SindhUniv. Res. Jour. (Sci. Ser.) Vol. 51 (03) 555-564 (2019)

http://doi.org/10.26692/sujo/2019.03.87

SINDH UNIVERSITYRESEARCH JOURNAL (SCIENCESERIES)

A Unique Spectral Spatial Bayesian Framework via Elastic Net Regression for the Classification of Hyperspectral images

B. N. SOOMRO[++], N. A. JAFFAR*, S. BHATTI, L. A. THEBO

Department of Computer Systems, Mehran University of Engineering and Technology

Received 13th February 2019 and Revised 28th July 2019

Abstract: This article presents not to simply a unique two stage regularization/shrinkage estimator for regression; rather, explicit to make the Bayesian framework connection to the Elastic Net procedure via the post-processed Edge preserving filtering which consist of two steps. We evaluated the quality of bands with pixel-based classifier associated with the Elastic Net based regularized regression. Next, spatial contextual information is used for refining the classification results obtained in the first step. This is achieved by means of a generic but powerful bilateral filtering post-processing, with a color guidance image retrieved from the principal components of the hyper-spectral image. Under the generalized Elastic Net framework, our proposed model showed the less time complexity. When comparing three widely used hyper-spectral data sets with the other classification methods, our method has shown the noticeable classification accuracy while the number of training samples is relatively small.

1. INTRODUCTION

Hyperspectral image classification has been acknowledged as the fundamental and challenging task of hyperspectral data processing. The advantage of hyperspectral imaging technique is that, considering that every element (water, tree, soil, etc) is defined by a precise spectrum (spectral signature), it should be possible to accurately classify every pixel of the image by considering their spectrum. The abundance of spectral and spatial information has provided great opportunities to effectively characterize and identify ground materials.

In recent days for supervised hyper-spectral data classifi-cation many methods have been used. Maximum likelihood are included in classic Techniques (ML) (Schowengerdt, 2007) neural networks, (Subramanian, *et al.*, 1997) NN classifiers (Samaniego, *et al.,* 2008) among many others. The quality of these pixel-wise classification methods is strongly related to the quality and number of training samples.

By blending the α_1 and α_2 penalties, Elastic Net (ELN) regression resolves these issues. Alike to the LASSO, the variable selection and continuous shrinkage is simultaneously done by the ELN, it is also selectable for the groups of correlated features. With the α_1 ELN generates the sparse features and α_2 penalty facilitates the stability and correlated characteristics along with the free parameter α establishes the penalties relative strength. Nowadays ELN emerged with logistic regression, (Tomioka *et al.*, 2011) multi-kernel learning (Hussain and Shawe-Taylor., (2011)". The effectiveness of this technique is popularized in multidimensional image classification (Balamurugan, 2013) and functional magnetic resonance imaging (FMRI) (John *et al.,* 2012) classification and also in devise of sparse classifiers Li (Wang, *et al.*, 2006).

It is a easy and effective way to regularize for spatial smoothness to meliorate the input space with features accounting for the neighborhood of the pixels. In recent days Edge-preserving filtering [EPF] (Friedman, *et al.*, 2010) (Bushra *et al.*, 2016) has shown the remarkable growth for the image processing.

To performs the weighted average of the neighborhood samples for HSI classification the EPF is well suited. EPF is a non-linear filter and it considers the spectral and spatial distances between the pixels, which enables the EPF to preserve the image details. The spatial variability presents due to the noise can be smoothed out with the help of EPF. Benefited with the weighted averaging approach of EPF, for the post-processing step it is well suited to add the geometric

[++]Corresponding: Author: bushra.naz@faculty.muet.edu.pk
Department of computer systems Engineering, Mehran University Of Engineering And Technology.

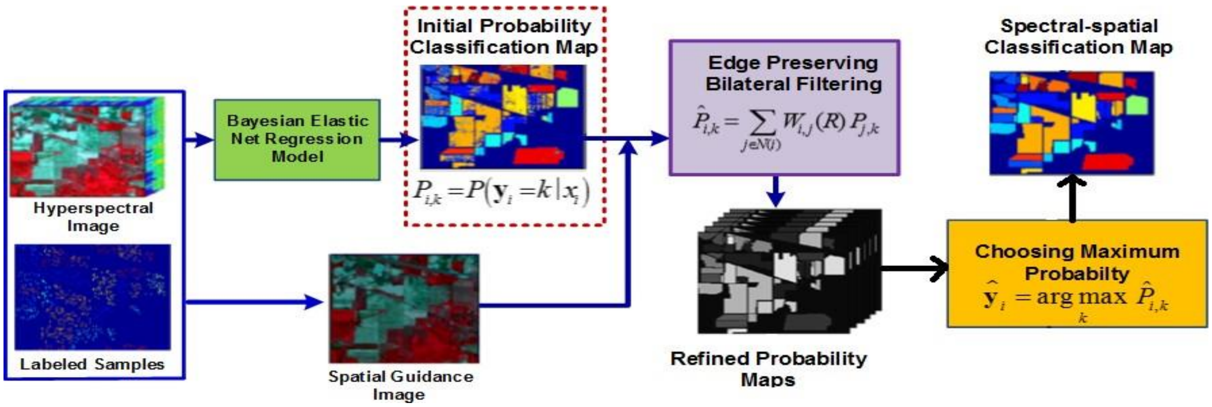


Fig. 1. A two stage post-processing spectral-spatial HSI classification model with Elastic Net regularized regresion and Bilateral Filetring.

information with the pixel-wise spectral-only classifier. Previous research (Bushra *et al.*, 2017) (Xu 2014) have indicated that post-processing classification is an important step in improving the quality of classifiers. In last decades researchers have been proposed the different kinds of EPFs, i.e., Joint bilateral, weighted least square (Farbman *et al.*, 2008), guided (He, *et al.*, 2013). Domain transform (Gastal, *et al.*, 2011), local linear Steins unbiased risk estimate and ℓ_0 -gradient (Xu *et al.*, 2011) filters. All filters have the joint filtering approach which smoothen the given image on the basis of a guidance image. During the filtering process the spatial information is well considered. Moreover, nowadays image processing researchers keenly noticed the Edge Preserving Filtering (EPF) (He, *et al.*, 2013). A lots of applications in image processing field is introduced such as high dynamic imaging (Farbman *et al.*, 2008)., stereo matching (Hosni *et al.*, 2013) image fusion (Li , *et al.*, 2013), (Li and Kang. 2012) enhancing (Zhang and Allebach. 2008), (He, *et al.*, 2011) and de- noising (Lin 2010) have been developed. Furthermore, for HSI visualization (Kotwal and Chaudhuri. 2010) and classification (Shahzad *et al.*, 2015) EPFs drawn the remarkable results.

For images an EPF is a non-linear, edge-preserving and noise-reducing smoothing filter. The intensity value at each pixel in an image is replaced by a weighted average of the classification results. Chosen the maximum probability we then finally classify the HSI. With the neighborhood pixel information, bilateral filtering with spatiao-spectral information is emerged for the refining the probabilities and must ensuring the refined probabilities are aligned with real object boundaries.

This research has proposed a post-processed bilateral filter logistic regression with Elatic Net regularization (ELNR- EPF). We have generalized a mathematical convenient approach to estimate the sparse coefficient by the Bayesian learning. In the regression termed as (ELNR) which gives the initial result as a multiple probability map and then at the second stage a post processing bilateral filter (a first stage we obtained the initial classification map with ELN logistic filtering approach which consider the spectral and spatial characteristics) is applied to refine and preserves the edges another pixel, it should not only occupy a nearby location but also have a similar value. The technique of EPF is as an extension of the bilateral filter.

2. PROPOSED FRAMEWORK

A. *Problem formulation*

For input an L band set of N pixel vector $\mathbf{X} \equiv [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N] \in \mathbb{R}^{L \times N}$ is given , where each $\mathbf{x}_i = [x_1, x_2, \ldots, x_L]^T \in \mathbb{R}^L$ is a L0Dimension vector. Consider the $S \equiv (1, \ldots, N)$ indexing HSI N pixel. Assume the K labels classes as $k \equiv (1, \ldots, K)$ and $\mathbf{Y} \equiv [\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_N] \in \mathbb{R}^{K \times N}$ as an image of class labels, where each $\mathbf{y}_i = \left[y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(K)} \right]^T \in \mathbb{R}^K$ as an image of class labels, where each is a 1-of-K encoding. Let the training set $T_\tau \equiv ((x_1, y_1), \ldots, (x_\tau, y_\tau)) \in (\mathbb{R}^L \times C)^\tau$ the total number of labeled samples is denoted by the τ . The major task of the HSI with the defines notation is to classify the label $y_i \in k$ for each pixel $i \in S$. This classification results in the thematic map of class labels $\mathbf{Y}$. The proposed framework is shown in **(Fig. 1).**

Bayesian Elastic Net Regression

In this article we broaden the Bayesian framework to the ELN penalty with complete characterization of ELN prior. The core elements of Bayesian ELN regression- the prior and posterior distributions. This places the ELN in the context of a model-based framework where point estimation, prediction, and model uncertainty can be addressed from a Bayesian learning perspective. Here, the posterior mode does not play a central role in the Bayesian paradigm; rather, inference about $\boldsymbol{\beta}$ is based on the entire posterior distribution $p(\boldsymbol{\beta}/\mathbf{y})$, the prediction of future observations $\tilde{\mathbf{y}}$ is based on the posterior predictive distribution $p(\tilde{\mathbf{y}}/\mathbf{y})$, and uncertainty about the specification of the regression model is addressed via the posterior distribution over the model space. A key advantage of casting ELN regression in a Bayesian framework is that uncertainty about the regularization parameters α and γ can be incorporated into the model through a prior distribution. Integrating over this uncertainty essentially creates an infinite mixture of ELN regression models, allowing for adaptive, data-based shrinkage of the regression coefficients.

Initially, we used the pixel wise classification model to build the class densities based on Bayesian framework with ELN prior (ELNR) to generate the pixel-wise classification map $P_{i,k} = P\left(\mathbf{y}_i = k \ \middle|\mathbf{x}_i\right)$ using probability map $P = \left(p_1, \ldots, p_k\right)$. Specifically, the probability is defined as follows:

$$P_{i,k} = \begin{cases} 1 & if \ \mathbf{x}_i \in k \\ 0 & otherwise \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

To refine the initial classification map, in probability optimization we will added spatial aware information via bilateral filtering with the color reference image $\mathbf{R}$. This research models the posterior class probabilities using the multinomial logistic regression (MLR) [33]. We have consider the linear regression in standard with ELN based regularized multino-mial logistic regression, initially the pixel-wise probability Pi,k = P (yi = k xi) which is related to a given sample belonging to the class k can be given as:

$$P_{i,k} = P\left(\mathbf{y}_i = k \ \middle| x_i; \boldsymbol{\beta}_{0k}, \boldsymbol{\beta}_k\right) = \frac{e^{\boldsymbol{\beta}_{0k} + \mathbf{x}_i^T \boldsymbol{\beta}_k}}{\sum_{j=1}^{K} e^{\boldsymbol{\beta}_{0j} + \mathbf{x}_i^T \boldsymbol{\beta}_j}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

where $k = 1, 2, \ldots, K-1, i \in S$, while the $\left(\boldsymbol{\beta}_{0k}, \boldsymbol{\beta}_k\right)^T \in \mathbb{R}^{L+1}$ is the regress or of coefficients corresponding to the class k for MLR model. taking the advantages of ELN penalty and assuming the sparsity and correlation of repressors $\left(\boldsymbol{\beta}_{0k}, \boldsymbol{\beta}_k\right)$, the general prior of $\left(\boldsymbol{\beta}_{0k}, \boldsymbol{\beta}_k\right)$ is modeled as ELN prior is modeled as ELN prior [34]. It is worth to mention here that under the Bayesian framework, the ELN penalty corresponds between the Gaussian and Laplacian priors. Then maximum a posterior (MAP) to estimate the regress or can be written as maximum penalized log-likelihood, given as:

$$\left(\boldsymbol{\beta}_{0k}, \boldsymbol{\beta}_k\right) = \max_{\{\boldsymbol{\beta}_{0k}, \boldsymbol{\beta}_k\}_1^K} \frac{1}{N} \sum_{i=1}^{N} \ell\left(\boldsymbol{\beta}_{0k}, \boldsymbol{\beta}_k\right) - \lambda \sum_{k=1}^{K} P_\alpha\left(\boldsymbol{\beta}_k\right) \quad \ldots\ldots\ldots (3)$$

where

$$\ell\left(\boldsymbol{\beta}_{0k}, \boldsymbol{\beta}_k\right) = \log \sum_{i=1}^{N} P\left(\mathbf{y}_i \ \middle| \mathbf{x}_i; \boldsymbol{\beta}_{0k}, \boldsymbol{\beta}_k\right) \quad \ldots (4)$$

and

$$P_\alpha(\boldsymbol{\beta}_k) = (1-\alpha)\left\|\boldsymbol{\beta}_k\right\|_{\ell 2}^2 + \alpha\left\|\boldsymbol{\beta}_k\right\|_{\ell 1} \quad \ldots\ldots\ldots (5)$$

is the ELN penalty. while the P_α measures the penalty strength of Lasso $(\alpha = 1)$ and Ridge-regression $(\alpha = 0)$. The penalization term is now a convex combination of the ℓ_1 norm and ℓ_2 -norm of the coefficient vector .The parameter $\alpha \in (0,1)$ is used for determining the proportions between the these two types of regularization.

For the $N \times k$ indicator response matrix $\mathbf{Y}$, with elements $y_{il} = I\left(g_i = l\right)$ log-likelihood function can be specified in more explicit form which is a concave function of the parameters.

$$L(\{\beta_{0\ell},\mathbf{\beta}_\ell\}_1^K) = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left[\sum_{l=1}^{K} y_{il}\left(\beta_{0l}+x_i^T\mathbf{\beta}_l\right) - \log\left(\sum_{l=1}^{K} e^{\beta_{0l}+x_i^T\mathbf{\beta}_l}\right)\right] \quad (6)$$

To avoid the complexity of Eq.(6) log likelihood can be approximated to a quadratic lower bound by performing the partial Newton steps, in the spirit of path wise coordinate descent algorithm while allowing only $\{\beta_{0l},\beta_l\}$ to vary for a single class at a time.

$$L_{Ql}(\beta_{0l},\mathbf{\beta}_l) = -\frac{1}{2\mathrm{N}}\sum_{i=1}^{N}\omega_{il}\left(z_{il}-\beta_{0l}-x_i^T\mathbf{\beta}_l\right)^2 + C\left(\left\{\tilde{\beta}_{0k}+\tilde{\mathbf{\beta}}_k\right\}_1^K\right) \quad .(7)$$

where,

$$z_{il} = \beta_{0l} + x_i^T\beta_l + \frac{y_i - P_l(x_i)}{P_l(x_i)\left(1-P_l(x_i)\right)} \quad \text{……...(8)}$$

$$\omega_{il} = P_l(x_i)\left(1-P_l(x_i)\right) \quad \text{…………………….....(9)}$$

Here z_{il} denotes the working response and ω_{il} represents the weights. using the cyclical coordinate descent, this proposed regularized problem is solved.

Post-processing Edge Preserving Bilateral Filtering Inspired by the HSI phenomenon, nowadays to concentrates the information in the spectral and spatial domains. Spatial hierarchies may consist of features, sizes, places, contexts, pyramids, or even ultrametrics based on the spatially perceptive relationships. One simple yet effective method for spatial spectral classification is by applying adaptive filters or moving windows to the spectral bands. In this letter, to improve the ELNR classification performance obtained through the spectral information alone via the Bayesian Elastic Net Regression, we integrate the spatial information with spectral information with a a very simple and generic Edge Preserving Filtering (EPF) approach known as bilateral filter (BF).

With only spectral information, the classification map contains the noise and does not aligned with the real object boundaries. The optimized probabilities are modeled as a weighted average of its neighborhood probabilities,

$$\hat{P}_{i,k} = \sum_{j\in N(i)} W_{i,j}(\mathbf{R})\, P_{j,k} \quad \text{………………………(10)}$$

where $\mathbf{R}$ and $\hat{P}_{i,k}$ refers to the guidance and input image, respectively. While i and j represent the ith and jth pixels. While $N(i)$ is the neighborhood of the ith pixel and the weight $W_{i,j}$ for the bilater filtering which preserves the edges of a specified guidance image can be defined using two Gaussian decreasing functions,

$$W_{i,j}(R) = 1/K_i^b \sum_{j\in w_i} G_{\delta_s}\left(\|i-j\|\right) G_{\delta_r}\left(\left|R_i - R_j\right|\right) \quad (11)$$

δ_s is used for local window of a pixel, and δ_r expounds the decreasing weight of a pixel via the intensity difference between the reference pixels, i.e., R_i and R_j . Here, w_i is a local window of size $(2\delta_s+1)\times(2\delta_s+1)$ around pixel i, k_i^b is a normalizing term of the bilateral filter.

Referring **Eq.12**, if the neighborhood pixels of pixel i in the guidance image have similar intensities or colors, i.e., $R_i \approx R_j$, the weight of pixel j , represented by neighboring pixel j , will be quite large, especially when it is very close to i , i.e., $\|i-j\|$ is very small. In contrast, if the neighboring pixels have quite different intensities in the guidance image $\mathbf{R}$, the situation will be the opposite. The pixels with the similar color or intensities present in the reference image must have the similar output.

For Guidance image we choose the PCA. Benefitted with the bilateral filter features, i.e., we can get the more accurate and effective classification accuracy by avoiding the blurring while removing the noise between the homogenous zones. Moreover, Bilateral Filter provides the δs and δr parameters modification in non-iterative way.

Once the initial probabilities map are refined with EPF approach, the classification of pixel i for label yi can be given by choosing the maximum probability as:

$$\mathbf{y}_i = \arg\max_k \hat{P}_{i,k} \quad \text{…………………………….(12)}$$

The step by step algorithm for the proposed framework is illustrated in **Algorithm 1**.

Algorithm 1: Elastic Net Regression & EP Bilateral Filtering

1. Input:

- $L \times N$ HSI $\mathbf{X} \equiv (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)$ $K * N$, spatial window size ω, ELN regularization α and λ.

2. Compute the pixel-wise ELN regularized regression for initial classification with PCD.

- For each pixel find initial probability map via the **Eq: 2**

3. Apply Edge Preserving Bilateral Filtering via **Eq: 10** - Choose maximum probability using the **Eq: 12**

4. Output Post-processed Spectral-spatial classification result.

3. EXPERIMENTS AND RESULTS.

A. Data Sets And Metrics

In order to evaluate the performance of the proposed method, three well known hyperspectral datasets are employed i.e., Indian Pines, University of Pavia and Centre of Pavia images.

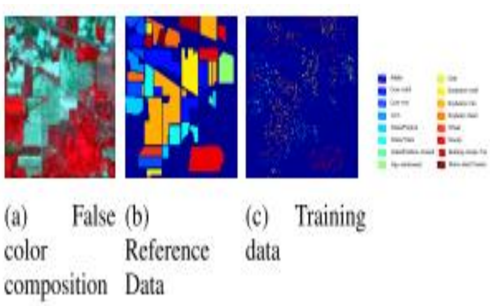


Fig. 2. AVIRIS Indian Pines dataset.

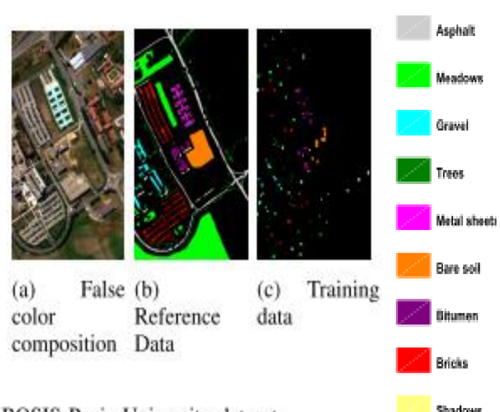


Fig. 3. ROSIS Pavia University dataset.

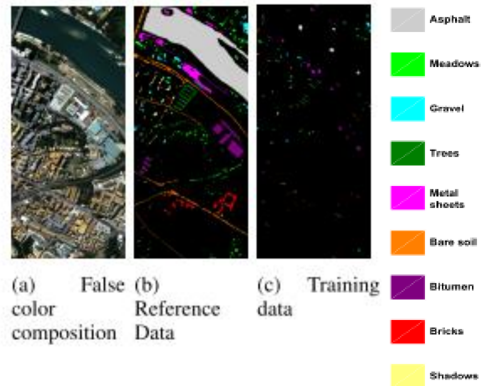


Fig. 4. ROSIS Pavia Center dataset.

The purpose of the experiments is to compare the performance of the proposed post-processed approach with other state-of-the-art analysis classifiers, such as Sparse multinomial logistic regression (SMLR), LORSAL with multilevel logistic spatial prior (denoted as LOR-MLL), Support vector machine (SVM) and SVMs with composite kernels (denoted by SVM-CK) that combine the spectral information and spatial information via a weighted kernel summation, which are well-established techniques in the machine learning community. Furthermore our previously proposed Bilayer Elastic Net Regression (ELN^2-RegMLR).All parameters of these methods are set according to the reference papers.

4. CLASSIFICATION RESULTS

In this experimental analysis we have represented the results in boldface that are significantly better than others. From the experimental results, we have the following outcomes.

1) AVIRIS Indian Pines Data Set.: This image is a classical benchmark to validate the accuracy of hyper-spectral image analysis algorithms. For comparison, we adopt the aforementioned state-of-the-art supervised classifiers. **(Fig. 5(a)−(h))** shows the ground-truth and classification results obtained by the different tested methods for the Indian Pines data. Moreover, **(Table I)** gives all comparable results of different classifiers. From Table I and Fig. 5, classifiers with spatial information (MLL prior, CK, EPF) have shown a clear advantage over the pixel-only counterparts. while our method ELNR-EPF gives the second best spectral spatial classification result, which is about 2.55% higher than the state-of-the-art SVM-CK classifier in OA. It is also noticeable that although our approach is linear but it produces comparable results with the well known SVM and LOR-MLL and ELN2-RegMLR methods in pixel-wise and spectral-spatial counterparts.

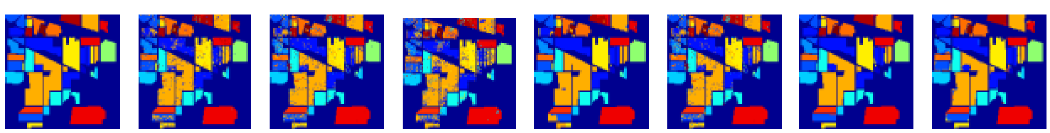
2) ROSIS University of Pavia Data Set.: In second experiment, we evaluate our method using the ROSIS University of Pavia data set while comparing with the other state-of-the-art methods mentioned previously. Fig. 6(a)−(h) illustrates the reference map and classification results of the classifiers listed in Table 3.

From **(Table 2 and Fig. 6),** we can conclude that proposed approach achieves the second highest accuracy among all of the other classifiers. It is worth to emphasize here that all classical spectral-spatial classification methods adopted the RBF kernel for mapping, indeed we used linear system without the RBF kernel and still get the Comparable results than RBF kernel mapping approaches. We get the second highest spatiao-spectral classification accuracy and get 1.57% higher than SVM-CK approach.

3) ROSIS Centre of Pavia Data Set.: In this experiment, we evaluate our method using the ROSIS center of Pavia data set while comparing with the other state-of-the-art methods mentioned previously. We have compared our previous proposed method and other well known classification methods: which obtain the best classification result in the center of Pavia data set. **(Fig. 4)** show the reference, false color map and training data. It is worth to emphasize that these training samples are out of the testing samples. **Fig.7(b)−(h)** illustrates the classification results while the Table 3. represents the numerical classification results in the form of OA, AA and κ coefficients. From the Fig. 7(b)−(h) it is observed that the all classifiers achieve good classification results, but our method still gets the highest result, which is which is 0.15% higher than the our previous proposed method ELN2- RegMLR.

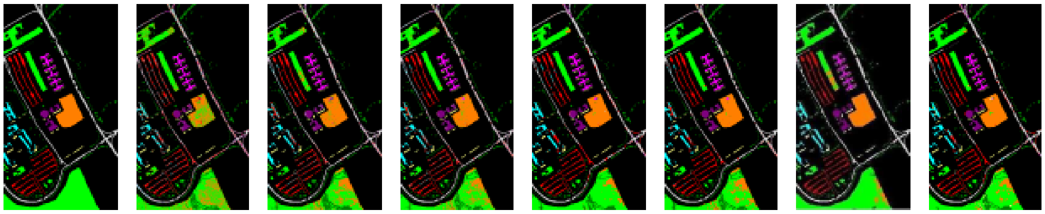
Table I: Classification Accuracy (%) For The Indian Pines Image Using Training Samples And Testing Samples

Class Type	SMLR	SVM	ELNR	LOR-MLL	SVM-CK	ELN^2-RegMLR	ELNR-EPF
Alfalfa	64.17	81.25	27.08	70.42	95.83	97.92	**100.0**
Corn-no-till	82.03	86.28	82.79	92.96	96.67	**96.78**	94.60
Corn-min-till	70.97	72.80	58.27	86.65	90.93	**97.07**	94.18
Corn	64.81	58.10	45.71	79.38	85.71	**100.0**	74.37
Grass/Pasture	91.07	92.39	84.56	94.25	93.74	**98.21**	97.82
Grass/Tree	96.46	96.88	91.37	98.66	97.32	99.40	**99.50**
Grass/Pasture-mowed	38.70	43.48	43.51	50.00	69.57	13.04	**100.0**
Hay-Windowed	99.30	98.86	99.09	99.39	98.41	**100.0**	**100.0**
Oats	28.89	0.00	17.33	50.00	55.56	22.27	**100.0**
Soybeans-no-till	76.61	71.53	61.08	90.20	93.80	**99.68**	88.10
Soybeans-min till	83.03	84.38	80.55	93.89	94.37	**98.55**	98.30
Soybeans-clean till	81.76	85.51	74.82	94.37	93.66	**98.42**	94.53
Wheat	99.63	100.0	99.47	99.58	99.47	98.97	**100.0**
Woods	96.41	93.30	95.56	97.66	99.14	**100.0**	99.50
Bldg-grass-tree-drives	66.61	64.91	60.23	79.30	87.43	**96.49**	87.53
Stone-steel towers	70.24	88.24	89.41	73.06	**100.0**	95.29	90.40
OA	83.94	84.52	79.17	92.70	94.86	**97.93**	97.41
AA	75.67	79.24	68.41	84.36	90.73	88.41	**94.96**
κ	0.746	0.823	0.761	0.917	0.941	**0.975**	0.947



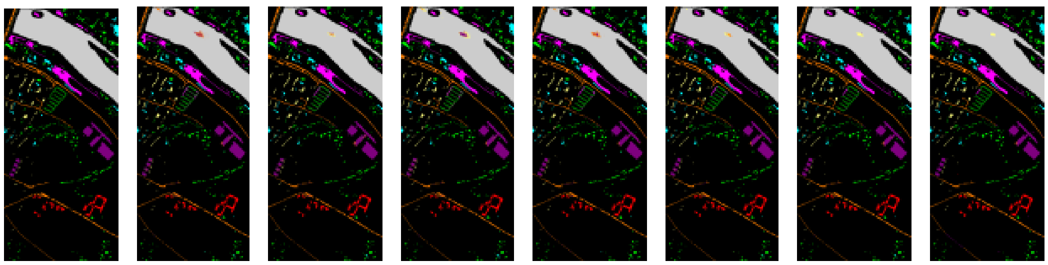
(a) Groundtruth (b) SMLR (c) SVM (d) ELNR (e) LOR-MLL (f) SVM-CK (g) ELN^2-RegMLR (h) ELNR-EPF

Fig. 5. Indian Pines image with OA. (a)ground truth of Image ; (b)SMLR(OA=83.94%); (c)SVM(OA=84.52%); (d)ELNR(OA = 79.17%); (e)LOR-MLL (OA = 92.70%); (f) SVM-CK(OA =94.86%); (g)ELN^2-RegMLR(OA =97.93%) (h)ELNR-EPF (OA = 97.41%); with about 10% training samples.



(a) Groundtruth (b) ELNR (c) SVM (d) SMLR (e) SVM-CK (f) LOR-MLL (g) ELN^2-RegMLR (h) ELNR-EPF

Fig. 6. University of Pavia image with OA. (a)ground truth of Image ; (b)SMLR(OA=81.63%); (c) SVM(OA=79.15%); (d)ELNR(OA = 78.90%); (e)LOR-MLL (OA = 85.69%); (f)SVM-CK (OA =87.18%); (g)ELN2-RegMLR(OA =96.03%) (h)ELNR-EPF(OA = 89.75%) with about 9% training samples.

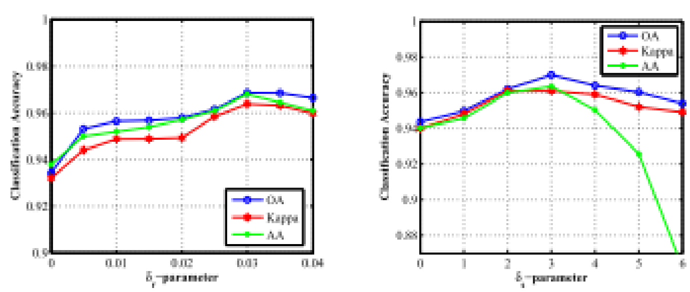


(a) Groundtruth (b) SMLR (c) SVM (d) ELNR (e) LOR-MLL (f) SVM-CK (g) ELN^2-RegMLR (h) ELNR-EPF

Fig. 7. University of Pavia image with OA. (a)ground truth of Image ; (b)SMLR(OA=97.98%); (c) SVM(OA=97.70%); (d)ELNR(OA = 97.78%); (e)LOR-MLL (OA = 98.73%); (f)SVM-CK (OA =98.47%); (g)ELN^2-RegMLR(OA=98.58%); (h)ELNR-EPF(OA = 98.63%) with about 9% training samples.

TABLE 3: CLASSIFICATION ACCURACY (%) FOR THE CENTRE OF PAVIA IMAGE USING TRAINING AND TESTING SAMPLES

Class Type	SMLR	SVM	ELNR	LOR-MLL	SVM-CK	ELN^2-RegMLR	ELNR-EPF
Asphalt	99.41	99.23	99.11	99.41	99.34	**99.98**	99.58
Meadows	94.04	92.64	93.04	94.04	94.39	95.04	**98.26**
Gravel	95.18	95.51	95.08	**96.80**	95.65	96.30	96.80
Trees	82.36	81.64	81.39	89.62	84.70	**93.62**	92.80
Metal Sheets	96.75	96.04	97.35	97.85	99.04	**99.58**	99.53
Bare-soil	94.51	94.19	92.89	97.49	96.39	**97.83**	97.05
Bitumen	9.45	94.61	95.95	98.12	**98.41**	94.12	98.10
Bricks	99.62	99.72	99.12	99.83	99.97	**100.0**	**100.0**
Shadows	99.95	99.85	**100.0**	**100.0**	98.73	99.28	99.34
OA	97.98	97.70	97.78	**98.73**	98.47	98.58	98.63
AA	96.32	94.93	95.99	97.20	96.43	**98.50**	98.41
κ	0.953	0.958	0.943	0.977	0.972	0.988	0.986



(a) (b)

Fig. 8. Impact of the Edge Preserving filtering parameters (a) the δ_r parameter (b) the δ_s parameter

C. Analysis of the δ_r and δ_s parameters

For the proposed method the blur degree (δr) and the filtering size (δs) are the key parameters. The influence of these parameters are depicted in (Fig. 8) using the Indian Pines image. We have chosen the 10% training samples. To test the impact of the δr parameter we have fixed the parameter δs = 3, while for testing the impact of the δs parameter the blur degree parameter δr is fixed to 0.03. From the Fig. 8(a), it is worth to notify that the setting the δr parameter is vitally very important for the proposed method. To analyzed the influence of δr parameter on classification accuracies, we set this parameter from 0.005 to 0.04 with the step size of and other parameters are same as used for the proposed method. We can observe in the Fig. 8(a) that the classification accuracies (OA, AA and κ) increases in the case when the δr 0.03 while for δr > 0.03 the tendency of accuracies have decreased gradually. For the δr = 0.03, we get the best classification result. 0.15% higher than the our previous proposed method ELN^2-RegMLR.results, but our method still gets the highest result, which is0.15% higher than the our previous proposed method ELN2-RegMLR.We also illustrated the influence of the blurring parameter visually as shown in Fig. 9. A classical data set Indian Pines we have chosen and we have set the δr=1, 2, 3 and 4. From the Fig. 9 we have observed that if this parameter is not chosen perfectly then the classification results will be deeply blurred.

To analyze the influence of the filtering size δs parameter, we set filtering size from 0 to 6 with step size of 1. It is also worth to recall that the spatial local window size is directly related with the filtering size as we define previously, i.e., $\Omega = (2\delta s + 1) \times (2\delta s + 1)$, thus setting the δs parameter is significantly important to the performance of our proposed BF based method. The influence of this parameter is illustrated in the Fig. 8(b). It can be easily observed from the Fig. 8(b) that the classification accuracies have the same tendency as we have observed in the analysis of the δr parameter. With the δs ≤ 3, the classification accuracies are increases gradually while for the filtering size parameter δs > 3, there is a dramatically reduction in the average accuracy, while the OA and κ decreases gradually. Although the AA for δs = 4 is not reduced so much but for the δs = 5 and δs = 6 it is showing the very low AA, the reason is that because for a small-scale class like "Oats" which contains only 20 samples can be totally mis-classified when the filtering size is very large. It can be concluded that δs should not be set to be too small or large. For the best result we set the filtering size δs = 3 which leads to the spatial window size of 7 × 7.

5. CONCLUSION

In this paper we have introduced the HSI classification generic but powerful post-processed Edge preserving bilateral filtering approach which includes both spatial and spectral Information. The proposed method aims at optimizing the pixel-wise classification maps in a local filtering framework. A grouping and sparsity promoting multinomial logistic regression with regularized Elastic Net is proposed to estimate the initial classification map. Apart from yielding better results, The proposed scheme provides computational efficiency with the local probabilities optimization. In addition, the proposed spatio-spectral ELNR-EPF opens a wide field for future developments in which filtering methods can be easily incorporated. This paper has shown that local smoothing is also able to achieve a high classification accuracy. For the future work we can extend our work by handling the adaptively filtering size and blur degree of EPF.

REFERENCES:

Benediktsson J., M. Pesaresi, and K. Amason (2003). "Classification and feature extraction for remote sensing images from urban areas based on mor- phological transformations". IEEE Trans. Geosci. Remote Sens., 41(9): 1940-1949.

Bushra N. S., L. Xiao L, Huang, S. Shahzad and M. Molaei. (2016) Bilayer Elastic Net regression model for supervised spectral-spatial hyperspectral image classification J. IEEE Journal of Selected Topics in Earth Observations and Remote Sensing,9(9): 4102-4116.

Bushra N. S., L. Xiao, S. Shahzad and M. Molaei (2016) "Spatial-aware supervised learning for hyperspectral image classification: Comprehensive assessment". Journalof Donghua University, 33(6): 954-960.

Bushra N. S., L. Xiao, M. Molaei, L. Huang, Z. Lian, S. Shahzad (2017) "Local and non-local Context-aware Elastic Net representation based classification for Hyperspectral images". IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(6), Digital Object Identifier: 10.1109/JSTARS.2017.2666118.

Balamurugan, P., (2013) "Large-Scale Elastic Net Regularized Linear Classi- fication SVMs and Logistic Regression.," Data Mining (ICDM), 2013 IEEE 13th International Conference on , vol., no. 949, 954, 7-10.

Camps-Valls G., N. Shervashidze and K. M. Borgwardt (2010) "Spatio-spectral remote sensing image classification with graph kernels". IEEE Geosci. Remote Sens. Lett., 7(4): 741-745.

Camps-Valls, G., L. Gomez-Chova, J. Mu˜noz-Mar´ı, J. Vila-Franc´es, and J. Calpe-Maravilla, (2006) “Composite kernels for hyperspectral image classification,” IEEE Geosci. Remote Sens. Lett., vol. 3, 93−97.

Fauvel M., J. Chanussot and J. A. Benediktsson (2012) “A spatial-spectral kernel based approach for the classification of remote sensing images”. Pattern Recognit., 45(1): 381 392.

Farbman Z., R. Fattal, D. Lischinski and R. Szeliski (2008). ‘Edge-preserving decompositions for multi-scale tone and detail manipulation”. ACM Trans. Graph., 2008, 27(3): 67-76.

Friedman, H., J. T. Hastie, and R. Tibshirani, (2010)”Regularization paths for generalized linear models via coordinate descent.”, Journal of Statistical Software., vol. 33, no. 1, 844−856

Gastal E. S. L. and M. M. Oliveira (2011). “Domain transform for edge-aware image and video processing”. ACM Trans. Graph., 30(4): 6978Pp.

He K., Sun J., and X. Tang (2013) “Guided image filtering”. IEEE Trans. Pattern Anal. Mach. Intell., 35(6): 1397 1409.

Hosni A., C. Rhemann M. Bleyer, C. Rother and M. Gelautz, (2013) “Fast cost-volume filtering for visual correspondence and beyond”. EEE Trans. Pattern Anal. Mach. Intel., 35(2): 504-511.

He K., J. Sun, and X. Tang. (2011) “Single image haze removal using dark channel prior”. IEEE Trans. Pattern Anal. Mach. Intell., 33(12): 2341-2353.

Hussain Z. and J. Shawe-Taylor, (2011) ”Metric learning analysis.”, PinView FP7- 216529 Project Deliverable Report D3, vol.3.

John W. K., D. A. Degenhart, P. Daniel Siewiorek, (2012) ”Sparse Linear Regression with Elastic Net Regularization for Brain-Computer Inter- faces.”,34th Annual International Conference of the IEEE EMBS San Diego, California USA,pp. 4275-4278,

Kang X., S. L. Li, M. Fang., and J. A. Benediktsson (2015) “Extended random walker-based classification of hyperspectral images”. IEEE Trans. Geosci. Remote Sens., 53(1): 144-153.

Kotwal K. and S. Chaudhuri. (2010) “Visualization of hyperspectral images using bilateral filtering”. IEEE Trans. Geosci. Remote Sens., 48(5): 2308-2316.

Liu J., Z. Wu, Z. Wei, L. Xiao L. Sun (2013) “Spatial-spectral kernel sparse representation for hyperspectral image classification”. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens., 6(6): 2462-2471.

Li Wang, Ji Zhu and Hui Zou, (2006) ”The doubly regularized support vectormachines.” , Statistica Sinica, vol. 16, no, 589−615.

Le S., Z. Wu J. Liu L. Xiao and Z. Wei, (2015)”Supervised Spectral-Spatial Hyperspectral Image Classification With Weighted Markov Random Fields.”, IEEE Transactions on Geoscience and Remote Sensing. vol. 53, no. 3, 1490−1503.

Li S., X. Kang and J. Hu (2013) “Image fusion with guided filtering”. IEEE Trans. Image Process., 22(7): 2864 2875.

Li S. and X. Kang. (2012) “Fast multi-exposure image fusion with median filter and recursive filter”. IEEE Trans. Consum. Electron., 58(2): 626-632.

Lin C. H., J. S. Tsai and C. T. Chiu. (2010) “Switching bilateral filter with a texture/noise detector for universal noise removal”. IEEE Trans. Image Process., 19(9): 2307 2320.

Li, J., J. M. Bioucas-Dias, and A. Plaza (2011), ”Hyperspectral image.segmentation using a new Bayesian approach with active learning.”,IEEE Trans. Geosci. Remote Sens., vol. 49, 3947−3960.

Naz B. S., L. Xiao L. Huang, S. Shahzad, M. Molaei. (2016) “Bilayer Elastic Net regression model for supervised spectral-spatial hyperspectral image classification”. IEEE Journal of Selected Topics in Earth Observations and Remote Sensing, 9(9): 4102−4116.

Qiu T., A. Wang, N. Yu, and A. Song. (2013) “LLSURE: Local linear sure- based edge-preserving image filtering”. IEEE Trans. Image Process., 22(1): 80-90.

Shahzad S., L. Xiao, and B. N. Soomro (2015) “Hyperspectral image clas- sification via Elastic Net regression and Bilateral filtering”. in Proc.IEEE Int. Prog. in Infor. and Computing, Nanjing, China, 1(1): 56-60.

Schowengerdt, R. A. (2007) ”Models and Methods for Image Processing.”, Remote Sensing,3rd ed. New York, NY, USA: Academic.

Samaniego, L., A. Bardossy, K.. Schulz, (2008) ”Supervised Classification of Remotely Sensed Imagery Using a Modified k -NN Technique, ” in

Geoscience and Remote Sensing, IEEE Transactions on vol. 46, no.7, 2112 2125.

Subramanian, S., N. Gat, M. Sheffield, J. Barhen, and N. Toomari-an, (1997) Methodology for hyperspectral image classification using novel neural network, in Proc. SPIE Algorithms Multispectr. Hyperspectr. Imagery III, vol. 3071, 128137Pp.

Tomioka, R., T. Suzuki and M. Sugiyama. (2011) "Super-Linear Convergence of Dual Augmented Lagrangian Algorithm for Sparse Learning.", Journal of Machine Learning Research, arXiv:0911.4046, vol. no. 1537-1586.

Tomasi C. and R. Manduchi (1998) "Bilateral filtering for gray and color images". Proc. Int. Conf. Comput. Vis., 839-846.

Xu L., C. Lu, Y. Xu and J. Jia (2011) "Image smoothing via A_0-gradient minimization". ACM Trans. Graph., 30(6): 174 185.

Xu L., J. Li (2014) "Bayesian classification of hyperspectral imagery based on probabilistic sparse representation and Markov random field". IEEE Trans. Geosci. Remote Sens11(4): 823-827.

Zou H. and T. Hastie, (2005) "Regularization and variable selection via the Elastic Net.", J. R. Stat. Soc. Ser. B Stat. Methodology, vol. 67, no. 2, .301–320.

Zhang B., J. P. Allebach. (2008) "Adaptive bilateral filter for sharpness en- hancement and noise removal". IEEE Trans. Image Process., 17(5): 664-678.