local minima as can be verified from the convergence plot in Fig. 2.

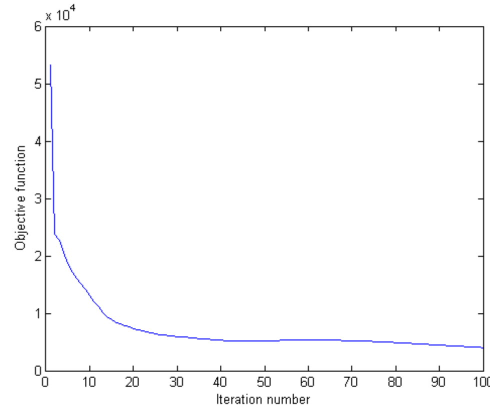

Fig. 2. Convergence of objective function

5. CONCLUSION

In this work we addressed the problem of hyper-spectral image denoising when the images are corrupted by impulse noise. To the best of our knowledge this is the first work that accounts for spatio-spectral correlation of the hyper-spectral datacube for removing impulse noise. Motivated by the Blind Compressed Sensing (BCS) framework, we learnt a dictionary to sparsify the images (accounting for intra-band spatial redundancy) and imposed a row-sparsity penalty to account for inter-band spectral correlations. Strictly speaking, the denoising problem is not a Compressed Sensing topic, since the full data is available.

Our formulation led to a $l_{2,1}$-norm BCS problem with l_1-norm data fidelity. We solved the optimization problem using the Split Bregman technique. We compared our method with previous techniques and showed that our method yields about 5dB improvement in PSNR on an average.

We studied the problem of hyper-spectral impulse denoising because it had not been addressed before. In practice hyper-spectral images are always corrupted by mixed Gaussian and impulse noise. In the future, we would like to extend our work so as to account for mixed noise.

We believe in reproducible research. The code for reproducing the results is currently available upon request, but will be available soon in Matlab central.

REFERENCES

[1] H. Zhang, W. He, L. Zhang, H. Shen and Q. Yuan, "Hyperspectral Image Restoration Using Low-Rank Matrix Recovery," IEEE Trans. Geoscience and Remote Sensing, Vol.52 (8), pp. 4729,4743, 2014.

[2] Y. Zhao and J. Yang, "Hyperspectral Image Denoising via Sparse Representation and Low-Rank Constraint," IEEE

Trans. Geoscience and Remote Sensing, Vol.53 (1), pp.296-308, 2015.

[3] Q. Yuan; L. Zhang; H. Shen, "Hyperspectral Image Denoising Employing a Spectral–Spatial Adaptive Total Variation Model," IEEE Trans. Geoscience and Remote Sensing, Vol.50 (10), pp.3660-3677, 2012.

[4] H. Hwang, R.A. Haddad, "Adaptive median filters: new algorithms and results", IEEE Trans. Image Process., Vol. 4, pp. 499–502, 1995.

[5] R. H. Chan, C. W. Ho and M. Nikolova, "Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization", IEEE Trans. Image Process. Vol. 14, 1479–1485, 2005.

[6] B. Wohlberg and P. Rodríguez, "An l1-TV algorithm for deconvolution with salt and pepper noise", IEEE ICASSP, pp. 1257-1260, 2009

[7] S. Wang, Q. Liu, Y. Xia, P. Dong, J. Luo, Q. Huang, D. D. Feng, "Dictionary learning based impulse noise removal via L1–L1 minimization", Signal Processing, Vol. 93 (9), pp. 2696-2708, 2013.

[8] X. Chen, Z. Du, J. Li, X. Li and H. Zhang, "Compressed sensing based on dictionary learning for extracting impulse components", Signal Processing, Vol. 96, pp. 94-109, 2014.

[9] M. Golbabaee, P. Vandergheynst, "Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery", IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 2741-2744, 2012.

[10] A. Gogna, A. Shukla, H. Agarwal and A. Majumdar, "Split Bregman Algorithms for Sparse / Joint-sparse and Low-rank Signal Recovery: Application in Compressive Hyperspectral Imaging", IEEE ICIP 2014 (accepted).

[11] A. Majumdar and R. K. Ward, "Compressed Sensing of Color Images", Signal Processing, Vol. 90 (12), 3122-3127, 2010.

[12] S. Gleichman, Y. C. Eldar, "Blind Compressed Sensing", IEEE Trans. on Information Theory, Vol. 57, pp. 6958-6975, 2011.

[13] S. G. Lingala, M. Jacob, "Blind Compressed Sensing Dynamic MRI", IEEE Trans. Med. Imag., Vol. 32, pp. 1132-1145, 2013.

[14] A. Majumdar, "Improving Synthesis and Analysis Prior Blind Compressed Sensing with Low-rank Constraints for Dynamic MRI Reconstruction", Magnetic Resonance Imaging, (accepted)

[15] A. Shukla and A. Majumdar, "Row-sparse Blind Compressed Sensing for Reconstructing Multi-channel EEG signals", Biomedical Signal Processing and Control, accepted.

[16] R. Chartrand, "Nonconvex splitting for regularized low-rank + sparse decomposition", IEEE Trans. Signal Process., Vol. 60, pp. 5810-5819, 2012.

[17] I. Daubechies and M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint", Communications on Pure and Applied Mathematics Vol. 4 (57), 1413–1457, 2004.

[18] A. Majumdar and R. K. Ward, "Synthesis and Analysis Prior Algorithms for Joint-Sparse Recovery", IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 3421-3424, 2012.

[19] Reno:http://www.spectir.com/services/airborne-hyperspectral/. Last accessed: 3rd September 2014.

[20] WDC:https://engineering.purdue.edu/%7ebiehl/MultiSpec/hyperspectral.html. Last accessed: 3rd September 2014.

[21] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems", SIAM Journal on Scientific Computing 14 (6), 1487-1503, 1993