Hybrid Bayesian Classifier for Improved Classification Accuracy

Uttam Kumar, *Student Member, IEEE*, S. Kumar Raja, Chiranjit Mukhopadhyay, and T. V. Ramachandra, *Senior Member, IEEE*

***Abstract*—The widely used Bayesian classifier is based on the assumption of equal prior probabilities for all the classes. However, inclusion of equal prior probabilities may not guarantee high classification accuracy for the individual classes. Here, we propose a novel technique—*Hybrid Bayesian Classifier* (HBC)—where the class prior probabilities are determined by unmixing a supplemental low spatial–high spectral resolution multispectral (MS) data that are assigned to every pixel in a high spatial–low spectral resolution MS data in Bayesian classification. This is demonstrated with two separate experiments—first, class abundances are estimated per pixel by unmixing Moderate Resolution Imaging Spectroradiometer data to be used as prior probabilities, while posterior probabilities are determined from the training data obtained from ground. These have been used for classifying the Indian Remote Sensing Satellite LISS-III MS data through Bayesian classifier. In the second experiment, abundances obtained by unmixing Landsat Enhanced Thematic Mapper Plus are used as priors, and posterior probabilities are determined from the ground data to classify IKONOS MS images through Bayesian classifier. The results indicated that HBC systematically exploited the information from two image sources, improving the overall accuracy of LISS-III MS classification by 6% and IKONOS MS classification by 9%. Inclusion of prior probabilities increased the average producer's and user's accuracies by 5.5% and 6.5% in case of LISS-III MS with six classes and 12.5% and 5.4% in IKONOS MS for five classes considered.**

***Index Terms*—Bayesian classifier, prior probability, unmixing.**

I. Introduction

IMAGE classification using conventional Bayesian classifier is a supervised method based on prior probabilities [1]. Prior probability resolves confusions among classes that are not well separable [2] and is therefore effective in improving classification accuracy. Bayesian classifier is generally used with the assumption that prior probabilities are equal, as reliable prior probabilities are not easily available. However, with the use of equal prior probabilities, the performance of the classifier is not optimal [3], which is evident from the misclassified pixels primarily due to spectral confusion between classes, apart from the increased computing and sampling requirements [4]. The theoretical analysis of the effect of prior probability has been discussed in detail by Mingguo *et al.* [5]. Earlier works used prior probabilities based on previous year crop statistics [6], geographical data [7], elevation data [8], and spatial characteristics specified through a Markov random field model at reference resolution [9], improving the overall accuracy and kappa coefficient. Therefore, it is desirable to obtain reliable prior probabilities for each spectral class and use them to classify the pixels that are likely to misclassify [1], even though they are difficult to determine within the same time period and for the same spectral classes.

In this context, a new classification method—*Hybrid Bayesian Classifier* (HBC)—based on linear unmixing and Bayesian classifier is proposed to assign a class label to each pixel in a *high spatial–low spectral resolution* (HS-LSR) multispectral (MS) data. The prior probabilities of the different classes in the Bayesian classifier to classify each pixel in the HS-LSR data, such as the Indian Remote Sensing Satellite Linear Imaging Self Scanner-III (IRS LISS-III) MS and IKONOS MS, are obtained from abundance estimates by unmixing *low spatial–high spectral resolution* (LS-HSR) MS data, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat ETM+, respectively. The terms HS-LSR and LS-HSR have been used in a relative sense here, depending upon the spatial resolution of the images. For example, in the first experiment, MODIS bands are referred as LS-HSR and the Indian Remote Sensing Satellite (IRS) LISS-III MS bands are HS-LSR data, while in the second experiment, Landsat ETM+ bands are LS-HSR, and IKONOS MS bands are HS-LSR data. The reason for selecting MODIS and Landsat Enhanced Thematic Mapper Plus (ETM+) images as LS-HSR supplement data is because of their economic viability and high temporal resolution that enables their procurement for any part of the globe, throughout the year, corresponding to any HS-LSR data. However, the technique, in general, can be applied on any other LS-HSR (such as hyperspectral bands) and HS-LSR (such as MS bands) data classification. The novelty of this approach lies in the fact that LS-HSR and HS-LSR MS data are combined to improve the classification results, which can be thought of as a fusion process, in the sense that information from two different sources

Manuscript received January 17, 2010; accepted September 23, 2010. This work was supported in part by the Indian Institute of Science and in part by the Natural Resources Data Management System, Ministry of Science and Technology, Department of Science and Technology, Government of India.

U. Kumar is with the Department of Management Studies and Centre for Sustainable Technologies, Indian Institute of Science, Bangalore 560 012, India (e-mail: uttam@ces.iisc.ernet.in).

S. K. Raja is with the Institut de Recherche en Informatique et Systemes Aleatoires, 35042 Rennes Cedex, France and also with the Technicolor Research and Innovation, 35576 Cesson Sévigné, France (e mail: s.kumar.raja@yahoo.com).

C. Mukhopadhyay is with the Department of Management Studies, Indian Institute of Science, Bangalore 560 012, India (e-mail: cm@mgmt.iisc.ernet.in).

T. V. Ramachandra is with the Centre for Ecological Sciences, Centre for Sustainable Technologies and Centre for Infrastructure, Sustainable Transport and Urban Planning, Indian Institute of Science, Bangalore 560 012, India (e-mail: cestvr@ces.iisc.ernet.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2010.2087006

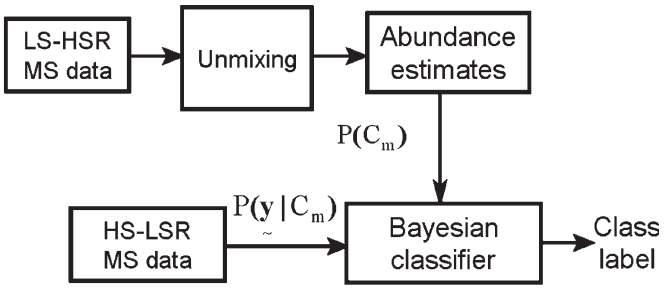

Fig. 1. HBC.

(sensors) are combined to arrive at an improved classified image by systematically exploiting the relevant information from both sources, as shown in Fig. 1. Zhukov *et al.* [10] attempted a similar multisensor multiresolution image fusion based on low spatial resolution (LSR) unmixing and high spatial resolution (HSR) classification. However, the end product was not a classified image but a set of MS fused images equivalent to the number of LSR bands at the HSR image dimension. In the following section, linear unmixing using orthogonal subspace projection (OSP) and Bayesian classifier are reviewed. The proposed HBC is discussed in Section III, while Section IV demonstrates the experimental results and validation followed by conclusion in Section V.

II. Review of Methods

A. Linear Unmixing

With K spectral bands and M classes $(C_1, \ldots, C_M)$, each pixel has an associated K-dimensional pixel vector $\boldsymbol{y} = (y_1, \ldots, y_K)'$ whose components are the gray values corresponding to the K spectral bands. Let $\boldsymbol{E} = [e_1, \ldots, e_M]$, where, for $m = 1, \ldots, M$, e_m is a $K \times 1$ column vector representing the endmember spectral signature of the mth target material. For a pixel, let α_m denote the fraction of the mth target-material signature and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)'$ denote the M-dimensional abundance column vector. The linear mixture model for $\boldsymbol{y}$ is given by

$$\boldsymbol{y} = \boldsymbol{E\alpha} + \boldsymbol{\eta} \tag{1}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K)'$ are independent identically distributed $N(0, \sigma^2)$ [11]. Equation (1) represents a standard signal-detection model where $\boldsymbol{E\alpha}$ is a desired signal vector to be detected. Since, OSP detects one signal (target) at a time, we divide a set of M targets into desired (C_m) and undesired $(C_1, \ldots, C_{m-1}, C_{m+1}, \ldots C_M)$ targets. A logical approach is to eliminate the effects of the undesired targets that are considered as "interferers" to C_m before the detection of C_m takes place. Now, in order to find α_m, the desired target material is $\boldsymbol{e}_m$. The term $\boldsymbol{E\alpha}$ in (1) can be rewritten to separate the desired spectral signature $\boldsymbol{e}_m$

$$\boldsymbol{y} = \boldsymbol{e}_m\alpha_m + \boldsymbol{Rr} + \boldsymbol{\eta} \tag{2}$$

where $\boldsymbol{r} = (\alpha_1, \ldots \alpha_{m-1}, \alpha_{m+1}, \ldots, \alpha_M)$ and $\boldsymbol{R} = [e_1, \ldots, e_{m-1}, e_{m+1}, \ldots, e_M]$. Thus the interfering signatures in $\boldsymbol{R}$ can be removed by the operator

$$\boldsymbol{P} = \left(\boldsymbol{I} - \boldsymbol{R}(\boldsymbol{R}^T\boldsymbol{R})^{-1}\boldsymbol{R}^T\right) \tag{3}$$

which is used to project $\boldsymbol{y}$ into a space orthogonal to the space spanned by the interfering spectral signatures, where $\boldsymbol{I}$ is a $K \times K$ identity matrix [12]. Operating on $\boldsymbol{y}$ with $\boldsymbol{P}$ and noting that $\boldsymbol{PR} = 0$

$$\boldsymbol{Py} = \boldsymbol{Pe}_m\alpha_m + \boldsymbol{P\eta}. \tag{4}$$

After maximizing the signal-to-noise ratio, an optimal estimate of α_m is

$$\widehat{\alpha}_m = \frac{\boldsymbol{y^T P^T P y}}{\boldsymbol{e_m^T P^T P e}_m}. \tag{5}$$

Note that $1 \geq \alpha_m \geq 0$, and thus, $(\alpha_1, \ldots, \alpha_M)$ can be taken as proportional probabilities of the M classes, i.e., $P(C_m) \propto \alpha_m$.

B. Bayesian Classifier

Associated with any pixel, there is an observation $\underset{\sim}{y}$. With M classes $(C_1, \ldots, C_M)$, Bayesian classifier calculates the posterior probability of each class conditioned on $\boldsymbol{y}$ [13]

$$P(C_m|\underset{\sim}{y}) = \frac{P(\underset{\sim}{y}|C_m)P(C_m)}{P(\underset{\sim}{y})}. \tag{6}$$

In (6), since $P(\underset{\sim}{y})$ is constant for all classes, only $P(\underset{\sim}{y}|C_m)P(C_m)$ is considered. $P(\underset{\sim}{y}|C_m)$ is computed assuming class conditional independence; therefore, $P(\underset{\sim}{y}|C_m)$ is given by

$$P(\underset{\sim}{y}|C_m) = \prod_{k=1}^{K} P(y_k|C_m). \tag{7}$$

III. New HBC

HBC uses the abundance of each class obtained from LS-HSR data by linear unmixing as prior probability while classifying the HS-LSR data using Bayesian classifier of the same geographical area and time frame. That is, given the observation vector $\underset{\sim}{y}$ for a pixel, it is classified to fall in class l if $l = \text{Arg Max}_m P(\underset{\sim}{y}|C_m)P(C_m)$, where $P(\underset{\sim}{y}|C_m)$ is as that in (7), calculated using the HS-LSR data, and $P(C_m) \propto \alpha_m$, calculated using the LS-HSR data. The assumptions in this method are the following: 1) If there are r HS-LSR pixels contained in one LS-HSR pixel, i.e., the resolution ratio is $(r:1)$, the prior probabilities for all the r HS-LSR pixels are equal corresponding to the same LS-HSR pixel and 2) the two data types have a common origin or upper left corner, i.e., the edges of the $r \times r$ HS-LSR pixels overlaps exactly with the corresponding LS-HSR pixel. The limitations are as follows: 1) (K-1) should be $\geq M$ in the LS-HSR data and 2) M in HS-LSR should be $\leq M$ in the LS-HSR data.

IV. Experimental Results

Two separate experiments were carried out. In the first case, LISS-III MS data (three bands of 23.5 m $\times$ 23.5 m spatial

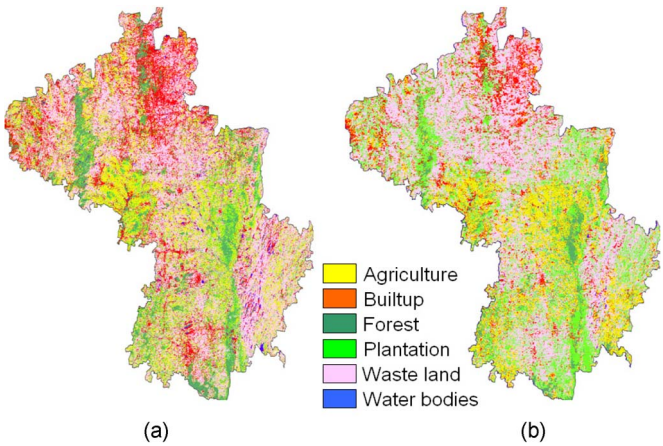

Fig. 2. LISS-III MS classified images: (a) Bayesian classifier. (b) HBC.

TABLE I
CLASS STATISTICS FROM BAYESIAN AND HBC FOR LISS-III DATA

Classifiers →	Bayesian classifier		HFC	
Class ↓	Ha	%	Ha	%
Agriculture	155, 451	19.04	142931	17.51
Builtup	139, 759	17.12	79280	9.71
Forest	93, 241	11.42	55721	6.83
Plantation	89, 493	10.96	176132	21.58
Waste land	329, 473	40.36	355587	43.56
Water bodies	8854	1.08	66.20	0.81

resolution resampled to 25 m, acquired on December 25, 2002) of 5320 × 5460 size and MODIS eight-day composite data (seven bands of 250 m × 250 m, acquired in December 19–26, 2002) of 532 × 546 dimension, were coregistered with known ground-control points (rmse − 0.11). Training data were collected from the ground representing approximately 10% of the study area covering the entire spectral gradient of the classes. Separate test data were collected for validation. An LISS-III MS classified image using conventional Bayesian classifier is shown in Fig. 2(a). Assuming that there are six (fixed) numbers of representative endmembers (pure pixels), the entire image was modeled in terms of those spectral components, extracted using N-FINDR algorithm [14] from the MODIS images. In the absence of pure pixels, alternative algorithms [15] can be used for endmember extraction, which is a limitation of N-FINDR. It may be noted that some objects (for example, buildings with concrete roofs, tiled roofs, asphalt, etc.) exhibit high degrees of spectral heterogeneity representing variable endmembers. This intraclass spectral variation with variable endmembers can be addressed through techniques discussed in [16]–[18] and is beyond the scope of this paper.

Abundance values were estimated for each pixel through unmixing and used as prior probabilities in HBC to classify LISS-III MS data [Fig. 2(b)]. Table I is the class statistics, and Table II indicates the producer's and user's accuracies. The overall accuracy and kappa for HBC (93.54% and 0.91) is higher than the Bayesian classifier (87.55% and 0.85). For any particular class, if the reference data has more pixels with correct label, the producer's accuracy is higher, and if the pixels with the incorrect label in classification result are less, the user's

TABLE II
ACCURACY ASSESSMENT FOR LISS-III DATA

Classifiers →	Bayesian classifier		HBC			
Class ↓	PA*	UA*	PA*		UA*	
Agriculture	87.54	87.47	90.15	↑	95.56	↑
Builtup	85.11	81.68	89.39	↑	98.33	↑
Forest	85.71	88.73	92.61	↑	96.36	↑
Plantation	84.44	91.73	95.95	↑	91.03	↓
Waste land	88.03	90.37	98.67	↑	89.66	↓
Water bodies	90.91	88.89	88.18	↓	97.00	↑
Average	86.96	88.15	92.49	↑	94.66	↑

* PA – Producer's Accuracy; UA – User's Accuracy.

accuracy is higher [5]. The Bayesian classifier wrongly classified many pixels belonging to waste/barren/fallow as built-up. The forest class was overestimated, and the plantation was underestimated by the Bayesian classifier. The only minority class in the study area is water bodies. Classified image using conventional Bayesian classifier had 1.08% (8854 ha) of water bodies in the study area. After the ground visit, we found that there are not many water bodies, and the extents of most individuals were < 2000 m^2. Therefore, only a few water bodies that had spatial extent $\geq 62\,500$ m^2 could be used as endmembers. The minimum detected water class was 5% using unmixing of LS-HSR data. It may have happened that the prior probability was unlikely for this class while classifying the LISS-III MS data using HBC. The classified image obtained from HBC showed 0.81% (66.20 ha) of water bodies. A few pixels were wrongly classified using HBC; therefore, the producer's accuracy decreased from 90.91% (in conventional Bayesian classifier) to 88.18% (in HBC). However, given the same set of training pixels for classification, the user's accuracy has increased from 88.89% to 97%. The producer's accuracy increased for agriculture (2.6%), built-up (4.3%), forest (7%), plantation (11.5%), and wasteland (10.6%), and the user's accuracy increased for agriculture (8%), built-up (16.6%), forest (7.6%), and water bodies (8%) in the HBC output. On the other hand, the producer's accuracy decreased (2.7%) for water bodies, and the user's accuracy decreased (∼0.7%) for plantation and wasteland classes, in agreement with similar observations reported in [5]. HBC was intended to improve classification accuracies by correctly classifying pixels which were likely to be misclassified by Bayesian classifier. Therefore, a cross comparison of the two classified images located the pixels that were assigned different class labels at the same location. These wrongly classified pixels, when validated with ground data, revealed a 6% (∼1 742 832 pixels) improvement in classification by HBC.

In the second experiment, IKONOS MS data (four bands of 4 m, acquired on November 24, 2004) of 700 × 700 size and Landsat ETM+ data (six bands, excluding thermal and panchromatic, of 30 m, acquired on November 22, 2004) of 100 × 100 dimension were coregistered (rmse−0.09). The Landsat pixels were resampled to 28 m so that 49 IKONOS pixels would fit in one Landsat pixel. The scenes correspond to Bangalore City, India, near the central business district, having race course, bus stand, railway lines, parks, built-up with concrete roofs, asbestos roofs, blue plastic roofs, coal tarred roads

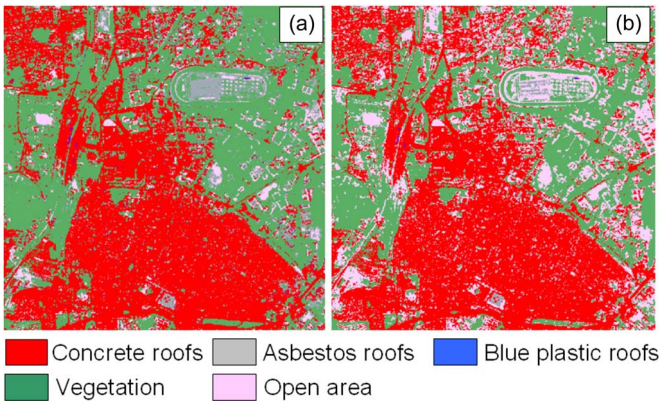

Fig. 3. IKONOS classified images: (a) Bayesian classifier. (b) HBC.

TABLE III
ACCURACY ASSESSMENT FOR IKONOS DATA

Classifiers →	Bayesian classifier		HFC			
Class ↓	PA	UA	PA		UA	
Concrete roofs	69.99	84.01	76.49	↑	93.89	↑
Asbestos roofs	84.77	87.77	91.89	↑	94.46	↑
Vegetation	94.21	87.55	87.24	↓	89.13	↑
Blue plastic roof	84.33	81.17	97.00	↑	85.60	↑
Open area	51.49	69.49	95.00	↑	74.22	↑
Average	76.96	81.99	89.52	↑	87.46	↑

TABLE IV
CLASS STATISTICS FOR IKONOS DATA

Classifiers →	Bayesian classifier		HFC	
Class ↓	Ha	%	Ha	%
Concrete roofs	346.67	44.34	356.18	45.56
Asbestos roofs	47.99	7.41	6.79	0.87
Blue plastic roof	5.83	0.75	0.21	0.03
Vegetation	329.72	42.18	260.01	33.26
Open area	41.60	5.32	158.58	20.28

with flyovers, and a few open areas (playground, walkways, vacant land, etc.). Class proportions from Landsat image pixels were used as prior probabilities to classify the IKONOS MS images using HBC [Fig. 3(b)]. The overall accuracy and kappa for HBC (89.53% and 0.87) is higher than the Bayesian classifier (80.46% and 0.69). The producer's accuracy increased by 17% for concrete, asbestos, blue plastic roof, and open area and decreased by 7% for vegetation in HBC (Table III). The user's accuracy increased by ∼5.5% for all the classes in HBC. The Bayesian classifier wrongly classified and overestimated many pixels belonging to the open area as asbestos roof (Table IV). Vegetation was overestimated by 9% using the Bayesian classifier, and open area was underestimated by ∼15%. A cross comparison of the two classified images showed that 95 733 pixels (0.33% of the study area) were differently classified by the two classifiers. Validation of these pixels showed an improvement of 9% by HBC, higher than the accuracies reported in [6]–[8].

V. CONCLUSION

The major contribution of this technique lies in the fact that abundance estimates from LS-HSR data were utilized as prior probabilities to classify HS-LSR data using a Bayesian classifier, improving the overall accuracy by 6% and 9% with IRS LISS-III MS and IKONOS MS data, respectively, as compared with conventional Bayesian classifier, demonstrating the robustness of the approach.

REFERENCES

[1] J. Ediriwickrema and S. Khorram, "Hierarchical maximum-likelihood classification for improved accuracies," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 810–816, Jul. 1997.
[2] D. K. McIver and M. A. Friedl, "Using prior probabilities in decision-tree classification of remotely sensed data," *Remote Sens. Environ.*, vol. 81, no. 2, pp. 253–261, Aug. 2002.
[3] C. Lee and D. A. Landgrebe, "Fast likelihood classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 4, pp. 509–517, Jul. 1991.
[4] L. Pedroni, "Improved classification of Landsat Thematic Mapper data using modified prior probabilities in large and complex landscapes," *Int. J. Remote Sens.*, vol. 24, no. 1, pp. 91–113, Jan. 2003.
[5] Z. Mingguo, C. Qiangguo, and Q. Mingzhou, "The effect of prior probabilities in the maximum likelihood classification on individual classes: A theoretical reasoning and empirical testing," *Photogramm. Eng. Remote Sens.*, vol. 75, no. 9, pp. 1109–1116, 2009.
[6] L. L. F. Janssen and H. Middelkoop, "Knowledge-based crop classification of a Landsat Thematic Mapper image," *Int. J. Remote Sens.*, vol. 13, no. 15, pp. 2827–2837, Oct. 1992.
[7] H. Cetin, T. A. Warner, and D. W. Levandowski, "Data classification, visualization, and enhancement using n-dimensional probability density functions (nPDF): AVIRIS, TIMS, TM, and geophysical applications," *Photogramm. Eng. Remote Sens.*, vol. 59, no. 12, pp. 1755–1764, Dec. 1993.
[8] A. H. Strahler, "The use of prior probabilities in maximum likelihood classification of remotely sensed data," *Remote Sens. Environ.*, vol. 10, no. 2, pp. 135–163, Sep. 1980.
[9] G. Storvik, R. Fjortoft, and A. H. S. Solberg, "A Bayesian approach to classification of multiresolution remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 539–547, Mar. 2005.
[10] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1212–1226, May 1999.
[11] U. Kumar, N. Kerle, and T. V. Ramachandra, "Constrained linear spectral unmixing technique for regional land cover mapping using MODIS data," in *Innovations and Advanced Techniques in Systems, Computing Sciences and Software Engineering*, K. Elleithy, Ed. Berlin, Germany: Springer-Verlag, 2008, pp. 87–95.
[12] C.-I. Chang, "Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, Mar. 2005.
[13] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco, CA: Morgan Kaufmann, 2003.
[14] M. E. Winter, "N-Findr: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE Imaging Spectrometry*, 1999, vol. 3753, pp. 266–275.
[15] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 650–663, Mar. 2004.
[16] C. A. Bateson, G. P. Asner, and C. A. Wessman, "Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1083–1094, Mar. 2000.
[17] C. Song, "Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability?" *Remote Sens. Environ.*, vol. 95, no. 2, pp. 248–263, Mar. 2005.
[18] G. M. Foody and H. T. X. Doan, "Variability in soft classification prediction and its implications for sub-pixel scale change detection and super resolution mapping," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 8, pp. 923–933, 2007.