



LR-HyClassify: A Low Rank Based Framework for Classification of Degraded Hyperspectral Images

Sourish Sarkar^(✉) and Rajiv R. Sahay

Department of Electrical Engineering, Indian Institute of Technology,
Kharagpur, India

sourish.sarkar@iitkgp.ac.in, rajiv@ee.iitkgp.ernet.in

Abstract. We propose a novel framework for the classification of hyperspectral data corrupted by severe degradation. We propose an optimization framework for extracting discriminative features from noisy hyperspectral data which are then passed onto a simple classifier which exploits both the spatial and spectral correlations in the data. Instead of directly extracting the features from the noisy input data, we learn a basis matrix from the underlying clean data using a combination of non-negative matrix factorization and nuclear-norm minimization. The input degraded data are then projected onto the basis vectors to obtain features. We use structural incoherence to maximize the discriminative ability of the features, which, to the best of our knowledge, is being used for the first time in hyperspectral image processing literature. We show the deterioration in the performance of classification algorithms with the progressive addition of noise while demonstrating the robustness of our algorithm. We evaluate the performance of our algorithm on three well known datasets, namely, Botswana, Salinas and Indian Pines, obtaining state-of-the-art results that validate the efficacy of our algorithm.

Keywords: Hyperspectral image classification · Structural incoherence · Non-negative matrix factorization · Nuclear-norm minimization

1 Introduction

Hyperspectral image (HSI) classification has been an area of active research since over 20 years where the task is to assign each pixel in a HSI to its respective class. HSI classification is a challenging task due to the high data dimensionality, availability of very few labelled samples and the presence of noise. The presence of noise damages both the spatial and spectral information available in the data, thereby degrading the performance of classification algorithms drastically. This motivates us to devise a novel technique which can achieve impressive classification accuracies even in the presence of significant amount of noise. Algorithms

used in the earlier days of research in HSI classification considered only the spectral signatures which did not provide satisfactory classification results. However, recent works [4, 8, 9] acknowledge the contribution of the spatial information contained in HSI in addition to the spectral information. The importance of spatial information stems from the simple hypothesis that nearby pixels in a HSI belong to the same class with high probability. In [12], the authors address the problem of robust face recognition in the presence of data corrupted due to occlusion and disguise. They incorporate a constraint of structural incoherence in order to make the bases learnt for different classes independent. This leads to the achievement of a higher discriminating ability resulting in an impressive classification performance. Inspired by their approach, in this work, we aim to exploit both the spectral and spatial information and classify a severely degraded HSI by learning discriminative features. Instead of learning these features from the corrupted data, we learn them from the underlying clean data which leads to state-of-the-art classification results.

2 Prior Work

Several methods which perform a pixelwise classification of HSIs have been proposed [10, 14, 23]. Sparse representation based classifiers have been used for HSI classification in [5, 6]. A combination of low rank and sparsity for HSI classification can be found in [8, 19, 21]. Collaborative representation based classifiers have been used in [9, 11]. In [1], extended morphological profiles were used for HSI classification. The authors in [22] used conditional random fields for HSI classification. The random forest framework was investigated for HSI classification in [7]. Owing to the excellent performance in classification of images and videos, deep learning based methods have been extensively investigated for HSI classification. Stacked auto encoders (SAEs) [16], convolutional neural networks (CNNs) [13], deep belief networks (DBNs) [4] and deep recurrent neural networks (RNNs) [15] have been used for HSI classification. Recently, generative adversarial networks (GANs) have been introduced for the purpose of HSI classification [24]. However, these models are data hungry and are prone to overfitting due to the scarcity of available labelled samples. Hence, most of the recent research with deep models deals with finding new techniques to mitigate this problem.

3 Proposed Methodology

In this work, we present a novel framework wherein we combine non-negative matrix factorization (NMF), nuclear norm minimization (NNM) and structural incoherence to learn discriminative features from a degraded HSI. Given a HSI of size $m \times n \times b$, the training samples are the b -dimensional spectral vectors. Let $\mathbf{Y} \in \mathcal{R}^{b \times N}$ denote a matrix formed by stacking together all the N available training samples. Hence, for a HSI with C classes, $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_C]$ where $\mathbf{Y}_i = [\mathbf{y}_1^i, \mathbf{y}_2^i, \dots, \mathbf{y}_{n_i}^i] \in \mathcal{R}^{b \times n_i}$ ($i = 1, 2, \dots, C$) denotes the training samples

belonging to class i and n_i is the number of available training samples from class i so that $N = Cn_i$. We model the data as:

$$\mathbf{Y} = \mathbf{L} + \mathbf{S} + \mathbf{N} \quad (1)$$

where $\mathbf{L} = [\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_C] \in \mathcal{R}^{b \times N}$ is a low rank matrix, $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_C] \in \mathcal{R}^{b \times N}$ is the matrix of sparse noise and $\mathbf{N} = [\mathbf{N}_1, \mathbf{N}_2, \dots, \mathbf{N}_C] \in \mathcal{R}^{b \times N}$ denotes additive Gaussian noise. We aim to combine non-negative matrix factorization and nuclear norm minimization to simultaneously denoise the available data and learn a basis matrix for classification. We wish to solve the following optimization problem:

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}, \mathbf{U} \geq \mathbf{0}, \mathbf{V} \geq \mathbf{0}} \quad & \alpha \|\mathbf{L}\|_* + \beta \|\mathbf{S}\|_1 + \gamma \sum_{j \neq i} \|\mathbf{L}_j^T \mathbf{L}_i\|_F^2 + \delta \|\mathbf{L} - \mathbf{U}\mathbf{V}\|_F^2 \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{L} - \mathbf{S}\|_F^2 \leq \epsilon \end{aligned} \quad (2)$$

where $\|\cdot\|_F$ and $\|\cdot\|_*$, respectively, denote the Frobenius norm and the nuclear norm of a matrix, $\alpha, \beta, \gamma, \delta$ are positive parameters and ϵ is a small constant whose value is fixed to 0.001. It is noteworthy that the basis matrix \mathbf{U} is being learnt from the underlying clean data \mathbf{L} and not directly from the degraded training data \mathbf{Y} . The third term in Eq. (2) corresponds to the structural incoherence [17] which measures the similarity between the derived low-rank matrices of different classes. Hence, minimizing this term separately for every class tends to make these derived matrices incoherent and enhances their discriminating ability. Hence, for the i^{th} class, we wish to solve:

$$\begin{aligned} \min_{\mathbf{L}_i, \mathbf{S}_i, \mathbf{U} \geq \mathbf{0}, \mathbf{V} \geq \mathbf{0}} \quad & \alpha \|\mathbf{L}_i\|_* + \beta \|\mathbf{S}_i\|_1 + \gamma \sum_{j \neq i} \|\mathbf{L}_j^T \mathbf{L}_i\|_F^2 + \delta \|\mathbf{L}_i - (\mathbf{U}\mathbf{V})_i\|_F^2 \\ \text{s.t.} \quad & \|\mathbf{Y}_i - \mathbf{L}_i - \mathbf{S}_i\|_F^2 \leq \epsilon \end{aligned} \quad (3)$$

We use the method of augmented Lagrangian multipliers (ALM) [18] to solve the above problem. To this end, we first introduce an auxiliary variable \mathbf{Z}_i as follows:

$$\begin{aligned} \min_{\mathbf{L}_i, \mathbf{S}_i, \mathbf{U} \geq \mathbf{0}, \mathbf{V} \geq \mathbf{0}, \mathbf{Z}_i} \quad & \alpha \|\mathbf{L}_i\|_* + \beta \|\mathbf{S}_i\|_1 + \gamma \sum_{j \neq i} \|\mathbf{L}_j^T \mathbf{Z}_i\|_F^2 + \delta \|\mathbf{Z}_i - (\mathbf{U}\mathbf{V})_i\|_F^2 \\ \text{s.t.} \quad & \|\mathbf{Y}_i - \mathbf{L}_i - \mathbf{S}_i\|_F^2 \leq \epsilon, \mathbf{Z}_i = \mathbf{L}_i \end{aligned} \quad (4)$$

The augmented Lagrangian is as follows:

$$\begin{aligned} \mathcal{L}(\mathbf{L}_i, \mathbf{S}_i, \mathbf{U} \geq \mathbf{0}, \mathbf{V} \geq \mathbf{0}, \mathbf{Z}_i \geq \mathbf{0}, \mathbf{M}_1, \mathbf{M}_2; \mu) = & \alpha \|\mathbf{L}_i\|_* + \beta \|\mathbf{S}_i\|_1 \\ & + \gamma \sum_{j \neq i} \|\mathbf{L}_j^T \mathbf{Z}_i\|_F^2 + \delta \|\mathbf{Z}_i - (\mathbf{U}\mathbf{V})_i\|_F^2 + \langle \mathbf{M}_1, \mathbf{Y}_i - \mathbf{L}_i - \mathbf{S}_i \rangle \\ & + \langle \mathbf{M}_2, \mathbf{Z}_i - \mathbf{L}_i \rangle + \frac{\mu}{2} \left[\|\mathbf{Y}_i - \mathbf{L}_i - \mathbf{S}_i + \frac{\mathbf{M}_1}{\mu}\|_F^2 + \|\mathbf{Z}_i - \mathbf{L}_i + \frac{\mathbf{M}_2}{\mu}\|_F^2 \right] \end{aligned} \quad (5)$$

where $\mathbf{M}_1, \mathbf{M}_2$ are Lagrange multipliers, μ is a positive parameter and $\langle \cdot, \cdot \rangle$ denotes the inner product. We now provide the updation rules for each of the variables.

- Computing \mathbf{L}_i with other variables fixed:
 With some algebraic manipulations, the \mathbf{L}_i -subproblem can be obtained from Eq. (5) as follows:

$$\min_{\mathbf{L}_i} \frac{\alpha}{2\mu} \|\mathbf{L}_i\|_* + \frac{1}{2} \|\mathbf{L}_i - (\mathbf{Y}_i - \mathbf{S}_i + \mathbf{Z}_i + \frac{\mathbf{M}_1}{\mu} + \frac{\mathbf{M}_2}{\mu})\|_F^2 \quad (6)$$

Eq. (6) can be solved using singular value thresholding [3].

- Computing \mathbf{S}_i fixing others:
 The \mathbf{S}_i -subproblem can be obtained from Eq. (5) as follows:

$$\min_{\mathbf{S}_i} \frac{\beta}{\mu} \|\mathbf{S}_i\|_1 + \frac{1}{2} \|\mathbf{S}_i - (\mathbf{Y}_i - \mathbf{L}_i + \frac{\mathbf{M}_1}{\mu})\|_F^2 \quad (7)$$

Eq. (7) can be solved using the soft shrinkage operation.

- Computation of \mathbf{Z}_i :
 In order to achieve factorization of \mathbf{Z}_i , the non-negativity constraint on \mathbf{Z}_i must be satisfied. To this end, we first introduce an indicator function as follows:

$$l_+(\mathbf{Z}_i) = \begin{cases} 0, & Z_{i_{m,n}} \geq 0 \quad \forall m, n \\ \infty, & \text{otherwise} \end{cases} \quad (8)$$

where, $Z_{i_{m,n}}$ denotes the $(m, n)^{th}$ element of \mathbf{Z}_i . The \mathbf{Z}_i sub-problem is then obtained from Eq. (5) as follows:

$$\min_{\mathbf{Z}_i} \gamma \sum_{j \neq i} \|\mathbf{L}_j^T \mathbf{Z}_i\|_F^2 + \delta \|\mathbf{Z}_i - (\mathbf{UV})_i\|_F^2 + \frac{\mu}{2} \|\mathbf{Z}_i - \mathbf{L}_i + \frac{\mathbf{M}_2}{\mu}\|_F^2 + l_+(\mathbf{Z}_i) \quad (9)$$

To solve this, we use the alternating direction method of multipliers (ADMM) [2]. To do so, we first introduce an auxiliary variable in order to make the objective function separable:

$$\begin{aligned} \min_{\mathbf{Z}_i} \gamma \sum_{j \neq i} \|\mathbf{L}_j^T \mathbf{Z}_i\|_F^2 + \delta \|\mathbf{Z}_i - (\mathbf{UV})_i\|_F^2 + \frac{\mu}{2} \|\mathbf{Z}_i - \mathbf{L}_i + \frac{\mathbf{M}_2}{\mu}\|_F^2 + l_+(\mathbf{C}_i) \\ \text{s.t. } \mathbf{C}_i = \mathbf{Z}_i \end{aligned} \quad (10)$$

The above equation is solved iteratively by updating one of the variables and keeping the other fixed until convergence. The update equations are:

$$\mathbf{Z}_i = [(\delta + \frac{\mu}{2} + \frac{\rho}{2})\mathbf{I} + \gamma \sum_{j \neq i} \mathbf{L}_j \mathbf{L}_j^T]^{-1} [\delta(\mathbf{UV})_i + \frac{\mu \mathbf{L}_i}{2} - \mathbf{M}_2 - \mathbf{T} + \rho \mathbf{C}_i] \quad (11)$$

$$\mathbf{C}_i \leftarrow \max(\mathbf{Z}_i + \frac{\mathbf{T}}{\rho}, 0) \quad (12)$$

$$\mathbf{T} \leftarrow \mathbf{T} + \rho(\mathbf{Z}_i - \mathbf{C}_i) \quad (13)$$

$$\rho \leftarrow \min(\kappa \rho, \rho_{max}) \quad (14)$$

where, \mathbf{I} denotes the identity matrix of appropriate size, \mathbf{T} is the Lagrange multiplier and $\rho > 0$, $\kappa > 0$, ρ_{max} are parameters.

– Computation of \mathbf{U} and \mathbf{V} :

$$\min_{\mathbf{U} \geq \mathbf{0}, \mathbf{V} \geq \mathbf{0}} \|\mathbf{Z}_i - (\mathbf{UV})_i\|_F^2 \quad (15)$$

The above optimization problem can be solved directly using any of the existing NMF solvers.

– The final step is to update the multipliers and μ :

$$\mathbf{M}_1 \leftarrow \mathbf{M}_1 + \mu(\mathbf{Y}_i - \mathbf{L}_i - \mathbf{S}_i) \quad (16)$$

$$\mathbf{M}_2 \leftarrow \mathbf{M}_2 + \mu(\mathbf{Z}_i - \mathbf{L}_i) \quad (17)$$

$$\mu \leftarrow \min(\psi\mu, \mu_{max}) \quad (18)$$

Once we obtain the basis matrix \mathbf{U} learnt from the underlying clean data \mathbf{L} , we project all the degraded training and testing data onto \mathbf{U} to obtain discriminative features which are subsequently used for classification. For any spectral vector \mathbf{y} , we obtain its feature vector as follows:

$$\mathbf{f} = \mathbf{U}^\dagger \mathbf{y} \quad (19)$$

where \mathbf{U}^\dagger denotes the pseudoinverse of \mathbf{U} . We propose to use a very simple classifier to classify these obtained features with high accuracy. For a test feature \mathbf{f}_{test} and the training features $\mathbf{f}_i, i = 1, 2, \dots, N$, let the spatial positions of the test and training features (corresponding to the test and training spectral vectors, respectively) be denoted by $\mathbf{p}_{test} = [x, y]^T$ and $\mathbf{p}_i = [x_i, y_i]^T$, respectively. Let $d_1 = dist(\mathbf{p}_{test}, \mathbf{p}_i)$ and $d_2 = dist(\mathbf{f}_{test}, \mathbf{f}_i)$, where $dist(\mathbf{a}, \mathbf{b})$ denotes the squared Euclidean distance between the vectors \mathbf{a} and \mathbf{b} . Then we propose to define

$$d_i \triangleq \alpha_1 d_1 + \alpha_2 d_2 \quad (20)$$

as the dissimilarity between the test feature \mathbf{f}_{test} and the training feature \mathbf{f}_i . The test sample is then classified to the class of the training sample with which this dissimilarity metric is minimum. d_1 simply takes into account the spatial correlation acknowledging the fact that pixels close to each other belong to the same class with a high probability while d_2 corresponds to the Euclidean distance between the test and training features. Note that setting α_1 to zero leads to the nearest neighbour classifier.

4 Experimental Results

In order to assess the performance of our algorithm, we use three HSI datasets. We synthetically corrupt these datasets by adding Gaussian noise of 0.05 standard deviation. Note that the spectral vectors are normalized to [0–1]. We also add impulse noise and stripes to bands 61 to 70 in all the three datasets. We randomly choose 10 per cent of labelled samples from each class for training and use the rest for testing. We compare our algorithm with SVM [14], SRC [20] and CRNN [11]. The classification performance is measured by the overall accuracy

Algorithm 1. Proposed algorithm for obtaining discriminative features

Input: Training dataset: $\{\mathbf{Y}_i\}$, $i=1,2, \dots, C$; parameters $\alpha, \beta, \gamma, \delta$ in Eq.(3) and number of basis vectors r **Output:** Learnt basis matrix \mathbf{U}

- 1: Initialise $\kappa > 0, \rho > 0, \mu_{max} = 10^5, \rho_{max} = 10^5, \epsilon = 0.001, \mathbf{U}$ =random matrix with positive elements of appropriate size.
 - 2: **for** $i = 1$ to C **do**
 - 3: Initialise $\mu > 0, \psi > 0, \mathbf{M}_1 = \mathbf{0}, \mathbf{M}_2 = \mathbf{0}, \mathbf{Z}_i = \mathbf{0}, \mathbf{S}_i > \mathbf{0}, \mathbf{V}$ =random matrix with positive elements of appropriate size.
 - 4: **while** $\|\mathbf{Y}_i - \mathbf{L}_i - \mathbf{S}_i\|_F^2 / \|\mathbf{Y}_i\|_F^2 < \epsilon$ **do**
 - 5: Update \mathbf{L}_i by Eq. (6)
 - 6: Update \mathbf{S}_i by Eq. (7)
 - 7: Update \mathbf{Z}_i by Eq. (11)
 - 8: Update \mathbf{C}_i by Eq. (12)
 - 9: Update \mathbf{T}_i by Eq. (13)
 - 10: Update ρ_i by Eq. (14)
 - 11: Update \mathbf{U} and \mathbf{V} by Eq. (15)
 - 12: Update multipliers and penalty parameter μ using Eqs. (16,17 and 18)
 - 13: **end while**
 - 14: **end for**
-

(OA) which is defined as the ratio between the number of correctly predicted pixels to the total number of test pixels. The robustness of our algorithm to noise is analysed by gradually increasing the amount of noise added and monitoring its effect on the class accuracies as compared to the other algorithms. The parameters $\alpha, \beta, \gamma, \delta, \alpha_1$ and α_2 are tuned to obtain the best results.

4.1 Datasets

The datasets used are Indian Pines, Botswana and Salinas. The Indian Pines dataset has a size of $145 \times 145 \times 200$ and has 16 classes. The Botswana dataset has a size of $1476 \times 256 \times 145$ and has 14 classes. The third dataset is the Salinas dataset which has a size of $512 \times 217 \times 204$ and 16 classes. The Botswana dataset has only 326 labelled training samples and hence is particularly more challenging among these datasets due to the scarcity of training samples.

4.2 Classification Performance

The performance of the classifiers on the three synthetically corrupted datasets is demonstrated in Tables 1, 2 and 3. We find that our algorithm achieves accuracies of 97.39%, 99.85% and 98.28% on the Botswana, Salinas and Indian Pines datasets respectively, outperforming the other methods. In the Salinas dataset, CRNN achieves the next best accuracy of 90.42% while SVM achieves the next best accuracies of 30.15% and 52.03% in the Botswana and the Indian Pines datasets, respectively. Note that our algorithm outperforms other state-of-the-art methods in these two datasets by a very large margin. Figure 1 depicts

Table 1. Classification accuracy (%) for the Botswana dataset

Class	SVM	SRC	CRNN	Proposed approach
Water	88.89	34.57	46.50	100
Hippo grass	3.30	4.40	5.49	100
Floodplain grasses 1	23.01	10.62	12.39	86.28
Floodplain grasses 2	10.36	2.60	7.25	100
Reeds 1	22.31	15.70	21.49	100
Riparian	26.86	12.40	6.20	98.76
Fireshcar 2	30.90	17.17	17.17	81.55
Island interior	37.70	12.02	24.04	99.45
Acacia woodlands	16.96	7.77	6.36	100
Acacia shrublands	19.73	9.42	8.07	100
Acacia grasslands	35.40	27.73	35.40	100
Short mopane	22.70	8.59	11.66	100
Mixed mopane	17.01	9.96	5.40	100
Exposed soils	74.12	24.70	72.94	100
OA (%)	30.15	14.54	18.41	97.39

Table 2. Classification accuracy (%) for the Salinas dataset

Class	SVM	SRC	CRNN	Proposed approach
Broccoli green weeds 1	97.34	98.34	99.56	100
Broccoli green weeds 2	97.67	95.68	99.85	100
Fallow	75.42	0	99.10	100
Fallow rough plow	98.72	98.88	98.96	98.17
Fallow smooth	96.47	99.34	97.80	99.05
Stubble	99.04	99.80	99.75	99.78
Celery	99.25	99.35	99.60	99.97
Grapes untrained	95.63	81.47	79.54	100
Soil vinyard develop	98.76	98.73	99.25	100
Corn senesced green weeds	75.38	79.02	95.73	99.97
Lettuce romaine 4wk	0	0	98.86	99.79
Lettuce romaine 5wk	95.50	93.54	99.60	99.71
Lettuce romaine 6wk	97.81	97.94	98.06	99.76
Lettuce romaine 7wk	90.13	90.34	93.87	99.27
Vinyard untrained	9.92	46.72	66.43	100
Vinyard vertical trellis	79.15	75.52	98.95	100
OA (%)	80.82	80.02	90.42	99.85

Table 3. Classification accuracy (%) for the Indian Pines dataset

Class	SVM	SRC	CRNN	Proposed approach
Alfalfa	0	0	4.88	100
Corn-notill	34.44	36.96	33.85	96.19
Corn-mintill	1.07	2.28	4.82	97.86
Corn	0.94	0.94	0.47	96.71
Grass-pasture	10.80	4.14	9.19	98.16
Grass-trees	88.28	86.15	75.19	99.70
Grass-pasture-mowed	0	0	0	96.00
Hay-windrowed	96.51	97.67	73.25	100
Oats	0	0	0	100
Soybean-notill	7.66	12.11	11.31	95.66
Soybean-mintill	89.09	85.56	75.37	98.73
Soybean-clean 5wk	0.19	4.31	5.06	98.50
Wheat	33.70	27.17	28.26	99.46
Woods	97.89	99.21	94.20	100
Buildings-Grass-Trees-Drives	6.92	1.73	2.59	99.71
Stone-Steel-Towers	80.95	85.71	69.05	100
OA (%)	52.03	51.77	46.68	98.28

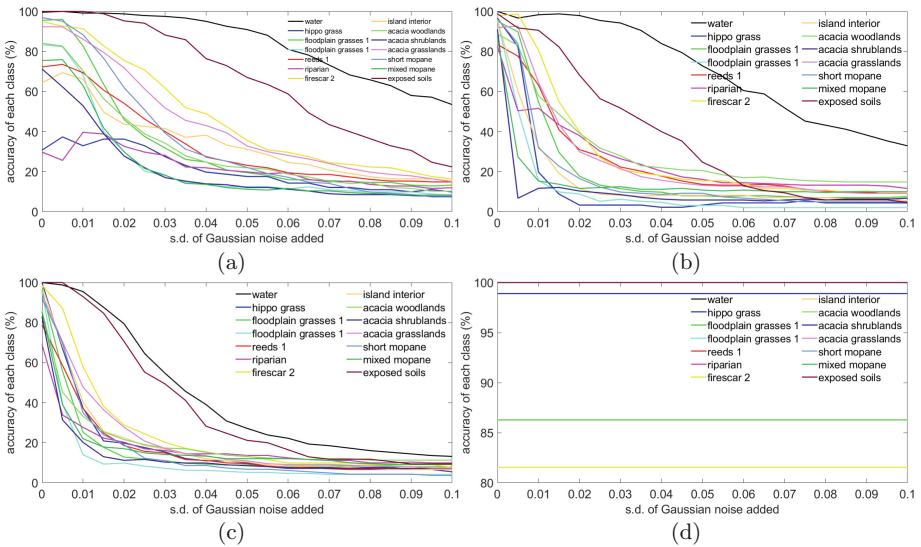


Fig. 1. Effect of progressive addition of noise on individual class accuracies on the Botswana dataset: (a) SVM, (b) SRC, (c) CRNN and (d) proposed approach

the sensitivity of the classifiers to noise. We progressively increase the standard deviation of Gaussian noise using increments of 0.005 upto 0.1 and examine the effect on class specific accuracies obtained by the algorithms. From Fig. 1, we infer that our algorithm is robust to noise in the data since the class specific accuracies do not drop and remain fairly constant while the performance of all the other methods deteriorates sharply with the increase in noise levels.

5 Conclusion

A novel algorithm for the classification of degraded hyperspectral data is proposed. A combination of nuclear norm minimization and non-negative matrix factorization is used to exploit the low rank nature of the data. A basis matrix is learnt from the underlying clean data which is used to extract features from the input degraded data. The discriminative ability of the underlying clean data is exploited using structural incoherence, which to the best of our knowledge, is being introduced for the first time in the hyperspectral image processing literature. Both the spatial and spectral information are exploited for classification which lead to state-of-the-art results.

References

1. Benediktsson, J.A., Palmason, J.A., Sveinsson, J.R.: Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **43**(3), 480–491 (2005)
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
3. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. *SIAM J. Opt.* **20**(4), 1956–1982 (2010)
4. Chen, Y., Zhao, X., Jia, X.: Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **8**(6), 2381–2392 (2015)
5. Chen, Y., Nasrabadi, N.M., Tran, T.D.: Hyperspectral image classification using dictionary-based sparse representation. *IEEE Trans. Geosci. Remote Sens.* **49**(10), 3973–3985 (2011)
6. Chen, Y., Nasrabadi, N.M., Tran, T.D.: Hyperspectral image classification via kernel sparse representation. *IEEE Trans. Geosci. Remote Sens.* **51**(1), 217–231 (2012)
7. Ham, J., Chen, Y., Crawford, M.M., Ghosh, J.: Investigation of the random forest framework for classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **43**(3), 492–501 (2005)
8. Jia, S., Zhang, X., Li, Q.: Spectral-spatial hyperspectral image classification using $\ell_{1/2}$ regularized low-rank representation and sparse representation-based graph cuts. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **8**(6), 2473–2484 (2015)
9. Jiang, J., Chen, C., Yu, Y., Jiang, X., Ma, J.: Spatial-aware collaborative representation for hyperspectral remote sensing image classification. *IEEE Geosci. Remote Sens. Lett.* **14**(3), 404–408 (2017)

10. Li, J., Bioucas-Dias, J.M., Plaza, A.: Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Trans. Geosci. Remote Sens.* **48**(11), 4085–4098 (2010)
11. Li, W., Du, Q., Zhang, F., Hu, W.: Collaborative-representation-based nearest neighbor classifier for hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **12**(2), 389–393 (2015)
12. Lu, Y., Yuan, C., Zhu, W., Li, X.: Structurally incoherent low-rank nonnegative matrix factorization for image classification. *IEEE Trans. Image Process.* **27**(11), 5248–5260 (2018)
13. Makantasis, K., Karantzalos, K., Doulamis, A., Doulamis, N.: Deep supervised learning for hyperspectral data classification through convolutional neural networks. In: *Proceedings IGARSS*, pp. 4959–4962. IEEE (2015)
14. Melgani, F., Bruzzone, L.: Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **42**(8), 1778–1790 (2004)
15. Mou, L., Ghamisi, P., Zhu, X.X.: Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **55**(7), 3639–3655 (2017)
16. Ozdemir, A., Gedik, B., Cetin, C.: Hyperspectral classification using stacked autoencoders with deep learning. In: *Proceedings WHISPERS*, pp. 1–4. IEEE (2014)
17. Ramirez, I., Sprechmann, P., Sapiro, G.: Classification and clustering via dictionary learning with structured incoherence and shared features. In: *Proceedings CVPR*, pp. 3501–3508. IEEE (2010)
18. Rockafellar, R.T.: Augmented lagrange multiplier functions and duality in nonconvex programming. *SIAM J. Contr.* **12**(2), 268–285 (1974)
19. Sun, W., Yang, G., Du, B., Zhang, L., Zhang, L.: A sparse and low-rank near-isometric linear embedding method for feature extraction in hyperspectral imagery classification. *IEEE Trans. Geosci. Remote Sens.* **55**(7), 4032–4046 (2017)
20. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009)
21. Zhao, Y., Yang, J.: Hyperspectral image denoising via sparse representation and low-rank constraint. *IEEE Trans. Geosci. Remote Sens.* **53**(1), 296–308 (2015)
22. Zhong, P., Wang, R.: Learning conditional random fields for classification of hyperspectral images. *IEEE Trans. Image Process.* **19**(7), 1890–1907 (2010)
23. Zhong, Y., Zhang, L.: An adaptive artificial immune network for supervised classification of multi/hyperspectral remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **50**(3), 894–909 (2012)
24. Zhu, L., Chen, Y., Ghamisi, P., Benediktsson, J.A.: Generative adversarial networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **56**(9), 5046–5063 (2018)