# Knowledge Based Stacking of Hyperspectral Data for Land Cover Classification

Yangchi Chen
Center for Space Research
3925 W. Braker Lane,
Austin, TX 78759
The University of Texas at Austin
Email: yanji@csr.utexas.edu

Melba M. Crawford
Laboratory for Applications of Remote Sensing
Lilly Hall of Life Sciences
Purdue University
Email: mcrawford@purdue.edu

Joydeep Ghosh
Department of Electrical
and Computer Engineering
The University of Texas at Austin
Email: ghosh@ece.utexas.edu

*Abstract*— **Hyperspectral data provide new capability for discriminating spectrally similar classes, but unfortunately such class signatures often overlap in multiple narrow bands. Thus, it is useful to incorporate reliable spatial information when possible. However, this can result in increased dimensionality of the feature vector, which is already large for hyperspectral data. Markov random field (MRF) approaches, such as iterated conditional modes (ICM), can provide evidence relative to the class of a neighbor through Gibbs' distribution, but suffer from computational requirements and curse of dimensionality issues when applied to hyperspectral data. In this paper, a new knowledge based stacking approach is presented to utilize spatial information within homogeneous regions and at class boundaries, while avoiding the curse of dimensionality. The approach learns the location of the class boundary and combines original bands with the extracted spectral information of a neighborhood to train a hierarchical support vector machine (HSVM) classifier. The new method is applied to hyperspectral data collected by the Hyperion sensor on the EO-1 satellite over the Okavango delta of Botswana. Classification accuracies are compared to those obtained by a pixel-wise HSVM classifier, majority filtering and ICM to demonstrate the advantage of the knowledge based stacking approach.**

## I. INTRODUCTION

With the launch of the Hyperion sensor on the NASA Earth Observation 1 (EO-1) satellite, researchers in the remote sensing community are now able to exploit data collected by hyperspectral sensors over extended areas and in multiple time periods at minimal cost.

The Hyperion sensor has the advantage of representing spectral signatures in much greater detail than traditional multispectral spaceborne sensors (which only have about 3-15 bands), and thus has much greater potential for providing improved characterization and discrimination of targets. Although Hyperion is able to collect hundreds of bands simultaneously, calibration is difficult because it is a pushbroom sensor, and the signal-to-noise ratio is low for certain wavelengths, resulting in "striped" columns in many bands. While normalization of statistics in local windows and application of low pass filters can mitigate the effect, these approaches are often inadequate and can even induce artificial effects in the data. Spatial neighborhood information, which is often more reliable but difficult to analyze, provides an alternative source of information which should be utilized in conjunction with spectral data to label the class pixels.

### A. Related Work

Although many algorithms have been developed to utilize spatial information, most methods are designed to reduce the impact of outliers in homogeneous areas or model texture patterns, and can only be used with low dimensional multispectral data. Most studies in land cover classification that use hyperspectral data are pixel-based. Because of the medium (30 m) spatial resolution of Hyperion images, pixels on the class boundaries often belong to multiple classes and pose a potential problem for these algorithms. To better distinguish these mixed pixels, contextual information should be incorporated into the classification process.

Previous studies that include spatial context have been members of four general categories of approaches. The first is a stacking vector approach [1], whereby the original or averaged bands or Fourier transform of neighboring features are concatenated with the original spectral vectors. While these approaches provide insight into the neighborhood, they are handicapped by the insufficient number of labeled samples and relatively high dimensional inputs. A stacking vector approach also has difficulty in classifying pixels at the image boundaries because of large changes in spectral signatures between a targeted pixel and its neighbors.

The second set of studies is based on image segmentation whereby images are divided into many homogeneous segments according to their spatial-spectral closeness [2]. Classification is performed by comparing the similarities of labeled samples to the means of each segment. These studies uncover segments that are spatially and spectrally homogeneous. Unfortunately, the classification accuracies of these algorithms are very sensitive to the initial segmentation settings. For example, the number of segments and the specifics of these segments are critical to achieving a good classified map.

A simple, commonly used approach performs majority filtering [3] after the image is first classified by a pixel-wise classifier. The majority filtering process assigns a pixel's label according to its 1st-order or 2nd-order neighbors. If the local neighborhood is dominated by one class, the label of the

targeted pixel is changed to reflect the majority. This process removes outliers in homogeneous areas, but the resulting classified maps are often blocky and do not properly identify class boundaries.

In recent years, Markov random fields (MRF) [4], [5] have been widely used for incorporating spatial-spectral information in the classification process. The general assumption of MRF is that $\pi(c)$, the prior probability of each class $c$, can be modeled as a discrete MRF.

$$\pi(c|c_i, \forall i \in I) = P(c|c_s, \forall s) \tag{1}$$

$I$ is the whole image, $S$ is the local neighborhood. $s \in S$ are pixels in the neighborhood. The isotropic behavior and the local dependencies make MRF an ideal algorithm for learning contextual information from $S$. According to the Hammersley-Clifford theorem, $\pi(c)$ is equivalent to a Gibbs distribution such that

$$\pi(c) = 1/Z \ \exp(-\sum V_s(c)) \tag{2}$$

$Z$ is a user defined constant and $V_s$ is the energy function of the Gibbs distribution for $\forall s \in S$. Setting the $V_s$ function is an important issue for MRF estimation.

To determine the unknown class label $c$ of each pixel of the image, spatial information is utilized by a Bayesian estimator; the maximum a posteriori (MAP) classifier selects the optimum $\hat{c}$, given by:

$$\hat{c} = \arg\min_c \{-\log P(X|c) + V_s(c)\} \tag{3}$$

where $X$ is the input space, and $P(X|c)$ is the conditional probability. This optimization problem is a non-convex nonlinear problem and can be solved by different heuristic approaches. The most common algorithm uses simulated annealing (SA) to estimate the best Gibbs distribution.

$$\hat{c} = \arg\min_c 1/Z \ \exp\left\{-\frac{1}{T}U(x)\right\} \tag{4}$$

with

$$U(x) = \sum_{s \in S} \log P(X|c) + V_s(c)$$

The method converges to its optimal solution if the temperature $T$ is slowly lowered to zero. Previous studies showed that $c$ will converge to $\hat{c}$ almost surely, but its convergence rate is very low.

Iterated conditional modes (ICM) [6] is the most widely used approach to determine the MRF parameters. ICM starts with a classified image and recursively reduces the total energy until it converges to a local minimum. The goodness of the final classified image is highly dependent on a good initial classified image. Both MRF using the SA heuristic and ICM have potential problems when applied to hyperspectral data. Because of the curse of dimensionality, the maximum a posteriori (MAP) classifier is often impacted by small training data sets, which result in near-singular covariance matrices. To overcome this problem, Jackson and Landgrebe proposed an adaptive Bayesian classifier [7] that uses a semi-supervised

approach to increase the number of labeled samples. The resultant covariance matrix is more stable, but excessive computational time is required to obtain the MAP solution

Camps-Valls *et. al.* [8] proposed an algorithm that learns kernel functions of spatial and spectral similarities of hyperspectral data separately. It then combines the two kernel functions to form a kernel machine that satisfies Mercer's conditions. Because of the variety of spatial textures, it is difficult to fit all scenarios in one spatial kernel. The results are obtained by experimenting with different combinations of kernels. Thus, the tuning process is typically time consuming.

In this paper, we take the view that data from the homogeneous areas and data from mixed areas should be treated differently. We propose a max-cut algorithm that extracts useful spatial information from the second order neighborhood by discovering the natural boundaries between classes on the image. An integrated, supervised classifier that merges selected spatial and spectral information to train a hierarchical support vector machine classifier [9] is described. The new algorithm is applied to hyperspectral data collected by the Hyperion sensor on the EO-1 satellite over the Okavango delta of Botswana. Classification accuracies and the resultant classified image are compared to those achieved by a pixel-wise classifier, majority filtering and MRF model using.

## II. METHODOLOGY

Most studies reviewed in the previous section tackle this relatively complex problem using a single approach applied uniformly throughout the image. In contrast, our experiments show improvements in classification accuracies and reduction in processing time by breaking the problem into two smaller problems, which are each easier to solve. Initial testing shows that stacking average bands of a pixel's neighbors into the vector of original bands improves the classification accuracy of samples of some classes (trees and grassland), but decreases the accuracy of samples from others (water-related areas). This is due to the complexity of the local neighborhood. We observed experimentally that samples of wetland classes tend to have more complex neighbors while bushes and flat areas tend to be more homogeneous. This indicates that we could potentially benefit by distinguishing these two types of neighborhoods, then providing associated input data that support both spectral and spatial information in a more logical way than simple vector stacking.

The significant difference in the spatial characteristics of homogeneous and mixed neighborhoods indicates that these two types of data should be treated differently in classification. In order to handle these two types of data in a straightforward, computationally efficient way, we develop a pre-processing approach that determines the dissimilarity among pixels and separates mixed neighborhoods into similar subsets. This requires an algorithm that is capable of recognizing the boundaries between different class labels in the user defined neighborhood. We accomplish this by defining the dissimilarity between pixels according to their Kullback-Leibler (KL) divergence [10]. Each neighborhood is separated

into more homogeneous subsets by maximizing the overall dissimilarity between these subsets; the problem is solved by graph optimization.

In the next section, we introduce the max-cut optimization used as a preprocessing algorithm to explore the homogeneity of the neighborhood. We then describe knowledge based stacking, and then briefly explain hierarchical support vector machines (HSVM).

### A. Max-Cut Problem

The max-cut problem is a nonlinear optimization problem whereby an undirected graph with nonnegative edge weights is partitioned into two groups such that the sum of the cut edges between these two groups have the maximum weight [11]. Define an undirected graph $G = (N, E)$ where $N$ represents nodes, $E$ represents edges of the graph, and $w_{ij} \geq 0$ represents the weight of an edge linking nodes $i$ and $j$. The objective is to find the best binary partition that has the cut $\delta(K^*)$ that $K^* \subseteq N$ and $\{ij \in E : i \in K^*, j \notin K^*\}$ that has the maximum weight:

$$w(\delta(K^*)) = \sum_{ij \in \delta(K^*)} w_{ij}. \tag{5}$$

The graph is assumed to be complete by setting $w_{ij} = 0$ for all non-edges $ij$.

The max-cut problem can be represented using an integer quadratic programming formulation with decision variables $X$: $x_i \in \{1, -1\} \ \forall i \in N$. For a cut $\delta(K)$; $x_i = 1 \iff i \in K$. If $ij \in \delta(K)$, $x_i x_j = -1$. Thus:

$$w(\delta(K)) = \frac{1}{2} \sum_{i<j} w_{ij} (1 - x_i x_j) \tag{6}$$

and the resulting max-cut integer quadratic problem is:

$$\begin{aligned} \max \quad & w(\delta(K)) \\ s.t. \quad & x_i \in \{+1, -1\}, i \in N \end{aligned} \tag{7}$$

The original max-cut problem can be relaxed to a constrained quadratic problem and solved using semi-definite programming [12]. An extension of the interior point method [13] provides a computationally efficient method for solving the semi-definite problem.

This binary split algorithm was tested locally in several areas of the image that are either homogeneous (Fig. 1) or mixed (Fig. 2 and 3). These examples show that max-cut can successfully detect the class boundaries by maximizing the total distance between subsets in the defined second order neighborhood. These images also show the complexity of local neighborhoods.

### B. Knowledge Based Stacking

Vector stacking improves the classification accuracy of homogeneous areas, but reduces the accuracy of the mixed areas. Thus, the goal of knowledge-based stacking is to obtain the right subset of information to support the classifier. We seek to utilize the class boundaries identified by the max-cut optimization to design a logical approach to incorporate
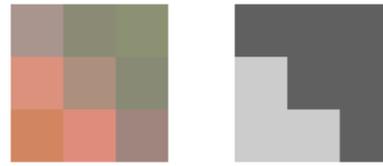


Fig. 1.   Island Interior: (Left) Original Image, (Right) Max-cut Result



Fig. 2.   Firescar: (Left) Original Image, (Right) Max-cut Result

spectral signatures of a second order neighborhood in classifier training. The logic is clarified by these two scenarios:

- If a pixel is located in a homogeneous area, the additional information that should be provided to a classifier should be the average of the bands of its neighbors, thereby leveraging spatial smoothing.
- If a pixel's spectral signature is different from its neighbors', using the average bands of all neighbors reduces the classification accuracy. Additional information that should be used in this case are the original bands of the pixel or the averaged bands of neighboring pixels that are similar to the targeted pixel.

In either scenario, we double the number of bands/features. At this preprocessing stage, we add data that increases the likelihood of a pixel belonging to the correct class by utilizing bands of neighbors that are similar to the target pixel. This approach corrects pixel values of a class that are in the tail of its distribution and tends to classify correctly near boundaries, which potentially contain mixed pixels. Max-cut optimization, which is capable of splitting neighboring pixels into more homogeneous subsets, is applied to the whole image to find suitable neighboring pixels for knowledge-based stacking.



Fig. 3.   Primary Floodplain: (Left) Original Image, (Right) Max-cut Result
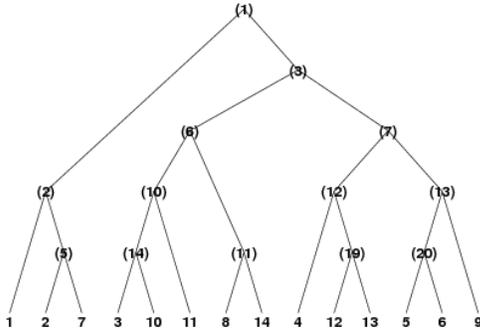
Fig. 4.   Typical HSVM hierarchical structure

TABLE I
BOTSWANA TRAINING DATA: INDIVIDUAL CLASS

| Class | Number of Pixels |
|---|---|
| Water | 158 |
| Primary Floodplain | 228 |
| Riparian | 237 |
| Firescar | 178 |
| Island Interior | 183 |
| Woodlands | 199 |
| Savanna | 162 |
| Short Mopane | 124 |
| Exposed Soils | 111 |

### C. Hierarchical Support Vector Machine

The increased number of features associated with this stacking approach makes a support vector machine an especially desirable classifier for this problem. In our previous study [9], a hierarchical support vector (HSVM) was developed to handle problems that involve not only high dimensional inputs but also complex land cover that are difficult to discriminate. The HSVM is similarly advantageous as it splits a complex multi-class problem into smaller binary classification problems.

To exploit the natural class groupings in combination with the SVM classifier, we also apply max-cut optimization in the hierarchical class decomposition framework by searching for the maximum total distance between two class partitions. The original class samples are treated as an undirected graph $G$ where node $n_i$ represents class $i$ and the non-negative weight:

$$w_{ij} = \frac{1}{2} \sum_{\forall x} \left( f_i(x) \log \frac{f_i(x)}{f_j(x)} + f_j(x) \log \frac{f_j(x)}{f_i(x)} \right) \quad (8)$$

is the average KL distance [10] between the density function of class $i$ and class $j$. The new HSVM approach solves this max-cut problem to achieve the required unsupervised class decomposition at each node of the binary hierarchical structure. The output space is hierarchically decomposed into pure leaf nodes that have only one class label at each node (see Fig. 4). Since this max-cut unsupervised decomposition uses total pairwise distance measures to investigate natural class groupings, the hierarchical structure results in a fast and intuitive SVM training process that requires little tuning. As demonstrated in our previous research [9], the method also has both high accuracy levels and good generalization.

### III. RESULTS

The NASA EO-1 satellite acquired a sequence of data over the Okavango Delta, Botswana in 2001-2003. The Hyperion sensor on EO-1 acquires data at $30m^2$ pixel resolution over a 7.7 km strip in 242 bands covering the 400-2500 nm portion of the spectrum in 10 nm windows. Preprocessing of the data was performed by the UT Center for Space Research to mitigate the effects of bad detectors, inter-detector miscalibration, and intermittent anomalies. Uncalibrated and noisy bands that cover water absorption features were removed, and the remaining

145 bands were included as candidate features: [10-55, 82-97, 102-119, 134-164, 187-220]. The data analyzed in this study, acquired May 31, 2001, consist of observations from 14 identified classes representing the land cover types in seasonal swamps, occasional swamps, and drier woodlands located in the distal portion of the Delta.

Ten randomly sampled partitions of the training data were sub-sampled such that 75% of the original data were used for training and 25% for testing. The numbers of labeled pixels for each individual class are presented in Table I. In order to investigate the impact of the quantity of training data on classifier performance, these training data were then sub-sampled to obtain ten samples comprised of 50%, 30%, and 15% of the original training data. All classifiers were evaluated using the ten test samples composed of 25% of the original training data.

Because the training and test data are spatially collocated and selected from the relatively homogeneous area, an extended test set was also acquired and used to evaluate the generalization of these classifiers to other areas that are often on the class boundaries. Note that this extended data may have substantially different characteristics as it is taken from a geographically separate and often mixed location. Its purpose here is to investigate the capability of the various methods for extending results obtained from using the contextual information in classification. Hereafter, these data are referred to as the test and edge test data, respectively.

Experiments were performed using pixel-wise hierarchical SVM (HSVM) [9], majority filtering (MF), iterated conditional modes (ICM), and the proposed max-cut stacking HSVM (MC-HSVM). Average classification accuracies for test data for the 10 experiments conducted with each classifier are listed in Table III. Classification accuracies on extended edge test set are presented in Table IV. The more detailed individual class accuracies are shown in Table II. Detailed pairwise comparisons are presented in the next three sections to demonstrate the performance gain due to properly stacking vectors.

### A. Comparing HSVM and MC-HSVM

From both Table III and Table IV, MC-HSVM is shown to be the clear winner over pixel-wise HSVM. Since both classifiers follow the same HSVM framework, the improvement in classification accuracy is unmistakably due to the knowledge-based stacking. In this pairwise comparison, MC-HSVM not

TABLE II

BOTSWANA EDGE TEST DATA: INDIVIDUAL CLASS ACCURACY (STD. DEV.)

| Class | Total Numbers | HSVM | MF | ICM(MRF) | MC-HSVM |
|---|---|---|---|---|---|
| Water | 145 | 99.2(0.54) | 96.0(2.43) | **99.2(0.22)** | 98.6(0.60) |
| Primary Floodplain | 145 | 91.9(3.62) | **94.3(3.56)** | 93.8(2.56) | 91.3(3.27) |
| Riparian | 140 | 70.1(6.92) | 77.3(7.23) | 72.2(7.81) | **80.0(3.05)** |
| Firescar | 152 | 95.7(2.45) | **98.8(0.37)** | 98.2(1.49) | 94.1(4.16) |
| Island Interior | 149 | 88.8(5.26) | **97.4(2.62)** | 94.0(2.23) | 86.7(9.58) |
| Woodlands | 133 | 90.5(1.46) | **98.5(0.35)** | 93.1(2.37) | 95.0(2.49) |
| Savanna | 141 | 89.0(2.86) | **99.1(0.58)** | 91.5(2.55) | 97.6(1.17) |
| Short Mopane | 143 | 79.4(6.47) | **82.4(5.33)** | 79.2(4.73) | 80.3(4.14) |
| Exposed Soils | 155 | 86.5(2.04) | 66.1(5.59) | 85.9(2.31) | **93.2(3.62)** |

TABLE III

BOTSWANA TEST DATA: ACCURACY (STD. DEV.)

| Training % | HSVM | MF | ICM(MRF) | MC-HSVM |
|---|---|---|---|---|
| 15% | 96.5(0.95) | **98.6(0.57)** | 97.5(0.71) | 97.5(0.77) |
| 30% | 97.3(1.14) | **99.3(0.11)** | 98.5(0.30) | 98.7(0.63) |
| 50% | 97.9(0.51) | **99.5(0.23)** | 98.9(0.12) | 98.7(0.58) |
| 75% | 97.7(0.52) | **99.7(0.13)** | 99.3(0.10) | 99.2(0.44) |

TABLE IV

BOTSWANA EDGE TEST DATA: ACCURACY (STD. DEV.)

| Training % | HSVM | MF | ICM(MRF) | MC-HSVM |
|---|---|---|---|---|
| 15% | 83.6(2.58) | 84.5(2.41) | 84.5(2.47) | **86.3(2.31)** |
| 30% | 87.4(1.53) | 89.4(2.14) | 88.2(1.59) | **89.9(1.49)** |
| 50% | 87.9(1.24) | 89.8(1.25) | 89.3(1.12) | **90.8(1.23)** |
| 75% | 88.5(0.86) | 90.9(0.94) | 89.8(0.89) | **91.9(0.66)** |

only achieves higher overall accuracy in both homogeneous areas and mixed areas, it also provides consistently better results for individual classes. (See Table II). The improved accuracy in small classes, such as exposed soils, shows that MC-HSVM can label these classes correctly by properly learning selected contextual information.

### B. Comparing MF and MC-HSVM

Although the overall accuracy tables indicate that majority filtering performs well under different settings, visual evaluation of the classified images does not support this conclusion. Fig. 5 and 6 show that MF results tend to yield very blocky results, and samples on the class boundary are often misclassified. From Fig. 5, the narrow river channel (blue pixels) is relatively small compared to the neighboring floodplain and riparian pixels. The MF result shows that these water pixels are dominated by their local neighbors. In addition, some detailed spatial textures are removed by the majority filtering process. A similar result is shown in Fig. 6. The individual class accuracy table (Table II) provides additional clarification. The table shows that MF performs very well on some spatially extensive classes but fails to recognize small classes with complex boundaries. For example, MF performs well on woodland and grasslands, two classes with homogeneous neighbors; however, the accuracy of MF drops on exposed soils due to the small size of the patches.. These pixels were reclassified as they are dominated by their more frequently occurring neighbors. Images classified by MC-HSVM (Fig. 5 and 6) do not have these problems. Overall

accuracies of the extended edge test set achieved by MC-HSVM are consistently higher.

### C. Comparing ICM and MC-HSVM

Algorithms based on MRF have been widely used for utilizing spatial-spectral information. Our experiments here show that MC-HSVM is slightly better than ICM in terms of overall classification accuracies and is competitive with ICM in individual class accuracies. MC-HSVM has slightly higher accuracies for bushes and flat areas, like woodlands, grasslands and exposed soils, while ICM performs well in wetland classes, such as floodplain, firescar and island interior. Because these wetland samples tend to have more complex geometric configurations, these results indicate that ICM exploits the complexity of local neighborhoods, resulting in higher accuracies. The MC-HSVM method yields higher accuracies in homogeneous and small classes because it incorporates properly selected contextual information in the classification model.

Average processing time for ICM and MC-HSVM are quite different. For these 40 Botswana experiments, each having 256*1465 pixels, 9 classes and 145 feature spaces, using a 3GHz Pentium 4 CPU, the processing times for all four classifiers are listed in Fig. 7. Because of the nature of MRF, ICM is the slowest classifier and averages 75 minutes per image. MC-HSVM took 25 minutes to classify an image which includes running the max-cut stacking, only one-third of the processing time required by ICM. Thus, MC-HSVM is not only competitive with ICM in terms of accuracy, but clearly beats ICM with respect to time complexity.

### IV. CONCLUSION AND FUTURE RESEARCH

The goal of the study is to fully exploit the spectral information provided by hyperspectral sensing, using contextual information to further improve classification and create a robust classifier. In this paper, we develop a knowledge-based stacking algorithm by pre-processing the image using max-cut optimization to recognize the spatial class boundaries. The proposed method is applied to data collected over the Okavango Delta of Botswana and compared to other very competitive and well studied approaches.

The classified map and classification accuracy tables presented in this paper indicate that the proposed max-cut stacking method is able to provide more accurate predictions on

Fig. 5.    Wetland Area, (Left) Original Image, (Center) MC-HSCM Result, (Right) MF Result



Fig. 6.    Wetland Area 2, (Left) Original Image, (Center) MC-HSCM Result, (Right) MF Result



Fig. 7.    Average Processing Time

correct assignments of classes that are spectrally similar but not geographically collocated. For example, island interior has very similar signatures to exposed soils but has quite different neighbors (Island interior regions are often located in the wetland area while exposed soils are often mixed with grassland or short mopane.) A post-processing filter, in conjunction with semi-supervised learning, should be able to learn such high-level information and improve the classification results.

## REFERENCES

[1] R. M. Haralick, K. Shanmugam, and D. I., "Textural features for image classification," *IEEE Trans. On SMC*, vol. 3, no. 6, pp. 610–621, 1973.

[2] L. O. Jimnez, J. L. Rivera-Medina, E. Rodrguez-Daz, E. Arzuaga-Cruz, and M. Ramrez-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, Apr 2005.

[3] W. A. Davis and F. G. Peet, "A method of smoothing digital thematic maps," *Remote Sensing of Environment*, vol. 6, pp. 45–49, 1977.

[4] 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.

[5] 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, Jan 1987.

[6] J. Besag, "On the statistical analysis of dirty pictures," *Journal of Royal Statistical Society*, vol. 48, no. 3, pp. 259–302, 1986.

[7] Q. Jackson and D. Landgrebe, "Adaptive bayesian contextual classification based on markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2454–2463, Nov 2002.

[8] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93 – 97, 2006.

both homogeneous areas and samples selected from mixed neighborhoods. It not only shows that it provides more details than the MF algorithm does, but also is better than the MRF ICM in both classification accuracy and speed.

This study focuses on intelligently fusing spatial and spectral information of the input space and exploits the HSVM classifier, which handles both a high dimensional input space and complex land cover types. Potential future research includes extending this study to incorporate semi-supervised learning and developing a post-processing approach which accounts for neighborhood configurations (similar to MRF) to

[9] Y. Chen, M. M. Crawford, and J. Ghosh, "Integrating support vector machines in a hierarchical output decomposition framework," in *2004 International Geosci. and Remote Sens. Symposium*, Anchorage, Alaska, Sept. 20-24 2004, pp. 949–953.

[10] S. Kullback, *Information Theory and Statistics*. New York: John Wiley and Sons., 1959.

[11] M. J. Todd, "Semidefinite Optimization," Cornell University, Ithaca, NY 14853, Tech. Rep., 2001.

[12] M. X. Goemans and D. P. Williamson, "Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming," *J. Assoc. Comput. Mach.*, vol. 42, pp. 1115–1145, 1995.

[13] Y. Nesterov and A. Nemirovskii, *Interior Point Polynomial Methods in Convex Programming: Theory and Applications*. Philadelphia: Society for Industrial and Applied Mathematics, 1994.