

Old Dominion University

## ODU Digital Commons

---

Electrical & Computer Engineering Faculty  
Publications

Electrical & Computer Engineering

---

2013

# Hyperspectral Image Classification Using a Spectral-Spatial Sparse Coding Model

Ender Oguslu  
*Old Dominion University*

Guoqing Zhou  
*Guilin University of Technology*

Jiang Li  
*Old Dominion University, jli@odu.edu*

Lorenzo Bruzzone (Ed.)

Follow this and additional works at: [https://digitalcommons.odu.edu/ece\\_fac\\_pubs](https://digitalcommons.odu.edu/ece_fac_pubs)



Part of the [Data Science Commons](#), [Electrical and Computer Engineering Commons](#), [Remote Sensing Commons](#), and the [Theory and Algorithms Commons](#)

---

### Original Publication Citation

Oguslu, E., Zhou, G., & Li, J. (2013) Hyperspectral image classification using a spectral-spatial sparse coding model. In L. Bruzzone (Ed.), *Image and Signal Processing for Remote Sensing XIX, Proceedings of SPIE Vol. 8892* (88920R). SPIE of Bellingham, WA. <https://doi.org/10.1117/12.2030261>

This Conference Paper is brought to you for free and open access by the Electrical & Computer Engineering at ODU Digital Commons. It has been accepted for inclusion in Electrical & Computer Engineering Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact [digitalcommons@odu.edu](mailto:digitalcommons@odu.edu).

# Hyperspectral image classification using a spectral-spatial sparse coding model

<sup>a,b</sup>Ender Oguslu, <sup>c\*</sup>Guoqing Zhou and <sup>a</sup>Jiang Li

<sup>a</sup>Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA 23529

<sup>b</sup>Air Force NCO Vocational Collage, Gaziemir, Izmir, Turkey

<sup>c</sup>Guangxi Key Laboratory for Spatial Information and Geomatics, Guilin University of Technology, Guilin, P. R. China 541004

## ABSTRACT

We present a sparse coding based spectral-spatial classification model for hyperspectral image (HSI) datasets. The proposed method consists of an efficient sparse coding method in which the  $l_1/l_q$  regularized multi-class logistic regression technique was utilized to achieve a compact representation of hyperspectral image pixels for land cover classification. We applied the proposed algorithm to a HSI dataset collected at the Kennedy Space Center and compared our algorithm to a recently proposed method, Gaussian process maximum likelihood (GP-ML) classifier. Experimental results show that the proposed method can achieve significantly better performances than the GP-ML classifier when training data is limited with a compact pixel representation, leading to more efficient HSI classification systems.

**Keywords:** Remote sensing, sparse coding, feature selection.

## 1. INTRODUCTION

Remote sensing images are widely used for terrain property determination. There are three popular remote sensing image types including panchromatic image, multi-spectral image and hyperspectral image (HSI). A panchromatic image consists of only one band and is usually displayed as a grey scale image and the radiometric information is the main information utilized for data analysis. Multi-spectral images record over several separate wavelength ranges at various spectral resolutions. It is cost efficient and is currently the most common type being used. Hyperspectral images contains ground reflectance measurements across hundreds of very narrow spectral bands throughout the visible, near-infrared and mid-infrared portions of the electromagnetic spectrum. HSI yields fine discrimination between different targets based on their spectral response in each of the narrow bands, making it a useful modality for fine differentiation between ground objects. It also has the potential for more accurate and detailed information extraction than any other types of remote sensing data for determining material classification, geological feature identification and environmental monitoring [1]. In HSI dataset, each pixel consists of a high-dimensional vector with each component in the vector as the ground reflectance in a specific band. Many techniques have been proposed for HSI classification including support vector machines (SVMs), neural networks and graph-based methods [19, 20]. In addition, there is useful spatial information reflected by the neighborhood structure in HSI data that could be potentially help HSI classification. The spatial information has been utilized for the classification such as the stacking vector approach [6], segmentation based techniques [7], majority filtering [8] and the Markov random field method (MRF) [9-10].

Sparse coding is a relatively new method recently developed for image classification. In sparse coding, a set of over-complete basis functions called "dictionary" is learned from training data sets. Each data point is then projected onto the dictionary to obtain a new representation. After that, learning and classification will be performed on the new representations. Recent advances in sparse coding achieved state-of-the-art performances in many applications including pixel-based HSI classification [4-5] and image classification [11-13]. In our recent study [14], we implemented a sparse coding method for HSI image classification and we achieved better results than the supervised local linear embedding-weighted  $k$ NN (SLLS-W $k$ NN) algorithm [3]. In this paper, we further improve our method by reducing the dimensionality of the new representation in the sparse coding framework and by utilizing spatial information in HSI data to boost the classification accuracy. We applied the proposed algorithm to a HSI dataset collected at the Kennedy Space Center (KSC) and compared our algorithm to the Gaussian process maximum likelihood (GP-ML) classifier [2]. Experimental results show that the proposed method can achieve better classification accuracies with a compact pixel representation, leading to more efficient HSI classification systems.

---

\*Corresponding author: gzhou@glut.edu.cn; phone: 1-86 -773 5896073. This project was funded by Guangxi Key Laboratory of Spatial Information and Geomatics under the grant number of 1207115-15.

## 2. PROPOSED METHOD

Figure 1 shows the flowchart of the proposed method for HSI image classification. The proposed method consists of four steps: (i) patch extraction, (ii) dictionary learning and reduction (iii) feature encoding and pooling and (iv) feature selection and classification. Each step is detailed in the following subsections.

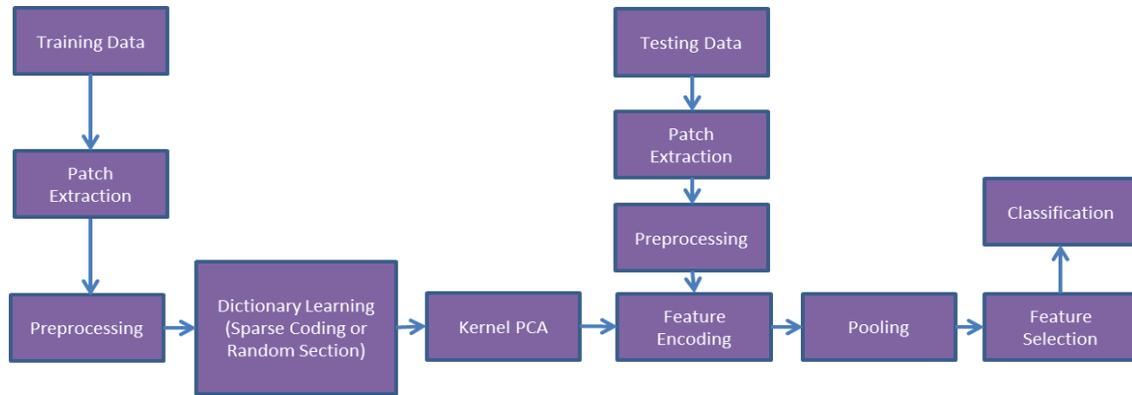


Figure 1. Flowchart of the proposed algorithm.

### 2.1 Patch extraction from HSI data

Existing sparse coding methods usually extract signal pixel segments along the spectral direction only as we did in our previous work [14]. The disadvantage of this simple method is that the spatial information in the HSI data set is not used that may help the classification task. Due to the point spread function (PSF) of the imaging device, material information in one pixel is spread over its neighbors, processing those pixels separately will degrade the discriminate capability of HSI data. In this paper, we first computed an average value of the eight neighbors of each pixel in each spectral band, resulting in another  $B$  values for each pixel, where  $B$  is the number of bands in HSI. These  $B$  values were then augmented to the original  $B$  values of the pixel, making the pixel represented by a  $2 \times B$ -dimensional vector.

After that, the patch extraction is the same as we did in our previous work except that the operation is operated on the  $2 \times B$ -dimensional pixel vector. In particular, we randomly selected patches with a dimension of  $b$  (each patch contains  $b$  bands) along the spectral direction as shown in Figure 2. Note that the extraction process is random and those extracted patches are possibly overlapped or some bands might not be picked. The parameter “ $b$ ” is often called “receptive field length”. If  $m$  is the total number of sampled patches for dictionary learning, we denote it as  $\mathbf{X} = \{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ , where  $x^{(i)} \in \mathbb{R}^b$ . Before the dictionary learning step described in the next section, each patch  $x^{(i)}$  was normalized to be zero mean and unit variance.  $\mathbf{X}$  was then whitened by the zero-phase components analysis (ZCA) [22]. It was shown in [23] that this process is critical for the quality of the learned feature representations and was utilized in this study.

### 2.2 Dictionary learning based on the constructed patches and dimension reduction

There are usually two different popular dictionary learning methods. One is called sparse coding in which the dictionary is learned based on the  $L_1$  norm optimization framework [15]. The second method is simple and straight-forward in which the dictionary is randomly picked from the extracted patches. The former is computationally very expensive and it usually does not provide extra benefits over the random selection method. We compared those two methods in our previous study and showed that a randomly selected dictionary is good enough for HSI classification [14]. In many applications such as image classification, a couple of thousands basis functions can lead to a good performance [12]. However, the final feature representation for an image is usually in a very high dimensionality of tens of thousands, requiring intensive computational resources. A recently study by Gkioulekas et al. showed that basis functions learned by the sparse coding framework are usually highly correlated and can be compressed by PCA or Kernel PCA without performance degradation [13]. In this paper, we utilized the Kernel PCA to reduce the size of the randomly picked dictionary, yielding a dictionary  $\mathbf{D}$  for the HSI data.

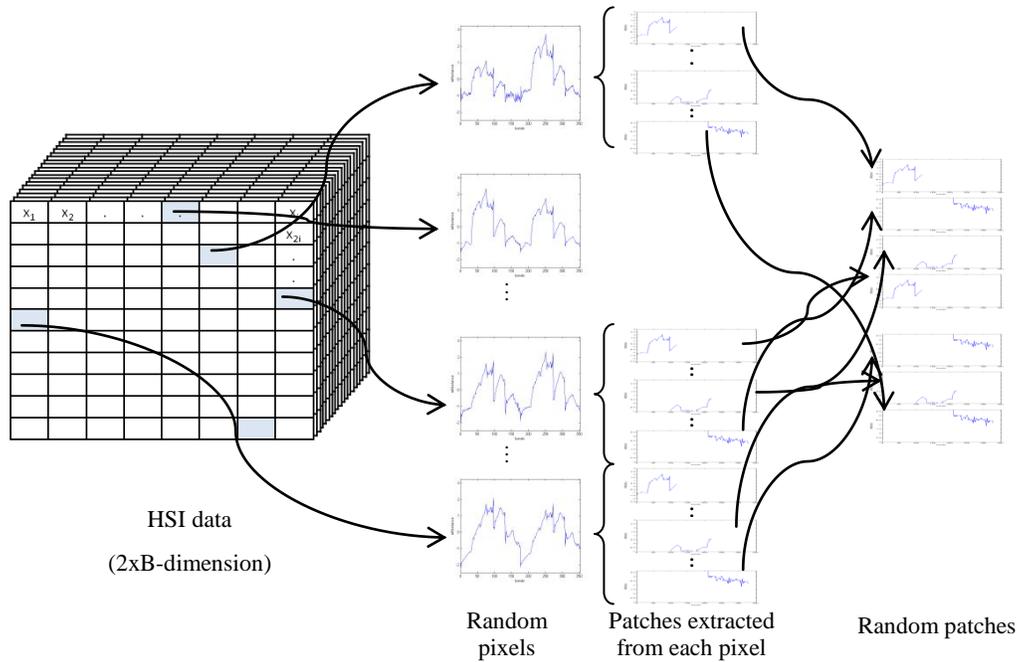


Figure 2. Extraction of patch blocks [14].

### 2.3 Encoding and pooling

With the dictionary,  $\mathbf{D}$ , we obtained a new representation for an HSI pixel as follows,

1. *Patch extraction.* We used a window of size  $b$  bands to divide the pixel to patches. The window was moved along the spectral direction with a step size of '1', resulting in  $2B-b+1$  patches for the pixel.
2. *Preprocessing.* These patches were preprocessed to make them zero mean, unit variance and whitened as described in the previous section.
3. *Encoding.* For each patch, we obtained a new representation by applying the soft thresholding technique as equation (1). Note that the dimension of the representation for the patch is  $2n$ .

$$\mathbf{a}^{(i)} = \text{sign}(z^{(i)}) \max(0, |z^{(i)}| - t) \quad (1)$$

where  $t$  is an adjustable threshold and  $z^{(i)} = \mathbf{D}^T \cdot x^{(i)}$ .

4. *Pooling.* We split all representations from the  $2B-b+1$  patches of the pixel into  $k$  equal-sized groups and sum them in each group to obtain  $k$  representations. The final representation for the HSI pixel is achieved by concatenating the  $k$  representations. The dimension of the final representation is  $2kn$ , where  $n$  is the number of basis function in the dictionary.

### 2.4 Feature selection and classification

After the encoding and pooling steps described above, the dimensionality of the new feature representation is  $2kn$ . In our application, the dimensionality of the representation is still at the order of hundreds that may be undesirable for real-time systems. We utilized the  $l_1/l_q$  regularized multi-class logistic regression [16-17] as shown in equation (2) to select effective features for HSI classification,

$$\min_x \sum_{l=1}^k \sum_{i=1}^m b_{il} \log(1 + \exp(-y_{il}(w_l^T a_{il} + c_l))) + \lambda \|w\|_{l_1/l_q} \quad (2)$$

where  $a_{il}^T$  is the  $i$ -th feature for the  $l$ -th class,  $b_{il}$  is the weight for  $a_{il}^T$ ,  $y_{il}$  is the response of  $a_{il}$ ,  $c_l$  is the intercept (scalar) for the  $l$ -th class and  $\lambda$  is a regularization parameter. In this multiclass problem,  $a_{il}$ ,  $\forall l$ ,  $A \in \mathbf{R}^{m \times n}$ ,

$x \in R^{n \times k}$ ,  $c \in R^{1 \times k}$ ,  $y \in R^{m \times k}$ . The  $l_1/l_q$  regularized regression optimization is a recently developed method that favors the group sparsity in the model [16-17]. Once we obtain the final representation vector for each pixel, we apply a linear support vector machine (SVM) classifier [21] to classify the HSI data to different land cover categories. The regularization parameters of the logistic regression and SVM classifier are determined by cross-validation.

### 3. DATA PREPARATION AND PERFORMANCE EVALUATION

#### 3.1 Data description

In this paper, we used the hyperspectral data set collected by NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over Kennedy Space Center in March 1996 [18] to evaluate our proposed method. This data set contains 224 bands with an 18-m spatial resolution and a 10-nm spectral resolution over the range of 400-2500 nm. After removing noisy and corrupted bands, there are 176 bands that can be used for classification. Among the available 314,368 (512x614) pixels, 5211 pixels were labeled to 13 classes by the KSC staff as summarized in Table 1. In this data, the spectral signatures of some of the classes are very similar making those classes difficult to discriminate [2]. Figure 4 shows one band from this data.



Figure 4. A sample band of Kennedy Space Center HSI data.

Table 1. Class names and number of labeled KSC data.

Class	Class	Number/Percentage of
1	Scrub	761 (14.6%)
2	Willow swamp	243 (4.66%)
3	Cabbage palm hummock	256 (4.92%)
4	Cabbage/oak hummock	252 (4.84%)
5	Slash pine	161 (3.07%)
6	Oak/broadleaf hummock	229 (4.38%)
7	Hardwood swamp	105 (2.0%)
8	Graminoid marsh	431 (8.27%)
9	Spartina marsh	520 (9.9%)
10	Cattail marsh	404 (7.76%)
11	Salt marsh	419 (8.04%)
12	Mud flats	503 (9.66%)
13	Water	927 (17.8%)
	TOTAL	5211

#### 3.2 Experiment setup and performance evaluation

In the experiment, we utilized the same training-testing configuration as that in [2] for a fair comparison. First, we divided the entire dataset to four equally sized subsets. The experiments were performed based on four-fold cross-validation. Three parts were used for training and the remaining part was used for testing. This process was repeated

four times such that each subset was used for testing once. For each fold, we also tested the system performance by using different number of training data samples. The training data was subsampled at the rate of 20%, 50%, 75% and 100%. A sampling rate of 20% means that 20% of the training data (or 15% of the entire dataset,  $20\% \times 75\% = 15\%$ ) was used for training. There were four parameters in the proposed method: (i) number of basis functions,  $n$ , (ii) group number in pooling,  $k$ , (iii) length of each patch called receptive field length,  $b$ , and (iv) threshold for encoding,  $t$ , in the coding step. In this study, we performed experiments many times and optimized those parameters. With those optimized parameters, we conducted our final experiment and compared our results with those obtained by a recent algorithm, Gaussian process maximum likelihood, for classifying this hyperspectral data [2].

## 4. RESULTS

### 4.1 Classification results

Table 2 shows the classification results using different number of training data points (without the feature selection step), where the threshold ' $t$ ' is set as 0.1. It is observed that the classification accuracy increases with the number of data points being used in training.

Table 2. Classification results using different no. of training samples without feature selection

The size of the dictionary ( $n$ )	Group number in pooling ( $k$ )	The size of the final representation of the pixels ( $2kn$ )	20% sampling rate		50% sampling rate		75% sampling rate		100% sampling rate	
			Sparse coding	Random selection	Sparse coding	Random selection	Sparse coding	Random selection	Sparse coding	Random selection
2	50	200	94.12	95.04	96.75	97.21	96.93	97.68	96.95	97.35
5	35	350	95.28	95.56	97.82	97.89	98.15	98.43	98.22	97.79
10	30	600	95.80	95.70	98.02	98.08	98.54	98.57	98.45	98.47
15	25	750	95.92	95.83	98.19	98.15	98.73	98.70	98.56	98.53
20	20	800	95.85	95.83	98.24	98.04	98.65	98.73	98.70	98.57
50	15	1500	95.32	95.25	98.31	98.10	98.54	97.68	98.45	98.23
100	12	2400	94.81	94.92	98.01	97.83	98.33	98.43	98.44	98.45

In addition, we compared our results with those obtained by a recently proposed method for HSI classification, Gaussian process maximum likelihood (GP-ML) for spatially adaptive classification of hyperspectral data [2], in terms of overall accuracy as shown in Table 3. In GP-ML model, first, each band of a given class is modeled by a Gaussian random process which indexed by spatial coordinates. Then, each land cover class at a given location is characterized by a multivariate Gaussian distribution with specific parameters adapted for that location. In the comparison, the parameters of the proposed algorithm were set  $n=20$  (reduced from 500 basis functions),  $k=20$ ,  $b=120$  and  $t=0.1$ . Before the feature selection step, there were 800 features and the feature selection algorithm kept 300 of them. It is clear that the proposed algorithm with both learning methods, random dictionary and sparse coding, has significantly better performance than the GP-ML method on limited number of training samples and competitive performance on sufficient number of training samples.

Table 3. Comparison of proposed algorithm and GP-ML.

Classifier	Classification Accuracy (%)			
	20%	50%	75%	100%
Linear SVM using sparse coding for dictionary learning	95.85	98.24	98.65	98.70
Linear SVM with random dictionary	95.83	98.04	98.73	98.57
GP-ML	91.78	97.89	98.89	98.87

## 5. CONCLUSION

We proposed a sparse coding framework for HSI classification by considering spatial-spectral structures of HSI pixels in this paper. The proposed algorithm benefited from the sparse coding framework, the spatial-spectral structures of HIS and feature selection to build a robust, compact classifier. We applied the feature selection algorithm twice in the proposed method. We first used it to reduce the number of basis function in the learned dictionary and then the feature

selection technique was reutilized to reduce the dimensionality of the final feature representation. Furthermore, we showed that randomly selected dictionaries can achieve good results making an efficient HSI data classification system possible.

## REFERENCES

- [1] C. Chang "Hyperspectral Data Exploitation: Theory and Applications", Wiley, New York, 2007.
- [2] G. Jun and J. Ghosh, "Spatially adaptive classification of land cover with remote sensing data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 7, pp. 2662-2673, 2011.
- [3] L. Ma, M. M. Crawford and J. Tian, "Local manifold learning-based k-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.* Vol. 48, no. 11, pp. 4099-4109, 2010.
- [4] A. S. Charles, B. A. Olshausen and C. J. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.*, vol. 5, no. 5, pp. 963-978, 2011.
- [5] A. Castrodad, Z. Xing, J. B. Greer, E. Bosch, L. Carin and G. Sapiro, "Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.* vol. 49, no. 11, pp. 4263-4281, 2011.
- [6] R. M. Haralick and K. D. L. Shanmugam, "Textural features for image classification," *IEEE Trans. On SMC*, vol. 3, no. 6, pp. 610-621, 1973.
- [7] L. O. Jimnez, J. L. Rivera-Medina, E. Rodriguez-Daz, E. Arzuaga-Cruz and M. Ramirez-Vlez, "Integration of spatial and spectral information by means of unsupervised extraction and classification for homogenous objects applied to multispectral and hyperspectral data," *IEEE Trans. Geosci. and Remote Sens.*, vol. 43, no. 4, pp. 844-851, 2005.
- [8] W. A. Davis and F. G. Peet, "A method of smoothing digital thematic maps," *Remote Sensing of Environment*, vol. 6, pp. 45-49, 1977.
- [9] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 6, pp. 721-741, 1982.
- [10] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using gibbs random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 1, pp. 39-55, 1987.
- [11] J. Yang, K. Yu, Y. Gong and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1794-1801, 2009.
- [12] A. Coates, A. Y. Ng and S. Mall, "The Importance of Encoding Versus Training with Sparse Coding and Vector Quantization," *Proceedings of the 28th International Conference on Machine Learning (ICML)*, 2011.
- [13] I. Gkioulekas and T. Zickler, "Dimensionality Reduction Using the Sparse Linear Model," *Twenty-fifth Annual Conference on Neural Information Processing Systems (NIPS)*, 2011.
- [14] E. Oguslu, K. Iftekharuddin and J. Li, "Sparse coding for hyperspectral images using random dictionary and soft thresholding", *Proc. SPIE 8399, 83990A*, 2012.
- [15] T. T. Wu and K. Lange, "Coordinate descent algorithms for lasso penalized regression," *Annals of Applied Statistics*, vol. 2, no. 1, 2008.
- [16] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411-430, 2000.
- [17] A. Coates, L. Honlak and A. Y. Ng, "An analysis of single-layer networks in unsupervised feature learning," In *International Conference on AI and Statistics*, 2011.
- [18] J. Liu, S. Ji and J. Ye, "Multi-task feature learning via efficient L2,1-norm minimization", *UAI*, 2009.
- [19] J. Liu, L. Yuan, S. Chen and J. Ye, "Multi-task feature learning via efficient L2,1-norm minimization", *Technical Report ASU*, 2009.
- [20] J. Ham, C. Yangchi, M. M. Crawford and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *Geoscience and Remote Sensing IEEE Transactions on*, vol. 43, no. 3, pp. 492-501, 2005.
- [21] F. Palsson, J. R. Sveinsson, J. A. Benediktsson and H. Aanaes, "Classification of Pansharpened Urban Satellite Images," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 1, pp. 281-297, 2012.
- [22] G. Camps-Valls, T. Bandos Marshava and D. Zhou, "Semi-Supervised Graph-Based Hyperspectral Image Classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 10, pp. 3044-3054, 2007.
- [23] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang and C. J. Lin, "LIBLINEAR: A Library for Large Linear Classification," *Journal of Machine Learning Research* 9, pp. 1871-1874, 2008.