

Class Information-Based Band Selection for Hyperspectral Image Classification

Meiping Song, Xiaodi Shang, *Student Member, IEEE*, Yulei Wang, Chunyan Yu, *Member, IEEE*,
and Chein-I Chang[✉], *Fellow, IEEE*

Abstract—This paper presents a class information (CI)-based band selection (BS) approach to hyperspectral image classification (HSIC). It introduces a new concept from an information theory point of view, CI which can be used to determine an appropriate weight imposed on each class of interest. Specifically, two types of criteria, intraclass information criterion (IC) and interclass IC are derived as CI probabilities to measure CI that can be used to determine the number of training samples required to be selected for each class. With such CI-calculated probabilities, another new concept called class self-information (CSI) is also defined for each class that can be further used to define the class entropy (CE) so that CSI and CE can be used to determine the number of bands required for BS, n_{BS} . In order to find desired n_{BS} bands, two types of BS methods based on CSI and CE are custom-designed, called single class signature-constrained BS (SCSC-BS) which utilizes the constrained energy minimization (CEM) to constrain each individual class signature to select bands for a particular class according to its CSI-determined n_{BS} and a multiple class signatures-constrained BS (MCSC-BS) which takes advantage of linearly constrained minimum variance (LCMV) to constrain all class signatures to select CE-determined n_{BS} bands for all classes.

Manuscript received November 1, 2018; revised January 19, 2019 and April 20, 2019; accepted May 27, 2019. Date of publication July 18, 2019; date of current version October 31, 2019. The work of M. Song was supported in part by the National Nature Science Foundation of China under Grant 61601077 and in part by the Fundamental Research Funds for the Central Universities under Grant 3132017124. The work of Y. Wang was supported in part by the Fundamental Research Funds for the Central Universities under Grant 3132017080 and in part by the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences, under Grant LSIT201707D. The work of C. Yu was supported by the National Nature Science Foundation of Liaoning Province under Grant 20170540095. The work of C.-I Chang was supported in part by the Fundamental Research Funds for Central Universities under Grant 3132016331 and a project under Grant ZD20180073. (*Corresponding author: Xiaodi Shang*.)

M. Song and Y. Wang are with the Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information and Technology College, Dalian Maritime University, Dalian 116026, China, and also with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710216, China (e-mail: smping@163.com; wangyulei.wyl@163.com).

X. Shang and C. Yu are with the Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: shangxiaodi@qq.com; yuchunyan1997@126.com).

C.-I Chang is with the Center of Hyperspectral Imaging in Remote Sensing, Information and Technology College, Dalian Maritime University, Dalian 116026, China, with the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan, is with the Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore, MD 21250 USA, and also with the Department of Computer Science and Information Management, Providence University, Taichung 02912, Taiwan (e-mail: cchang@umbc.edu).

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

Digital Object Identifier 10.1109/TGRS.2019.2920891

These SCSC-BS and MCSC-BS selected bands are then used to perform classification and evaluated by CI-weighted classification measures by real image experiments. The results show that HSIC using judiciously selected partial bands as well as CI-weighted measures can improve HSIC with using full bands.

Index Terms—Band selection (BS), between class distance (BCD), class entropy (CE), class information (CI), class self-information (CSI), class Fisher's ratio (CFR), constrained energy minimization (CEM), information criterion (IC), linearly constrained minimum variance (LCMV), multiple class signatures-constrained BS (MCSC-BS), single class signature-constrained BS (SCSC-BS), within class distance (WCD).

NOMENCLATURE

BS	Band selection.
BCD	Between class distance.
CD	Class density.
CE	Class entropy.
CI	Class information.
CSI	Class self-information.
CFR	Class Fisher's ratio.
CEM	Constrained energy minimization.
IC	Information criterion.
LCMV	Linearly constrained minimum variance.
MCSC-BS	Multiple class signatures-constrained BS.
SB-MCSC-BS	Sequential backward MCSC-BS.
SF-MCSC-BS	Sequential feed forward MCSC-BS.
SB-SCSC-BS	Sequential backward SCSC-BS.
SF-SCSC-BS	Sequential feed forward SCSC-BS.
SCSC-BS	Single class signature-constrained BS.
SR	Sample ratio.
WCD	Within class distance.

LIST OF SYMBOLS

Symbol	Definitions
C_i	i th class.
\hat{C}_j	Classified j th class by a classifier.
M	Total number of classes to be classified.
$\{C_i\}_{i=1}^M$	All classes of interest.
N	Total number of data samples.
n_i	Total number of data samples in C_i .
n_{ii}	Total number of data samples in C_i correctly classified in C_i .
n_{ji}	Total number of data samples in C_i but classified into C_j .
N^{training}	Total number of training samples for all classes.

n_i^{training}	Total number of training samples used for the i th class, C_i .
\hat{N}	Total number of classified data samples.
\hat{n}_j	Total number of data samples classified into C_j .
\hat{n}_{ij}	Total number of data samples classified in C_j which are supposed to in C_i .
n_{BS}	Number of bands to be selected.
$n_{\text{CSI}}^{\text{CSI}}$	Number of bands determined by CSI.
$n_{\text{BS}}^{\text{CSI}}(C_i)$	Number of bands determined by CSI for the i th class, C_i .
$n_{\text{BS}}^{\text{CE}}$	Number of bands determined by CE for $\{C_i\}_{i=1}^M$.
$I^{\text{CSI}}(C_i)$	CSI for C_i .
p_i^{CI}	Generic CI probability calculated for class C_i using CI as a criterion.
$P_A(C_i)$	Accuracy of C_i , see (18).
P_{AA}	Average accuracy, see (19).
$P_{\text{CI-OA}}(C_i)$	Accuracy using CI as a criterion, see (20).
P_{OA}	Overall accuracy, see (16).
$P_{\text{PR}}(C_i)$	Precision of C_i , see (21).
$P_{\text{CI-PR}}$	Precision using CI as a criterion, see (22).
$\vartheta_{\text{BCD}}^{\text{CI}}(C_i)$	BCD measure for C_i see (4).
$\vartheta_{\text{CD}}^{\text{CI}}(C_i)$	CD for C_i , see (2).
$\vartheta_{\text{CFR}}^{\text{CI}}(C_i)$	CFR for C_i , see (5).
$\vartheta_{\text{WCD}}^{\text{CI}}(C_i)$	WCD measure for C_i , see (1).
$\vartheta_{\text{SR}}^{\text{CI}}(C_i)$	SR for C_i , see (3).
$\Omega_{\text{BS}}^{\text{CSI}}(C_i)$	Band subset determined by CSI for the i th class, C_i .
$\Omega_{\text{BS}}^{\text{CSI}}$	Band subset determined by CSI for all classes, $\{C_i\}_{i=1}^M$, see (26).
$V(\Omega_{\text{BS}})$	$(\mathbf{d}_{\Omega_{\text{BS}}}^T \mathbf{R}_{\Omega_{\text{BS}}}^{-1} \mathbf{d}_{\Omega_{\text{BS}}})^{-1}$, see (36).
$V(\mathbf{b}_l)$	$(\boldsymbol{\mu}_{\mathbf{b}_l}^T \mathbf{R}_{\mathbf{b}_l}^{-1} \boldsymbol{\mu}_{\mathbf{b}_l})^{-1}$, see (38).

I. INTRODUCTION

IN RECENT years, BS for hyperspectral image classification (HSIC) has received considerable interest, for example, [1]–[26]. In general, BS can be performed by several approaches. One is band prioritization (BP) [1] which uses a custom-designed criterion to calculate a priority score of every single band for its ranking. BS selects bands according to their assigned priority scores [1]–[5]. Such BP-based BS is generally unsupervised regardless of a specific application. It makes use of a BP criterion to select bands according to data characteristics or statistics such as variance, signal-to-noise ratio (SNR), entropy, Jeffries–Matusita (JM) distance [16], [18], [27], and information divergence (ID) [3], [6], [18]. Unfortunately, a BP criterion suffers from at least four drawbacks. First of all, it does not provide a means of determining how many bands needed to be selected, n_{BS} . Second, it must prioritize all bands because it does not know n_{BS} . Third, it requires band decorrelation to remove bands highly correlated with already selected bands.

However, how to select an appropriate threshold for band decorrelation is a challenging issue. Fourth, the bands selected by a BP criterion such as variance, SNR, JM measure and ID are completely characterized by data statistics regardless of applications. In other words, such BP-selected bands are fixed and cannot vary with different applications. To address this issue, another approach is to use an application-based criterion to generate band features that can be used to select bands. Consequently, it requires a feature selection algorithm to find an optimal set of features that determine bands to be selected. BS of this type is generally supervised because it usually requires prior knowledge about the data to be processed such as training samples, the number of classes of interest to be classified [10]–[15]. In this case, the main focus of BS is placed on design and development of strategies for searching bands. On the other hand, BS can be also categorized into three groups. One group is band clustering, which clusters bands into a finite number of clusters [13]–[15]. Another group is sequential multiple band selection (SQMBS) which selects multiple bands “sequentially” [5], [16], [17]. Specifically, SQMBS generally starts off with either one band or two, and then begins to grow the selected band sets by adding one band at a time according to a searching strategy in which case the well-known sequential floating forward selection (SFFS) [11] is generally used to select bands such as [16] and [17]. A third group is simultaneous multiple band selection (SMMBS) which selects multiple bands “simultaneously” [18]–[25] in the sense that all bands must be selected at the same time, not one after another as does SQMBS.

However, two major issues arising from BS generally have a significant impact on classification results and in [1]–[26]. One is determining the number of bands to be selected, n_{BS} . The other is selecting desired bands for classification once n_{BS} is determined. As for the first issue, one commonly used approach is to take advantage of virtual dimensionality (VD) developed in [27]–[32], target specified VD (TSVD) in [33] and [34] and band specified VD (BSVD) [35]. Since determining n_{BS} is very challenging, over the past years BS has mainly focused on the second issue that is to develop algorithms for finding desired bands with n_{BS} being determined empirically or other criteria such as VD. Interestingly, to the authors’ best knowledge there is no work reported in the literature on how to determine n_{BS} , particularly for the classification. For example, VD in [27]–[32] was specifically developed for determining the number of spectrally distinct signatures using correlation-covariance eigenvalues analysis by the Harsanyi–Farrand–Chang (HFC) method [36]. On the other hand, TSVD in [33] and [34] was designed for targets of interest specified by a particular application, whereas BSVD [35] was specifically developed for selecting bands of interest according to the mutual orthogonality of bands regardless of applications.

This paper takes an interesting twist by looking into the information provided by each of classes of interest which can be used to determine n_{BS} specifically for classification. Its idea is to introduce CI that can be used to measure the information contained in each class of interest for classification.

The concept of CI is new and derived from information theory [37] which has never been explored in the classification literature. To make this approach work, two key issues need to be addressed: 1) to design a criterion to measure CI of each class and; 2) to utilize CI to determine n_{BS} for each of M classes, $\{C_i\}_{i=1}^M$.

In order to resolve the first issue, five criteria are particularly designed to measure CI, which are WCD, CD, SR, BCD, and CFR, all of which are derived from a classification perspective [38], viz. between-class variance, within-class variance, SNR and FR. Despite the fact that WCD, BCD, and FR have been used for pattern classification, they are particularly used as criteria to design and develop classifiers but not to measure CI as proposed in this paper. Specifically, these five criteria can be grouped into two categories, intraclass IC which measures class variability of data samples within a class and interclass IC which measures class separability of data samples in a class from other classes. The concepts of intraclass variability and interclass separability are new and cannot be found in the existing literature. With this interpretation, WCD and CD can be considered as intraclass IC, while BCD and CFR can be regarded as interclass IC. Interestingly, the commonly used SR used in classification, defined as the ratio of class size to the entire data size, can also be viewed as an intraclass IC. These defined ICs can be further used to calculate CI probabilities for two purposes. One is to determine the number of training samples required to be selected for each class of interest. The other is to weigh the significance of each class when it comes to evaluating classification performance.

To take up the second issue, another new concept, called CSI borrowed from information theory [36], is also introduced to measure self-information contained in each of the classes, $\{C_i\}_{i=1}^M$ using CI-calculated probabilities from intraclass ICs and interclass ICs. In other words, the higher the CI probability is, the less the information is contained in a class and also the less uncertainty the class to be characterized. Such CI probabilities can be further used to define CSI that determines how many bands should be selected to classify each class. With different values of CSI, the number of bands selected for each class, referred to as $n_{BS}^{CSI}(C_i)$ for the i th class C_i , is expected to be different; accordingly, in order to classify M classes, it will require a total of bands, denoted by n_{BS}^{CSI} no greater than $\sum_{i=1}^M n_{BS}^{CSI}(C_i)$ since there may have the same bands selected for multiple classes. If a class contains no uncertainty, its CSI value is zero, which indicates no information contained in the class. Consequently, no band is needed to classify this class. On the other hand, if the CSI of a class has a large value, this implies that the class has greater uncertainty, in which case more bands will be needed for BS to classify the class.

As an alternative to n_{BS}^{CSI} , the self-information of each class, $n_{BS}^{CSI}(C_i)$, can be also used to define CE in the same way as self-information used to define entropy in information theory [37]. The resulting CE calculated from $n_{BS}^{CSI}(C_i)$ can be considered as the ensemble average CSI per class, denoted by n_{BS}^{CE} . As a result, the number of bands, n_{BS} , required to classify for all M classes would be $n_{BS} = n_{BS}^{CE} \times M$. Obviously, the n_{BS}^{CSI} derived from CSI is different from $n_{BS}^{CE} \times M$ obtained from CE with the latter assuming that $n_{BS}^{CSI}(C_i) = n_{BS}^{CE}$ for

$1 \leq i \leq M$ without calculating $n_{BS}^{CSI}(C_i)$ for an individual class, i th class C_i .

Once the value of $n_{BS}^{CSI}(C_i)$ or n_{BS}^{CE} is determined, a follow-up task is to select desired bands for classification. When $n_{BS}^{CSI}(C_i)$ is used for each of M classes, $\{C_i\}_{i=1}^M$, an SCSC-BS is developed by constraining each of class mean vectors, $\{\mu_i\}_{i=1}^M$ to find $n_{BS}^{CSI}(C_i)$ bands, denoted by $\Omega_{BS}^{CSI}(C_i)$, and then fuses the obtained $\{\Omega_{BS}^{CSI}(C_i)\}_{i=1}^M$ by finding their union $\Omega_{BS}^{CSI} = \bigcup_{i=1}^M \Omega_{BS}^{CSI}(C_i)$ as the final band subset to be used to classify M classes, $\{C_i\}_{i=1}^M$. In this case, the total number of bands to be used to classify all M classes, $\{C_i\}_{i=1}^M$, denoted by $|\Omega_{BS}^{CSI}|$, is bounded below and above by $\max_{1 \leq i \leq M} n_{BS}^{CSI}(C_i) \leq |\Omega_{BS}^{CSI}| \leq \sum_{i=1}^M n_{BS}^{CSI}(C_i)$. On the other hand, when n_{BS}^{CE} is used, another BS algorithm different from SCSC-BS is developed, called MCSC-BS by constraining all the M class mean vectors, $\{\mu_i\}_{i=1}^M$, together simultaneously with no need of fusion to find a final band subset Ω_{BS}^{CE} to classify M classes, $\{C_i\}_{i=1}^M$. To evaluate performance of SCSC-BS and MCSC-BS for HSIC, two recently developed iterative classifiers, called iterative constrained energy minimization (ICEM) in [39] and iterative linearly constrained minimum variance (ILCMV) classifier developed HSIC in [40] and [41] are used due to the fact that both ICEM and ILCMV are also designed by constraining individual class signature and all class signatures, respectively, as SCSC-BS and MCSC-BS are designed.

As a summary, there are several unique novelties derived from this paper which cannot be found in the literature. First and foremost is the introduction of CI to calculate the information of each of M classes of interest in terms of probability. Such CI-calculated probabilities give rise to three new ideas in applications. One is to determine the number of training samples required to be selected for each class. Another is to weigh the significance of each class in classification measures. A third one is to define CSI of each of the classes that can be used to determine the number of bands required to be selected, $n_{BS}^{CSI}(C_i)$ for the i th class C_i with $1 \leq i \leq M$. Such CSI can be then used to define CE to determine the number of bands, $n_{BS} = n_{BS}^{CE} \times M$ for all classes where the CE is calculated by averaging CSI over all the classes. Unlike $n_{BS}^{CSI}(C_i)$, which requires fusing $\{\Omega_{BS}^{CSI}(C_i)\}_{i=1}^M$ to find final $\Omega_{BS}^{CSI} = \bigcup_{i=1}^M \Omega_{BS}^{CSI}(C_i)$ bands for all the M classes, the obtained CE determines n_{BS} for all the classes $\{C_i\}_{i=1}^M$ all together without band fusion. Finally, to find appropriate bands for classification two types of BS methods, SCSC-BS and MCSC-BS are custom-designed according to the determined values of n_{BS}^{CSI} and $n_{BS} = n_{BS}^{CE} \times M$, respectively. These BS methods constrain class signatures as band features to find desired bands, a new idea which has not been investigated in the literature.

II. MEASURES OF CLASS INFORMATION

Assume that there are M classes of interest to be classified, $\{C_i\}_{i=1}^M$. Let μ_i be the sample mean of the i th class, C_i . Obviously, not all M classes provide the same amount of information. According to information theory [37], the information of a class describes how much certainty contained in

the class. It is this uncertainty that determines how many bands required to classify the class. Interestingly, by far, no work reported in the literature has been directed to the exploration of CI in BS. The five criteria presented in the following are particularly designed to measure CI from a classification point of view. These five criteria can be grouped into two categories, i.e., intraclass IC and interclass IC.

A. Intraclass Information Criteria

The criteria in this category are solely based on independent classes with no interaction from other classes and can be used to measure the variability of data samples within a single individual class.

1) *Class Variability IC*: A simple means of measuring the variability of a class is its class variance, σ^2 . In other words, if a class has zero variance, it indicates that the class provides no CI at all and thus, it has no uncertainty. This means that the class contains all data samples with the same identical spectral signature. Hence, no band is needed to classify this class. Conversely, if a class has a large variance, it implies that the data samples in the class are widely spread in terms of spectral variability, in which case its CI described as uncertainty is large. This indicates that this class requires a large number of bands for its classification. Hence, class variability can be used to measure CI contained in a class. The WCD, which serves the purpose, is defined as follows:

$$\vartheta_{\text{WCD}}^{\text{CI}}(C_i) = \sum_{\mathbf{r} \in C_i} (\mathbf{r} - \boldsymbol{\mu}_i)^T (\mathbf{r} - \boldsymbol{\mu}_i) = \sigma_i^2 \quad (1)$$

2) *Class Density IC*: The class variability, WCD in (1) only focuses on class variance as the second order statistics without including the first order statistics, i.e., the sample mean. The CD defined here combines sample mean and sample variance (1) into a criterion. It can be also considered as SNR derived from communications/signal processing. It is the ratio of the class mass considered as signal energy specified by $\|\boldsymbol{\mu}\|^2$ with $\boldsymbol{\mu}$ as the sample means to the class area considered as noise energy measured by σ^2 . The CD is defined as

$$\vartheta_{\text{CD}}^{\text{CI}}(C_i) = \frac{\boldsymbol{\mu}_i^T \boldsymbol{\mu}_i}{\sigma_i^2} \quad (2)$$

which can be used to address this need. Hence, according to (2), the higher the density of a class is, better clustered the data samples in the class are, thus the less CI the class contains and the less uncertainty the class has.

3) *Sample Ratio IC*: In the traditional HSIC, a commonly used criterion is SR which can also be used as a CI measure defined by

$$\vartheta_{\text{SR}}^{\text{CI}}(C_i) = \frac{n_i}{N} \quad (3)$$

where n_i is the number of data samples in the i th class, C_i and N is the total number of data samples in the data set. Since SR is calculated only based on the size of an individual single class relative to the entire data size, N which is fixed, it has nothing to do with other classes, thus it can be considered as an intraclass IC. More specifically, higher the SR of a class, more likely a data sample to be selected from the class, which leads to the less uncertainty of the class.

B. Interclass Information Criteria

Unlike intraclass IC which describe the variability of data samples within a single class, the interclass IC presented in this section measure the separability of data samples among classes.

1) *Class Separability IC*: The simplest criterion of CI to measure class separability is to consider a class represented by its class center specified by its sample mean. Afterward, the separability of a class from all other classes can be measured by BCD among all class centers (sample means) defined by

$$\vartheta_{\text{BCD}}^{\text{CI}}(C_i) = \min_{1 \leq j \neq i \leq M} \|\boldsymbol{\mu}_j - \boldsymbol{\mu}_i\| \quad (4)$$

which indicates that the larger the BCD is, the better the separability of the class from other classes is; and thus, the class has less information and the less uncertainty.

2) *Class Fisher's Ratio IC*: Since BCD and WCD are developed independent intraclass IC and interclass IC without referring one to the other, the following criterion is actually designed by combining the strengths of both BCD and WCD into a single criterion. Its idea is originated from the well-known FR widely used for classification [38]. By taking advantage of FR, we can define CFR as

$$\vartheta_{\text{CFR}}^{\text{CI}}(C_i) = \frac{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i^*}\|^2}{\sigma_i^2 + \sigma_{i^*}^2} \quad (5)$$

where

$$i^* = \arg \left\{ \min_{1 \leq j \neq i \leq M} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\| \right\} \quad (6)$$

and

$$\sigma_k^2 = \frac{1}{|C_k|} \sum_{\mathbf{r} \in C_k} (\mathbf{r} - \boldsymbol{\mu}_k)^T (\mathbf{r} - \boldsymbol{\mu}_k) \quad (7)$$

is the variance of class C_k where $|C_k|$ is the number of elements in C_k . That is, $\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i^*}\|$ is the minimal distance of $\boldsymbol{\mu}_i$ between other class centers, $\{\boldsymbol{\mu}_j\}_{j=1, j \neq i}^M$.

Interestingly, JM distance [16], [18], [27] can also be used as an alternative to CFR since it also takes care of intraclass variability and interclass separability as CFR does. However, the JM distance makes an underlying assumption that it works when class data samples are Gaussian distributed, which is not required by CFR. This Gaussian class data distribution assumption runs into several issues. First of all, it requires an effective estimation technique to reliably find Gaussian statistics. Most importantly, it is generally not applicable to small classes. For example, classes 1, 7, 9, and 16 of the Purdue data (see Fig. 2), all of which have less than 100 data samples (see Table I). Hence, compared to the total number of data samples, 10 249, the class data samples of each of these four classes have less than 0.5% of entire data samples. Accordingly, the class data distribution of each of these four classes may not be appropriately described by a Gaussian distribution, especially class 9 and class 7, which have only 20 and 28 data samples, respectively. Hence, the JM distance is inapplicable to these four classes. On the other hand, a second issue is that even though the Gaussian class data distribution assumption may be true for large classes,

for example, classes 2, 11, and 14 of the Purdue data, all of which have more than 1000 data samples, how do we know that the given Gaussian class distribution is a single Gaussian distribution or a Gaussian mixture distribution by a number of Gaussian distributions? Therefore, in the latter case, in addition to estimating Gaussian statistics, we also run into a third issue of how to estimate the number of Gaussians used for such Gaussian mixture. Most recently, when the JM distance is used for HSIC, it has been shown to be ineffective in [42]. Finally, despite the fact that the JM distance measures separability of two classes on a more convenient scale [0-2] in terms of Bhattacharyya distance [16], [18], [27]. However, it is also known that the Bhattacharyya distance used as a measure of separability has the disadvantage that it continues to grow even after the classes have become so well separated that any classification procedure could distinguish them perfectly. In contrast to the JM distance, our proposed CFR does not have all of the above-mentioned issues. Even though the idea of CFR is originated from the FR used by the Fisher linear discriminant analysis, CFR is defined differently based on one class against the rest of classes instead of FR defined between two classes. From this point of view, CFR is indeed a new concept modified from FR.

It should be noted that according to (5) the criteria, WCD defined in (1) and BCD defined in (4), seem special cases of (5). As a matter of fact, it is generally not true and is demonstrated by experiments in Table IX(a), (d), and (e). This is because CFR in (5) constraints both (1) and (4) in one equation, while WCD in (1) and BCD in (4) can stand alone by themselves as individual and separate criteria without constraining one on another.

By virtue of CI criteria defined in (1)–(5), we can calculate CI probabilities associated with each of classes in terms of probabilities by

$$p_i^{\text{WCD}} = p^{\text{WCD}}(C_i) = \frac{\vartheta_{\text{WCD}}^{\text{CI}}(\boldsymbol{\mu}_i)}{\sum_{j=1}^M \vartheta_{\text{WCD}}^{\text{CI}}(\boldsymbol{\mu}_j)} \quad (8)$$

$$p_i^{\text{CD}} = p^{\text{CD}}(C_i) = \frac{\vartheta_{\text{CD}}^{\text{CI}}(\boldsymbol{\mu}_i)}{\sum_{j=1}^M \vartheta_{\text{CD}}^{\text{CI}}(\boldsymbol{\mu}_j)} \quad (9)$$

$$p_i^{\text{SR}} = p^{\text{SR}}(C_i) = \vartheta_{\text{SR}}^{\text{CI}}(C_i) = \frac{n_{ii}}{n_i} \quad (10)$$

$$p_i^{\text{BCD}} = p^{\text{BCD}}(C_i) = \frac{\vartheta_{\text{BCD}}^{\text{CI}}(\boldsymbol{\mu}_i)}{\sum_{j=1}^M \vartheta_{\text{BCD}}^{\text{CI}}(\boldsymbol{\mu}_j)} \quad (11)$$

$$p_i^{\text{CFR}} = p^{\text{CFR}}(C_i) = \frac{\vartheta_{\text{CFR}}^{\text{CI}}(\boldsymbol{\mu}_i)}{\sum_{j=1}^M \vartheta_{\text{CFR}}^{\text{CI}}(\boldsymbol{\mu}_j)} \quad (12)$$

all of which can be considered as CI probability assigned to class C_i . However, it is worth noting that a higher p_i^{WCD} in (8) indicates more information on class C_i . In order to be consistent with other criteria, we take its reciprocal $(p_i^{\text{WCD}})^{-1}$ to ensure that the probability $(p_i^{\text{WCD}})^{-1}$ is proportional to the information of class C_i . In this case, (8) is replaced by

$$\tilde{p}_i^{\text{WCD}} = \frac{(p^{\text{WCD}}(C_i))^{-1}}{\sum_{j=1}^M (p^{\text{WCD}}(C_j))^{-1}}. \quad (13)$$

Now, using (9)–(13), higher the $p^{\text{CI}}(C_i)$ is, less uncertainty the class C_i has, and thus, fewer the bands required to identify the class.

To simplify notations, let p_i^{CI} be a generic CI probability calculated for class C_i using a CI criterion which can be one of the five criteria, WCD, CD, SR, BCD and CFR, specifically, $\text{CI} = \text{WCD}$ with $p^{\text{CI}=\text{WCD}}(C_i) = \tilde{p}_i^{\text{WCD}}$. Using p_i^{CI} , we can further define CSI of each class as

$$I^{\text{CSI}}(C_i) = I_i^{\text{CSI}} = -\log_2 p_i^{\text{CI}}. \quad (14)$$

Two comments on CI and CSI are noteworthy.

- 1) The values of CI and CSI are inversely proportional to each other. That is, for a class, the large value of its CI is, less uncertainty of the class has and the smaller the value of its CSI is. For example, if $p_i^{\text{CI}} = 1/2$, then $I_i^{\text{CSI}} = -\log_2 p_i^{\text{CI}} = 1$ bit, in which case it only needs one bit to clarify the class. On the other hand, if a class has $p_i^{\text{CI}} = 1/4$, then $I_i^{\text{CSI}} = -\log_2 p_i^{\text{CI}} = 2$ bits, in which case it requires two bits to specify the class. In summary, smaller the CI probability of a class, more uncertainty the class has, and thus, higher its CSI value.
- 2) If a class has a higher CI probability, i.e., p_i^{CI} , it indicates that it has a lower CSI value, $I^{\text{CSI}}(C_i)$, calculated from (14). Interpreting one bit in (14) as one band implies that it requires fewer bands classifying this particular class and vice versa. Thus, if a class contains no uncertainty at all, its CI probability is one, $p^{\text{CI}}(C_i) = p_i^{\text{CI}} = 1$, which indicates that its CSI value is zero, i.e., $I^{\text{CSI}}(C_i) = 0$. In this case, the entire data set contains only one single class. This means that no information contained in the data set because it can be specified by a single class and thus, no band is needed to classify this data set. Conversely, if a data set contains all classes which have equal CI probabilities, i.e., $p_i^{\text{CI}} = 1/M$. This indicates that all classes have the same value of CSI that will require the same number of bands to classify each individual class. By virtue of CSI defined in (14), a class has a lower value of CI, p_i^{CI} , it will have a higher CSI value, $I^{\text{CSI}}(C_i)$, which will require more resources in the sense of using more bands to identify the class.

Furthermore, we can also introduce a new concept of CI, the entropy of CI defined as the averaged CSI over all the classes of interest by

$$H^{\text{CE}}(\{C_i\}_{i=1}^M) = \sum_{i=1}^M p_i^{\text{CI}} I^{\text{CSI}}(C_i) = -\sum_{i=1}^M p_i^{\text{CI}} \log_2 p_i^{\text{CI}}. \quad (15)$$

III. CLASSIFICATION MEASURES

Assume that

M = the number of classes;

n_i = the number of data samples in the i th class according to ground truth, i.e., $n_i = \sum_{j=1}^M n_{ji}$;

N = total number of data samples, $N = \sum_{i=1}^M n_i$;

n_{ji} = the number of data samples in the i th class to be classified into the j th class;

n_{ii} = the number of data samples in the i th class correctly classified into the i th class;

\hat{C}_j = the set of data samples being classified in the j th class C_j ;

\hat{n}_{ji} = the number of data samples classified in the j th class, which are supposed to in the i th class, C_i ;

\hat{n}_j = the number of data samples classified in the j th class, C_j , i.e., $\hat{n}_j = \sum_{i=1}^M \hat{n}_{ji}$;

\hat{N} = total number of classified data samples in M classes = $\sum_{i=1}^M \hat{n}_i$;

$p(\hat{C}_j) = (\hat{n}_j / \hat{N})$ = probability of \hat{C}_j = SR of \hat{C}_j ;

In the traditional HSI classification, a widely used criterion is overall accuracy (OA). P_{OA} , defined as

$$P_{OA} = \frac{1}{N} \sum_{i=1}^M n_{ii} \quad (16)$$

where n_{ii} is the number of signal samples in the i th class, C_i correctly to be classified into the i th class, C_i and N is the total number of data samples in the data set. Interestingly, we can re-express (16) as

$$P_{OA} = \frac{1}{N} \sum_{i=1}^M n_{ii} = \sum_{i=1}^M \left(\frac{n_i}{N} \right) \frac{n_{ii}}{n_i} = \sum_{i=1}^M p_i^{\text{SR}} \frac{n_{ii}}{n_i} \quad (17)$$

which shows that P_{OA} actually uses SR as CI in (3) to measure the information contained in each class. Now, if we further define

$$P_A(C_i) = \text{accuracy of the } i\text{th class} = \frac{n_{ii}}{n_i} \quad (18)$$

then a criterion called average accuracy (AA) can be defined as

$$\begin{aligned} P_{AA} &= p_{\text{average-accuracy}}(\{C_i\}_{i=1}^M) \\ &= \frac{1}{M} \sum_{i=1}^M n_{ii}/n_i = \frac{1}{M} \sum_{i=1}^M P_A(C_i). \end{aligned} \quad (19)$$

By virtue of (14), P_{OA} in (17) can be further extended to

$$P_{\text{CI-OA}} = \sum_{i=1}^M p_i^{\text{CI}} P_A(C_i) \quad (20)$$

where CI can be specified by one of the five CI IC specified by (9)–(13).

For an M -classes classification, an M -class confusion matrix can be constructed in Fig. 1.

As shown in Fig. 1, new classification measures will be used for experiments are summarized as follows [43]:

$$\begin{aligned} P_{\text{PR}}(\hat{C}_j) &= \text{precision rate of the } j\text{th class, } C_j \\ &= p(\hat{C}_j | \text{classification}) = \frac{\hat{n}_{jj}}{\hat{n}_j}. \end{aligned} \quad (21)$$

Using (21), we can further derive

$$P_{\text{CI-PR}} = \sum_{i=1}^M p_i^{\text{CI}} P_{\text{PR}}(\hat{C}_i) \quad (22)$$

which will be used to calculate CI-weighted precision rate (PR).

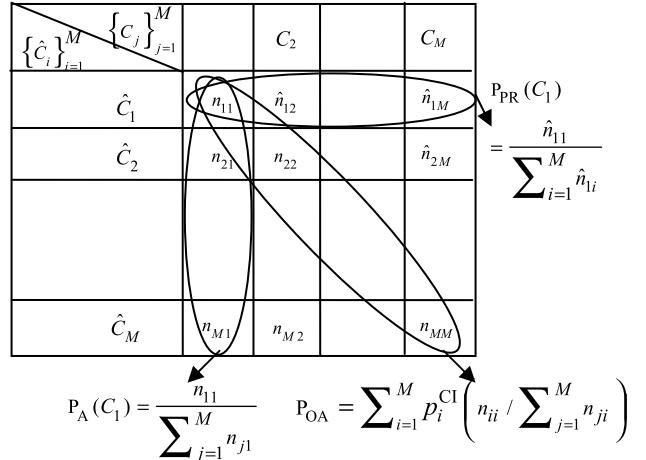


Fig. 1. M -class confusion matrix.

As also noted in (3), $p^{\text{CI}}(C_i) = p_i^{\text{CI}} = (n_i/N)$, i.e., SR, (20) is reduced to traditional OA specified by (17). Also, when $p^{\text{CI}}(C_i) = p_i^{\text{CI}} = (1/M)$, (20) is reduced P_{AA} in (19).

As noted above, CI probabilities are calculated solely based on the classes of interest with exclusion of BKG. We can also calculate CI probabilities by including BKG as a single class. In this case, as shown in the following experiments, it turns out that except SR BKG class is one of the classes that yield small $p_{\text{BKG}}^{\text{CI}}$ probabilities. This explains why many works on HSIC reported in the literature on HSIC have excluded BKG from consideration. However, it does not mean that the BKG class should not be included for classification, since it has been shown in [39]–[41] that BKG has a significant impact on misclassification.

It should also be noted that $P_A(C_i)$ in (18) and $P_{\text{PR}}(\hat{C}_j)$ in (21) are referred to as producer's accuracy (PA) and user's accuracy (UA), respectively, in traditional remote sensing community [44]–[46].

IV. DETERMINATION OF TRAINING SAMPLES FOR EACH OF THE CLASSES OF INTEREST BY CI

A commonly used approach to selecting training samples is random sampling, particularly for cross-validation. In this case, it assumes that the information of all classes is equally likely. It seems that no work is devoted to taking CI into account for determining how many training samples should be selected for each class, n_i^{training} . This section develops an effective means of determining n_i^{training} by taking advantage of CI probabilities.

Suppose that there are a total of N^{training} training samples required for classification. A conventional approach is to randomly select training samples according to a certain percentile such as 5% and 10% from each of the class. There are two major problems associated with this approach. One is the determination of how much percentile needs to be used for training sample selection in the first place. The other is the same chosen percentile be applied to each of classes regardless of their information. Particularly, if a class with a small number of data samples, it runs into an issue that there will be no or very few training samples can be selected.

By contrast, if a class with a large number of data samples, it may result in over-sampling with too many training samples selected than it needs.

The proposed approach presented in this section resolves this issue. It first determines the total number of training samples, N^{training} required for classification. Instead of using percentile, it selects the number of training samples in accordance with their CSI probabilities. Specifically, for each class, C_i , the number of training samples n_i^{training} , can be defined as the number of training samples required to be sampled for class C_i by

$$n_i^{\text{training}} = \lceil N^{\text{training}} \times p^{\text{CI}}(C_i) \rceil \quad (23)$$

where $\lceil x \rceil$ is the smallest integer equal to or greater than x , i.e., $x \leq \lceil x \rceil < x + 1$ and $p^{\text{CI}}(C_i)$ is the CI probability of C_i . For example, if CFR in (5) is used, $p^{\text{CFR}}(C_i)$ in (12) is used to calculate (23).

As a result of (23), the total number of training samples required to be used for classification is

$$N^{\text{training}} \leq \sum_{i=1}^M n_i^{\text{training}} \leq N^{\text{training}} + M. \quad (24)$$

However, on some occasions, $n_i^{\text{training}} \geq n_i/2$. In this case, we set the upper bound, $n_i^{\text{upper}} = n_i/2$. On the other hand, if $n_i^{\text{training}} < n_i/100$, then we set the lower bound, $n_i^{\text{lower}} = n_i/100$. Nevertheless, these upper and lower bounds can be chosen empirically to avoid over-sampling and under-sampling training data. They can be adapted if there is a need.

It is known that determining the number of training samples for classes of interest is a very challenging issue. Interestingly, the CI in (23) indeed provides a guideline for this purpose. Specifically, it shows by experiments conducted in Section IX that the number of training samples selected for the four smallest classes 9, 7, 1 and 16 in the Purdue data by EPF-based methods in [46] must be at least 50% of total class samples according to Table V(a). Unfortunately, such selection was done empirically with no justification given in [46]. The CI in (23) offers the explanation for their selection.

V. CLASS ENTROPY-DETERMINED NBS

While CI of each class determines the number of training samples required to be selected for the class, CSI of each class can also determine the number of bands required to be selected for classifying the class.

According to (14), we can interpret $I^{\text{CSI}}(C_i)$, which is the CSI of the i th class, C_i , as the band rate of C_i and determine the number of bands required for classification of the i th class C_i as

$$n_{\text{BS}}^{\text{CSI}}(C_i) = \lceil I^{\text{CSI}}(C_i) \rceil = \lceil -\log_2 p_i^{\text{CSI}} \rceil. \quad (25)$$

That is, it will require at least $n_{\text{BS}}^{\text{CSI}}(C_i)$ bands to accommodate information of C_i . Let $\Omega_{\text{BS}}^{\text{CSI}}(C_i)$ be the set of bands selected for the i th class, C_i , by a BS method with its size determined by $n_{\text{BS}}^{\text{CSI}}(C_i)$. Then,

$$\Omega_{\text{BS}}^{\text{CSI}} = \bigcup_{i=1}^M \Omega_{\text{BS}}^{\text{CSI}}(C_i) \quad (26)$$

will be the set of desired bands selected for the processed hyperspectral image for classification and its size can be bounded below and above by

$$\max_{1 \leq i \leq M} n_{\text{BS}}^{\text{CSI}}(C_i) \leq |\Omega_{\text{BS}}^{\text{CSI}}| \leq \sum_{i=1}^M n_{\text{BS}}^{\text{CSI}}(C_i). \quad (27)$$

where $|A|$ is the size of set A .

On the other hand, the entropy defined by (15), $H^{\text{CE}}(\{C_i\}_{i=1}^M)$ is the CE which can be considered as band rate/per class required to the classification of M classes and defined by

$$n_{\text{BS}}^{\text{CE}} = H^{\text{CE}}(\{C_i\}_{i=1}^M). \quad (28)$$

With this interpretation, the number of bands, n_{BS} , required to perform M -class classification will be

$$n_{\text{BS}} = \lceil n_{\text{BS}}^{\text{CE}} \times M \rceil. \quad (29)$$

It should be noted that n_{BS} in (29) is different from $n_{\text{BS}}^{\text{CSI}}(C_i)$ in (25) in the sense that (29) does not require calculating $n_{\text{BS}}^{\text{CSI}}(C_i)$ for each class with $n_{\text{BS}}^{\text{CE}} = n_{\text{BS}}^{\text{CSI}}(C_i)$ for $1 \leq i \leq M$.

Once $n_{\text{BS}}^{\text{CSI}}(C_i)$ or $n_{\text{BS}}^{\text{CE}}$ is determined by (25) or (29) a follow-up task is to develop a BS method to find an appropriate band subset with its size, n_{BS} to perform classification.

Although many BS methods have been reported and available in the literature, the two BS approaches to be presented in the following two sections are particularly designed to take advantage of class signatures CSI probabilities. Both are derived from adaptive beamforming arising array signal processing. Their idea considers each class signature vector, denoted by \mathbf{d} , as a desired signal and then constrains the \mathbf{d} by locking in its direction via a constraint. Two different rationales can be designed along with this line. One is developed from target detection perspective by considering each class signature vector as a desired target signal source arriving from its specific direction via a scalar constraint. The other is from a target classification point of view which considers all the M class signature vectors as multiple target signal sources arriving from M desired detections simultaneously and makes use of a vector or matrix constraint to classify M target signal sources as separate classes. Each of these two approaches is described in the following two sections.

VI. SINGLE CLASS SIGNATURE-CONSTRAINED BS

The first approach is derived from the well-known sub-target detection algorithm, called CEM developed in [27] and [48]–[50].

Suppose that a hyperspectral image is represented by a collection of image pixel vectors, denoted by $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$ where $\mathbf{r}_i = (r_{i1}, r_{i2}, \dots, r_{iL})^T$ for $1 \leq i \leq N$ is an L -dimensional pixel vector, N is the total number of pixels in the image and L is the total number of spectral channels. Furthermore, assume that $\mathbf{d} = (d_1, d_2, \dots, d_L)^T$ is specified by a desired signature of interest to be used for target detection. The goal is to find a target detector that can detect data samples specified by the desired target signal \mathbf{d} via a finite impulse response (FIR) linear filter with L filter coefficients, $\{w_1, w_2, \dots, w_L\}$, denoted by an L -dimensional vector

$\mathbf{w} = (w_1, w_2, \dots, w_L)^T$ which minimizes the filter output energy subject to the constraint $\mathbf{d}^T \mathbf{w} = \mathbf{w}^T \mathbf{d} = 1$. Let y_i denote the output of the designed FIR filter resulting from the input \mathbf{r}_i . Then y_i can be expressed by

$$y_i = \sum_{l=1}^L w_l r_{il} = (\mathbf{w})^T \mathbf{r}_i = \mathbf{r}_i^T \mathbf{w} \quad (30)$$

and the average energy of the filter output is given by

$$\begin{aligned} (1/N) \sum_{i=1}^N y_i^2 &= (1/N) \sum_{i=1}^N (\mathbf{r}_i^T \mathbf{w})^2 \\ &= \mathbf{w}^T [(1/N) \sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^T] \mathbf{w} = \mathbf{w}^T \mathbf{R}_{L \times L} \mathbf{w} \end{aligned} \quad (31)$$

where $\mathbf{R} = (1/N) \sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^T$ is the sample auto-correlation matrix of the image. The goal is to solve the following linearly constrained optimization problem:

$$\min_{\mathbf{w}} \{\mathbf{w}^T \mathbf{R} \mathbf{w}\} \quad \text{s.t. } \mathbf{d}^T \mathbf{w} = \mathbf{w}^T \mathbf{d} = 1 \quad (32)$$

where $\mathbf{w}^T \mathbf{R} \mathbf{w}$ can be considered as either variance resulting from signals not passing through the filter. The optimal solution to (32) is shown in [27] and [48]–[50] to be

$$\mathbf{w}^{\text{CEM}} = (\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d})^{-1} \mathbf{R}^{-1} \mathbf{d} \quad (33)$$

and

$$\min_w \mathbf{w}^T \mathbf{R}^{-1} \mathbf{w} = (\mathbf{w}^{\text{CEM}})^T \mathbf{R}^{-1} \mathbf{w}^{\text{CEM}} = (\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d})^{-1} \quad (34)$$

which is the minimum variance resulting from unwanted signal sources impinging upon an array of sensors [51], referred to as CEM error. With the optimal weight, \mathbf{w}^{CEM} specified by (33) a filter called CEM, denoted by $\delta^{\text{CEM}}(\mathbf{r})$ was derived [48] specified by

$$\delta^{\text{CEM}}(\mathbf{r}) = (\mathbf{w}^{\text{CEM}})^T \mathbf{r} = (\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d})^{-1} (\mathbf{R}^{-1} \mathbf{d})^T \mathbf{r}. \quad (35)$$

Now, assume that $\{\mathbf{b}_l\}_{l=1}^L$ is a set of band images representing a hyperspectral image cube where \mathbf{b}_l is the l th spectral band represented by a column vector, $\mathbf{b}_l = (b_{l1}, b_{l2}, \dots, b_{lN})^T$ and $\{b_{li}\}_{i=1}^N$ is the set of all N pixels in the l th band image, \mathbf{b}_l . In addition, we also assume that $\{\mu_i\}_{i=1}^M$ are class signature vectors where μ_i is the signature vector representing the i th class, for example, class mean vector as the desired signature \mathbf{d} , i.e., $\mathbf{d} = \mu_i$.

By taking advantage of the CEM error derived from (31), we can define the following new measure that can be used as a criterion for SCSC-BS:

$$V(\Omega_{\text{BS}}) = (\mathbf{d}_{\Omega_{\text{BS}}}^T \mathbf{R}_{\Omega_{\text{BS}}}^{-1} \mathbf{d}_{\Omega_{\text{BS}}})^{-1} \quad (36)$$

which is the minimum variance specified by (34) but uses only those bands in a band subset Ω_{BS} . Most importantly, we can prove by the following theorem that

$$\{V(\Omega_j)\}_{j=1}^L \quad (37)$$

is a monotonically decreasing sequence where $\Omega_j = \{\mathbf{b}_{l_1}, \mathbf{b}_{l_2}, \dots, \mathbf{b}_{l_j}\}$ is any j -band subset containing j bands, $\mathbf{b}_{l_1}, \mathbf{b}_{l_2}, \dots, \mathbf{b}_{l_j}$.

A. Sequential Feed Forward SCSC-BS

For each single band \mathbf{b}_l , we can replace the full band set Ω_{BS} in (33) with the single band \mathbf{b}_l to yield

$$V(\mathbf{b}_l) = (\boldsymbol{\mu}_{\mathbf{b}_l}^T \mathbf{R}_{\mathbf{b}_l}^{-1} \boldsymbol{\mu}_{\mathbf{b}_l})^{-1} \quad (38)$$

which can be used as a criterion to measure the variance caused by data sample vectors not specified by μ using only one single band, \mathbf{b}_l . Now, if we consider a band as a feature vector, SFFS developed in [11] can be used to develop an SFFS-based BS to augment bands to be selected one at a time sequentially by (38). The resulting BS is referred to as SF-SCSC-BS.

More specifically, SF-SCSC-BS selects the first band, denoted by $\mathbf{b}_{l_1}^*$

$$\mathbf{b}_{l_1}^* = \arg \min_{\mathbf{b}_l \in \Omega} V(\mathbf{b}_l) = \arg \left\{ \min_{\mathbf{b}_l \in \Omega} (\mathbf{d}_{\mathbf{b}_l}^T \mathbf{R}_{\mathbf{b}_l}^{-1} \mathbf{d}_{\mathbf{b}_l})^{-1} \right\}. \quad (39)$$

Then it selects the second band, denoted by $\mathbf{b}_{l_2}^*$ which yields the minimal variance by removing $\mathbf{b}_{l_1}^*$ from the full band set Ω as follows:

$$\begin{aligned} \mathbf{b}_{l_2}^* &= \arg \left\{ \min_{\mathbf{b}_l \in \Omega - \{\mathbf{b}_{l_1}\}} V(\mathbf{b}_l) \right\} \\ &= \arg \left\{ \min_{\mathbf{b}_l \in \Omega - \{\mathbf{b}_{l_1}\}} (\mathbf{d}_{\mathbf{b}_l}^T \mathbf{R}_{\mathbf{b}_l}^{-1} \mathbf{d}_{\mathbf{b}_l})^{-1} \right\}. \end{aligned} \quad (40)$$

The same process is repeated over and over again by continuously removing selected bands from the full band set Ω . The details of implementing SF-SCSC-BS step by step are summarized below.

SF-SCSC-BS

1. Initial condition:

Determine n_{BS} .

Find

$$\mathbf{b}_{l_1} = \arg \left\{ \min_{\mathbf{b}_l \in \Omega} (\mathbf{d}_{\mathbf{b}_l}^T \mathbf{R}_{\mathbf{b}_l}^{-1} \mathbf{d}_{\mathbf{b}_l})^{-1} \right\} \quad (39)$$

$$\Omega_1 = \{\mathbf{b}_{l_1}\}.$$

2. Band augmentation

$$\mathbf{b}_{l_j} = \arg \left\{ \min_{\mathbf{b}_l \in \Omega_{j-1}^c} (\mathbf{d}_{\Omega_{j-1} \cup \{\mathbf{b}_l\}}^T \mathbf{R}_{\Omega_{j-1} \cup \{\mathbf{b}_l\}}^{-1} \mathbf{d}_{\Omega_{j-1} \cup \{\mathbf{b}_l\}})^{-1} \right\} \quad (41)$$

where $\Omega_{j-1} = \{\mathbf{b}_{l_1}, \mathbf{b}_{l_2}, \dots, \mathbf{b}_{l_{j-1}}\}$ and $\Omega_{j-1}^c = \Omega - \Omega_{j-1}$.

3. If $j < n_{\text{BS}}$,

$$\Omega_j = \{\mathbf{b}_{l_1}, \mathbf{b}_{l_2}, \dots, \mathbf{b}_{l_j}\} = \Omega_{j-1} \cup \{\mathbf{b}_{l_j}\} \quad (42)$$

and go step 2. Otherwise, BS is terminated. The final set of selected bands is given by $\{\mathbf{b}_{l_1}, \mathbf{b}_{l_2}, \dots, \mathbf{b}_{l_{n_{\text{BS}}}}\}$ where n_{BS} is the number of bands needed to be selected.

It should be pointed out that SF-SCSC-BS does not have to run through all bands. It can terminate the augmentation process as long as the number of bands selected, $n_{\text{BS}}(C_i)$ is reached.

B. Sequential Backward SCSC-BS

In contrast to SF-SCSC-BS, we can also develop a sequential backward search (SBS), which uses leave-one-out as a technique to select optimal feature vectors. The resulting BS is called SB-SCSC-BS. That is, for each single band, say l th band, \mathbf{b}_l , we consider the band subset $\Omega - \{\mathbf{b}_l\}$ by removing \mathbf{b}_l from the full band set Ω and then replace the band set Ω_{BS} in (36) with $\Omega - \{\mathbf{b}_l\}$ to derive

$$\mathbf{V}(\Omega - \{\mathbf{b}_l\}) = (\mathbf{d}_{\Omega - \{\mathbf{b}_l\}}^T \mathbf{R}_{\Omega - \{\mathbf{b}_l\}}^{-1} \mathbf{d}_{\Omega - \{\mathbf{b}_l\}})^{-1}. \quad (43)$$

Specifically, the band which yields the maximal variance

$$\begin{aligned} \mathbf{b}_{l_1}^* &= \arg \{ \max_{\mathbf{b}_l \in \Omega} \mathbf{V}(\Omega - \{\mathbf{b}_l\}) \} \\ &= \arg \{ \max_{\mathbf{b}_l \in \Omega} (\mathbf{d}_{\Omega - \{\mathbf{b}_l\}}^T \mathbf{R}_{\Omega - \{\mathbf{b}_l\}}^{-1} \mathbf{d}_{\Omega - \{\mathbf{b}_l\}})^{-1} \} \end{aligned} \quad (44)$$

will be selected as the first band, denoted by $\mathbf{b}_{l_1}^*$ to be the most significant band since the variance in (44) produced the maximal variance if $\mathbf{b}_{l_1}^*$ is removed from Ω . Now let $\Omega_1 = \Omega - \{\mathbf{b}_{l_1}\}$. Then the second band which yields the maximal variance again

$$\begin{aligned} \mathbf{b}_{l_2}^* &= \arg \{ \min_{\Omega_1 - \{\mathbf{b}_l\}} \mathbf{V}(\Omega_1 - \{\mathbf{b}_l\}) \} \\ &= \arg \{ \min_{\Omega_1 - \{\mathbf{b}_l\}} (\mathbf{d}_{\Omega_1 - \{\mathbf{b}_l\}}^T \mathbf{R}_{\Omega_1 - \{\mathbf{b}_l\}}^{-1} \mathbf{d}_{\Omega_1 - \{\mathbf{b}_l\}})^{-1} \} \end{aligned} \quad (45)$$

then selected as the second band, $\mathbf{b}_{l_2}^*$. The same process is repeated over and over again by continuously removing selected bands from the full band set Ω . The resulting algorithm is called SB-SCSC-BS described as follows.

SB-SCSC-BS

1. Initial Condition:

Determine n_{BS} .

Find

$$\mathbf{b}_{l_1} = \arg \{ \max_{\mathbf{b}_l \in \Omega} (\mathbf{d}_{\Omega - \{\mathbf{b}_l\}}^T \mathbf{R}_{\Omega - \{\mathbf{b}_l\}}^{-1} \mathbf{d}_{\Omega - \{\mathbf{b}_l\}})^{-1} \} \quad (44)$$

$$\Omega_1 = \{l_1\}.$$

2. Band Reduction

$$\mathbf{b}_{l_j} = \arg \{ \max_{\mathbf{b}_l \in \Omega_{j-1}^c} (\mathbf{d}_{\Omega - (\Omega_{j-1} \cup \{\mathbf{b}_l\})}^T \mathbf{R}_{\Omega - (\Omega_{j-1} \cup \{\mathbf{b}_l\})}^{-1} \mathbf{d}_{\Omega - (\Omega_{j-1} \cup \{\mathbf{b}_l\})})^{-1} \} \quad (46)$$

where $\Omega_{j-1} = \{\mathbf{b}_{l_1}, \mathbf{b}_{l_2}, \dots, \mathbf{b}_{l_{j-1}}\}$ and $\Omega_{j-1}^c = \Omega - \Omega_{j-1}$.

3. If $j < n_{BS}$,

$$\Omega_j = \{\mathbf{b}_{l_1}, \mathbf{b}_{l_2}, \dots, \mathbf{b}_{l_j}\} = \Omega_{j-1} \cup \{\mathbf{b}_{l_j}\} \quad (47)$$

and go step 2. Otherwise, BS is terminated. The final set of selected bands is given by $\{\mathbf{b}_{l_1}, \mathbf{b}_{l_2}, \dots, \mathbf{b}_{l_{n_{BS}}}\}$ where n_{BS} is the number of bands needed to be selected.

It should be noted that SB-SCSC-BS is different from SF-SCSC-BS in two different aspects. First of all, SB-SCSC-BS ranks all bands individually compared to SF-SCSC-BS which augments selected bands one at a time using (39). Second, SB-SCSC-BS uses the sample correlation matrix formed by data sample vectors using all bands except the bands already selected, $\mathbf{R}_{\Omega - \Omega_j}$, while SF-SCSC-BS only

uses the selected bands in Ω_j to form the sample correlation matrix \mathbf{R}_{Ω_j} .

VII. MULTIPLE CLASS SIGNATURES-CONSTRAINED BS

The SCSC-BS presented in Section VI is designed to select an optimal subset of bands, $\Omega_{BS}^{CSI}(C_i)$ for each individual class C_i according to the number of bands, $n_{BS}^{CSI}(C_i)$ determined by its CSI of the i th class signature μ_i . As a result, for any two different classes, C_i and C_j with their class signatures μ_i and μ_j , their $n_{BS}^{CSI}(C_i)$ and $n_{BS}^{CSI}(C_j)$ will be different and so are their selected band subsets, Ω_i and Ω_j . The final band subset used for classification will be determined by the union of their selected band subsets. Unlike SCSC-BS, this section develops a MCSC-BS which selects bands for all classes not custom-designed for a particular class. Two key features that differentiate MCSC-BS from SCSC-BS are: 1) criterion to determine the number of bands, which is CE compared to CSI used by SCSC-BS and 2) constrains all class signatures simultaneously as opposed to SCSC-BS which constrains single class signature one for each class. Accordingly, CEM used to derive SCSC-BS cannot be directly applicable to MCSC-BS. Fortunately, another well-known target detection algorithm, called target-constrained interference-minimized filter (TCIMF) developed in [52] which has also been recently developed as multiple-class classification algorithm [40], [41] can be used to replace CEM for this purpose.

Suppose that $\mu_1, \mu_2, \dots, \mu_M$ are M specific class signatures of interest, which can be either provided by *a priori* knowledge or obtained by *a posteriori* knowledge, training samples, etc., from class knowledge. Suppose that $\{\mathbf{r}_i\}_{i=1}^N$ is the set of data sample vectors in a hyperspectral image where $\mathbf{r}_i = (r_{i1}, r_{i2}, \dots, r_{iL})^T$ is the i th L -dimensional data sample vector and L is the total number of spectral bands.

Now, we interpret each of $\mu_1, \mu_2, \dots, \mu_M$ as a desired signal arrival direction in adaptive beamforming in [50] and let $\mathbf{D}^M = [\mu_1 \mu_2 \dots \mu_M]$ be an $L \times M$ class signature matrix where L is the total number of band. Assume that an FIR linear filter is specified by L filter coefficients $L \times M$, denoted by an L -dimensional vector $L \times M$. The LCMV problem considered in [27], [50], and [51] can be later reformulated as the following M -class constrained optimization problem:

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \mathbf{R} \mathbf{w} \} \quad (48)$$

$$\text{s.t. } (\mathbf{D}^M)^T \mathbf{w} = \mathbf{c}^M \quad (49)$$

where $\mathbf{c}^M = (c_1, c_2, \dots, c_M)^T$ is an M -dimensional constraint vector for a general purpose and $\mathbf{R} = (1/N) \sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^T$ is the sample correlation matrix of size $L \times L$. It should be noted that each of M constraints, c_1, c_2, \dots, c_M , in (49) is used to impose on a particular class, i.e., the j th constraint, c_j is imposed on the j th class, C_j and chosen to be any arbitrary M -dimensional vector. For example, in adaptive beamforming of array signal processing \mathbf{c}^M is only used to lock signal arrivals in M desired directions. For simplicity, we can choose \mathbf{c}^M to be the M -dimensional unity vector with ones in all the M components, $\mathbf{1}^M = (1, \dots, \underbrace{1}_j, \dots, \underbrace{1}_M)^T$. In LCMV, these M signal arrival directions are specified by

M class signatures. Hence, we can also use the M -dimensional unity vector to constrain the class signature matrix \mathbf{D}^M . In this case, the optimal solution to (48) and (49) with \mathbf{c}^M replaced by $\mathbf{1}^M$ is solved by

$$\mathbf{w}^* = \mathbf{R}^{-1} \mathbf{D}^M ((\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M)^{-1} \mathbf{1}^M \quad (50)$$

where $(\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M$ is an $M \times M$ matrix. Substituting (50) into (48) yields

$$(\mathbf{w}^*)^T \mathbf{R} \mathbf{w}^* = (\mathbf{1}^M)^T ((\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M)^{-1} \mathbf{1}^M \quad (51)$$

which can be considered as a generalization of (34).

Let $\{\mathbf{r}_l^i\}_{i=1}^N$ be the total number of data sample vectors with the first l bands. Now, we consider to reformulate (48) and (49) using a partial band subset, $\Omega_l = \{\mathbf{B}_{b_1}, \mathbf{B}_{b_2}, \dots, \mathbf{B}_{b_l}\}$ instead of the full band set $\Omega = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_L\}$ and

$$\mathbf{R}_{\Omega_l} = \frac{1}{N} \sum_{i=1}^N \mathbf{r}_l^i (\mathbf{r}_l^i)^T \quad (52)$$

is the sample correlation matrix using a partial band subset consisting of the first l band images. For simplicity of notation, let $\Omega_l = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_l\} = \{\mathbf{B}_{b_1}, \mathbf{B}_{b_2}, \dots, \mathbf{B}_{b_l}\}$ (50) becomes

$$(\mathbf{w}_{\Omega_l}^*)^T \mathbf{R}_{\Omega_l} \mathbf{w}_{\Omega_l}^* = (\mathbf{1}^M)^T ((\mathbf{D}_{\Omega_l}^M)^T \mathbf{R}_{\Omega_l}^{-1} \mathbf{D}_{\Omega_l}^M)^{-1} \mathbf{1}^M. \quad (53)$$

By virtue of (48) and (49), two MCSC-BS algorithms can also be derived in a similar manner that SF-SCSC-BS and SB-SCSC-BS are derived in Sections VI-A and VI-B.

A. Sequential Feed Forward MCSC-BS

Similar to SF-SCSC-BS the following LCMV-based SF-MCSC-BS augment selected bands one at a time sequentially based on (51) in a feed-forward manner.

SF-MCSC-BS

1. Initial condition

Determine n_{BS} .

Find the first band

$$\mathbf{B}_{l_1} = \arg \left\{ \min_{\mathbf{B}_l \in \Omega} (\mathbf{1}^M)^T ((\mathbf{D}_{\mathbf{B}_l}^M)^T \mathbf{R}_{\mathbf{B}_l}^{-1} \mathbf{D}_{\mathbf{B}_l}^M)^{-1} \mathbf{1}^M \right\} \quad (54)$$

where $\mathbf{R}_{\mathbf{B}_l} = \frac{1}{N} \sum_{i=1}^N (\mathbf{r}_l^i)^2$

Let $\Omega_1 = \{l_1\}$.

2. Band Augmentation by SFBS:

$$\mathbf{B}_{l_j} = \arg \left\{ \min_{\mathbf{B}_l \in \Omega_{j-1}^c} (\mathbf{1}^M)^T ((\mathbf{D}^M)^T \mathbf{R}_{\Omega_{j-1} \cup \{\mathbf{B}_l\}}^{-1} \mathbf{D}^M)^{-1} \mathbf{1}^M \right\} \quad (55)$$

where $\Omega_{j-1}^c = \Omega - \Omega_{j-1}$.

3. If $j < n_{BS}$,

$$\Omega_j = \{\mathbf{B}_{l_1}, \mathbf{B}_{l_2}, \dots, \mathbf{B}_{l_j}\} = \Omega_{j-1} \cup \{\mathbf{B}_{l_j}\} \quad (56)$$

and go to step 2. Otherwise, BS is terminated. The final set of selected bands is given by $\Omega_{n_{BS}}$.

B. Sequential Backward MCSC-BS

In analogy with SB-SCSC-BS, we can replace Ω_l in (53) with $\Omega_l^c = \Omega - \Omega_l$ to derive

$$(\mathbf{w}_{\Omega_l^c}^*)^T \mathbf{R}_{\Omega_l^c} \mathbf{w}_{\Omega_l^c}^* = (\mathbf{1}^M)^T ((\mathbf{D}_{\Omega_l^c}^M)^T \mathbf{R}_{\Omega_l^c}^{-1} \mathbf{D}_{\Omega_l^c}^M)^{-1} \mathbf{1}^M \quad (57)$$

which measures the maximal variance caused by removing the band image \mathbf{B}_l from Ω . The resulting BS is referred to as SB-MCSC-BS described in the following algorithm.

SB-MCSC-BS

1. Initial condition:

Determine n_{BS} .

Find

$$\mathbf{B}_{l_1} = \arg \left\{ \max_{\mathbf{B}_l \in \Omega} (\mathbf{1}^M)^T ((\mathbf{D}_{\Omega - \{\mathbf{B}_l\}}^M)^T \mathbf{R}_{\Omega - \{\mathbf{B}_l\}}^{-1} \mathbf{D}_{\Omega - \{\mathbf{B}_l\}}^M)^{-1} \mathbf{1}^M \right\} \quad (58)$$

$\Omega_1 = \{\mathbf{B}_{l_1}\}$.

2. BS by SBBS

$$\mathbf{B}_{l_j} = \arg \left\{ \max_{\mathbf{B}_l \in \Omega_{j-1}^c} (\mathbf{1}^M)^T ((\mathbf{D}_{\Omega_{j-1} \cup \{\mathbf{B}_l\}}^M)^T \mathbf{R}_{\Omega_{j-1} \cup \{\mathbf{B}_l\}}^{-1} \mathbf{D}_{\Omega_{j-1} \cup \{\mathbf{B}_l\}}^M)^{-1} \mathbf{1}^M \right\} \quad (59)$$

where $\Omega_{j-1} = \{\mathbf{B}_{l_1}, \mathbf{B}_{l_2}, \dots, \mathbf{B}_{l_{j-1}}\}$ and $\Omega_{j-1}^c = \Omega - \Omega_{j-1}$.

3. If $j < n_{BS}$,

$$\Omega_j = \{\mathbf{B}_{l_1}, \mathbf{B}_{l_2}, \dots, \mathbf{B}_{l_j}\} = \Omega_{j-1} \cup \{\mathbf{B}_{l_j}\} \quad (60)$$

and go to step 2. Otherwise, BS is terminated. The final set of selected bands is given by $\Omega_{n_{BS}}$.

Finally, the discussions on the difference between SB-SCSC-BS and SF-SCSC-BS made at the end of Section VI can also be applied to SB-MCSC-BS and SF-MCSC-BS.

VIII. REAL IMAGE EXPERIMENTS

Three popular and widely used real hyperspectral images, available on the website http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes, were used for experiments, Purdue Indiana Indian Pines, Salinas, University of Pavia, Italy.

A. Purdue's Indiana Indian Pines

The first real image to be used for experiments is an agriculture scene. It is a well-known Airborne Visible Infrared Imaging Spectrometer (AVIRIS) image scene, Purdue Indiana Indian Pine test site shown in Fig. 2(a), its ground truth map in Fig. 2(b) along with different classes highlighted by various colors in Fig. 2(c). Table I also tabulates all the specific types of 16 classes with the number of data samples in parentheses collected for each class. It has a size of 145×145 pixel vectors taken from an area of mixed agriculture and forestry in Northwestern Indiana, USA, with details of band and wavelength is given in the caption. The data set to be used for experiments is obtained from

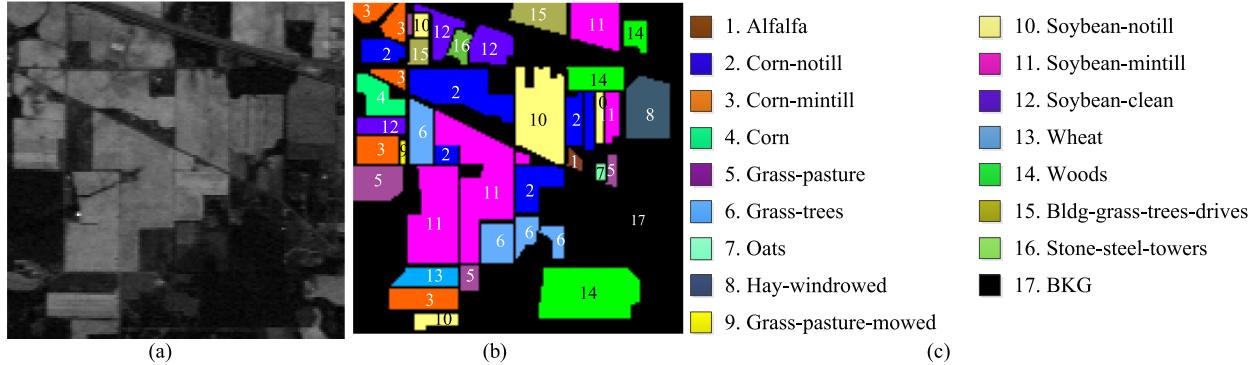


Fig. 2. AVIRIS image scene: Purdue Indiana Indian Pines test site. (a) Band 186 (2162.56 nm). (b) Ground truth map. (c) Classes by colors.

TABLE I
LABELS OF THE PURDUE INDIANA INDIAN PINES

class 1 (46)	alfalfa	class 7 (28)	grass/pasture-mowed	class 13 (205)	wheat
class 2 (1428)	corn-notill	class 8 (478)	hay-windrowed	class 14 (1265)	woods
class 3 (830)	corn-min	class 9 (20)	oats	Class 15 (386)	bldg-grass green-drives
class 4 (237)	corn	class 10 (972)	soybeans-notill	class 16 (93)	stone-steel towers
class 5 (483)	grass/pasture	class 11 (2455)	soybeans-min	class 17 (10249)	BKG
class 6 (730)	grass/trees	class 12 (593)	soybeans-clean		

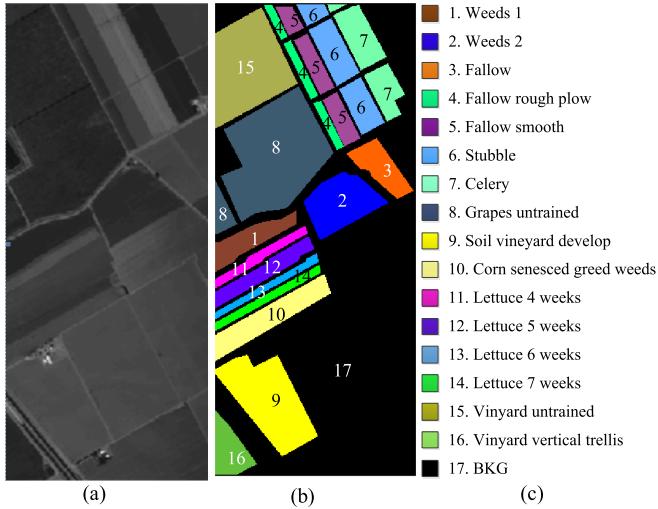


Fig. 3. Ground truth of the Salinas scene with 16 classes. (a) Salinas scene. (b) Ground truth image. (c) Classes by colors.

the website <http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/documentation.html>. It was recorded in June 1992 with 220 bands with including 20 water absorption bands (bands 104–108 and 150–163, 220).

B. Salinas

The Salinas image shown in Fig. 3 is also an AVIRIS image collected over an agriculture area in Salinas Valley, California, and with a spatial resolution of 3.7 m per pixel with a spectral resolution of 10 nm. The image cube has size a of $512 \times 217 \times 224$. This scene is very similar to the Purdue Indiana Indian Pines scene which also includes 20 water absorption bands, 108–112, 154–167, and 224. Fig. 3(b) and (c) shows the color composite of the Salinas image and the corresponding ground truth map shown

in Fig. 3(b) along with color class labels in Fig. 3(c). Table II tabulates the number of data samples (in parentheses) collected for each class among all the 16 classes.

C. University of Pavia

The third hyperspectral image data shown in Fig. 4(a) was collected by the ROSIS-03 satellite sensor over an urban area surrounding the University of Pavia, Italy. It is the size of $610 \times 340 \times 115$ with a spatial resolution of 1.3 m per pixel and a spectral coverage ranging from 0.43 to $0.86 \mu\text{m}$ with a spectral resolution of 4 nm (12 most noisy channels were removed before experiments). Nine classes of interest are considered for this image. Fig. 4(b) shows its ground-truth map along with color class labels in Fig. 4(c). Table III also tabulates the number of data samples in parentheses collected for each of the nine classes.

IX. DETERMINING NUMBER OF BANDS AND FINDING BANDS BY CLASS INFORMATION

Table IV(a)–(c) tabulates CI probabilities and $n_{BS}^{\text{CSI}}(C_i)$ of all the classes including BKG that were calculated by WCD, CD, SR, BCD and CFR for Purdue's Indian Pines, Salinas and University of Pavia, respectively, where CSI probabilities greater than 0.1 are highlighted for comparison. As shown in Table IV(a)–(c), small classes are generally ranked very high by CI probabilities compared to SR which ranks large classes with high probabilities. For example, among 16 classes in the Purdue data each of the five smallest classes, class 9 (20 samples), class 7 (28 samples), class 1 (46 samples), class 16 (93 samples) and class 13 (205 samples) is ranked with CSI probabilities greater than 0.1 (boldfaced) by at least two of four CI measures as opposed to class 11 (2455 samples), class 2 (1428 samples) and class 14 (1265 samples) which are ranked by SR with probabilities greater than 0.1 but

TABLE II
LABELS OF SALINAS

class 1 (2009)	Brocoli_green_weeds_1	class 10 (3278)	Corn_senesced_green_weeds
class 2 (3726)	Brocoli_green_weeds_2	class 11 (1068)	Lettuce_romaine_4wk
class 3 (1976)	Fallow	class 12 (1927)	Lettuce_romaine_5wk
class 4 (1394)	Fallow_rough_plow	class 13 (916)	Lettuce_romaine_6wk
class 5 (2678)	Fallow_smooth	class 14 (1070)	Lettuce_romaine_7wk
class 6 (3959)	Stubble	Class 15 (7268)	Vinyard_untrained
class 7 (3579)	Celery	class 16 (1807)	Vinyard_vertical_trellis
class 8 (11271)	Grapes_untrained	class 17 (56975)	BKG
class 9 (6203)	Soil_vinyard Develop		

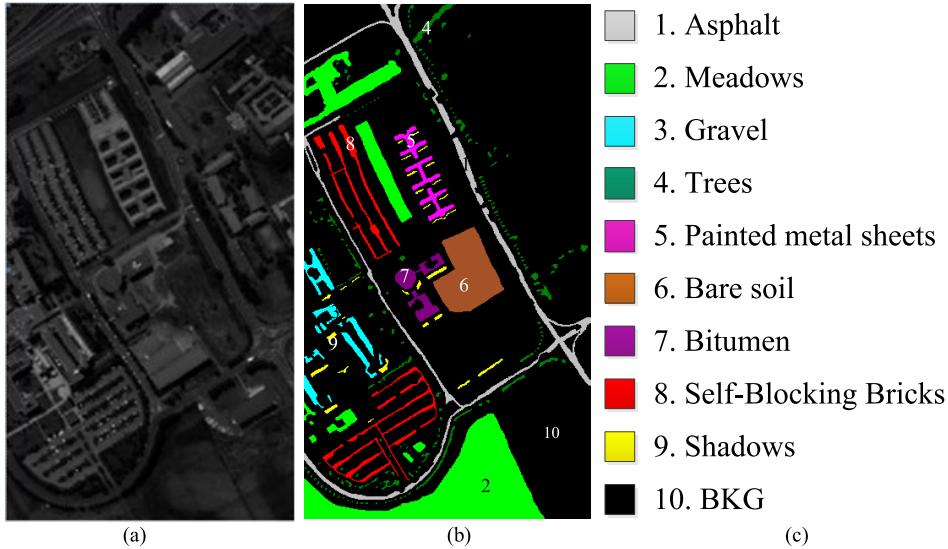


Fig. 4. Ground truth of the University of Pavia scene with nine classes. (a) University of Pavia scene. (b) Ground truth map. (c) Classes by colors.

TABLE III
LABELS OF PAVIA

class 1 (6631)	Asphalt	class 5 (1345)	Painted metal sheets	Class 9 (947)	Shadows
class 2 (18649)	Meadows	class 6 (5029)	Bare Soil	Class 10 (164624)	BKG
class 3 (2099)	Gravel	class 7 (1330)	Bitumen		
class 4 (3064)	Trees	class 8 (3682)	Self-Blocking Bricks		

very low probabilities by all the four CI measures. Using the CI probabilities in Table IV(a)–(c) and (19) and (20), we can further find the number of training samples required for each class.

Table V(a)–(c) tabulates the number of training and test samples for each of the three image scenes used for experiments. Interestingly, CI IC except SR provides a guideline for how to determine the number of training samples for each class. For example, according to Table V(a), the number of training samples selected for the four smallest classes in the Purdue data (i.e., classes 9, 7, 1, 16) with less than 100 data samples was determined by CI IC, WCD, CD, BCD and CFR as 50% of each class size. Interestingly, the same numbers of training samples selected for these four classes by EPF in [47] happened to be also 50% of each class size. But unfortunately, there was no explanation given in [47]. Our proposed CI IC in Table V(a) indeed offer such justification.

Table VI calculates class entropies of Purdue’s Indian Pines, Salinas and University of Pavia. Table VII calculates the number of bands, n_{BS} , by class entropies in Table VI with

BKG excluded and included where n_{BS} is determined by $n_{BS}^{CSI}(C_i)$ in Table IV(a)–(c) and $n_{BS} = \lceil n_{BS}^{CE} \times M \rceil$ with $\lceil x \rceil$ defined as the smallest integer $\geq x$. According to $n_{BS}^{CSI}(C_i)$ found in Table IV(a)–(c) and $n_{BS} = \lceil n_{BS}^{CE} \times M \rceil$ in Table VII with BKG included, Table VIII(a)–(c) list bands selected by SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS and SB-MCSC-BS using three intraclass ICs, WCD, CD, SR and two interclass ICs, BCD, CFR all the three image scenes with and without BKG.

X. EXPERIMENTS CONDUCTED ON PURDUE’S INDIAN PINES SCENE

The experiments conducted in this section constitute a very important part of this paper since the results provide many interesting findings that have never been explored in the past and reported in the literature. It is also noted that CI is completely determined by data itself, not by a classifier. Hence, the conclusions drawn from the conducted experiments can also be applied to any classifier. In addition, for the three image scenes in Section VIII, the same experiments

TABLE IV

- (a) CI PROBABILITIES AND $n_{BS}^{CSI}(C_i)$ CALCULATED BY BCD, WCD, CD, CFR, AND SR USING (25) FOR PURDUE'S INDIAN PINES.
 (b) CI PROBABILITIES AND $n_{BS}^{CSI}(C_i)$ CALCULATED BY BCD, WCD, CD, CFR, AND SR USING (25) FOR SALINAS.
 (c) CI PROBABILITIES AND $n_{BS}^{CSI}(C_i)$ CALCULATED BY BCD, WCD, CD, CFR, AND SR
 USING (25) FOR UNIVERSITY OF PAVIA

(a)

class	CI probabilities					$n_{BS}^{CSI}(C_i)$ in (25)				
	Intra-class IC			Inter-class IC		Intra-class IC			Inter-class IC	
	WCD	CD	SR	BCD	CFR	WCD	CD	SR	BCD	CFR
class 1 (46)	0.1890	0.2059	0.0045	0.0414	0.0504	2	2	6	4	3
class 2 (1428)	0.0122	0.0139	0.1393	0.0613	0.0165	5	5	2	3	5
class 3 (830)	0.0175	0.0187	0.0810	0.0192	0.0020	5	4	3	4	7
class 4 (237)	0.0071	0.0079	0.0231	0.0754	0.0145	5	5	4	3	5
class 5 (483)	0.0087	0.0096	0.0471	0.0821	0.0244	5	5	4	3	4
class 6 (730)	0.0315	0.0296	0.0712	0.0466	0.0320	4	4	3	4	4
class 7 (28)	0.2065	0.2166	0.0027	0.0495	0.0739	2	2	6	4	3
class 8 (478)	0.0709	0.0764	0.0466	0.0414	0.0504	3	3	4	4	3
class 9 (20)	0.1407	0.1319	0.0020	0.0466	0.0320	2	3	7	4	4
class 10 (972)	0.0254	0.0261	0.0948	0.0201	0.0014	4	4	3	4	7
class 11 (2455)	0.0208	0.0222	0.2395	0.0192	0.0020	4	4	2	4	7
class 12 (593)	0.0080	0.0083	0.0579	0.0201	0.0014	5	5	3	4	7
class 13 (205)	0.2095	0.1766	0.0200	0.0930	0.4159	2	2	4	3	1
class 14 (1265)	0.0231	0.0247	0.1234	0.0821	0.0244	4	4	3	3	4
class 15 (386)	0.0141	0.0132	0.0377	0.0492	0.0135	5	5	4	4	5
class 16 (93)	0.0150	0.0184	0.0091	0.2529	0.2454	5	4	5	2	2
sum	N/A					57	62	61	57	71

(b)

class	CI probabilities					$n_{BS}^{CSI}(C_i)$ in (25)				
	Intra-class IC			Inter-class IC		Intra-class IC			Inter-class IC	
	WCD	CD	SR	BCD	CFR	WCD	CD	SR	BCD	CFR
class 1 (2009)	0.1476	0.3269	0.0371	0.0631	0.0549	2	2	4	3	3
class 2 (3726)	0.0225	0.0433	0.0688	0.0631	0.0549	4	4	3	3	3
class 3 (1976)	0.0677	0.0459	0.0365	0.0279	0.0244	3	4	4	4	4
class 4 (1394)	0.0796	0.0245	0.0258	0.1011	0.3156	3	4	4	3	2
class 5 (2678)	0.0966	0.0541	0.0495	0.0419	0.0495	3	3	4	4	4
class 6 (3959)	0.0173	0.0292	0.0731	0.1867	0.2024	5	4	3	2	2
class 7 (3579)	0.0308	0.0406	0.0661	0.1253	0.1442	4	4	3	3	2
class 8 (11271)	0.0248	0.0169	0.2082	0.0140	0.0021	4	5	2	5	7
class 9 (6203)	0.1656	0.1452	0.1146	0.0251	0.0189	2	2	3	4	4
class 10 (3278)	0.0114	0.0094	0.0606	0.0479	0.0170	5	5	3	4	5
class 11 (1068)	0.0571	0.0545	0.0197	0.0251	0.0189	3	3	4	4	4
class 12 (1927)	0.1272	0.0879	0.0356	0.0279	0.0244	3	3	4	4	4
class 13 (916)	0.0946	0.0760	0.0169	0.0641	0.0295	3	3	5	3	4
class 14 (1070)	0.0156	0.0176	0.0198	0.1002	0.0262	5	5	4	3	4
class 15 (7268)	0.0369	0.0241	0.1343	0.0140	0.0021	4	4	3	5	7
class 16 (1807)	0.0048	0.0039	0.0334	0.0724	0.0150	6	6	4	3	5
sum	N/A					59	61	59	57	64

(c)

class	CI probabilities					$n_{BS}^{CSI}(C_i)$ in (25)				
	Intra-class IC			Inter-class IC		Intra-class IC			Inter-class IC	
	WCD	CD	SR	BCD	CFR	WCD	CD	SR	BCD	CFR
class 1 (6631) (asphalt)	0.0489	0.0462	0.1550	0.0353	0.0149	4	4	2	4	5
class 2 (18649) (meadows)	0.0273	0.0409	0.4360	0.0677	0.0137	4	4	1	3	5
class 3 (2099) (gravel)	0.0887	0.1503	0.0491	0.0237	0.0086	3	2	4	4	5
class 4 (3064) (trees)	0.0229	0.0529	0.0716	0.1366	0.0658	4	3	3	2	3
class 5 (1345) (metal sheets)	0.0060	0.0381	0.0314	0.3834	0.2384	6	4	4	1	2
class 6 (5029) (bare soil)	0.0173	0.0263	0.1176	0.0677	0.0137	5	4	3	3	5
class 7 (1330) (bitumen)	0.3066	0.3644	0.0311	0.0353	0.0149	2	2	4	4	5
class 8 (3682) (bricks)	0.1399	0.2669	0.0861	0.0237	0.0086	2	2	3	4	5
class 9 (947) (shadow)	0.3425	0.0141	0.0221	0.2264	0.6213	2	5	4	2	1
sum	N/A					32	30	28	27	36

conducted for one image scene can also be applied to the other two image scenes. Hence, in this section, we will only focus our experiments and discussions on the Purdue Indiana

Indian Pines scene. For those who are interested in Salinas and University of Pavia scenes the results in Tables IV(b) and (c), V(b) and (c), and VI and VII and the bands selected

TABLE V

(a) NUMBERS OF TRAINING AND TEST SAMPLES FOR PURDUE'S DATA. (b) NUMBERS OF TRAINING AND TEST SAMPLES FOR SALINAS.
 (c) NUMBERS OF TRAINING AND TEST SAMPLES FOR UNIVERSITY OF PAVIA

class	# of class samples	# of training samples					
		EPF [47]	Intra-class IC			Inter-class IC	
			w/o CI	WCD	CD	SR	BCD
class 1 (46)	46	25	23	23	5	23	23
class 2 (1428)	1428	83	43	47	142	98	73
class 3 (830)	830	78	61	63	83	31	9
class 4 (237)	237	68	25	27	24	119	64
class 5 (483)	483	79	31	33	49	132	108
class 6 (730)	730	78	110	100	73	75	142
class 7 (28)	28	14	14	14	3	14	14
class 8 (478)	478	66	239	239	48	67	223
class 9 (20)	20	10	10	10	2	10	10
class 10 (972)	972	81	89	88	97	33	10
class 11 (2455)	2455	99	73	75	245	31	25
class 12 (593)	593	73	28	28	59	33	7
class 13 (205)	205	70	103	103	21	103	103
class 14 (1265)	1265	90	80	84	126	132	108
class 15 (386)	386	65	50	45	39	79	60
class 16 (93)	93	46	47	47	10	47	47
Total	10249	1025	1026	1026	1026	1027	1026

class	# of class samples	# of training samples					
		EPF in [47]	Intra-class IC			Inter-class IC	
			w/o CI	WCD	CD	SR	BCD
class 1 (2009)	2009	67	126	285	40	56	46
class 2 (3726)	3726	67	38	38	74	56	46
class 3 (1976)	1976	67	58	40	40	25	20
class 4 (1394)	1394	69	68	22	28	89	259
class 5 (2678)	2678	67	83	47	54	37	41
class 6 (3959)	3959	67	40	40	79	163	166
class 7 (3579)	3579	68	36	36	71	110	119
class 8 (11271)	11271	69	113	113	224	113	113
class 9 (6203)	6203	68	141	127	123	63	63
class 10 (3278)	3278	68	33	33	65	42	33
class 11 (1068)	1068	68	49	48	22	22	16
class 12 (1927)	1927	67	109	77	39	25	20
class 13 (916)	916	67	81	67	19	56	25
class 14 (1070)	1070	67	14	16	22	88	22
class 15 (7268)	7268	70	73	73	145	73	73
class 16 (1807)	1807	67	19	19	36	64	19
Total	54129	1083	1081	1081	1081	1082	1081

class	# of class samples	# of training samples					
		EPF in [47]	Intra-class IC			Inter-class IC	
			w/o CI	WCD	CD	SR	BCD
class 1 (6631)	6631	286	188	133	398	129	145
class 2 (18649)	18649	286	187	187	1117	246	187
class 3 (2099)	2099	285	340	433	126	87	84
class 4 (3064)	3064	285	88	153	184	496	640
class 5 (1345)	1345	285	23	110	81	673	673
class 6 (5029)	5029	285	66	76	302	246	134
class 7 (1330)	1330	285	665	665	80	129	145
class 8 (3682)	3682	285	535	768	221	87	84
class 9 (947)	947	285	474	41	57	474	474
Total	42776	2567	2566	2566	2566	2567	2566

TABLE VI

CLASS ENTROPIES CALCULATED BY WCD, CD, SR, BCD, AND CFR FOR PURDUE'S INDIAN PINES, SALINAS, AND UNIVERSITY OF PAVIA

CI criterion	BKG excluded					BKG included				
	Intra-class criteria			Inter-class criteria		Intra-class criteria			Inter-class criteria	
	WCD	CD	SR	BCD	CFR	WCD	CD	SR	BCD	CFR
Indian Pines	2.164	2.189	2.326	2.500	1.835	2.191	2.216	1.827	2.575	1.872
Salinas	2.456	2.269	2.512	2.534	2.122	2.465	2.276	1.917	2.579	2.083
Pavia	1.653	1.714	1.752	1.754	1.142	1.685	1.752	0.870	1.798	1.084

TABLE VII

 $\lceil n_{BS}^{CE} \times M \rceil$ DETERMINED BY CLASS ENTROPIES CALCULATED BY WCD, CD, SR, BCD, AND CFR FOR PURDUE'S INDIAN PINES, SALINAS, AND UNIVERSITY OF PAVIA

CI criterion	BKG excluded					BKG included				
	Intra-class criteria			Inter-class criteria		Intra-class criteria			Inter-class criteria	
	WCD	CD	SR	BCD	CFR	WCD	CD	SR	BCD	CFR
Indian Pines	$\lceil 34.62 \rceil$	$\lceil 35.02 \rceil$	$\lceil 37.22 \rceil$	40	$\lceil 29.36 \rceil$	$\lceil 37.24 \rceil$	$\lceil 37.66 \rceil$	$\lceil 31.06 \rceil$	$\lceil 43.77 \rceil$	$\lceil 31.83 \rceil$
Salinas	$\lceil 39.30 \rceil$	$\lceil 36.30 \rceil$	$\lceil 40.19 \rceil$	$\lceil 40.54 \rceil$	$\lceil 33.96 \rceil$	$\lceil 41.90 \rceil$	$\lceil 38.68 \rceil$	$\lceil 32.58 \rceil$	$\lceil 43.84 \rceil$	$\lceil 35.41 \rceil$
U. of Pavia	$\lceil 14.87 \rceil$	$\lceil 15.42 \rceil$	$\lceil 15.76 \rceil$	$\lceil 15.79 \rceil$	$\lceil 10.28 \rceil$	$\lceil 16.85 \rceil$	$\lceil 17.52 \rceil$	$\lceil 8.70 \rceil$	$\lceil 17.98 \rceil$	$\lceil 10.84 \rceil$

in Table VIII(b) and (c) should provide sufficient information for them to carry out all details of necessary experiments without any difficulty. Since similar conclusions can also be drawn, their results are not included here due to limited space.

According to [47], a comprehensive and comparative analysis was conducted among the most existing spectral-spatial techniques and the four EPF-based techniques, EPF-B-c, EPF-G-c, EPF-B-g, and EPF-G-g were shown to be the best classification techniques where "B" and "G" are used to specify bilateral filter and guided filter, respectively, and "g" and "c" indicate that the first principal component and color composite of three principal components are used as reference images [47]. Accordingly, in the following experiments the four EPF-based methods in [47], which can be considered as classification without CI (w/o CI) will be evaluated in comparison with the five proposed CI IC in terms of CI-overall accuracy (CI-OA). P_{CI-OA} in (20) along with the commonly used AA, P_{AA} in (19) OA, P_{OA} in (17) and P_{CI-PR} in (22). The computer environment used for experiments was specified by Intel Xeon E5-2650 2.6 GHz, 64 GB, 1600 MHz.

As for BS, there are many techniques have been proposed in the literature. A detailed comparative study and analysis was conducted in [25] where most updated and recent BS methods were discussed and compared for HSIC using P_{OA} and P_{PR} as criteria to measure the classification performance. To avoid unnecessary redundant experiments, the results in [25] were used as references to compare with the results obtained by our proposed methods. Nevertheless, it should be noted that the results of P_{PR} in [39] and [41] were calculated with BKG excluded, in which case, the values of P_{PR} in [39] and [40] is generally higher than P_{PR} with BKG included calculated in the following experiments. The prime reason is that the two classifiers (ICEM and ILCMV) and two BS methods (SCSC-BS and MSCS-BS) are all designed by inverting the sample correlation matrix \mathbf{R} to effectively suppress BKG compared to EPF-based which cannot be shown in the experiments.

Now, we implemented ICEM and ILCMV using bands selected in Table VIII(a)–(c) to perform classification. Table IX tabulates the P_A , P_{OA} , P_{AA} , P_{PR} , P_{CI-OA} , and P_{CI-PR} calculated by ICEM and ILCMV using bands selected by SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, and SB-MCSC-BS for Purdue's data where the three intraclass ICs, WCD, CD, SR and two interclass ICs, BCD, CFR were used as CI measures to select bands and calculate P_{CI-OA} and P_{CI-PR} . Table X also tabulates the P_A , P_{OA} , P_{AA} , P_{PR} , P_{CI-OA} , and P_{CI-PR} calculated by four EPF-based methods [47], ICEM and ILCMV using full bands for the Purdue data. Comparing the results in Table IX(a)–(e) to that in Table X, ICEM and ILCMV using bands selected by SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, SB-MCSC-BS in Table VIII(a) generally performed better than their counterparts using full bands in all the three measures, P_{AA} , P_{OA} , and P_{PR} . From Table X, ICEM and ILCMV clearly outperformed four EPF-based methods in terms of all the three measures, P_{AA} , P_{OA} , and P_{PR} , specifically, P_{OA} (ranging from 5% to 7% improvements) and P_{PR} (nearly 37% improvement). For further comparison, the results in [25, Table 3] were used for comparison where the highest values of P_{OA} obtained by using bands selected by various BS methods ranged from 94.91 to 95.89. Table IX(a)–(e) shows that the values of P_{CI-OA} generally performed better than P_{OA} without factoring CI into P_{OA} calculation, i.e., with CI = SR. Specifically, the P_{CI-OA} produced by the four CI IC, WCD, CD, BCD, and CFR were higher than 97% and were better than P_{OA} obtained in [25, Table 3] with 1.5% to 2.5% improvements. It is also noted from Table X that the performance of the EPF-based methods was much worse than that reported in [47] because the results in Table X were obtained by including all 20 water bands which were removed in [46]. As for P_{PR} its values in Table IX(a)–(e) were lower than the values in [25, Table 3]. This is also because the P_{PR} calculated in Table III was only based on 16 classes according to the ground truth without including BKG to account for its effect on 16 classes.

TABLE VIII

(a) BANDS SELECTED BY SF-SCSC-BS AND SB-SCSC-BS USING CSI IN TABLE IV(a) AND SF-MCSC-BS AND SB-MCSC-BS USING CE IN TABLE VII FOR PURDUE'S DATA. (b) BANDS SELECTED BY SF-SCSC-BS AND SB-SCSC-BS USING CSI IN TABLE IV(b) AND SF-MCSC-BS AND SB-MCSC-BS USING CE IN TABLE VII FOR SALINAS. (c) BANDS SELECTED BY SF-SCSC-BS AND SB-SCSC-BS USING CSI IN TABLE IV(c) AND SF-MCSC-BS AND SB-MCSC-BS USING CE IN TABLE VII FOR UNIVERSITY OF PAVIA

		(a)																			
CI measures	CI-based BS Methods	Selected bands																			
WCD	SF-SCSC-BS	3 11 12 17 19 20 23 24 25 26 28 29 31 33 34 35 36 37 38 42 44 53 55 58 61 62 75 85 86 90 94 97 98 102 107 111 121 122 135 154 157 168 169 179 181 191 197 203 212 215 216																			
	SB-SCSC-BS	3 10 12 14 15 16 17 18 20 21 28 29 31 34 35 36 37 38 56 57 59 60 108 121 126 139 140 167 168 169 173 174 189 190 191 192 193 201 202 203 204 205 211																			
	SF-MCSC-BS	29 30 25 33 24 23 40 203 26 86 15 46 116 65 173 31 93 88 77 3 192 121 115 101 191 96 42 4 59 22 123 172 155 12 213 165 124																			
	SB-MCSC-BS	5 3 29 35 4 96 11 12 158 42 121 101 213 95 186 31 165 6 87 93 28 180 178 65 1 194 210 13 77 145 202 211 172 141 99 100 105 24																			
CD	SF-SCSC-BS	3 11 12 17 19 20 23 24 25 26 28 29 31 33 34 35 36 37 38 42 44 53 55 58 61 62 75 85 90 94 97 98 107 121 122 135 154 168 169 179 181 191 197 203 212 215 216																			
	SB-SCSC-BS	3 10 12 14 15 16 17 18 20 21 28 29 31 34 35 36 37 38 56 57 59 60 121 139 140 167 168 169 173 174 189 190 191 192 193 201 202 203 204 205 211																			
	SF-MCSC-BS	29 30 25 33 24 23 40 203 26 86 15 46 116 65 173 31 93 88 77 3 192 121 115 101 191 96 42 4 59 22 123 172 155 12 213 165 124																			
	SB-MCSC-BS	5 3 29 35 4 96 11 12 158 42 121 101 213 95 186 31 165 6 87 93 28 180 178 65 1 194 210 13 77 145 202 211 172 141 99 100 105 24																			
SR	SF-SCSC-BS	1 3 11 12 17 19 23 24 25 26 28 29 33 34 36 37 38 42 44 55 58 61 62 75 77 85 86 93 94 97 98 102 107 111 121 122 123 141 154 157 168 173 179 181 203 212 216																			
	SB-SCSC-BS	3 12 14 15 16 17 18 20 21 28 29 31 34 35 36 37 38 56 57 60 108 121 126 139 140 167 168 169 174 183 184 189 190 191 193 201 202 203 204 205 206 208																			
	SF-MCSC-BS	29 30 25 33 24 23 40 203 26 86 15 46 116 65 173 31 93 88 77 3 192 121 115 101 191 96 42 4 59 22 123																			
	SB-MCSC-BS	5 3 29 35 4 96 11 12 158 42 121 101 213 95 186 31 165 6 87 93 28 180 178 65 1 194 210 13 77 145 202 211																			
BCD	SF-SCSC-BS	3 12 17 19 20 23 24 25 26 28 29 34 36 37 38 42 44 55 58 62 75 85 86 90 94 97 98 102 107 111 121 122 154 157 168 179 181 191 212 216																			
	SB-SCSC-BS	3 12 15 16 17 18 20 21 28 29 31 34 35 36 37 38 56 57 60 108 121 126 139 140 167 68 69 173 174 189 190 191 192 193 201 202 203 211																			
	SF-MCSC-BS	29 30 25 33 24 23 40 203 26 86 15 46 116 65 173 31 93 88 77 3 192 121 115 101 191 96 42 4 59 22 123 172 155 12 213 165 124 178 35 130 194 76 180																			
	SB-MCSC-BS	5 3 29 35 4 96 11 12 158 42 121 101 213 95 186 31 165 6 87 93 28 180 178 65 1 194 210 13 77 145 202 211 172 141 99 100 105 24 34 36 59 81 120 130																			
CFR	SF-SCSC-BS	3 5 11 17 19 20 23 24 25 26 28 29 31 34 35 36 37 38 42 44 53 55 58 61 62 75 86 90 94 97 98 107 111 114 121 122 147 154 157 168 169 173 179 181 190 191 197 201 212 215 216																			
	SB-SCSC-BS	3 12 14 15 16 17 18 20 21 28 29 31 34 35 36 37 38 56 57 59 60 121 126 139 140 166 167 168 169 172 173 174 175 187 188 189 190 191 192 193 201 202 203 211																			
	SF-MCSC-BS	29 30 25 33 24 23 40 203 26 86 15 46 116 65 173 31 93 88 77 3 192 121 115 101 191 96 42 4 59 22 123																			
	SB-MCSC-BS	5 3 29 35 4 96 11 12 158 42 121 101 213 95 186 31 165 6 87 93 28 180 178 65 1 194 210 13 77 145 202 211																			

TABLE VIII

(Continued.) (a) BANDS SELECTED BY SF-SCSC-BS AND SB-SCSC-BS USING CSI IN TABLE IV(a) AND SF-MCSC-BS AND SB-MCSC-BS USING CE IN TABLE VII FOR PURDUE'S DATA. (b) BANDS SELECTED BY SF-SCSC-BS AND SB-SCSC-BS USING CSI IN TABLE IV(b) AND SF-MCSC-BS AND SB-MCSC-BS USING CE IN TABLE VII FOR SALINAS. (c) BANDS SELECTED BY SF-SCSC-BS AND SB-SCSC-BS USING CSI IN TABLE IV(c) AND SF-MCSC-BS AND SB-MCSC-BS USING CE IN TABLE VII FOR UNIVERSITY OF PAVIA

CI measures	CI-based BS Methods	(b)																			
		Selected bands																			
WCD	SF-SCSC-BS	4 5 13 14 15 23 29 33 37 38 39 40 42 44 48 49 54 65 75 76 79 80 81 88 106 114 124 138 141 151 171 173 176 187 190 193 197 203 220 221 223																			
	SB-SCSC-BS	4 8 9 10 11 23 24 26 27 36 37 38 39 40 41 42 56 58 59 60 65 85 152 174 176 177 178 181 195 196 197																			
	SF-MCSC-BS	44 45 25 33 82 97 156 104 223 202 168 43 108 187 111 93 183 5 99 119 207 48 106 173 22 196 120 17 90 219 192 175 21 12 132 37 144 138 70 52 220 195																			
	SB-MCSC-BS	175 174 173 176 120 121 117 220 219 84 106 178 177 85 83 170 218 223 12 10 11 6 9 94 65 64 44 221 214 217 216 215 222 179 224 180 169 115 203 135 137 136																			
CD	SF-SCSC-BS	4 5 13 14 15 23 29 33 35 37 38 39 40 42 44 48 49 54 63 65 75 76 79 80 81 88 106 114 124 138 141 151 171 172 173 176 187 190 193 197 203 220 221 223																			
	SB-SCSC-BS	4 9 10 11 23 24 26 27 36 37 38 39 40 41 42 56 57 58 59 60 65 85 152 178 181 193 195 196 197																			
	SF-MCSC-BS	44 45 25 33 82 97 156 104 223 202 168 43 108 187 111 93 183 5 99 119 207 48 106 173 22 196 120 17 90 219 192 175 21 12 132 37 144 138 70																			
	SB-MCSC-BS	175 174 173 176 120 121 117 220 219 84 106 178 177 85 83 170 218 223 12 10 11 6 9 94 65 64 44 221 214 217 216 215 222 179 224 180 169 115 203																			
SR	SF-SCSC-BS	4 5 13 14 15 29 33 37 38 39 40 41 42 44 48 49 54 63 64 65 75 76 79 80 81 84 106 107 114 117 124 141 151 171 172 173 176 187 188 190 197 203 220 221 223																			
	SB-SCSC-BS	4 9 10 11 26 27 36 37 38 39 40 41 56 60 65 85 152 172 174 175 176 177 178 181 187 193 195 196 197 198																			
	SF-MCSC-BS	44 45 25 33 82 97 156 104 223 202 168 43 108 187 111 93 183 5 99 119 207 48 106 173 22 196 120 17 90 219 192 175 21																			
	SB-MCSC-BS	175 174 173 176 120 121 117 220 219 84 106 178 177 85 83 170 218 223 12 10 11 6 9 94 65 64 44 221 214 217 216 215 222																			
BCD	SF-SCSC-BS	3 12 14 15 16 17 18 20 21 28 29 31 34 35 36 37 38 56 57 60 108 121 126 139 140 167 168 169 174 183 184 189 190 191 193 201 202 203 204 205 206 208																			
	SB-SCSC-BS	4 8 9 10 11 24 26 27 36 37 38 39 40 41 56 57 58 59 60 65 85 117 152 172 174 175 176 177 178 181 187 193 195 196 197 198																			
	SF-MCSC-BS	44 45 25 33 82 97 156 104 223 202 168 43 108 187 111 93 183 5 99 119 207 48 106 173 22 196 120 17 90 219 192 175 21 12 132 37 144 138 70 52 220 195 32 28																			
	SB-MCSC-BS	175 174 173 176 120 121 117 220 219 84 106 178 177 85 83 170 218 223 12 10 11 6 9 94 65 64 44 221 214 217 216 215 222 179 224 180 169 115 203 135 137 136 134 138																			
CFR	SF-SCSC-BS	4 6 10 13 14 15 18 23 29 32 33 35 37 38 39 40 42 44 48 49 58 63 65 75 76 79 80 81 83 84 88 106 107 114 117 124 141 151 171 172 173 176 187 188 190 193 197 203 220 221 223																			
	SB-SCSC-BS	4 8 9 10 11 12 23 24 26 27 36 37 38 39 40 41 49 54 55 56 57 58 59 60 65 85 117 152 172 174 175 176 177 181 187 193 195 196 197 198																			
	SF-MCSC-BS	44 45 25 33 82 97 156 104 223 202 168 43 108 187 111 93 183 5 99 119 207 48 106 173 22 196 120 17 90 219 192 175 21 12 132 37																			
	SB-MCSC-BS	175 174 173 176 120 121 117 220 219 84 106 178 177 85 83 170 218 223 12 10 11 6 9 94 65 64 44 221 214 217 216 215 222 179 224 180																			

Several intriguing and interesting conclusions can be made on observations from Tables IX(a)–(e) and X.

- 1) In general, ICEM performed better than ILCMV with two major reasons. One is that ICEM is designed to

classify one class at a time by constraining a single class signature while suppressing all other class signatures compared to ILCMV which classify all classes of interest simultaneously by constraining all the class

TABLE VIII

(Continued.) (a) BANDS SELECTED BY SF-SCSC-BS AND SB-SCSC-BS USING CSI IN TABLE IV(a) AND SF-MCSC-BS AND SB-MCSC-BS USING CE IN TABLE VII FOR PURDUE'S DATA. (b) BANDS SELECTED BY SF-SCSC-BS AND SB-SCSC-BS USING CSI IN TABLE IV(b) AND SF-MCSC-BS AND SB-MCSC-BS USING CE IN TABLE VII FOR SALINAS. (c) BANDS SELECTED BY SF-SCSC-BS AND SB-SCSC-BS USING CSI IN TABLE IV(c) AND SF-MCSC-BS AND SB-MCSC-BS USING CE IN TABLE VII FOR UNIVERSITY OF PAVIA

CI measures	CI-based BS Methods	(c)
		Selected bands
WCD	SF-SCSC-BS	1 3 4 6 15 21 27 30 32 40 48 56 58 59 60 61 65 67 70 72 75 85 87 98 102 103
	SB-SCSC-BS	1 25 28 30 43 45 58 59 64 65 69 78 82 83 84 85 86 100 102 103
	SF-MCSC-BS	16 59 20 63 1 65 66 56 11 64 12 13 18 58 67 72 38
	SB-MCSC-BS	84 85 86 1 66 2 65 87 3 83 101 67 68 12 14 11 61
CD	SF-SCSC-BS	1 3 4 6 15 27 30 32 48 58 59 60 61 65 67 70 72 75 84 85 87 89 98 102 103
	SB-SCSC-BS	1 2 3 25 28 30 43 45 65 69 78 82 83 84 85 86 100 102 103
	SF-MCSC-BS	16 59 20 63 1 65 66 56 11 64 12 13 18 58 67 72 38 69
	SB-MCSC-BS	84 85 86 1 66 2 65 87 3 83 101 67 68 12 14 11 61 47
SR	SF-SCSC-BS	1 3 4 6 15 27 30 32 37 48 56 5 58 59 60 62 65 66 67 72 75 85 87 89 102
	SB-SCSC-BS	1 2 3 25 28 30 43 45 64 65 67 78 84 85 86 102 103
	SF-MCSC-BS	16 59 20 63 1 65 66 56 11
	SB-MCSC-BS	84 85 86 1 66 2 65 87 3
BCD	SF-SCSC-BS	1 3 4 15 27 30 32 37 56 57 58 59 62 65 66 72 75 85 87 98 102 103
	SB-SCSC-BS	1 25 28 30 45 64 65 67 78 83 84 85 100
	SF-MCSC-BS	16 59 20 63 1 65 66 56 11 64 12 13 18 58 67 72 38 69
	SB-MCSC-BS	84 85 86 1 66 2 65 87 3 83 101 67 68 12 14 11 61 47
CFR	SF-SCSC-BS	1 3 4 15 27 30 37 40 48 56 57 58 59 60 61 62 65 66 67 70 72 75 82 85 87 98 102 103
	SB-SCSC-BS	1 25 28 30 43 45 58 64 65 67 69 76 78 82 83 84 85 86 100 102
	SF-MCSC-BS	16 59 20 63 1 65 66 56 11 64 12
	SB-MCSC-BS	84 85 86 1 66 2 65 87 3 83 101

signatures and only suppressing BKG. As a result, ICEM suppressed effects resulting from BKG and all classes other than the particular class which is currently being classified. The other is that the bands selected by SCSC-BS are specifically designed for a particular class to be classified, whereas the bands selected by MCSC-BS are designed for all the classes of interest, not a particular class as SCSC-BS does.

- 2) There are three different versions of implementing ICEM. One is implemented in [39] which updates the mean of the class currently being classified and feeds back its Gaussian filtered classification map after each iteration. In order to improve its performance, it must use BS and nonlinear expansion (BSNE) as it was done in [39]. A second version is to retain the spatial locations of ground truth class samples and update the mean of the class currently being classified and feeds back its Gaussian filtered classification map after each iteration. A third version is the same as the second version but does not feed back the Gaussian filtered classification map for each class after each iteration as the second version does. Instead, it waits until all the classes are processed by CEM and feeds back all of the Gaussian filtered CEM-classification maps after each iteration.

In this case, the ICEM works like ILCMV except that the ILCMV classifies all the classes together simultaneously compared to ICEM which classifies one class at a time. Since both versions use the ground truth to update class means, it does not need BSNE. It also turns out that the best version of ICEM is the 3rd version which implements CEM in a similar manner as ILCMV does. In this case, this version of ICEM can be viewed as a class-independent version of ILCMV. The worst version is the 1st version which updates classified samples not ground truth samples after each iteration. Accordingly, this version of ICEM can be considered as *a posteriori* ICEM as opposed to the 2nd and 3rd versions which can be regarded as *a priori* ICEM.

- 3) Interestingly, if we examine Tables IX(a)–(e) and X, the values of P_{OA} via all the four CI measures along with SR for all classes are very high. But this is not true for P_{PR} . As a matter of fact, P_{PR} using the five CI measures performed comparably for larger classes such as the three largest classes, 2, 11, and 14 but completely in an opposite manner for smaller classes such as the three smallest classes, 1, 7, and 9, all of which have less than 50 data samples, that is, the smaller the class is, the lower the P_{PR} is. These results demonstrate that P_{PR}

TABLE IX

(a) P_A , P_{OA} , P_{CI-OA} , P_{AA} , P_{PR} AND P_{CI-PR} CALCULATED BY ICEM AND ILCMV USING BANDS SELECTED BY SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, AND SB-MCSC-BS WITH WCD USED AS THE CI MEASURE FOR PURDUE'S DATA. (b) P_A , P_{OA} , P_{CI-OA} , P_{AA} , P_{PR} AND P_{CI-PR} CALCULATED BY ICEM AND ILCMV USING BANDS SELECTED BY SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, AND SB-MCSC-BS WITH CD USED AS THE CI MEASURE FOR PURDUE'S DATA. (c) P_A , P_{OA} , P_{CI-OA} , P_{AA} , P_{PR} AND P_{CI-PR} CALCULATED BY ICEM AND ILCMV USING BANDS SELECTED BY SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, AND SB-MCSC-BS WITH SR USED AS THE CI MEASURE FOR PURDUE'S DATA. (d) P_A , P_{OA} , P_{CI-OA} , P_{AA} , P_{PR} AND P_{CI-PR} CALCULATED BY ICEM AND ILCMV USING BANDS SELECTED BY SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, AND SB-MCSC-BS WITH BCD USED AS THE CI MEASURE FOR PURDUE'S DATA. (e) P_A , P_{OA} , P_{CI-OA} , P_{AA} , P_{PR} AND P_{CI-PR} CALCULATED BY ICEM AND ILCMV BANDS SELECTED BY SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, AND SB-MCSC-BS WITH CFR USED AS THE CI MEASURE FOR PURDUE'S DATA

ICEM				ILCMV			
	SF-SCSC-BS	SB-SCSC-BS		SF-MCSC-BS	SB-MCSC-BS		
class	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$
class 1 (46)	0.9348	0.8958	0.9565	0.7333	0.9565	0.7333	0.9348
class 2 (1428)	0.9811	0.9441	0.9741	0.9573	0.9615	0.9033	0.9538
class 3 (830)	0.9747	0.9385	0.9783	0.9333	0.9651	0.9030	0.9687
class 4 (237)	0.9831	0.9873	0.9789	0.9707	0.9578	0.9116	0.9831
class 5 (483)	0.9731	0.9671	0.9607	0.9317	0.9275	0.8836	0.9317
class 6 (730)	0.9822	0.9076	0.9753	0.9001	0.9521	0.8720	0.9726
class 7 (28)	1.0000	0.6667	1.0000	0.6829	1.0000	0.6829	1.0000
class 8 (478)	0.9958	0.9917	0.9958	0.9917	0.9833	0.9476	0.9895
class 9 (20)	1.0000	0.5000	1.0000	0.4878	1.0000	0.5714	1.0000
class 10 (972)	0.9660	0.9418	0.9537	0.9307	0.9516	0.9177	0.9403
class 11 (2455)	0.9821	0.9538	0.9772	0.9512	0.9527	0.9420	0.9564
class 12 (593)	0.9798	0.9603	0.9747	0.9570	0.9629	0.9021	0.9528
class 13 (205)	0.9805	0.9437	0.9902	0.9575	0.9805	0.8855	0.9854
class 14 (1265)	0.9747	0.9817	0.9644	0.9721	0.9486	0.9419	0.9423
class 15 (386)	0.9845	0.9719	0.9767	0.9520	0.9456	0.9288	0.9378
class 16 (93)	0.9570	0.8900	0.9892	0.8762	0.9892	0.8288	0.9892
P_{OA}	0.9787		0.9733		0.9561		0.9560
P_{AA}	0.9781		0.9779		0.9647		0.9649
P_{CI-OA}	0.9789		0.9843		0.9784		0.9760
P_{PR}	0.9026		0.8866		0.8597		0.8684
P_{CI-PR}	0.8774		0.8571		0.8235		0.8501

(a)

ICEM				ILCMV			
	SF-SCSC-BS	SB-SCSC-BS		SF-MCSC-BS	SB-MCSC-BS		
class	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$
class 1 (46)	0.9348	0.8776	0.9783	0.7377	0.9565	0.7333	0.9348
class 2 (1428)	0.9741	0.9463	0.9783	0.9327	0.9615	0.9033	0.9538
class 3 (830)	0.9807	0.9378	0.9831	0.9241	0.9651	0.9030	0.9687
class 4 (237)	0.9789	0.9748	0.9789	0.9431	0.9578	0.9116	0.9831
class 5 (483)	0.9627	0.9470	0.9607	0.9393	0.9275	0.8836	0.9317
class 6 (730)	0.9849	0.8844	0.9822	0.9111	0.9521	0.8720	0.9726
class 7 (28)	1.0000	0.6829	1.0000	0.6829	1.0000	0.6829	1.0000
class 8 (478)	0.9937	0.9875	0.9958	0.9855	0.9833	0.9476	0.9895
class 9 (20)	1.0000	0.5000	1.0000	0.4348	1.0000	0.5714	1.0000
class 10 (972)	0.9660	0.9660	0.9619	0.9416	0.9516	0.9177	0.9403
class 11 (2455)	0.9837	0.9538	0.9768	0.9588	0.9527	0.9420	0.9564
class 12 (593)	0.9815	0.9510	0.9663	0.9518	0.9629	0.9021	0.9528
class 13 (205)	0.9854	0.9758	0.9854	0.9484	0.9805	0.8855	0.9854
class 14 (1265)	0.9763	0.9740	0.9589	0.9767	0.9486	0.9419	0.9423
class 15 (386)	0.9689	0.9517	0.9741	0.9377	0.9456	0.9288	0.9378
class 16 (93)	0.9785	0.8922	0.9892	0.8762	0.9892	0.8288	0.9892
P_{OA}	0.9782		0.9742		0.9561		0.9560
P_{AA}	0.9781		0.9794		0.9647		0.9649
P_{CI-OA}	0.9790		0.9876		0.9779		0.9750
P_{PR}	0.9002		0.8801		0.8597		0.8684
P_{CI-PR}	0.8855		0.8479		0.8229		0.8456

(b)

ICEM				ILCMV			
	SF-SCSC-BS	SB-SCSC-BS		SF-MCSC-BS	SB-MCSC-BS		
class	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$
class 1 (46)	0.9348	0.8776	0.9783	0.7377	0.9565	0.7333	0.9348
class 2 (1428)	0.9741	0.9463	0.9783	0.9327	0.9615	0.9033	0.9538
class 3 (830)	0.9807	0.9378	0.9831	0.9241	0.9651	0.9030	0.9687
class 4 (237)	0.9789	0.9748	0.9789	0.9431	0.9578	0.9116	0.9831
class 5 (483)	0.9627	0.9470	0.9607	0.9393	0.9275	0.8836	0.9317
class 6 (730)	0.9849	0.8844	0.9822	0.9111	0.9521	0.8720	0.9726
class 7 (28)	1.0000	0.6829	1.0000	0.6829	1.0000	0.6829	1.0000
class 8 (478)	0.9937	0.9875	0.9958	0.9855	0.9833	0.9476	0.9895
class 9 (20)	1.0000	0.5000	1.0000	0.4348	1.0000	0.5714	1.0000
class 10 (972)	0.9660	0.9660	0.9619	0.9416	0.9516	0.9177	0.9403
class 11 (2455)	0.9837	0.9538	0.9768	0.9588	0.9527	0.9420	0.9564
class 12 (593)	0.9815	0.9510	0.9663	0.9518	0.9629	0.9021	0.9528
class 13 (205)	0.9854	0.9758	0.9854	0.9484	0.9805	0.8855	0.9854
class 14 (1265)	0.9763	0.9740	0.9589	0.9767	0.9486	0.9419	0.9423
class 15 (386)	0.9689	0.9517	0.9741	0.9377	0.9456	0.9288	0.9378
class 16 (93)	0.9785	0.8922	0.9892	0.8762	0.9892	0.8288	0.9892
P_{OA}	0.9782		0.9742		0.9561		0.9560
P_{AA}	0.9781		0.9794		0.9647		0.9649
P_{CI-OA}	0.9790		0.9876		0.9779		0.9750
P_{PR}	0.9002		0.8801		0.8597		0.8684
P_{CI-PR}	0.8855		0.8479		0.8229		0.8456

(c)

ICEM				ILCMV			
	SF-SCSC-BS	SB-SCSC-BS		SF-MCSC-BS	SB-MCSC-BS		
class	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$	$P_{PR}(C_i)$	$P_A(C_i)$
class 1 (46)	0.9565	0.7458	0.9565	0.7586	0.9130	0.7241	0.9348
class 2 (1428)	0.9804	0.9340	0.9762	0.9464	0.9608	0.9111	0.9559
class 3 (830)	0.9783	0.9291	0.9711	0.9307	0.9759	0.9132	0.9602
class 4 (237)	0.9747	0.9706	0.9705	0.9544	0.9705	0.9237	0.9831
class 5 (483)	0.9793	0.9149	0.9420	0.9100	0.9193	0.8988	0.9255
class 6 (730)	0.9767	0.9141	0.9740	0.8989	0.9712	0.8742	0.9658
class 7 (28)	1.0000	0.6222	1.0000	0.6087	1.0000	0.6364	1.0000
class 8 (478)	0.9916	0.9834	0.9958	0.9855	0.9812	0.9552	0.9812
class 9 (20)	1.0000	0.5000	1.0000	0.4000	0.9000	0.5455	1.0000
class 10 (972)	0.9619	0.9590	0.9486	0.9584	0.9362	0.9037	0.9660
class 11 (2455)	0.9841	0.9535	0.9654	0.9511	0.9568	0.9359	0.9483
class 12 (593)	0.9848	0.9589	0.9629	0.9454	0.9494	0.9037	0.9595
class 13 (205)	0.9707	0.9755	0.9902	0.9144	0.9561	0.9116	0.9805
class 14 (1265)	0.9739	0.9536	0.9597	0.9775	0.9542	0.9664	0.9542
class 15 (386)	0.9767	0.9150	0.9637	0.9394	0.9378	0.9476	0.9456
class 16 (93)	0.9892	0.8932	0.9892	0.8288	0.9892	0.8288	0.9892
P_{OA}	0.9782		0.9742		0.9561		0.9560
P_{AA}	0.9781		0.9794		0.9647		0.9649
P_{CI-OA}	0.9790		0.9876		0.9779		0.9750
P_{PR}	0.9002		0.8801		0.8597		0.8684
P_{CI-PR}	0.8855		0.8479		0.8229		0.8456

(c)

TABLE IX

(Continued.) (a) P_A , P_{OA} , P_{CI-OA} , P_{AA} , P_{PR} AND P_{CI-PR} CALCULATED BY ICEM AND ILCMV USING BANDS SELECTED BY SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, AND SB-MCSC-BS WITH WCD USED AS THE CI MEASURE FOR PURDUE'S DATA. (b) P_A , P_{OA} , P_{CI-OA} , P_{AA} , P_{PR} AND P_{CI-PR} CALCULATED BY ICEM AND ILCMV USING BANDS SELECTED BY SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, AND SB-MCSC-BS WITH CD USED AS THE CI MEASURE FOR PURDUE'S DATA. (c) P_A , P_{OA} , P_{CI-OA} , P_{AA} , P_{PR} AND P_{CI-PR} CALCULATED BY ICEM AND ILCMV USING BANDS SELECTED BY SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, AND SB-MCSC-BS WITH SR USED AS THE CI MEASURE FOR PURDUE'S DATA. (d) P_A , P_{OA} , P_{CI-OA} , P_{AA} , P_{PR} AND P_{CI-PR} CALCULATED BY ICEM AND ILCMV USING BANDS SELECTED BY SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, AND SB-MCSC-BS WITH BCD USED AS THE CI MEASURE FOR PURDUE'S DATA. (e) P_A , P_{OA} , P_{CI-OA} , P_{AA} , P_{PR} AND P_{CI-PR} CALCULATED BY ICEM AND ILCMV BANDS SELECTED BY SF-SCSC-BS, SB-SCSC-BS, SF-MCSC-BS, AND SB-MCSC-BS WITH CFR USED AS THE CI MEASURE FOR PURDUE'S DATA

CI MEASURE FOR PURDUE'S DATA				
P_{OA}	0.9786	0.9667	0.9563	0.9570
P_{AA}	0.9799	0.9729	0.9545	0.9656
P_{CI-OA}	0.9786	0.9667	0.9563	0.9570
P_{PR}	0.8827	0.8693	0.8612	0.8669
P_{CI-PR}	0.9425	0.9432	0.9204	0.9171

(d)

ICEM				ILCMV				
class	SF-SCSC-BS	SB-SCSC-BS	SF-MCSC-BS	SB-MCSC-BS	SF-SCSC-BS	SB-SCSC-BS	SF-MCSC-BS	SB-MCSC-BS
class 1 (46)	0.9565	0.7458	0.9565	0.8302	0.9565	0.7719	0.9565	0.7458
class 2 (1428)	0.9755	0.9438	0.9818	0.9454	0.9608	0.9117	0.9622	0.9178
class 3 (830)	0.9832	0.9129	0.9867	0.9328	0.9663	0.9022	0.9639	0.8939
class 4 (237)	0.9831	0.9628	0.9789	0.9748	0.9747	0.9390	0.9747	0.8817
class 5 (483)	0.9400	0.9209	0.9545	0.9389	0.9275	0.8750	0.9379	0.8865
class 6 (730)	0.9767	0.8924	0.9781	0.8804	0.9767	0.8857	0.9808	0.8785
class 7 (28)	1.0000	0.6667	1.0000	0.6364	1.0000	0.6512	1.0000	0.6087
class 8 (478)	0.9958	0.9958	0.9937	0.9855	0.9854	0.9573	0.9728	0.9470
class 9 (20)	1.0000	0.4255	1.0000	0.4167	0.9000	0.5455	0.9000	0.5455
class 10 (972)	0.9475	0.9475	0.9753	0.9442	0.9290	0.8932	0.9342	0.9135
class 11 (2455)	0.9796	0.9551	0.9768	0.9654	0.9515	0.9374	0.9585	0.9424
class 12 (593)	0.9798	0.9447	0.9865	0.9466	0.9529	0.8956	0.9612	0.8796
class 13 (205)	0.9805	0.9526	0.9805	1.0000	0.9854	0.9309	0.9854	0.8783
class 14 (1265)	0.9692	0.9847	0.9794	0.9880	0.9344	0.9586	0.9320	0.9508
class 15 (386)	0.9767	0.9449	0.9819	0.9547	0.9637	0.9514	0.9404	0.9478
class 16 (93)	0.9892	0.8679	0.9570	0.8812	0.9892	0.8440	0.9892	0.8214
P_{OA}	0.9738	0.9790			0.9542			0.9558
P_{AA}	0.9771	0.9792			0.9596			0.9594
P_{CI-OA}	0.9789	0.9737			0.9664			0.9660
P_{PR}	0.8790	0.8888			0.8657			0.8524
P_{CI-PR}	0.8702	0.8801			0.8556			0.8364

(e)

ICEM				ILCMV				
class	SF-SCSC-BS	SB-SCSC-BS	SF-MCSC-BS	SB-MCSC-BS	SF-SCSC-BS	SB-SCSC-BS	SF-MCSC-BS	SB-MCSC-BS
class 1 (46)	0.9565	0.7333	0.9565	0.8000	0.9130	0.7241	0.9348	0.7679
class 2 (1428)	0.9762	0.9312	0.9657	0.9504	0.9608	0.9111	0.9559	0.9155
class 3 (830)	0.9676	0.9212	0.9783	0.9113	0.9759	0.9132	0.9602	0.8955
class 4 (237)	0.9747	0.9706	0.9747	0.9352	0.9705	0.9237	0.9831	0.9283
class 5 (483)	0.9586	0.9527	0.9752	0.9058	0.9193	0.8988	0.9255	0.9049
class 6 (730)	0.9822	0.9030	0.9767	0.9071	0.9712	0.8742	0.9658	0.8835
class 7 (28)	1.0000	0.6222	1.0000	0.6222	1.0000	0.6364	1.0000	0.7000
class 8 (478)	0.9937	0.9794	0.9937	0.9855	0.9812	0.9552	0.9812	0.9324
class 9 (20)	1.0000	0.4082	1.0000	0.4444	0.9000	0.5455	1.0000	0.5556
class 10 (972)	0.9619	0.9444	0.9455	0.9504	0.9362	0.9037	0.9660	0.9038
class 11 (2455)	0.9720	0.9614	0.9752	0.9557	0.9568	0.9359	0.9483	0.9286
class 12 (593)	0.9730	0.9537	0.9663	0.9393	0.9494	0.9037	0.9595	0.9032
class 13 (205)	0.9805	0.9054	0.9902	0.9486	0.9561	0.9116	0.9805	0.9054
class 14 (1265)	0.9660	0.9768	0.9708	0.9816	0.9542	0.9664	0.9542	0.9587
class 15 (386)	0.9741	0.9307	0.9715	0.9146	0.9378	0.9476	0.9456	0.9505
class 16 (93)	0.9892	0.8762	0.9785	0.8750	0.9892	0.8288	0.9892	0.8364
P_{OA}	0.9722	0.9714			0.9563			0.9570
P_{AA}	0.9766	0.9762			0.9545			0.9656
P_{CI-OA}	0.9830	0.9846			0.9644			0.9790
P_{PR}	0.8732	0.8767			0.8612			0.8669
P_{CI-PR}	0.8359	0.8387			0.8058			0.8085

TABLE X

 P_A , P_{OA} , P_{AA} , AND P_{PR} CALCULATED BY FOUR EPF-BASED METHODS, ICEM, AND ILCMV USING FULL BANDS FOR PURDUE'S INDIAN PINES

	EPF-B-g		EPF-B-c		EPF-G-g		EPF-G-c		ICEM		ILCMV	
Class	$P_A(C_i)$	$P_{PR}(C_i)$										
class 1 (46)	1.0000	0.3680	0.9783	0.3689	1.0000	0.3802	1.0000	0.3866	0.9565	0.7333	0.9565	0.7719
class 2 (1428)	0.8396	0.6031	0.8487	0.5968	0.8326	0.5918	0.8305	0.5777	0.9734	0.9398	0.9601	0.9196
class 3 (830)	0.7193	0.6769	0.7301	0.6779	0.7096	0.6693	0.7024	0.6565	0.9807	0.9346	0.9699	0.8974
class 4 (237)	0.9916	0.4927	0.9916	0.5043	1.0000	0.4807	1.0000	0.4907	0.9831	0.9628	0.9873	0.9286
class 5 (483)	0.9545	0.2826	0.9586	0.2912	0.9627	0.2808	0.9607	0.2873	0.9441	0.9383	0.8944	0.9290
class 6 (730)	0.9986	0.3699	0.9986	0.3660	0.9973	0.3534	0.9973	0.3724	0.9836	0.8831	0.9712	0.8997
class 7 (28)	0.9286	0.5417	0.9286	0.5652	0.9643	0.5745	0.9643	0.6279	1.0000	0.6512	1.0000	0.7568
class 8 (478)	1.0000	0.6685	1.0000	0.6676	1.0000	0.7124	1.0000	0.7050	0.9895	0.9834	0.9874	0.9574
class 9 (20)	1.0000	0.4348	1.0000	0.4762	1.0000	0.5714	1.0000	0.6897	1.0000	0.5405	1.0000	0.5263
class 10 (972)	0.9722	0.5662	0.9640	0.5584	0.9702	0.5541	0.9578	0.5558	0.9527	0.9344	0.9393	0.9185
class 11 (2455)	0.8151	0.8248	0.8122	0.8233	0.8069	0.8278	0.8098	0.8259	0.9744	0.9565	0.9470	0.9417
class 12 (593)	0.9410	0.5036	0.9427	0.5082	0.9444	0.4978	0.9578	0.4926	0.9663	0.9409	0.9545	0.8943
class 13 (205)	0.9951	0.6296	0.9951	0.6435	0.9951	0.6000	0.9951	0.5514	0.9756	0.9662	0.9854	0.9352
class 14 (1265)	0.9723	0.3144	0.9755	0.3093	0.9731	0.3130	0.9747	0.3097	0.9747	0.9770	0.9352	0.9689
class 15 (386)	1.0000	0.1097	1.0000	0.1121	1.0000	0.1147	1.0000	0.1148	0.9767	0.9642	0.9067	0.9669
class 16 (93)	1.0000	0.5082	1.0000	0.5284	1.0000	0.4794	1.0000	0.4346	0.9785	0.8667	0.9892	0.8598
P_{AA}	0.9455		0.9452		0.9473		0.9469		0.9756		0.9615	
P_{OA}	0.8984		0.8997		0.8954		0.8949		0.9726		0.9509	
P_{PR}	0.4934		0.4998		0.5001		0.5049		0.8858		0.8795	

can be used in conjunction with CI measures and SR to dictate performance of a classifier on smaller classes, while P_{OA} cannot.

- 4) Furthermore, by taking advantage of CI probabilities, P_{CI-OA} can improve P_{OA} . In contrast, P_{CI-PR} is worse than P_{PR} by adding more weights to the smaller classes and less weights to larger classes compared to P_{PR} which adds more weights to larger classes than that to smaller classes. This implies that P_{CI-PR} makes the difference between larger classes and smaller classes even more pronounced, a task which P_{OA} cannot do.
- 5) It should be noted that the performance of the EPF-based methods was much worse than reported in [47] because full bands were in our experiments compared to [46] which removed water bands. Also, it has been shown in [39]–[41] that EPF-based methods performed very well in P_{OA} but very poorly in P_{PR} . In addition, EPF-based methods cannot be further improved by BS due to their use of principal component analysis (PCA).

XI. CONCLUSION

This paper presents a CI-based BS approach to HSIC which is quite different from conventional BS approaches reported in the literature. It takes advantage of information theory to define two new concepts. One is CI which can be used to determine the number of training samples required to be selected for each class as well as to weigh classes of interest. The other is CSI which can be used to determine the number of bands to be selected for each of the classes. In order to measure CI, five CI-based IC are also introduced, WCD, CD, SR, BCD, and CFR, all of which can be grouped into two categories, intraclass IC and interclass IC. These five ICs are then used to calculate CSI for each class that determines how many bands required to be selected $n_{BS}^{CSI}(C_i)$ for each individual class and n_{BS}^{CE} for all classes. To select desired bands for various classes, two types of BS methods are custom-designed, SCSC-BS to select $n_{BS}^{CSI}(C_i)$ bands for each class, the i th class C_i

and MCSC-BS to select $n_{BS} = \lceil n_{BS}^{CE} \times M \rceil$ bands for all classes. Finally, to evaluate classification performance five classification measures, AA, OA, PR along with two CI-based criteria, CI-OA and CI-PR are used for performance analysis. The experimental results show that CI-based BS can improve classification using CSI-selected bands without using full bands.

As a conclusion, we summarize important contributions made in this paper as follows.

- 1) New concepts of CI and CSI are introduced for classification.
- 2) New class information measures, BCD, WCD, CD, and CFR are developed for classification despite that BCD, WCD and FR have been used in pattern classification. However, their use to measure CI is new and has never been explored in the classification literature. In particular, these five class measures can be grouped into two categories, intraclass IC in the sense of class variability and interclass IC in the sense of class separability.
- 3) Using CI to allocate the number of training samples for each class is completely new.
- 4) Using CSI to determine the number of bands to be selected for each class and using CE to determine the number of bands for all classes is also completely new.
- 5) Based on CSI- and CE-determined number of bands, two new separate BS algorithms, SCSC-BS and MCSC-BS are particularly designed for classification.
- 6) Using CI-calculated probabilities as weights can extend the traditional AA and OA to CI-weighted OA (P_{CI-OA}) which can improve classification accuracy. No similar work has been reported in the classification literature.

REFERENCES

- [1] C.-I Chang, *Hyperspectral Data Processing: Algorithm Design and Analysis*. Hoboken, NJ, USA: Wiley, 2013.
- [2] C. Conese and F. Maselli, "Selection of optimum bands from TM scenes through mutual information analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 48, no. 3, pp. 2–11, 1993.

- [3] C.-I Chang, Q. Du, T.-S. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.
- [4] R. Huang and M. He, "Band selection based on feature weighting for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 156–159, Apr. 2005.
- [5] C.-I Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.
- [6] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1552–1565, Jul. 2004.
- [7] P. W. Mousel, W. J. Kramber, and J. K. Lee, "Optimum band selection for supervised classification of multispectral data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 1, pp. 55–60, Jan. 1990.
- [8] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection," *Photogramm. Eng. Remote Sens.*, vol. 10, no. 7, pp. 793–802, Jul. 2004.
- [9] S. Jia, Z. Ji, Y. Qian, and L. Shen, "Unsupervised band selection for hyperspectral imagery classification without manual band removal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 531–543, Apr. 2012.
- [10] S. D. Stearns, B. E. Wilson, and J. R. Peterson, "Dimensionality reduction by optimal band selection for pixel classification of hyperspectral imagery," *Proc. SPIE*, vol. 2028, pp. 118–127, Oct. 1993.
- [11] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, 1119–1125, Nov. 1994.
- [12] S. Backer, P. Kempenaar, W. Debruyn, and P. Scheunders, "Band selection for hyperspectral remote sensing," *Pattern Recognit.*, vol. 2, no. 3, pp. 319–323, 2005.
- [13] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.
- [14] H. Su, H. Yang, Q. Du, and Y. Sheng, "Semisupervised band clustering for dimensionality reduction of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1135–1139, Nov. 2011.
- [15] H. Su and Q. Du, "Hyperspectral band clustering and band selection for urban land cover classification," *Geocarto Int.*, vol. 27, no. 5, pp. 395–411, 2012.
- [16] H. Yang, Q. H. D. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138–142, Jan. 2011.
- [17] C. Yu, Y. Wang, M. Song, and C.-I Chang, "Class signature-constrained background-suppressed approach to band selection for classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 14–31, Jan. 2019.
- [18] H. Su, Q. Du, G. Chen, and P. Du, "Optimized hyperspectral band selection using particle swarm optimization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2659–2670, Jun. 2014.
- [19] H. Su, B. Yong, and Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68–72, Jan. 2016.
- [20] Y. Yuan, G. Zhu, and Q. Wang, "Hyperspectral band selection by multitask sparsity pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 631–644, Feb. 2015.
- [21] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 51–64, Jan. 2017.
- [22] G. Zhu, Y. Huang, J. Lei, Z. Bi, and F. Xu, "Unsupervised hyperspectral band selection by dominant set extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 227–239, Jan. 2016.
- [23] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.
- [24] C.-I Chang, L. C. Lee, B. Xue, M. Song, and J. Chen, "Channel capacity approach to band subset selection for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4630–4644, Oct. 2017.
- [25] C. Yu, M. Song, and C.-I Chang, "Band subset selection for hyperspectral image classification," *Remote Sens.*, vol. 10, no. 1, p. 113, 2018. doi: [10.3390/rs10010113](https://doi.org/10.3390/rs10010113).
- [26] J. Bai, S. Xiang, L. Shi, and C. Pan, "Semisupervised pair-wise band selection for hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2798–2813, Jul. 2015.
- [27] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York, NY, USA: Academic, 2003.
- [28] C.-I Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
- [29] C.-I Chang, "Utility of virtual dimensionality in hyperspectral signal/image processing," in *Recent Advances in Hyperspectral Signal and Image Processing*, C.-I Chang, Eds. New Delhi, India: Research Signpost, 2006, pp. 1–27.
- [30] C.-I Chang, "Virtual dimensionality for hyperspectral imagery," *Proc. SPIE*, Sep. 2009. [Online]. Available: <http://newsroom.spie.org/x37123.xml>. doi: [10.1117/12.1200909.1749](https://doi.org/10.1117/12.1200909.1749).
- [31] C.-I Chang, "A review of virtual dimensionality for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1285–1305, Apr. 2018.
- [32] C.-I Chang, "A unified theory for target-specified virtual dimensionality of hyperspectral imagery," *Proc. SPIE*, vol. 8539, Oct. 2012, Art. no. 85390J.
- [33] C.-I Chang, *Real-Time Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation*. New York, NY, USA: Springer, 2017.
- [34] C.-I Chang, L.-C. Lee, and D. Paylor, "Virtual dimensionality analysis for hyperspectral imagery," *Proc. SPIE*, vol. 9501, pp. 95010R-1–95010R-11, May 2015.
- [35] C. Yu, L.-C. Lee, C.-I Chang, B. Xue, M. Song, and J. Chen, "Band-specified virtual dimensionality for band selection: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2822–2832, May 2018.
- [36] J. C. Harsanyi, W. Farrand, and C.-I Chang, "Detection of subpixel spectral signatures in hyperspectral image sequences," in *Proc. Annu. Meeting Amer. Soc. Photogram. Remote Sens.*, Reno, NV, USA, 1994, pp. 236–247.
- [37] R. M. Fano, *Transmission of Information*. Cambridge, MA, USA: MIT Press, 1961.
- [38] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [39] B. Xue *et al.*, "A subpixel target detection approach to hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5093–5114, Sep. 2017.
- [40] C. Yu *et al.*, "Multi-Class constrained background suppression approach to hyperspectral image classification," in *Proc. IEEE/GRSS Int. Geosci. Remote Sens. Symp. (IGARSS)*, Fort Worth, TX, USA, Jul. 2017, pp. 3357–3360.
- [41] C. Yu, B. Xue, M. Song, Y. Wang, S. Li, and C.-I Chang, "Iterative target-constrained interference-minimized classifier for hyperspectral classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1095–1117, Apr. 2018.
- [42] M. A. Hossain, X. Jia, and M. Pickering, "Subspace detection using a mutual information measure for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 424–428, Feb. 2014.
- [43] C.-I Chang, "A statistical detection theory approach to hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2057–2074, Apr. 2019.
- [44] S. Aronoff, "Classification accuracy: A user approach," *Photogramm. Eng. Remote Sens.*, vol. 48, no. 8, pp. 1299–1307, Aug. 1982.
- [45] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2009.
- [46] N. Patel and B. Kaushal, "Improvement of user's accuracy through classification of principal component images and stacked temporal images," *Geo-Spatial Inf. Sci.*, vol. 13, no. 4, pp. 243–248, Dec. 2010.
- [47] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
- [48] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Dept. Elect. Eng., Univ. Maryland, Baltimore, MD, USA, Aug. 1993.
- [49] W. H. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailings in the Coeur d'Alene River Valley, Idaho, through the use of a constrained energy minimization technique," *Remote Sens. Environ.*, vol. 59, pp. 64–76, Jan. 1997.
- [50] C.-I Chang, "Target signature-constrained mixed pixel classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 5, pp. 1065–1081, May 2002.

- [51] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- [52] H. Ren and C.-I Chang, "Target-constrained interference-minimized approach to subpixel target detection for hyperspectral images," *Opt. Eng.*, vol. 39, no. 12, pp. 3138–3145, Dec. 2000.



Meiping Song received the Ph.D. degree from the College of Computer Science and Technology, Harbin Engineering University, Harbin, China, in 2006.

From 2013 to 2014, she was a Visiting Associate Research Scholar with the Remote Sensing Signal and Image Processing Laboratory, University of Maryland, Baltimore, MD, USA. She is currently an Associate Professor with the College of Information Science and Technology, Dalian Maritime University, Dalian, China. Her research interests include remote sensing and hyperspectral image processing.



Xiaodi Shang (S'19) received the B.S. degree in software engineering from Qingdao University, Qingdao, China, in 2016. She is currently pursuing the Ph.D. degree in computer application technology with Dalian Maritime University, Dalian, China.

Her research interests include hyperspectral image classification, band selection, and applications.



Yulei Wang received the B.S. and Ph.D. degrees in signal and information processing from Harbin Engineering University, Harbin, China, in 2009 and 2015, respectively.

She was awarded by the China Scholarship Council in 2011 as a joint Ph.D. Student to study at the Remote Sensing Signal and Image Processing Laboratory, University of Maryland, Baltimore, MD, USA. She is currently a Lecturer with the College of Information Science and Technology, Dalian Maritime University, Dalian, China. Her research interests include hyperspectral image processing and vital sign signal processing.



Chunyan Yu (M'17) received the B.Sc. and Ph.D. degrees in environment engineering from Dalian Maritime University, Dalian, China, in 2004 and 2012, respectively.

In 2004, she joined the College of Computer Science and Technology, Dalian Maritime University, where she was a Post-Doctoral Fellow with the Information Science and Technology College, from 2013 to 2016. From 2014 to 2015, she was a Visiting Scholar with the College of Physicians and Surgeons, Columbia University, New York, NY, USA. She is currently an Associate Professor with the Information Science and Technology College, Dalian Maritime University. Her research interests include image segmentation, hyperspectral image classification, and pattern recognition.



Chein-I Chang (S'81–M'87–SM'92–F'10–LF'17) received the B.S. degree in mathematics from Soochow University, Taipei, Taiwan, in 1973, the M.S. degree in mathematics from the Institute of Mathematics, National Tsing Hua University, Hsinchu, Taiwan, in 1975, the M.A. degree in mathematics from the State University of New York at Stony Brook, Stony Brook, NY, USA, in 1977, the M.S. and M.S.E.E. degrees from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 1980 and 1982, respectively, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 1987.

He has been with the University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA, since 1987, where he is currently a Professor with the Department of Computer Science and Electrical Engineering. He was a Visiting Research Specialist with the Institute of Information Engineering, National Cheng Kung University, Tainan, Taiwan, from 1994 to 1995, and also a Distinguished Visiting Fellow/Fellow Professor, both of which were sponsored by the National Science Council in Taiwan, from 2009 to 2010. He was also a Distinguished Lecturer Chair with the National Chung Hsing University, Taichung, Taiwan, sponsored by the Ministry of Education in Taiwan, from 2005 to 2006, and a Distinguished Chair Professor with National Chung Hsing University, from 2014 to 2017. He is currently holding the Chang Jiang Scholar Chair Professorship and the Director of the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Dalian Maritime University, Dalian, China, since 2016. In addition, he has also been a Chair Professor with Providence University, Taichung, since 2012, and with National Chiao Tung University, Hsinchu, since 2019. He has authored four books, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Kluwer Academic Publishers, 2003), *Hyperspectral Data Processing: Algorithm Design and Analysis* (John Wiley & Sons, 2013), *Real Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection* (Springer, 2016), and *Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation* (Springer, 2017). He has also edited two books, *Recent Advances in Hyperspectral Signal and Image Processing*, in 2006, and *Hyperspectral Data Exploitation: Theory and Applications* (John Wiley & Sons, 2007) and co-edited with A. Plaza a book on *High Performance Computing in Remote Sensing* (CRC Press, 2007). His research interests include multispectral/hyperspectral image processing, automatic target recognition, and medical imaging. He holds seven patents on hyperspectral image processing.

Dr. Chang is a fellow of SPIE. He was a plenary speaker for the Society for Photo-Optical Instrumentation Engineers (SPIE) Optics+Applications, Remote Sensing Symposium in 2009. He was also a keynote speaker for the User Conference of Hyperspectral Imaging in 2010, Industrial Technology Research Institute, Hsinchu, the 2009 Annual Meeting of the Radiological Society of the Republic of China, Taichung, the 2008 International Symposium on Spectral Sensing Research, and the Conference on Computer Vision, Graphics, and Image Processing in 2003, Kimen, and 2013, Nantou, Taiwan. He was a Guest Editor of a special issue of the *Journal of High Speed Networks on Telemedicine and Applications* in 2000 and a Co-Guest Editor of another special issue of the same journal on *Broadband Multimedia Sensor Networks in Healthcare Applications* in 2007. He was also the Co-Guest Editor of special issues on *High Performance Computing of Hyperspectral Imaging* for the *International Journal of High Performance Computing Applications* in 2007, *Signal Processing and System Design in Health Care Applications* for the EURASIP *Journal on Advances in Signal Processing* in 2009, *Multippectral, Hyperspectral, and Polarimetric Imaging Technology* for the *Journal of Sensors* in 2016, and *Hyperspectral Imaging and Applications for Remote Sensing* in 2018. He was an Associate Editor in the area of hyperspectral signal processing for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2001 to 2007. He is currently an Associate Editor of the *Artificial Intelligence Research* and also on the editorial boards of the *Journal of High Speed Networks*, *Recent Patents on Mechanical Engineering*, the *International Journal of Computational Sciences and Engineering*, the *Journal of Robotics and Artificial Intelligence*, and the *Open Remote Sensing Journal*. He received a National Research Council Senior Research Associateship Award from 2002 to 2003 sponsored by the U.S. Army Soldier and Biological Chemical Command, Edgewood Chemical and Biological Center, Aberdeen Proving Ground, Maryland.