8-31-2015

Low-dimensional Representations of Hyperspectral Data for Use in CRF-based Classification

Yang Hu
yxh4745@rit.edu

Nathan D. Cahill
nathan.cahill@rit.edu

Sildomar Monteiro
stmeee@rit.edu

Eli S. Saber
esseee@rit.edu

David W. Messinger
messinger@cis.rit.edu

Recommended Citation
Yang Hu, Nathan D. Cahill, Sildomar T. Monteiro, Eli Saber, David W. Messinger, "Low-dimensional representations of hyperspectral data for use in CRF-based classification" Proc. SPIE 9643, Image and Signal Processing for Remote Sensing XXI, 96430L (October 15, 2015); doi:10.1117/12.2195229.

Low-dimensional Representations of Hyperspectral Data for Use in CRF-based Classification

Yang Hu[a] Nathan D. Cahill[b] Sildomar T. Monteiro[a,c] Eli Saber[a,c] and David W. Messinger[a]

[a]Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, NY, USA;
[b]Department of Electrical and Microelectronic Engineering, Rochester Institute of Technology, 77 Lomb Memorial Drive, Rochester, NY, USA;
[c]School of Mathematical Sciences, Rochester Institute of Technology, 85 Lomb Memorial Drive, Rochester, NY, USA

ABSTRACT

Probabilistic graphical models have strong potential for use in hyperspectral image classification. One important class of probabilisitic graphical models is the Conditional Random Field (CRF), which has distinct advantages over traditional Markov Random Fields (MRF), including: no independence assumption is made over the observation, and local and pairwise potential features can be defined with flexibility. Conventional methods for hyperspectral image classification utilize all spectral bands and assign the corresponding raw intensity values into the feature functions in CRFs. These methods, however, require significant computational efforts and yield an ambiguous summary from the data. To mitigate these problems, we propose a novel processing method for hyperspectral image classification by incorporating a lower dimensional representation into the CRFs. In this paper, we use representations based on three types of graph-based dimensionality reduction algorithms: Laplacian Eigemaps (LE), Spatial-Spectral Schroedinger Eigenmaps (SSSE), and Local Linear Embedding (LLE), and we investigate the impact of choice of representation on the subsequent CRF-based classifications.

Keywords: Hyperspectral image classification, conditional random fields, dimensionality reduction, manifold learning

1. INTRODUCTION

An increasing demand for efficient and precise hyperspectral image analysis algorithms is accompanied by rapid development in satellite and sensor technologies. Hyperspectral image classification, utilized in object detection in defense systems and change detection of agricultural fields, is an intensively researched area. In order to generate accurate classification results, low-dimensional representations of hyperspectral image with the aid of Conditional random fields (CRFs) is a selected choice due to their respective advantages listed below.

Low-dimensional representations, based on non-linear dimensionality reductions, are favorable for a number of reasons: First of all, more informative data structures are exposed under the manifold subspace, yieldling more useful features for classification. Secondly, thanks to the 'curse of dimensionality',[1] less number of dimensions and more samples, not the vice versa, are expected for classification task. There are currently many well-established dimensionality reduction techniques available. Laplacian Eigenmaps (LE)[2] retains the intrinsic geometry of the data by exploiting spectral characteristics of the graph Laplacian operator. Schroedinger Eigenmaps (SE)[3] extends Laplacian Eigenmaps by incorporating a potential matrix capable of encoding expert knowledge about the manifold structure. Local Linear Embedding (LLE)[4] describes each point as the linear combination of its neighbors in both original and projected data.

Proper probabilistic graphical models are meanwhile desired when considering the spatial content in hyperspectral image classification. Conditional random fields satisfy this demand since a) they require no independence assumption for the observed data, b) they provide flexibility in defining local and pairwise potentials, and c) they allow independent procedures between the modules of feature function calculation and parameter learning.

In this paper, we propose a classification algorithm combining non-linear dimensionality reduction techniques and CRFs. The paper is organized as following: the proposed algorithm is introduced in Section 2, testing classification results are presented in Section 3, and conclusions and future work are discussed in Section 4.

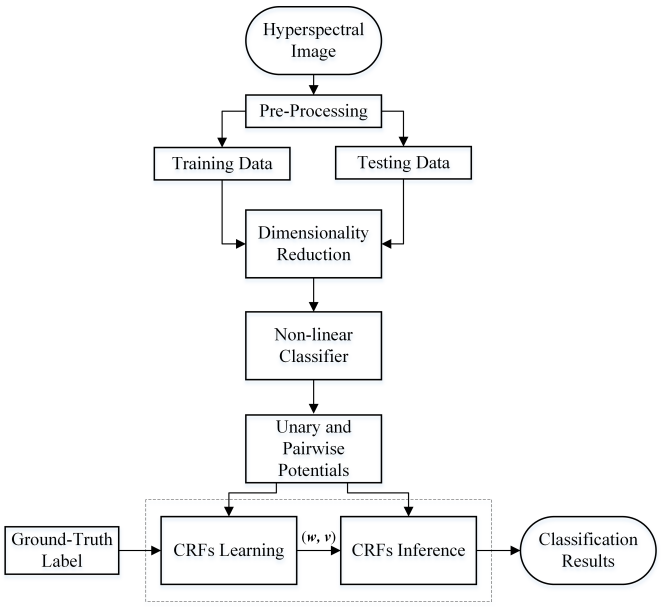


Figure 1: Proposed classification scheme.

2. PROPOSED ALGORITHM

The scheme to perform dimensionality reduction to the CRFs is shown in Fig.1. A raw hyperspectral dataset is first registered to the system, then the pre-processing module (data normalization, bands removal, etc.) follows when necessary. Since the access of hyperspectral image data is extremely limited, a leave-one-out cross validation strategy is employed to provide K folds training and testing datasets. Next, dimensionality reduction techniques are taking place on both training and testing data to generate a reasonable low-dimension representations of the original data. The CRFs are linear models, therefore, a non-linear classier is chosen as a compensation to generate unary and pairwise features which are served as the input to the CRFs. The classification is finally obtained after learning and inference processes.

2.1 Nonlinear Dimensionality Reduction

The ultimate goal of dimensionality reduction is to map the high dimensional data $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_n\} \subset \mathbb{R}^l$ into a lower space $\mathcal{Z} = \{\mathbf{z}_1, \mathbf{z}_2, .., \mathbf{z}_n\}$ in $\mathbb{R}^d$, where $d << l$, and where $\mathcal{Z}$ exhibits a more informative data structure that can be efficiently used in classification. Assuming the original high dimensional data lies on a smooth manifold, the graph based non-linear dimensionality reduction methods have been proven as powerful tools to serve this purpose. Three different approaches are investigated in this paper.

2.1.1 Laplacian Eigenmaps (LE)

LE preserves the local neighboring relationship within the low-dimensional representation after transformation. The major steps are listed as follows:

1. Constructing a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the vertex set and $\mathcal{E}$ denotes the edge set connecting two vertices if they are in proximity to each other. In this application, each pixel in the image is considered a vertex in $\mathcal{V}$ and proximity is defined by either ϵ-neighborhoods or k-nearest neighbors.

2. Determining the $n \times n$ weight matrix with element $W(i,j) = \exp(-\frac{\|\mathbf{x_i} - \mathbf{x_j}\|^2}{\sigma^2})$, if v_i and v_j are connected, 0 otherwise.

3. Building a diagonal matrix $D_{i,i} = \sum_i W(i,j)$ and calculating the matrix $L = D - W$, that is known as the Laplacian matrix. By computing the generalized eigen-decomposition of $L\mathbf{v} = \lambda D\mathbf{v}$, the ordered eigenvalues and their corresponding eigenvectors are: $0 = \lambda_0, \lambda_1, ...\lambda_k$ and $\mathbf{v}_0, \mathbf{v}_1, ...\mathbf{v}_k$. The projected data representation is $\mathbf{z}_i = [\mathbf{v}_1(i), \mathbf{v}_2(i), ..., \mathbf{v}_k(i)]^T$.

2.1.2 Spatial-spectral Schroedinger Eigenmaps

Schroedinger Eigenmaps (SE) extends LE by incorporating a potential matrix $\mathbf{V}$ that encoding expert information about how different graph vertices should be related. The low-dimensional data representation can be obtained by solving the generalized eigenvector problem $(\mathbf{L} + \alpha\mathbf{V})\mathbf{v} = \lambda D\mathbf{v}$, where $\mathbf{V}$ is denoted as the potential matrix and α is the weight governing the impacts of the Laplacian matrix and the potential matrix. Spatial-Spectral Schroedinger Eigenmaps (SSSE)[5] encodes the spatial and spectral information in the potential matrix and its procedure can be briefed as following:

1. Constructing a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the vertices representing each pixel in the image, and $\mathcal{E}$ denotes the edge set connecting two vertices if they are *spectrally* close to each other.

2. Determining the $n \times n$ weight matrix with element $W(i,j) = \exp(-\frac{\|\mathbf{x_i}^f - \mathbf{x_j}^f\|^2}{\sigma_f^2})$, if v_i and v_j are *spectrally* connected, 0 otherwise.

3. Building the potential matrix $\mathbf{V}$ which encodes the *spatial* closeness of two vertices:

$$\mathbf{V} = \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in \mathcal{N}_p(\mathbf{x}_i)} \mathbf{V}^{i,j} \exp(-\frac{\|\mathbf{x}_i^p - \mathbf{x}_j^p\|^2}{\sigma_p^2} - \frac{\|\mathbf{x_i}^f - \mathbf{x_j}^f\|^2}{\sigma_f^2}), \tag{1}$$

   where $\mathcal{N}_p(\mathbf{x}_i)$ denotes the *spatial* neighbors of $\mathbf{x}_i$ and $\mathbf{V}^{i,j}$ is defined as :

$$V_{k,l}^{(i,j)} = \begin{cases} 1, & (k,l) \in (i,i),(j,j) \\ -1, & (k,l) \in (i,j),(j,i) \\ 0, & otherwise \end{cases} . \tag{2}$$

   Then solve the generalized eigen-decomposition of $(\mathbf{L} + \alpha\mathbf{V})\mathbf{v} = \lambda D\mathbf{v}$. Same as in LE, $D_{i,i} = \sum_i W(i,j)$ and Lapacian matrix $L = D - W$. The resulting ordered eigenvalues are $0 = \lambda_0, \lambda_1, ...\lambda_k$ with eigenvectors $\mathbf{v}_0, \mathbf{v}_1, ...\mathbf{v}_k$, the projected data representation is $\mathbf{z}_i = [\mathbf{v}_1(i), \mathbf{v}_2(i), ..., \mathbf{v}_k(i)]^T$.

2.1.3 Local Linear Embedding (LLE)

LLE assumes each point can be linearly reconstructed from its neighbors. The procedure is summarized as follows:

1. Finding the k-nearest neighbors $\mathbf{x}_{i_j}$ of each points $\mathbf{x}_i$.

2. Estimating the weights $W(i,j)$ by minimizing the cost function $\underset{W}{\operatorname{argmin}} \sum_{i=1}^{N} \|\mathbf{x}_i - \sum_{j=1}^{k} W(i,j)\mathbf{x}_{i_j}\|$ with the constrain $\sum_{j=1}^{k} W(i,j) = 1$.

3. Calculating the $\mathbf{z}_i$ by minimizing the cost function $\underset{\mathbf{z}_i}{\operatorname{argmin}} \sum_{i=1}^{N} \|\mathbf{z}_i - \sum_{j=1}^{k} W(i,j)\mathbf{z}_{i_j}\|$ once the weights $W(i,j)$ are obtained,. It can be solved by performing eigen decomposition of $(I - W)^T(I - W)$ and discarding the eigenvector $\mathbf{v}_0$ corresponding to smallest eigenvalue λ_0, the projected data representation is $\mathbf{z}_i = [\mathbf{v}_1(i), \mathbf{v}_2(i), ..., \mathbf{v}_k(i)]^T$.

2.2 Conditional Random Fields

Assume a hyperspectral image $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ with each pixel $\mathbf{x}_i$ can be labeled as $\mathbf{y} = \{y_1, y_2, ..., y_n\}$. The probabilistic model is an efficient approach to solve the problem of assigning a label or tag $\mathbf{y}$ to an observed dataset $\mathbf{x}$. This task can be tackled by selecting the most likely class $\mathbf{y}$ to $\mathbf{x}$ according to the probability distribution:

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|\mathbf{x}). \tag{3}$$

According to Bayes' theorem,

$$p(\mathbf{y}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{y})p(\mathbf{y})}{p(\mathbf{y})} \propto p(\mathbf{x}|\mathbf{y})p(\mathbf{y}), \tag{4}$$

where $p(\mathbf{x}|\mathbf{y})$ and $p(\mathbf{y})$ denote likelihood and prior probability respectively.

Lafferty et al.[6] introduced conditional random fields without calculating the joint probability, which involves estimations for both prior and likelihood terms. Assuming an input data $\mathbf{x} \in \{x_1, x_2, ..., x_n\}$ is obtained, CRFs are capable of computing the probability of an output classification $\mathbf{y} \in \{y_1, y_2, ..., y_n\}$ given the observed $\mathbf{x}$, $p(\mathbf{y}|\mathbf{x})$ directly, which is derived as:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x}, \boldsymbol{\lambda})} \prod_{c \in C} \Psi_c(\mathbf{y}_c, \mathbf{x}, \boldsymbol{\lambda}), \tag{5}$$

where $Z(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{\mathbf{y}'} \prod_{c \in C} \Psi_c(\mathbf{y}'_c, \mathbf{x}, \boldsymbol{\lambda})$ is the partition function which assures the probabilities sum to 1 and Ψ_c is known as potential function. Equation (5) can be rearranged as:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x}, \boldsymbol{\lambda})} \exp(\sum_{i \in S} \phi_i(y_i, \mathbf{x}, \mathbf{w}) + \sum_{i \in S} \sum_{j \in \eta_i} \xi_{ij}(y_i, y_j, \mathbf{x}, \mathbf{v})), \tag{6}$$

where i indicates the site of the data and $j \in \eta_i$ denotes the neighboring sites of i. $\phi(.)$ and $\xi(.)$ are unary and pairwise potentials, respectively, with their corresponding weights $\mathbf{w}$ and $\mathbf{v}$ ($\boldsymbol{\lambda} = (\mathbf{w}, \mathbf{v})$).

A linear CRF model is adopted in our proposed algorithm to define $\phi_i(y_i, \mathbf{x}, \mathbf{w})$ and $\xi_{ij}(y_i, y_j, \mathbf{x}, \mathbf{v})$ as follows:

$$\phi_i(y_i, \mathbf{x}, \mathbf{w}) = \sum_{k=1}^{K} \delta(y_i = k) \mathbf{w}_k \boldsymbol{\varphi}_i(\mathbf{x}), \tag{7}$$

$$\xi_{ij}(y_i, y_j, \mathbf{x}, \mathbf{v}) = \sum_{k,l \in (1,2,...,K)} \delta(y_i = k)\delta(y_j = l)\mathbf{v}_{kl} \boldsymbol{\psi}_{ij}(\mathbf{x}), \tag{8}$$

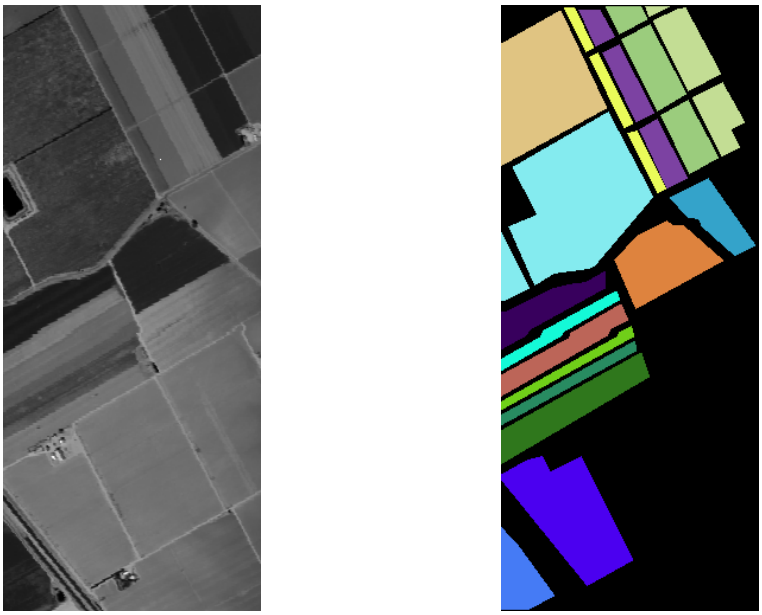
where K is the number of classes. The objective is to train the model to learn the corresponding weights $\mathbf{w}_k$ and $\mathbf{v}_{kl}$, then use these weights to infer the probability of $p(\mathbf{y}|\mathbf{x})$ for the new coming data.

3. EXPERIMENTAL RESULTS

We use the Salinas scene collected by AVIRIS sensor over Salinas Valley to evaluate the performance of the proposed algorithm, as shown in Fig.2. The dataset contains 512 by 217 pixels with 224 bands, 20 of which are discarded due to water absorption. There are 16 classes with a total of 54,129 labeled pixels.

The training and testing samples are arranged in the form of 'Leave-One-Out' 5-fold cross-validation. Training samples cover 75% of the whole data, with the same proportion from each class, and the remaining 25% is reserved for testing. We set $k = 20$ in k-nearest neighbors and $d = 20$ as the reduced dimension for all three dimensionality reduction techniques. We select Support Vector Machine (SVM)[7] with polynomial kernel as the non-linear classifier. The output of SVM s_i provides the unary and pairwise potentials for the CRFs. The unary potential $\boldsymbol{\varphi}_i$ in Eq.7 is constructed with s_i and its spatial location X_{s_i} and Y_{s_i}, the pairwise potential $\boldsymbol{\psi}_{ij}$ in Eq. 8 is generated by $\|s_i - s_{N(i)}\|^2$, where $N(i)$ denotes the 1^{st} order neighbors of i. In CRFs, tree reweighted belief propagation is adopted in the inference module and a marginal-based loss function cooperating with truncated fitting method[8] is utilized in the learning process.

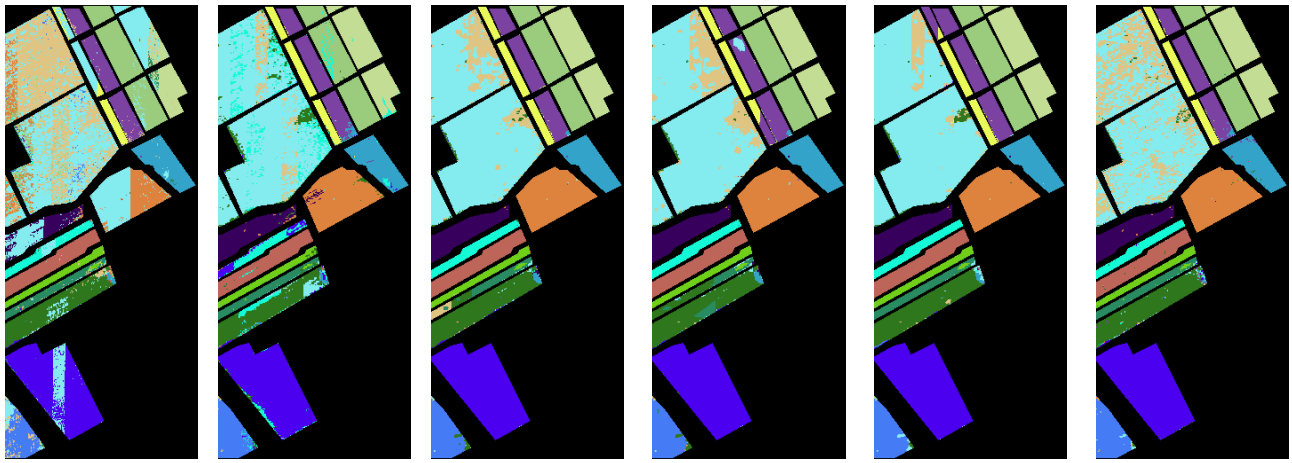
The test was run on a machine with Matlab R2014b having an Intel (R) Xeon (R) CPU processor, 3.50 GHz with 16 GB RAM.



(a) (b)

Figure 2: (a) Example of original Sanilas data in band 120 (b) Classification ground truth with 16 classes.

Table 2 and Fig.3 demonstrate the results with and without dimensionality reduction process and the former outperform the latter both visually and statistically. From (b) - (c), gradually increased values are selected to determine the optimal α in SSSE, in which $\alpha = 10$ provides the best micro-average precision (calculated as the total number of true positive pixels in all classes divided by the number of all pixels) at 86.67%. LE and LLE with SVM generate the micro-average precision at 79.67% and 87.25%, respectively. On the contrary, raw data processed by SVM alone produce a micro-average precision at 71.68% only. Please note, considering the embedding spatial information into the potential matrix, classified regions in the SE results are more homogeneous than in LE and LLE, where misclassifications are scattered in some class regions. The statistical evaluation is shown in Table 3. Comparing to results generated without dimensionality reduction process, better classification outcomes in terms of accuracy, precision and sensitivity are produced with the help of low-dimensional representations. Moreover, the processing time of using SVM for 5-fold cross-validation is significant shortened by using dimension reduced data, 7,040 sec for SSSE for example, in comparison to 111,964 sec without dimensionality reduction.



(a) (b) (c) (d) (e) (f)

Figure 3: SVM results based on (a)raw data without dimensionality reduction, (b) LE, (c) SSSE with $\alpha = 10$, (d) SSSE with $\alpha = 15$, (e) SSSE with $\alpha = 20$, and (f) LLE.

Table 1: Micro-Ave. Precision of SVM classifiers for raw data and using different low-dimensional representations

	Raw Data	LE	SSSE $\alpha = 10$	SSSE $\alpha = 15$	SSSE $\alpha = 20$	LLE
Micro-Ave. Precision	71.68%	79.67%	86.67%	84.61%	86.01%	87.25%

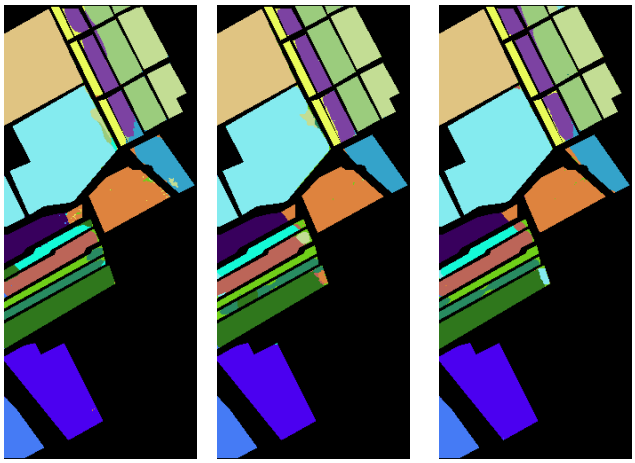
The CRF classification results are shown in Fig.4 with the statistical metric tabulated in Table.4. The classification results are further improved for all three dimensionality reduction techniques. This is due to the fact that CRFs is capable of rectifying the errors and smoothing the scatted regions with spatial relationship encoded in the pairwise potentials (upper left region for example).

Table 2: Performance evaluation for classification using SVM with raw data, LE, SSSE and LLE

		Accuracy				Precision				Sensitivity			
No. of Sample		RAW	LE	SSSE	LLE	RAW	LE	SSSE	LLE	RAW	LE	SSSE	LLE
Class 1	2009	98.07%	99.69%	99.94%	99.96%	48.03%	95.77%	99.80%	99.20%	100.00%	95.86%	98.57%	99.65%
Class 2	3726	93.64%	99.64%	99.96%	99.88%	39.94%	96.30%	99.73%	99.49%	55.27%	98.49%	99.68%	98.83%
Class 3	1976	99.73%	99.42%	99.81%	99.73%	96.41%	95.90%	99.80%	97.62%	96.16%	89.05%	95.17%	95.12%
Class 4	1394	98.61%	99.91%	99.95%	99.95%	46.48%	97.85%	99.64%	99.64%	98.63%	98.63%	98.44%	98.58%
Class 5	2678	99.68%	99.67%	99.88%	99.70%	95.29%	93.91%	97.95%	96.34%	98.15%	99.37%	99.54%	97.47%
Class 6	3959	98.38%	99.56%	99.97%	99.97%	96.08%	94.32%	99.77%	99.77%	84.01%	99.65%	99.87%	99.85%
Class 7	3579	96.59%	99.79%	99.92%	99.94%	48.90%	97.37%	99.53%	99.50%	98.98%	99.51%	99.19%	99.61%
Class 8	11271	75.39%	85.80%	88.29%	88.37%	58.65%	83.29%	92.68%	78.06%	43.29%	61.80%	65.45%	69.73%
Class 9	6203	98.13%	98.72%	99.89%	99.85%	84.48%	96.05%	99.95%	99.77%	99.09%	92.98%	99.10%	98.90%
Class 10	3278	99.01%	96.40%	98.63%	99.03%	84.50%	73.89%	93.26%	92.31%	99.07%	68.87%	85.51%	91.78%
Class 11	1068	99.36%	94.53%	99.98%	99.93%	77.34%	64.70%	99.16%	98.60%	88.53%	21.11%	99.72%	97.77%
Class 12	1927	99.94%	99.84%	99.98%	99.93%	99.79%	98.13%	100.00%	99.17%	98.67%	97.37%	99.48%	98.86%
Class 13	916	99.84%	99.82%	99.87%	99.86%	98.36%	93.67%	98.03%	98.36%	92.70%	95.55%	94.63%	93.66%
Class 14	1070	99.19%	99.75%	99.07%	99.77%	84.67%	89.44%	55.05%	92.34%	76.91%	97.85%	95.93%	96.02%
Class 15	7268	89.08%	87.02%	88.41%	88.80%	71.49%	10.32%	24.96%	48.51%	57.51%	59.67%	68.95%	60.28%
Class 16	1807	98.71%	99.77%	99.79%	99.84%	72.66%	94.24%	94.41%	95.68%	86.67%	98.72%	99.30%	99.37%
MEAN		96.46%	97.46%	98.33%	98.41%	75.19%	85.95%	90.86%	93.40%	85.85%	85.91%	93.66%	93.47%
STD.		6.28%	4.56%	3.91%	3.84%	20.80%	22.25%	20.74%	13.16%	18.25%	22.10%	10.96%	11.48%

4. CONCLUSION

In this paper, we investigated the usage of low-dimensional representation of hyperspectral images coupled with conditional random fields for classification purpose. After probing three different dimensionality reduction methods, the low-representation of hyperspectral image is proven to be efficient in capturing the informative structure of the raw data which can be effectively utilized in the subsequential CRFs. This framework is tested



(a) (b) (c)

Figure 4: Classification results of CRF on (a) LE, (b) SSSE with $\alpha = 10$, (c) LLE.

Table 3: Performance evaluation for classification using CRFs with LE, SSSE and LLE

		Accuracy			Precision			Sensitivity		
No. of Sample		LE	SSSE	LLE	LE	SSSE	LLE	LE	SSSE	LLE
Class 1	2009	99.50%	99.68%	99.62%	87.46%	91.44%	90.00%	98.82%	99.95%	99.67%
Class 2	3726	99.38%	99.45%	99.40%	98.55%	99.73%	99.52%	92.94%	92.85%	92.31%
Class 3	1976	99.43%	99.97%	99.69%	94.84%	99.70%	94.79%	90.14%	99.49%	96.75%
Class 4	1394	99.75%	99.80%	99.86%	99.00%	100.00%	100.00%	91.82%	92.69%	94.77%
Class 5	2678	99.24%	99.77%	99.54%	89.66%	95.44%	94.88%	94.79%	99.92%	95.81%
Class 6	3959	99.31%	99.80%	99.76%	96.11%	99.85%	96.69%	94.56%	97.53%	100.00%
Class 7	3579	99.33%	99.18%	99.92%	97.04%	97.18%	99.13%	93.06%	91.00%	99.69%
Class 8	11271	99.29%	99.39%	99.57%	96.57%	97.41%	99.66%	100.00%	99.66%	98.28%
Class 9	6203	99.99%	99.99%	100.00%	99.89%	99.92%	99.97%	100.00%	100.00%	100.00%
Class 10	3278	99.56%	99.22%	99.67%	99.88%	93.47%	94.63%	93.38%	93.61%	100.00%
Class 11	1068	99.46%	99.77%	99.71%	78.46%	89.14%	85.30%	93.01%	99.27%	99.89%
Class 12	1927	99.90%	99.47%	99.74%	97.09%	88.48%	98.18%	100.00%	96.44%	94.69%
Class 13	916	99.47%	99.39%	99.45%	82.86%	99.78%	90.72%	85.28%	73.59%	79.67%
Class 14	1070	99.65%	99.27%	99.71%	96.82%	71.59%	93.64%	86.84%	89.38%	92.01%
Class 15	7268	99.99%	99.96%	100.00%	99.93%	99.67%	99.97%	100.00%	100.00%	100.00%
Class 16	1807	99.96%	99.99%	99.99%	99.83%	100.00%	99.89%	98.85%	99.61%	99.67%
MEAN		99.57%	99.63%	99.73%	94.62%	95.17%	96.06%	94.59%	95.31%	96.45%
STD.		0.26%	0.29%	0.19%	6.55%	7.47%	4.41%	4.74%	6.85%	5.30%

on the Salina scene with convincing results. The future work includes but is not limited to: a) exploring other means for generating potential matrix in SSSE; b) investigating super-pixel in the CRFs learning ; c) testing the proposed algorithm on other accessible hyperspectral dataset.

ACKNOWLEDGMENTS

The authors would like to thank the U.S Department of Defense for funding this work.

REFERENCES

[1] Friedman, J. H., "On bias, variance, 0/1loss, and the curse-of-dimensionality," *Data mining and knowledge discovery* **1**(1), 55–77 (1997).

[2] Belkin, M. and Niyogi, P., "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation* **15**(6), 1373–1396 (2003).

[3] Czaja, W. and Ehler, M., "Schroedinger eigenmaps for the analysis of biomedical data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **35**, 1274–1280 (May 2013).

[4] Kim, D. H. and Finkel, L. H., "Hyperspectral image processing using locally linear embedding," in [*Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on*], 316–319, IEEE (2003).

[5] Cahill, N. D., Czaja, W., and Messinger, D. W., "Schroedinger eigenmaps with nondiagonal potentials for spatial-spectral clustering of hyperspectral imagery," *Proc. SPIE* **9088**, 908804–1–13 (2014).

[6] Lafferty, J. D., McCallum, A., and Pereira, F. C. N., "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in [*Proceedings of the Eighteenth International Conference on Machine Learning*], *ICML '01*, 282–289, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2001).

[7] Cristianini, N. and Shawe-Taylor, J., [*An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*], Cambridge University Press, New York, New York, USA (2000).

[8] Domke, J., "Learning graphical model parameters with approximate marginal inference," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(10), 2454–2467 (2013).