

Class Signature-Constrained Background-Suppressed Approach to Band Selection for Classification of Hyperspectral Images

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

Abstract—In hyperspectral image classification (HSIC), background (BKG) is generally excluded from consideration due to the fact that obtaining complete knowledge of BKG is nearly impossible in reality. Unfortunately, BKG has significant impact on classification and band selection (BS). This paper investigates both issues and presents a novel approach called class signature-constrained BKG suppression (CSCBS) approach to BS for HSIC, where class signatures can be obtained either by a priori or a posteriori knowledge or training samples, and BKG suppression can be accomplished by taking the inverse of the sample correlation matrix R . Its idea takes advantage of the concept of the linearly constrained minimum variance (LCMV)

Manuscript received February 13, 2018; revised April 27, 2018; accepted June 11, 2018. Date of publication November 12, 2018; date of current version December 24, 2018. The work of C. Yu was supported in part by the National Nature Science Foundation of Liaoning Province under Grant 20170540095 and in part by the Fundamental Research Funds for the Central Universities under Grant 3132018196. The work of Y. Wang was supported by the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences, under Grant LSIT201707D. 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 C.-I Chang was supported in part by the Fundamental Research Funds for Central Universities under Grant 3132016331 and in part under Grant ZD20180073.

(Corresponding author: Meiping Song.)

C. Yu is with the Center for Hyperspectral Imaging in Remote Sensing, Information and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: yuchunyan1997@126.com).

Y. Wang is with the Center of Hyperspectral Imaging in Remote Sensing, Information and Technology College, Dalian Maritime University, Dalian 116026, China, also with the State Key Laboratory of Integrated Services Networks, Xi'an 710000, China, and also with the Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences, Xi'an 710000, China (e-mail: wangyulei@dlmu.edu.cn).

M. Song is with the Center for Hyperspectral Imaging in Remote Sensing, Information and Technology College, Dalian Maritime University, Dalian 116026, China, and also with the State Key Laboratory of Integrated Services Networks, Xi'an 710000, China (e-mail: smping@163.com).

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

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

Digital Object Identifier 10.1109/TGRS.2018.2850152

developed from adaptive beamforming by constraining class signatures of interest while minimizing the effect caused by the unknown BKG so as to enhance the classification performance. There are two immediate applications of CSCBS. One is its application to HSIC, in which it becomes a CSCBS classifier. The other is its use of the LCMV-suppressed BKG as a measure to derive the band prioritization (BP) criteria and BS. Experimental results demonstrate that generally CSCBS does not need the full-band set for HSIC since a partial band subset selected by CSCBS-BP/BS can actually improve the classification results using full-band information.

Index Terms—Backward class signature-constrained background (BKG) suppression band prioritization (BCSCBS-BP), band selection (BS), class signature-constrained BKG suppressed approach (CSCBS), CSBS band selection (CSCBS-BS), CSCBS search backward BS (CSCBS-SBBS), CSCBS search feedforward BS (CSCBS-SFBS), forward CSCBS-BP (FCSCBS-BP), hyperspectral image classification (HSIC), iterative CSCBSC (ICSCBSC), linearly constrained minimum variance (LCMV).

NOMENCLATURE

AA	Average accuracy.
BCSCBS-BP	Backward class signature-constrained background suppression band prioritization.
BKG	Background.
BP	Band prioritization.
BS	Band selection.
CBS	Constrained BS.
CSCBS	Class signature-constrained BKG suppression.
CSCBSC	Class signature-constrained BKG suppressed classifier.
CSCBS-SBBS	CSCBS search backward BS.
CSCBS-SFBS	CSCBS search forward BS.
FCSCBS-BP	Forward CSCBS-BP.
HSIC	Hyperspectral image classification.
ICSCBSC	Iterative CSCBSC.
LCMV	linearly constrained minimum variance.
OA	Overall accuracy.
PR	Precision rate.
VD	Virtual dimensionality.

I. INTRODUCTION

HSIC has received considerable interests in recent years, where most of the works reported in the literature are supervised with an assumption that the information of all the classes of interest is provided *a priori* [1]–[29]. Unfortunately, due to the significant improvement on spectral resolution hyperspectral imaging sensors can uncover many unknown material substances that are usually embedded in the BKG and cannot be revealed by visual inspection. Since such BKG is generally unknown, it is very difficult to obtain, if not impossible. Accordingly, the BKG has been excluded from consideration for HSIC. This BKG issue has not received much attention mainly because the used classification criteria, such as OA, AA, or kappa coefficient, only measure accuracy of classes of interest and do not take into account the misclassification errors resulting from BKG. In HSIC, BKG has never been considered as a factor in the past. However, in practical applications, BKG is indeed a very important and integral part of the image data to be processed. Therefore, it cannot be simply ignored or discarded. This is because, BKG has significant impact on classification and in fact plays a critical role in misclassification, as shown in [30]–[32]. In this case, how to deal with BKG and how to come up classification criteria other than the above-mentioned criteria to account for misclassification errors pose great challenges in HSIC. Another issue is interband correlation resulting from very high spectral resolution which also raises an issue that many spectral bands may provide overlapped spectral information for classification. Thus, how to deal with redundant spectral information is also a crucial issue as well. This paper develops a novel approach, to be called CSCBS approach to BS for HSIC to address the above-mentioned issues.

The idea of CSCBS originates from the LCMV developed by Frost [33] for adaptive beamforming in array signal processing. It is a finite impulse response (FIR) filter to lock in desired arrival directions of signal sources impinging upon an array while minimizing the filter output energy of the array that results from undesired signal sources arriving from other directions. By taking advantage of LCMV, we reinterpret LCMV as CSCBS by specifying desired signal arrival directions and the number of desired signal arrival directions as the signatures of classes of interest and the number of classes, respectively, and all other directions coming from undesired signal sources are specified by the BKG signatures. With this interpretation, CSCBS works exactly in the same way that LCMV is designed for adaptive beamforming by constraining class signatures, and in the mean time, minimizing effects caused by BKG signatures. In other words, CSCBS constrains class signatures to perform classification, while at the same time, suppressing BKG to reduce the misclassification errors resulting from data sample vectors from BKG. Thus, when CSCBS is applied to HSIC, it is called CSCBSC.

In regard to the third issue of interband spectral correlation, BS has been widely used for this purpose in many applications in hyperspectral data exploitation [34]–[70]. Two significant benefits can be gained from BS. One is to help identify those bands that may be redundant and can be removed without the loss of significant spectral information. The other is due to

the fact that various material substances in many applications generally respond quite differently to different spectral bands. In this case, it requires judiciously selecting the appropriate bands that correctly respond to what are interesting in different applications such as endmember finding, anomaly detection, target detection, linear spectral unmixing, and specifically, HSIC.

In general, BS can be performed by a BP criterion to select bands according to their priorities ranked by the BP criterion. The other is to develop a band search strategy to select the desired bands. Therefore, the proposed CSCBS can be also used to derive BP criteria for BS, referred to as CSCBS-BP, as well as band search strategies for BS, referred to as CSCBS-BS. Three salient differences between a CSCBS-BP and a CSCBS-BS are worth noting. One is that in order for CSCBS-BP to perform BS, it must prioritize all the bands. By contrast, CSCBS-BS requires a band search strategy to select bands up to a desired number of bands without BP. Another is that CSCBS-BP performs a single band one at a time sequentially without referring to other bands as opposed to CSCBS-BS that generally selects bands based on the previously selected bands. A third one is that CSCBS-BS requires a search algorithm to find desired bands, whereas CSCBS-BP does not.

Combining CSCBS and LCMV derives two CSCBS-BP criteria and two CSCBS-BS methods. The first CSCBS-BP criterion is called FCSCBS-BP that augments bands by finding the bands with CBCBS-minimal variance in a forward manner. The second CSCBS-BP criterion is called BCSCBS-BP that removes bands via the leave one-out rule by finding the bands with CBCBS-maximal variance in a backward manner. Correspondingly, two CSCBS-BS methods can be also developed in a similar way that FCSCBS-BP and BCSCBS-BP are developed, called CSCBS-SFBS, derived from the sequential forward search (SFS) feature selection [71], and CSCBS-SBBS, derived from the sequential backward search (SBS) feature selection [71], respectively. The proposed CSCBS-based BP and CSCBS-based BS for HSIC are evaluated by experiments conducted on three popular real hyperspectral images available in the website [72] for performance analysis. Experimental results show that the bands selected by CSCBS-BP/BS can actually improve the classification performance over their counterparts with using full bands.

II. CLASS SIGNATURE-CONSTRAINED BACKGROUND SUPPRESSION APPROACH

Suppose that $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_M$ are M specific class signatures of interest, which can be either provided by *a priori* knowledge or obtained by *a posteriori* knowledge, training samples 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 $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_M$ as a desired signal arrival direction in adaptive beamforming in [33] and let $\mathbf{D}^M = [\mathbf{d}_1 \mathbf{d}_2 \dots \mathbf{d}_M]$ be an $L \times M$ class signature matrix. 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 [33] can be then reformulated as the following M -class constrained optimization problem:

$$\min_{\mathbf{w}} \{\mathbf{w}^T \mathbf{R} \mathbf{w}\} \quad (1)$$

$$\text{s.t. } (\mathbf{D}^M)^T \mathbf{w} = \mathbf{c}^M \quad (2)$$

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 (1) 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 can be 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 unit vector with ones in all the M components, $\mathbf{1}^M = (1, \dots, \underbrace{1}_j, \dots, \underbrace{1}_M)^T$. In CSCBS,

these M signal arrival directions are specified by M class signatures. Therefore, 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 (1) and (2) with \mathbf{c}^M replaced by $\mathbf{1}^M$ is solved by

$$\mathbf{w}^* = \mathbf{R}^{-1} \mathbf{D}^M ((\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M)^{-1} \mathbf{1}^M \quad (3)$$

where $(\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M$ is an $M \times M$ matrix. An approach using (1) and (2) to derive the optimal solution (3) is called CSCBS.

Now, applying CSCBS to HSIC results in a linear classifier specified by

$$\delta^{\text{CSCBSC}}(\mathbf{r}) = (\mathbf{w}^*)^T \mathbf{r} \quad (4)$$

which is also called CSCBSC with the M -class signature matrix \mathbf{D}^M constrained by $\mathbf{1}^M$ and BKG suppressed by \mathbf{R}^{-1} .

It should be noted that $\min_{\mathbf{w}} \{\mathbf{w}^T \mathbf{R} \mathbf{w}\}$ in (3) was referred to as minimum variance (MV) in [33], which will be used throughout this paper.

Now, substituting (3) into (1) yields

$$\begin{aligned} (\mathbf{w}^*)^T \mathbf{R} \mathbf{w}^* &= [\mathbf{R}^{-1} \mathbf{D}^M ((\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M)^{-1} \mathbf{1}^M]^T \mathbf{R} \\ &\quad \times [\mathbf{R}^{-1} \mathbf{D}^M ((\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M)^{-1} \mathbf{1}^M] \\ &= [(\mathbf{1}^M)^T ((\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M)^{-1} (\mathbf{D}^M)^T] (\mathbf{R}^{-1} \mathbf{R} \mathbf{R}^{-1}) \\ &\quad \times [\mathbf{D}^M ((\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M)^{-1} \mathbf{1}^M] \\ &= (\mathbf{1}^M)^T ((\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M)^{-1} (\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M \\ &\quad \times ((\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M)^{-1} \mathbf{1}^M \\ &= (\mathbf{1}^M)^T ((\mathbf{D}^M)^T \mathbf{R}^{-1} \mathbf{D}^M)^{-1} \mathbf{1}^M \end{aligned} \quad (5)$$

which is MV caused by operating $\delta^{\text{CSCBSC}}(\mathbf{r})$ on the entire image cube.

Let $\{\mathbf{r}_l^i\}_{i=1}^N$ be the total number of data sample vectors of the first l bands. Now, we consider to reformulate CSCBSC described by (1) and (2) 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, $\Omega = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_L\}$

$$\mathbf{R}_{\Omega_l} = \frac{1}{N} \sum_{i=1}^N \mathbf{r}_l^i (\mathbf{r}_l^i)^T \quad (6)$$

is the sample correlation matrix using a partial band subset consisting of the first l band images. For simplicity of notation, we let $\Omega_l = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_l\} = \{\mathbf{B}_{b_1}, \mathbf{B}_{b_2}, \dots, \mathbf{B}_{b_l}\}$. Then, we can prove that (4) with \mathbf{R}^{-1} replaced by (6) is indeed a monotonically decreasing sequence as follows.

Theorem 1

$$\{(\mathbf{c}^M)^T ((\mathbf{D}_{\Omega_l}^M)^T \mathbf{R}_{\Omega_l}^{-1} \mathbf{D}_{\Omega_l}^M)^{-1} \mathbf{c}^M\}_{l=1}^L \quad (7)$$

is a monotonically decreasing sequence in the sense that

$$\begin{aligned} &(\mathbf{c}^M)^T ((\mathbf{D}_{l+1}^M)^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1}^M)^{-1} \mathbf{c}^M \\ &< (\mathbf{c}^M)^T ((\mathbf{D}_l^M)^T \mathbf{R}_{\Omega_l}^{-1} \mathbf{D}_l^M)^{-1} \mathbf{c}^M \end{aligned} \quad (8)$$

where \mathbf{c}^M defined in (1) for a general purpose is an M -dimensional vector independent of the band subset Ω_l , $\mathbf{D}_{\Omega_l}^M = [\mathbf{d}_l^1 \mathbf{d}_l^2 \dots \mathbf{d}_l^M]$, with $\mathbf{d}_l^j = (d_l^j, d_l^j, \dots, d_l^j)^T$, $\mathbf{D}_{\Omega_{l+1}}^M = \begin{bmatrix} \mathbf{D}_l^M \\ \mathbf{d}_{l+1}^T \end{bmatrix}$, and $\mathbf{d}_{l+1} = (d_{l+1}^1, d_{l+1}^2, \dots, d_{l+1}^M)^T$. In this theorem, the constrained vector \mathbf{c}^M is used instead of $\mathbf{1}^M$ for a general formulation. Also, although $(\mathbf{D}_{\Omega_l}^M)^T \mathbf{R}_{\Omega_l}^{-1} \mathbf{D}_{\Omega_l}^M$ varies with the band subset, Ω_l , its matrix size is $M \times M$ and independent of the band subset, Ω_l ,

Proof: Proof of Theorem 1 is given in the Appendix.

III. CSCBS-BASED BAND PRIORITIZATION

A simple approach to BS is to design an appropriate BP criterion on the basis of data characteristics or statistics, such as variance, signal-to-noise ratio, entropy, and information divergence, to calculate a priority score of each of individual bands for its ranking. The band is then selected according to its calculated priority score. The measure (7) provides a new criterion different from many BP criteria reported in the literature [35].

A. Forward Band Prioritization by CSCBS

For any given band image, \mathbf{B}_l , we define a CSCBS-BP criterion by minimizing (7) as follows:

$$\text{MinV}(\mathbf{B}_l) = (\mathbf{1}^M)^T ((\mathbf{D}_{\mathbf{B}_l}^M)^T \mathbf{R}_{\mathbf{B}_l}^{-1} \mathbf{D}_{\mathbf{B}_l}^M)^{-1} \mathbf{1}^M \quad (9)$$

where \mathbf{c}^M and $\Omega_l = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_l\}$ in (7) are replaced by $\mathbf{1}^M$ and a singleton set $\Omega_l = \{\mathbf{B}_l\}$ that contains only one single band, \mathbf{B}_l , respectively.

Since (9) measures the minimal variance caused by all data sample vectors that are not specified by target signatures in $\mathbf{D}_{\mathbf{B}_l}^M$, this implies that the smaller the $\text{MinV}(\mathbf{B}_l)$ is, the higher priority the l th band, \mathbf{B}_l , is. Accordingly, using (9), a CSCBS-BP criterion, called FCSCBS-BP, can be developed as a BP criterion to rank all bands.

B. Backward Band Prioritization by CSCBS

In contrast to (9) which prioritizes bands by letting $\Omega_l = \{\mathbf{B}_l\}$ in (7), we can look into (7) using the so-called leave-one-out policy by replacing \mathbf{c}^M with $\mathbf{1}^M$ and $\Omega_l = \{\mathbf{B}_l\}$ with $\Omega_l^c = \Omega - \{\mathbf{B}_l\}$. That is, for each band image \mathbf{B}_l ,

Algorithm 1 FCSCBS-BP

1. Initial condition:
For each band, \mathbf{B}_l , find minimal priority score, $\text{MinV}(\mathbf{B}_l)$, according to (9).
2. Prioritize bands according to $\{\text{MinV}(\mathbf{B}_l)\}_{l=1}^L$

$$\mathbf{B}_{l_j} \succ \mathbf{B}_{l_k} \Leftrightarrow \text{MinV}(\mathbf{B}_{l_j}) < \text{MinV}(\mathbf{B}_{l_k}) \quad (10)$$

i.e., $\mathbf{B}_{l_1} \succ \mathbf{B}_{l_2} \succ \dots \succ \mathbf{B}_{l_L}$ where the notation “ \succ ” is used to indicate “superior to.”

3. 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.

we calculate (7) by removing \mathbf{B}_l from the full-band set Ω as follows:

$$\text{MaxV}(\mathbf{B}_l) = (\mathbf{1}^M)^T ((\mathbf{D}_{\Omega_l^c}^M)^T \mathbf{R}_{\Omega_l^c}^{-1} \mathbf{D}_{\Omega_l^c}^M)^{-1} \mathbf{1}^M \quad (11)$$

which measures the maximal variance caused by removing the band image \mathbf{B}_l from Ω . This implies that the larger the $\text{MaxV}(\mathbf{B}_l)$ is, the higher priority the l th band \mathbf{B}_l is. Using (11), we can further develop another BP criterion, called BCSCBS-BP, to rank all bands.

Algorithm 2 BCSCBS-BP

1. Initial condition:
For each band, \mathbf{B}_l , find priority score, $\text{MaxV}(\mathbf{B}_l)$, according to (11).
2. Prioritize bands according to $\{\text{MaxV}(\mathbf{B}_l)\}_{l=1}^L$

$$\mathbf{B}_{l_j} \succ \mathbf{B}_{l_k} \Leftrightarrow \text{MaxV}(\mathbf{B}_{l_j}) > \text{MaxV}(\mathbf{B}_{l_k}) \quad (12)$$

i.e., $\mathbf{B}_{l_1} \succ \mathbf{B}_{l_2} \succ \dots \succ \mathbf{B}_{l_L}$ where the notation “ \succ ” is used to indicate “superior to.”

3. 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.

IV. CSCBS-BAND SELECTION

BS is a commonly used approach to remove redundant spectral information by only keeping a selective band subset and has received considerable interest lately in HSIC. It can be performed in a supervised or an unsupervised manner. When it is implemented as a supervised BS method, it generally specifies its application first, such as target detection, classification, endmember extraction, and spectral unmixing, which determines features for BS via prior knowledge, for example, training samples. It is then followed by a feature extraction algorithm to prioritize features that are used to select desired bands. As a result, the selected bands vary with different applications. On the other hand, when it comes to an unsupervised BS method, no prior knowledge is available to be used for BS.

In this section, we extend the two CSCBS-based BP criteria designed in Section III to develop two CSCBS-based BS methods corresponding to FCSCBS-BP and BCSCBS-BP.

A. Band Augmentation by CSCBS-SFBS

The FCSCBS-BP designed in Section III-A ranks all bands by their priority scores, $\{\text{MinV}(\mathbf{B}_l)\}_{l=1}^L$. The following LCMV-based sequential forward BS (CSCBS-SFBS) takes advantage of FCSCBS-BP to augment selected bands one at a time sequentially based on (10) in a feedforward manner.

Algorithm 3 CSCBS-SFBS

1. Initial condition:
Determine n_{BS} .
Find

$$\mathbf{B}_{l_1} = \arg \left\{ \min_{\mathbf{B}_l \in \Omega} (\mathbf{1}^M)^T ((\mathbf{D}_{\mathbf{B}_l}^M)^T \mathbf{R}_{\mathbf{B}_l}^{-1} \mathbf{D}_{\mathbf{B}_l}^M)^{-1} \mathbf{1}^M \right\} \quad (13)$$

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

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

2. Band augmentation by SFBS

$$\mathbf{B}_{l_j} = \arg \left\{ \min_{\mathbf{B}_l \in \Omega_{j-1}^c} (\mathbf{1}^M)^T ((\mathbf{D}^M)^T \mathbf{R}_{\Omega_{j-1} \cup \{\mathbf{B}_l\}}^{-1} \mathbf{D}^M)^{-1} \mathbf{1}^M \right\} \quad (14)$$

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

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

$$\Omega_j = \{\mathbf{B}_{l_1}, \mathbf{B}_{l_2}, \dots, \mathbf{B}_{l_j}\} = \Omega_{j-1} \cup \{\mathbf{B}_{l_j}\} \quad (15)$$

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

As noted in the introduction, three key features distinguish CSCBS-SFBS from FCSCBS-BP. The first one is that FCSCBS-BP ranks all bands individually according to (14) compared to CSCBS-SFBS that augments the selected bands one after another sequentially. A second one is that CSCBS-SFBS takes into account by only using the selected bands in Ω_j , while FCSCBS-BP considers only the band currently being evaluated for ranking. A third one is that CSCBS-SFBS does not have to run through all bands as opposed to FCSCBS-BP that must run through all the bands in Ω . Specifically, CSCBS-SFBS can terminate its augmentation process as long as the number of bands to be selected, n_{BS} , is reached.

B. Sequential Backward Band Selection by CSCBS

If we consider a band as a feature vector, CSCBS-SFBS is actually a feature vector-based sequential forward floating search [71] which has been used in BS, called SFS [45]. As a complete opposite, we can also develop a feature vector-based SBS that selects the bands backward by starting from the full-band set. The search strategy used by SBS is generally referred to as leave-one-out search policy. It removes one band at a time from the currently remaining unselected bands using (11). The process is continued on in a backward manner. The resulting BS is referred to as sequential backward BS (CSCBS-SBBS), as described in Algorithm 4.

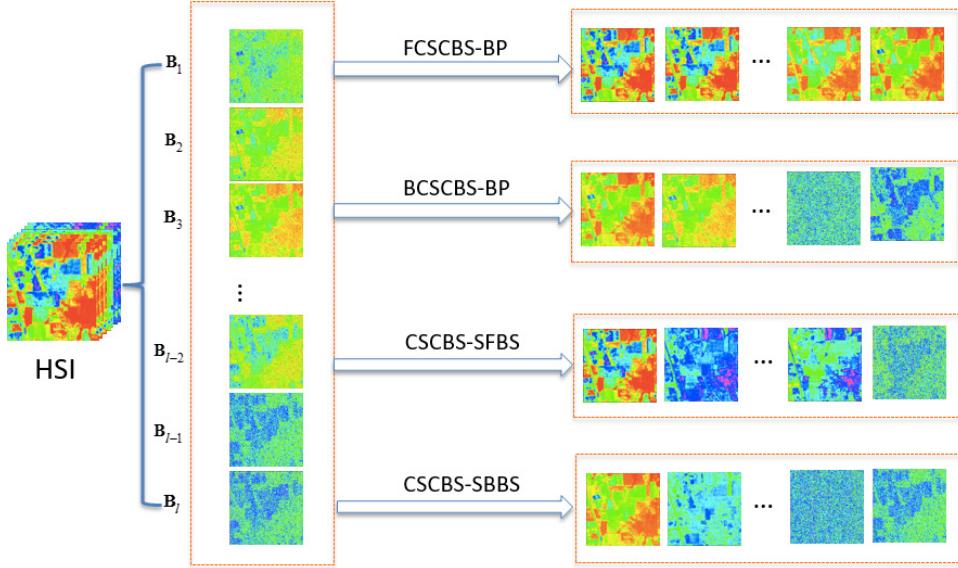


Fig. 1. CSCBS-based BP criteria and BS methods.

Algorithm 4 CSCBS-SBBS

- Initial condition:

Determine n_{BS} .
Find

$$\mathbf{B}_{l_1} = \arg \left\{ \max_{\mathbf{B}_l \in \Omega} (\mathbf{1}^M)^T ((\mathbf{D}_{\Omega - \{\mathbf{B}_l\}}^M)^T \mathbf{R}_{\Omega - \{\mathbf{B}_l\}}^{-1} \mathbf{D}_{\Omega - \{\mathbf{B}_l\}}^M)^{-1} \mathbf{1}^M \right\} \quad (16)$$

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

- BS by SBBS

$$\mathbf{B}_{l_j} = \arg \left\{ \max_{\mathbf{B}_l \in \Omega_{j-1}^c} (\mathbf{1}^M)^T ((\mathbf{D}_{\Omega - \{\mathbf{B}_l\}}^M)^T \mathbf{R}_{\Omega - (\Omega_{j-1} \cup \{\mathbf{B}_l\})}^{-1} \times \mathbf{D}_{\Omega - \{\mathbf{B}_l\}}^M)^{-1} \mathbf{1}^M \right\} \quad (17)$$

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

- If $j < n_{BS}$,

$$\Omega_j = \{\mathbf{B}_{l_1}, \mathbf{B}_{l_2}, \dots, \mathbf{B}_{l_j}\} = \Omega_{j-1} \cup \{\mathbf{B}_{l_j}\} \quad (15)$$

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

Fig. 1 describes a diagram that shows two types of LCMV-based BP criteria, FCSCBS-BP and BCSCBS-BP, and two types of LCMV-based BS methods, CSCBS-SFBS and CSCBS-SBBS.

There are three important perspectives between CSCBS-BS and CBS. One is that CSCBS-BS constrains class signatures in contrast to CBS that constrains bands. Another is that CSCBS-BS constrains multiple class signatures simultaneously compared to CBS that can only constrain one single band at a time. The most significant difference is that

CSCBS-BS is developed for a multiclass classification problem, a task that cannot be accomplished by CBS. Specifically, CSCBS-BS uses a class signature to specify a particular class. Finally, we would like to emphasize that although CBS is also designed for BS [37], the design rationale of CSCBS-BS deviates from that used in CBS in the sense that CSCBS-BS constrains class signatures using partial bands compared to CBS that constrains single individual bands with no respect to the target signal sources. Accordingly, CSCBS-BS is particularly designed for BS for HSIC, whereas CBS is designed for BP not BS without specific applications.

V. ITERATIVE CSCBSC

The classification performance is generally determined by two key factors, classifier and classification measures. In this section, we discuss the first issue and, then, the second issue in Section VI.

The concept of LCMV in [33] has been proven to be very useful in various applications. As noted in [73] and [74], it has been applied to hyperspectral mixed pixel classification (MPC). However, there are three significant differences between CSCBS-based BS and LCMV-based MPC (LCMV-MPC). The first and foremost difference is that CSCBS is particularly designed for BS used for HSIC, whereas LCMV-MPC is designed as an abundance estimator to be used for linear spectral unmixing. Consequently, CSCBS-BS only uses partial bands as opposed to LCMV-MPC that uses full bands. A second difference is that CSCBS-BS makes use of the MV resulting from CSCBS as its BP criterion to develop BS methods, but LCMV-MPC is only interested in abundance fractions estimated by MV, not MV itself. A third difference is that for MV to be used as a BP criterion CSCBS-BS proves that MV is monotonically decreasing as the number of bands increases. This fact was not shown in [73] and [74], because LCMV-MPC does not need this fact for MPC. However, it is this monotonically decreasing

sequence of MV to make the CSCBS-produced MV be capable of measuring effectiveness of selected bands. This is quite different from LCMV-MPC in the sense that the latter uses full bands and has no regard to BS. Most interestingly, LCMV-MPC does not take advantage of MV, which is the key measure to design the BP/BS criterion used for CSCBS-BS. Thus, CSCBS-BS is in fact completely different from CBS and LCMV-MPC in many ways as mentioned previously.

Algorithm 5 ICSCBSC

1. Initial condition: Let $\Omega^{(0)}$ be the original band set, $T^{(0)}$ be the original desired class signatures. Set an initial value for the error threshold ε . Let $k = 1$.
2. Implement δ_k^{CSCBSC} on $\Omega^{(k)}$ using $T^{(k)}$ and $R^{(k)}$ to produce B_k^{LCMV} which is the classification map.
3. Use a Gaussian filter to blur $|B_k|^{\text{CSCBSC}}$ where $|B_k|^{\text{CSCBSC}}$ is the absolute value of B_k^{CSCBSC} . The resulting image is denoted by Gaussian-filtered CSCBSC-classification map, $|GB_k|^{\text{CSCBSC}}$.
4. Form a new hyperspectral image cube $\Omega^{(k)}$ by augmenting the hyperspectral image cube, $\Omega^{(k-1)}$ obtained at $k - 1$ iteration with adding new $|GB_k|^{\text{CSCBSC}}$ obtained at the k th iteration, i.e., $\Omega^{(k)} = \Omega^{(k-1)} \cup |GB_k|^{\text{CSCBSC}}$ where the desired signature matrix $T^{(k)}$ is obtained by finding new LCMV-classified class sample means from $\Omega^{(k)}$.
5. Calculate the following stopping rule,

$$\text{TI}^{(k)} = \frac{\text{size}(|B_k|^{\text{CSCBSC}} \cap |B_{k-1}|^{\text{CSCBSC}})}{\text{size}(|B_k|^{\text{CSCBSC}} \cup |B_{k-1}|^{\text{CSCBSC}})} \quad (18)$$

where TI is Tanimoto index (TI) defined in [75] and size(S) is size of a set S, $|B_k|^{\text{CSCBSC}}$ and $|B_{k-1}|^{\text{CSCBSC}}$ are the binary image of the k th CSCBSC classification map and the binary image of the $k - 1$ st CSCBSC classification map thresholded by Ostu's method in [76] respectively.

6. If $\text{TI}^{(k)}$ is greater than a threshold ε , go to step 7. Otherwise, continue.
7. Form $\Omega^{(k+1)} = \Omega^{(k)} \cup \{ |GB_k|^{\text{CSCBSC}} \}$. Let $k \leftarrow k + 1$ and go to step 2.
8. CSCBSC is terminated and apply Ostu's method to threshold $|B_k|^{\text{CSCBSC}}$ into multi-class classification map.

As noted, LCMV can be also used for MPC. Unfortunately, it is ineffective when it is directly applied to classification. This is because, LCMV works on a single pixel basis and does not account for spatial correlation. To address this issue, a recent work [30] extended LCMV to an iterative multiclass BKG suppression classification (IMCBSC) that includes Gaussian filters to obtain spatial correlation of LCMV-MPC abundance fractional maps and feeds back the captured spatial information to the input image data in an iterative manner. Consequently, the more the iterations are carried out, the more the spatial information is included. As a matter of fact, IMCBSC is essentially an iterative version of CSCBSC, (ICSCBSC) to be presented in this section. Accordingly, the terminology of

ICSCBSC is used in this paper, and ICSCBSC will be used as a classifier throughout all our experiments.

For the purpose of visual illustration, Fig. 2 provides a graphic flowchart of implementing ICSCBSC.

VI. CLASSIFICATION MEASURES FOR PERFORMANCE ANALYSIS

As noted in Section I, CSCBS is developed to address the BKG issue; in which case, the commonly used OA and AA are not effective measures to take care of BKG. This section looks into a new classification measure, called PR, by extending binary classification to a multiclass classification problem in terms of a multiple hypotheses testing formulation as

$$\begin{aligned} H_1 : \text{class } 1 &\approx p(y|H_1) \\ &\vdots \\ H_M : \text{class } M &\approx p(y|H_M) \end{aligned} \quad (19)$$

where $p(y|H_i)$ is the probability density function of the observation y under hypothesis H_i to determine y belonging to the i th class H_i . By virtue of (19), an M -class confusion matrix can be constructed in Fig. 3.

Based on Fig. 3, we can derive several classification measures as follows.

M	Number of classes.
n_i	Number of data samples in the i th class.
n_{ji}	Number of signal samples in the i th class to be classified into the j th class.
n_{ii}	Number of signal samples in the i th class correctly classified into the i th class.
\hat{n}_{ij}	Number of data samples being classified in the j th class but actually in the ground-truth i th class.
\hat{n}_{jj}	Number of signal samples being classified in the j th class and also actually in the j th class.
$\tilde{n}_i = \sum_{j=1, j \neq i}^M n_{ji}$	Number of signal samples in the i th class misclassified into the i th class with $j \neq i$. = $\sum_{j=1}^M n_{ji}$.
N	Total number of data samples, $N = \sum_{i=1}^M n_i$.
C_i	Number of data samples in the i th class.
\tilde{C}_i	Number of data samples from classes $\{C_j\}_{j=1, j \neq i}^M$ classified into C_j . = (n_i / N) .
p_i	

$$\begin{aligned} p_{\text{accuracy}}(C_i) &= \text{accuracy of the } i\text{th class} \\ &= p(C_{ii}|C_i) = \frac{n_{ii}}{\sum_{j=1}^M n_{ji}} \end{aligned} \quad (20)$$

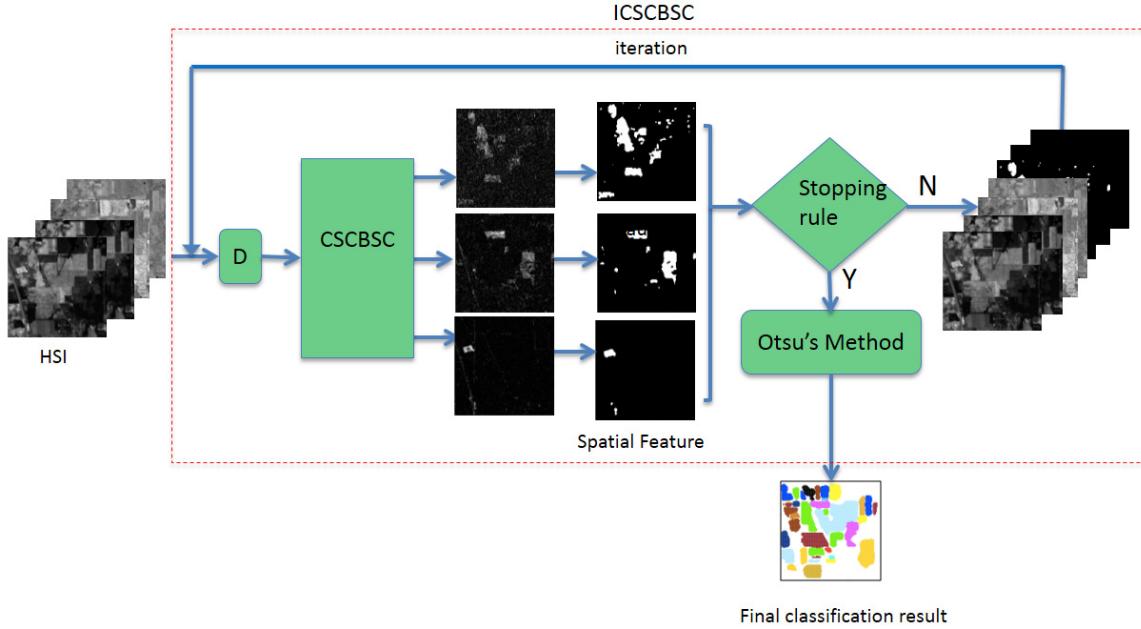


Fig. 2. Graphic implementation of ICSCBSC.

$$\begin{aligned}
 p(\tilde{C}_i | C_i) &= \text{misclassification rate of the } i\text{th class, } C_i \\
 &= \sum_{j=1}^M p(C_{ij} | C_i) = \frac{\sum_{j=1, j \neq i}^M n_{ij}}{\sum_{j=1}^M n_{ij}} \\
 &= 1 - p(C_{ii} | C_i) = 1 - p_{\text{accuracy}}(C_i) \\
 P(\tilde{C}) &= \sum_{i=1}^M p_i(C_i) p(\tilde{C}_i | C_i) \\
 &= \text{misclassification rate}
 \end{aligned} \tag{21}$$

$$P_{\text{PR}}(C_i) = p(C_i | \text{classification}) = \frac{\hat{n}_{ii}}{n_i} \tag{22}$$

$$P_{\text{PR}}(C) = \sum_{j=1}^M p_j \frac{\hat{n}_{jj}}{n_j} \tag{23}$$

$$\begin{aligned}
 P_{\text{OA}} &= p_{\text{overall-accuracy}}(\{C_i\}_{i=1}^M) \\
 &= \sum_{i=1}^M p_i p(C_{ii} | C_i) = \sum_{i=1}^M p_i \frac{n_{ii}}{n_i}
 \end{aligned} \tag{24}$$

$$P_{\text{AA}} = p_{\text{average-accuracy}}(\{C_i\}_{i=1}^M) = \frac{1}{M} \sum_{i=1}^M n_{ii}. \tag{25}$$

As a final remark, there is a critical difference between P_{OA} and P_{PR} . P_{OA} is calculated to measure the accuracy of any classifier given that the ground truth of class knowledge is provided *a priori*. By contrast, P_{PR} is calculated to measure the precision or efficacy of a given classifier. As a result, P_{OA} can be considered as *a priori* classification rate with provided ground truth of class knowledge as opposed to P_{PR} that can be considered as *a posteriori* classification rate, which uses classified knowledge obtained by *a posteriori* knowledge. Unfortunately, to the best of our knowledge, most works reported in the past only calculated P_{OA} and P_{AA} or kappa coefficients to evaluate the classification performance and none

true classification	C_1	C_2	\dots	C_M
C_1	n_{11}	n_{12}	\dots	n_{1M}
C_2	n_{21}	n_{22}	\dots	n_{2M}
C_M	n_{M1}	n_{M2}	\dots	n_{MM}

$p_{\text{precision}}(C_1) = \frac{\hat{n}_{11}}{\sum_{j=1}^M \hat{n}_{1j}}$
 $P_{\text{OA}} = \sum_{i=1}^M p_i \left(\frac{n_{ii}}{\sum_{j=1}^M n_{ji}} \right)$

Fig. 3. M -class confusion matrix.

of them except [30]–[32] use the precision rate to measure the effectiveness of a given classifier in classification performance.

VII. REAL IMAGES TO BE STUDIED

Three popular real hyperspectral images, Purdue University's Indiana Indian Pines, Salinas, and the University of Pavia, Italy, were used for experiments. As for these three hyperspectral image scenes, their MATLAB data files can be downloaded from the webpage [72].

A. Purdue Indian Pines Scene

The first image scene used for experiments is an airborne visible/infrared imaging spectrometer (AVIRIS) hyperspectral data set that is the Purdue Indiana Indian Pines test site shown in Fig. 4(a) with its ground truth of 17 class maps in Fig. 4(b). 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 are

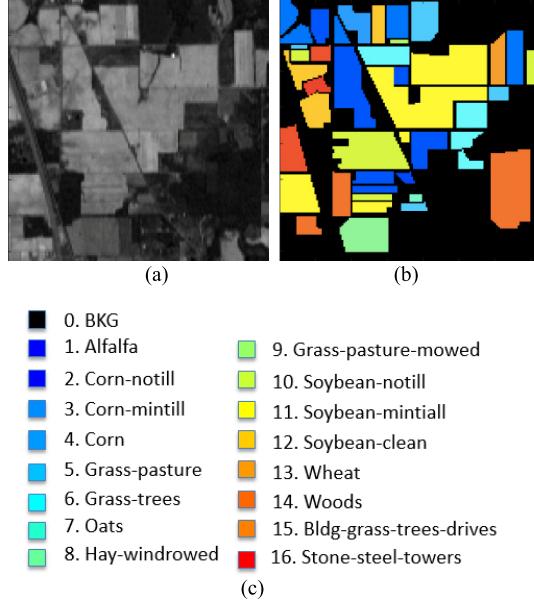


Fig. 4. Purdue's Indiana Indian Pines scene with 16 classes. (a) Purdue's scene. (b) Color ground-truth map. (c) Ground-truth class labels.

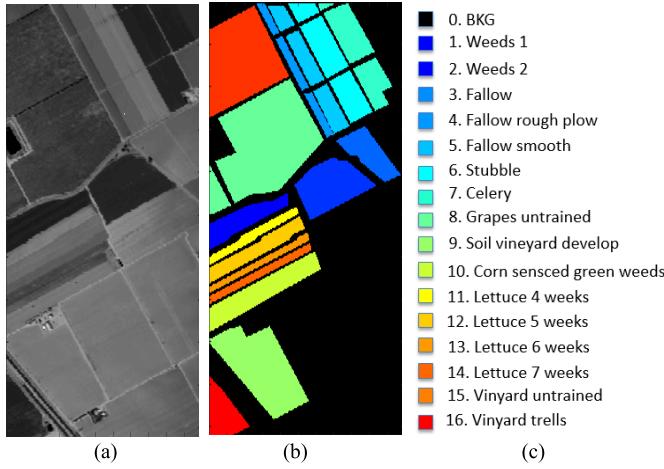


Fig. 5. Salinas scene with 16 classes. (a) Salinas scene. (b) Color ground-truth map. (c) Ground-truth class labels.

given in caption. The data set is available at the website, <https://purr.purdue.edu/publications/1947/serve/1?el=1>. It was recorded in June 1992 with 220 bands, which include water absorption bands (bands 104–108 and 150–163, 220).

B. Salinas Data

A second AVIRIS data used for experiments were Salinas scene shown in Fig. 5(a), which was captured by the AVIRIS sensor over Salinas Valley, California, USA, and with a spatial resolution of 3.7 m/pixel with spectral resolution of 10 nm. It has a size of $512 \times 217 \times 224$. Fig. 5(b) and (c) shows the color composite of the Salinas image along with the corresponding ground-truth class labels.

C. ROSIS Data

The last hyperspectral image data used for experiments were the University of Pavia image shown in Fig. 6, which is an

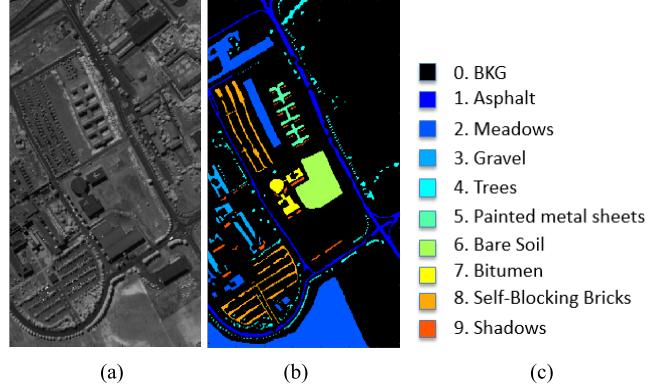


Fig. 6. University of Pavia scene with nine classes. (a) University of Pavia scene. (b) Color ground-truth map. (c) Class labels.

urban area surrounding the University of Pavia. It was recorded by the ROSIS-03 satellite sensor. It is of size $610 \times 340 \times 115$ with a spatial resolution of 1.3 m/pixel and a spectral coverage ranging from 0.43 to $0.86 \mu\text{m}$ with spectral resolution of 4 nm (12 most noisy channels were removed before experiments). Nine classes of interest plus BKG class, class 0 is considered for this image.

VIII. EXPERIMENTS

In order to conduct a decent comparative analysis, selecting an appropriate classifier to evaluate the classification performance is crucial. In the following experiments, the ICSCBSC was selected as the desired classifier for two main reasons. First of all, since both CSCBS-BP criteria and CSCBS-BS methods are derived from LCMV, ICSCBSC is a perfect candidate for this purpose. Second, ICSCBSC has been shown in [30] to perform well compared to the edge-preserving filter-based methods in [8]. Therefore, its selection is creditable. Also, for evaluation of CSCBS-based BP criteria and BS methods, the following four categories of BS methods are selected for comparison.

- 1) *Uniform Band Selection*: According to our extensive experiments, uniform BS (UBS) is a reasonably good BS method. It is the simplest BS method and does not require any prior knowledge or BS criterion.
- 2) *CBS*: There are two reasons to choose CBS for comparison. The most important one is that the CBS implemented in this paper used the LCMV-BCM version in [37, eq. (14)]. A second one is that CBS was compared and shown to perform better than many other BS methods in [37]; in that case, there is no need to include these methods here for comparison as long as we can show that the LCMV-BCM indeed performs better than CBS, which is the case according to Tables III–V.
- 3) *Minimum Estimated Abundance Covariance* [45]: It also uses the minimum covariance derived from the estimated abundance matrix, which is similar to the MV in (1).
- 4) *Dominant Set Extraction BS* in [56] and *Multigraph Determinantal Point Process* in [59]: These two are most recently developed BS methods for HSIC, and they also compared their results with constrained energy minimization (CEM)/LCMV-based methods that are also based on the LCMV formulation specified by (1).

TABLE I
 n_{BS} ESTIMATED BY HYSIME AND HFC/NWHFC

	HySime	$P_F=10^{-1}$	$P_F=10^{-2}$	$P_F=10^{-3}$	$P_F=10^{-4}$	$P_F=10^{-5}$
Purdue	13	73/21	49/19	35/18	27/18	25/17
Salinas	22	32/33	28/24	25/21	21/21	20/20
University of Pavia	58	25/34	21/27	16/17	14/14	13/12

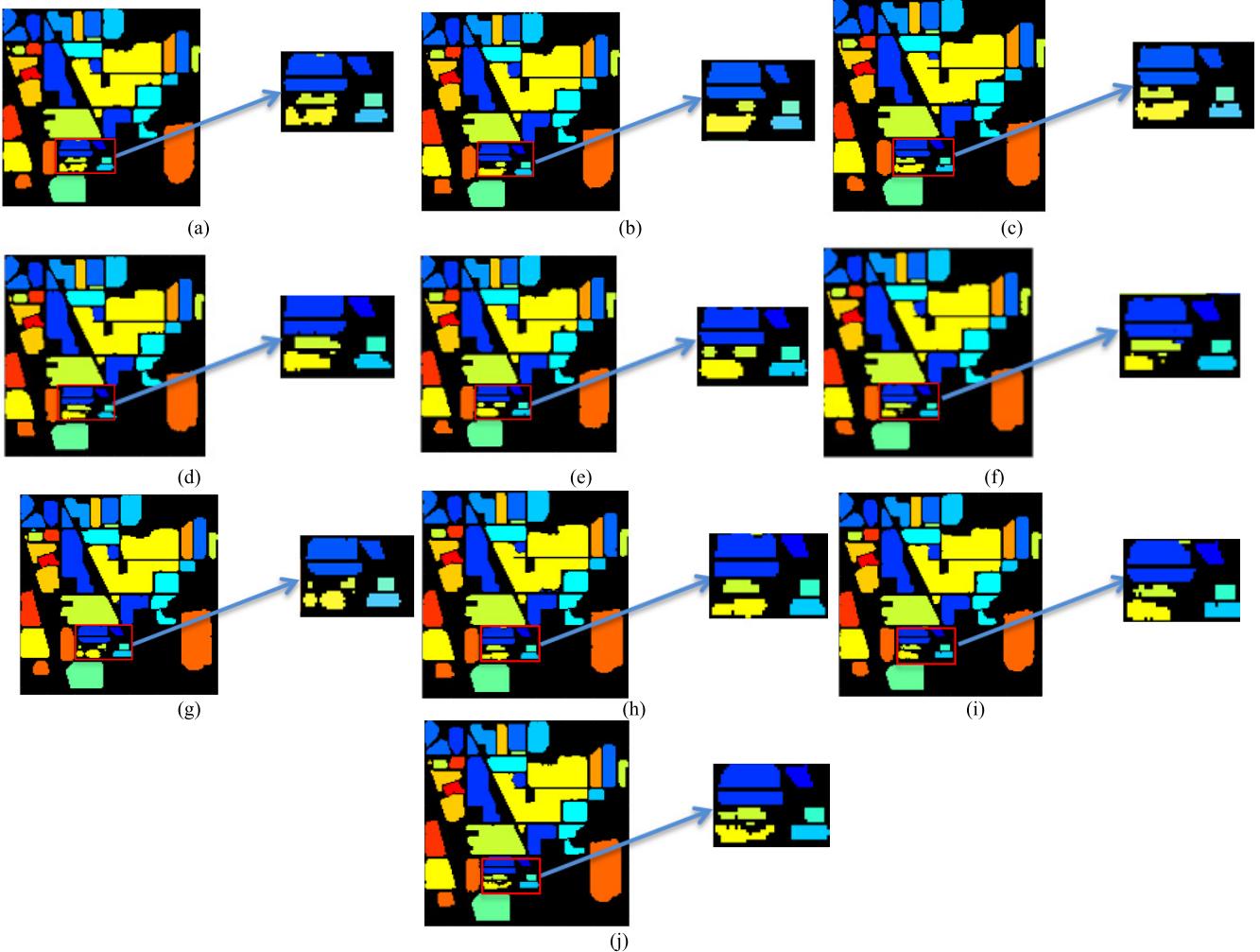


Fig. 7. Classification results produced by ICSCBSC with $\sigma = 0.5$ and $TI = 0.99$ for Purdue's data using the bands selected in Table II. (a) Full bands. (b) UBS. (c) MEAC. (d) MDPP. (e) DSEBS. (f) CBS. (g) BCSCBS-BP. (h) FCSCBS-BP. (i) CSCBS-SFBS. (j) CSCBS-SBBS.

Regarding the number of bands required for BS, n_{BS} , the VD developed in [74] and [77]–[79] was used, where two popular methods, Harsanyi–Farrand–Chang (HFC) method [73] and noise whitened HFC (NWHFC) methods [65], [70], and Hyperspectral Subspace identification minimum error (HySime) [81] were implemented to determine the value of n_{BS} . Table I tabulates n_{BS} estimated by HySime and HFC/NWHFC methods. Interestingly, compared to two other scenes, Salinas and the University of Pavia, the values of n_{BS} in Table I for Purdue data varied in a wide range when P_F was set from 10^{-1} to 10^{-3} . However, when P_F was set to smaller values below 10^{-3} , the values of n_{BS} became stable. This was mainly due to the fact that among the 16 classes in the Purdue data, there are cornlike, grasslike, and soybeanlike classes with very similar signatures. As a result,

P_F must be set to a smaller value to avoid false alarms caused by different signatures from the same class. For our experiments conducted in this paper, n_{BS} was chosen with $P_F = 10^{-4}$ in Table I, where $n_{BS} = 18$ for Purdue's data, $n_{BS} = 21$ for Salinas, and $n_{BS} = 14$ for the University of Pavia. Table II lists bands selected by UBS, CBS, minimum estimated abundance covariance (MEAC), FCSCBS-BP, BCSCBS-BP, CSCBS-SFBS, CSCBS-SBBS, DSEBS, and MDPP.

Figs. 7–9 show the classification results produced by ICSCBSC using the bands selected in Table II for three image scenes, where the parameters used by ICSCBSC were the same as those in [23], that is, $\sigma = 0.5$ and $TI \geq \varepsilon = 0.99$ for Purdue's data, $\sigma = 1$ and $TI \geq \varepsilon = 0.99$ for Salinas, and $\sigma = 0.5$ and $TI \geq \varepsilon = 0.98$ for the University

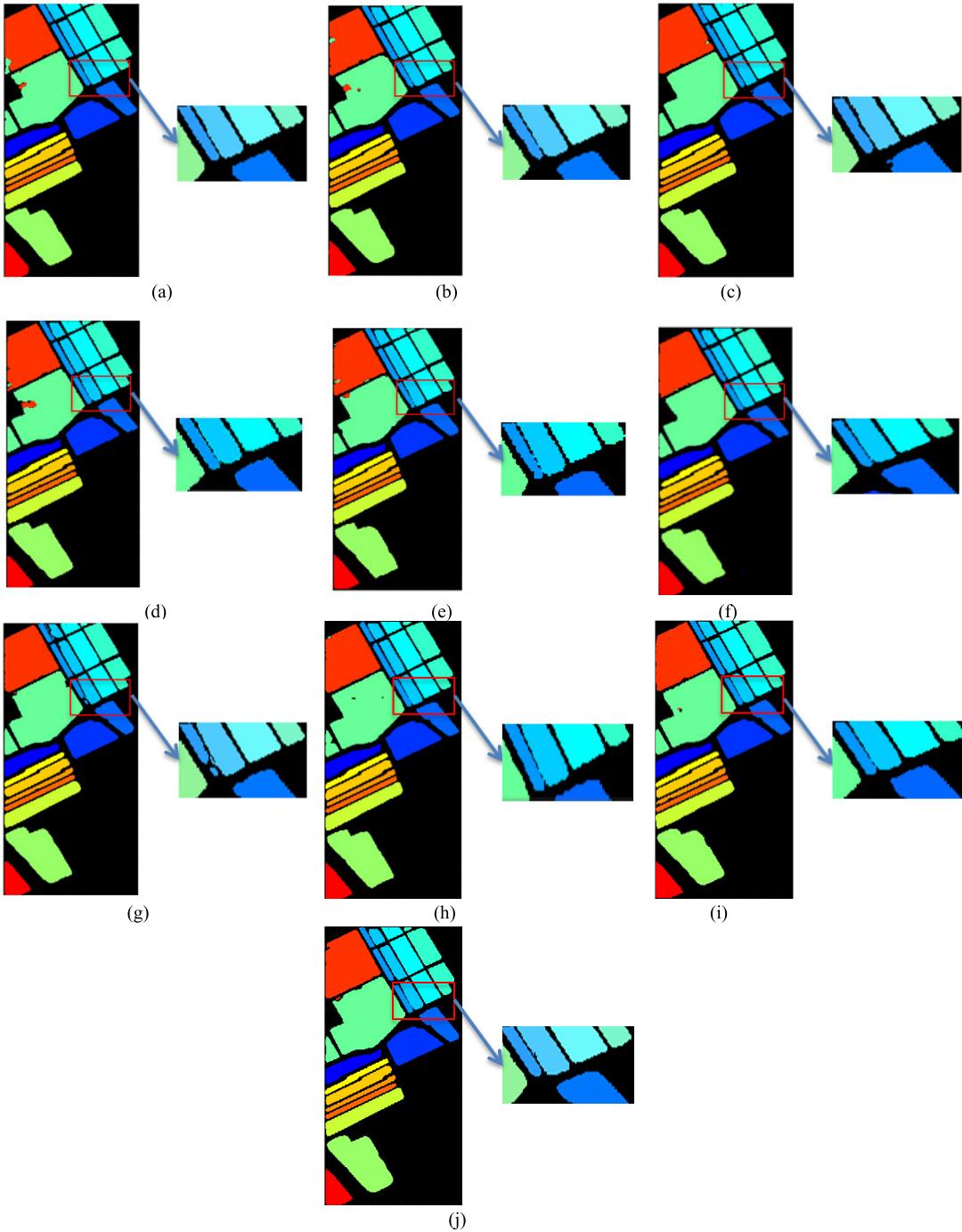


Fig. 8. Classification results produced by ICSCBSC with $\sigma = 1$ and $TI = 0.99$ for Salinas using the bands selected in Table II. (a) Full bands. (b) UBS. (c) MEAC. (d) MDPP. (e) DSEBS. (f) CBS. (g) BCSCBS-BP. (h) FCSCBS-BP. (i) CSCBS-SFBS. (j) CSCBS-SBBS.

of Pavia. In order to see the classification results more clearly, we selected a particular region of interest from each of three scenes to zoom the classification result of each of test BS methods. For example, by examining the classification results of class 10 (soybean-notill) and class 11 (soybean-mintall) colored by light green and yellow in the extracted zoomed-in areas of Fig. 7 for the Purdue data, the best performance was the one produced by CSCBS-SBBS. All these improvements were resulting from BKG suppression. Similarly, the same improvements can be witnessed in Figs. 8 and 9.

Tables II–V also tabulate their corresponding values of P_{OA} in (21) and P_{PR} in (23) from Figs. 7–9, where the best results are boldfaced and the accuracies of each of classes in (20) are also tabulated for detailed data analysis. It should be noted that the PR was calculated based on classes of interest without accounting for BKG in its PR calculation for the purpose of comparison. If the BKG is included in PR calculation, its values would be very low as shown in [30]–[32]. For completeness, the results of ICSCBSC using bands selected by UBS, CBS, and MEAC are also given for further comparison, where CSCBS-

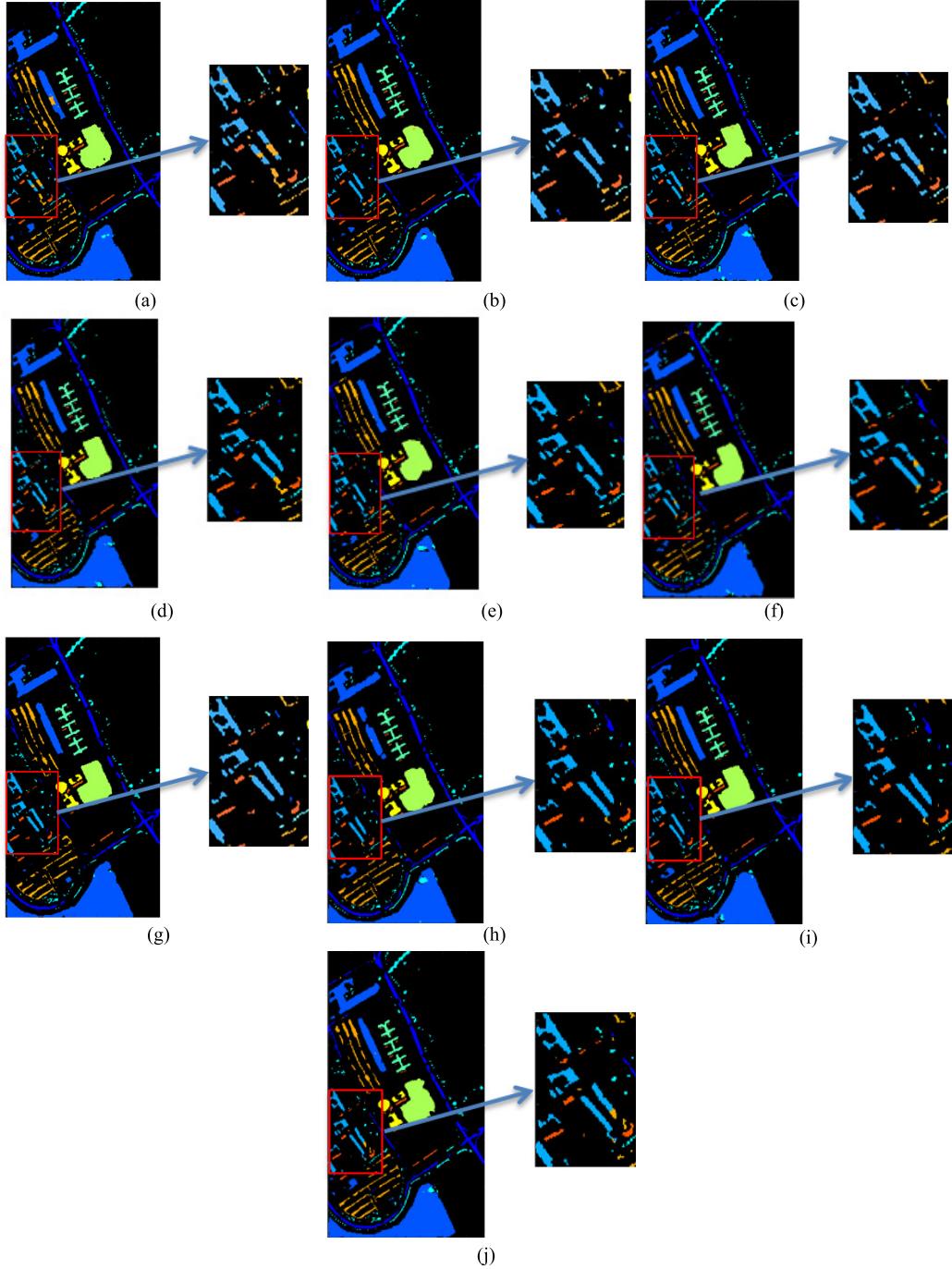


Fig. 9. Classification results produced by ICSCBSC with $\sigma = 0.5$ and $TI = 0.98$ for Salinas using the bands selected in Table II. (a) Full bands. (b) UBS. (c) MEAC. (d) MDPP. (e) DSEBS. (f) CBS. (g) BCSCBS-BP. (h) FCSCBS-BP. (i) CSCBS-SFBS. (j) CSCBS-SBBS.

based methods generally performed better among all test BS methods.

Once again, using the extracted zoomed-in areas of Fig. 7 for the Purdue data, as an illustrative example, the best accurate rates of class 10 and class 11 in Table III were produced by using CBS and UBS, respectively. However, if both P_{OA} and P_{PR} are taken into account, ICSCBSC using bands selected by MDPP, DSBES, and CSCBS-SBBS were among the best and performed better than ICSCBSC using full bands.

For Salinas P_{OA} and P_{PR} produced by ICSCBSC using bands selected by BCSCBS-BP were all higher than that produced by ICSCBSC using full bands with CSCBS-SBBS producing the highest P_{OA} and precision rates among all other BS methods.

Interestingly, conclusions for the University of Pavia were little bit different from those drawn for Purdue's data and Salinas, where ICSCBSC using bands selected by CBS, BCSCBS-BP, CSCBS-SFBS, and CSCBS-SBBS produced better P_{OA} and P_{PR} with second lowest P_{PR} pro-

TABLE II
BANDS SELECTED BY UBS, CBS, MEAC, MDPP, DSEBS, FCSCBS-BP, BCSCBS-BP, CSCBS-SFBS, AND CSCBS-SBBS

Data set	Method	Selected bands																				
		1	14	27	40	53	66	79	92	105	118	131	144	157	170	183	196	209	220			
Purdue Indian (18 bands)	UBS	5	36	157	90	158	103	45	4	6	3	162	97	41	99	104	106	88	154			
	CBS	159	92	3	96	82	36	39	55	41	1	2	33	206	38	163	17	204	9			
	MEAC	10	39	59	75	79	85	92	130	140	146	147	149	150	152	160	164	175	193			
	MDPP	42	129	97	131	174	16	176	177	172	43	192	193	98	171	99	132	40	33			
	DSEBS	29	28	30	32	27	26	25	24	33	31	23	22	21	20	19	34	18	120			
	FCSCBS-BP	5	35	29	11	3	42	12	96	121	4	101	158	65	31	213	99	165	26			
	BCSCBS-BP	29	30	25	33	24	23	40	203	26	86	15	46	116	65	173	31	93	88			
	CSCBS-SFBS	5	3	29	35	4	96	11	12	158	42	121	101	213	95	186	31	165	6			
Salinas (21 bands)	UBS	1	12	23	34	45	56	67	78	89	100	111	122	133	144	155	166	177	188	199	210	224
	CBS	47	46	32	51	53	50	52	29	44	31	55	28	25	33	54	27	9	56	21	42	48
	MEAC	107	148	203	149	5	8	105	3	28	12	18	10	44	36	25	17	51	32	110	68	58
	MDPP	1	8	11	22	27	28	50	57	58	65	90	99	105	119	123	134	142	157	175	191	204
	DSEBS	99	101	16	119	177	112	44	46	120	47	131	175	196	121	17	102	174	180	187	135	42
	FCSCBS-BP	44	45	43	46	47	48	42	49	50	51	52	53	54	55	56	57	58	59	41	60	61
	BCSCBS-BP	175	120	173	174	106	176	121	32	197	220	84	38	44	196	105	192	12	195	10	117	156
	CSCBS-SFBS	44	45	25	33	82	97	156	104	223	202	168	43	108	187	111	93	183	5			
U. of Pavia (14 bands)	CSCBS-SBBS	175	174	173	176	120	121	117	220	219	84	106	178	177	85	83	170	218	223	12	10	11
	UBS	1	9	17	25	33	41	49	57	65	73	81	89	97	103							
	CBS	91	90	93	92	89	88	80	94	97	81	98	87	96	95							
	MEAC	1	23	24	40	42	58	56	59	48	31	47	83	25	54							
	MDPP	2	23	44	46	50	62	66	73	89	91	92	93	96	102							
	DSEBS	86	102	64	20	21	63	65	6	19	22	7	66	95	67							
	FCSCBS-BP	16	15	14	17	18	13	19	12	20	11	21	10	22	9							
	BCSCBS-BP	84	83	66	67	77	82	1	81	78	79	68	76	101	69							
	CSCBS-SFBS	16	59	20	63	1	65	66	56	11	64	12	13	18	58							
	CSCBS-SBBS	84	85	86	1	66	2	65	87	3	83	101	67	68	12							

TABLE III

P_{OA} AND P_{PR} CALCULATED BY THE CLASSIFICATION RESULTS IN FIG. 7 USING THE BANDS SELECTED IN TABLE II FOR PURDUE'S DATA

Class	Full bands	UBS	CBS	MEAC	MDPP	DSEBS	FCSCBS-BP	BCSCBS-BP	CSCBS-SFBS	CSCBS-SBBS
1	95.65	95.65	97.83	93.48	95.65	95.65	97.83	95.65	97.83	97.83
2	96.01	97.13	95.52	93.07	96.08	96.99	96.36	93.14	96.08	96.85
3	96.99	96.51	97.11	96.27	97.35	97.23	95.90	95.78	96.27	97.35
4	98.73	98.73	99.16	98.31	99.58	98.31	97.89	97.89	97.89	98.31
5	89.44	90.68	94.00	91.51	92.34	93.58	95.45	92.96	92.13	92.55
6	97.12	97.67	97.12	97.40	96.71	97.12	95.89	97.26	96.16	97.95
7	100	100	96.43	100	100	100	100	100	100	96.43
8	98.78	98.54	98.33	99.16	97.49	97.91	99.58	98.95	97.91	99.58
9	100	100	100	90.00	100	100	100	100	100	100
10	93.93	91.98	95.58	93.31	94.65	93.00	92.70	94.03	94.44	94.96
11	94.70	96.13	94.30	94.55	95.48	95.85	95.23	94.79	95.40	95.11
12	95.45	94.94	95.62	96.29	96.80	97.30	96.29	95.45	95.45	96.46
13	98.54	98.54	97.56	99.02	97.56	96.59	96.59	98.05	97.56	98.54
14	93.52	94.15	96.21	94.78	94.70	94.55	95.65	96.44	93.52	94.86
15	90.67	95.60	97.56	92.49	96.89	93.52	95.34	93.52	95.08	95.85
16	98.92									
P_{OA}	95.09	95.69	95.76	94.91	95.89	95.88	95.74	95.25	95.42	96.09
P_{PR}	99.93	99.92	97.93	99.49	99.94	99.99	99.93	99.59	99.89	99.89

duced by ICSCBSC using full bands next to the worst by CBS.

It is worthwhile and beneficial by looking into the results in zoomed-in areas of Figs. 7–9 and the results in Tables III–V side by side carefully to see comparative performance among all the test BS methods. For example, the region selected for the Purdue data in Fig. 7 was an area consisting of soybeans-min and soybeans-clean highlighted by yellow regions and grass/pasture/grass-pasture-mowed highlighted by light blue

regions. According to Table III, the BS methods to produce 3 highest P_{OA} values were CSCBS-SBBS, MDPP, and DSEBS. By referring to Fig. 7, CSCBS-SBBS classified both soybeans and grass very well in Fig. 7(j) compared to CSCBS-SFBS that classified grass nearly perfectly but not soybeans in Fig. 7(i).

Finally, according to Figs. 7–9 and Tables III–V, the experimental results clearly showed that HSIC can be benefited by BS as long as n_{BS} and bands are appropriately determined

TABLE IV

 P_{OA} AND P_{PR} CALCULATED BY THE CLASSIFICATION RESULTS IN FIG. 8 USING THE BANDS SELECTED IN TABLE II FOR SALINAS

Class	Full bands	UBS	CBS	MEAC	MDPP	DSEBS	FCSCBS-BP	BCSCBS-BP	CSCBS-SFBS	CSCBS-SBBS
1	95.52	97.16	96.96	97.71	97.76	97.16	97.16	97.21	97.46	97.46
2	98.42	98.85	98.98	98.44	97.99	99.17	98.95	99.01	98.82	98.68
3	93.78	95.50	93.52	94.03	93.98	95.65	93.42	97.12	95.85	94.43
4	95.62	94.69	96.34	94.33	97.49	94.74	88.24	96.99	96.99	96.20
5	96.90	96.45	95.86	95.19	95.22	96.90	90.96	97.13	95.83	95.75
6	98.79	98.59	98.51	98.56	98.79	98.56	94.79	99.01	98.59	98.64
7	98.63	98.21	97.12	98.18	97.99	97.65	98.11	96.93	97.23	96.48
8	96.69	95.81	97.50	97.40	95.23	96.11	98.27	97.53	97.30	96.77
9	95.87	95.60	94.78	94.74	95.29	95.73	96.68	96.63	95.12	96.68
10	96.67	96.37	95.64	96.34	96.46	97.25	96.83	97.44	96.16	97.83
11	97.75	97.85	92.60	91.10	97.75	98.31	96.25	97.75	96.91	99.16
12	97.15	96.16	96.22	95.54	97.46	97.66	96.26	97.25	95.85	95.64
13	96.51	96.94	95.09	93.35	96.40	95.63	94.98	96.94	93.34	96.94
14	95.89	98.14	97.85	97.66	97.01	98.04	96.76	97.76	98.04	94.86
15	94.00	95.27	96.41	96.52	95.42	95.25	96.71	96.66	96.24	96.53
16	93.30	96.07	93.47	93.86	95.07	95.02	94.41	95.41	95.13	94.41
P_{OA}	96.37	96.49	96.49	96.45	96.25	96.63	96.40	97.33	96.72	96.81
P_{PR}	99.45	99.58	99.95	99.90	99.46	99.70	99.82	99.97	99.97	99.93

TABLE V

 P_{OA} AND P_{PR} CALCULATED BY THE CLASSIFICATION RESULTS IN FIG. 9 USING THE BANDS SELECTED IN TABLE II FOR THE UNIVERSITY OF PAVIA

Class	Full bands	UBS	CBS	MEAC	MDPP	DSEBS	FCSCBS-BP	BCSCBS-BP	CSCBS-SFBS	CSCBS-SBBS
1	77.24	75.86	75.03	76.32	74.01	76.17	72.21	80.25	76.23	77.44
2	86.42	87.67	86.49	86.44	87.97	87.71	86.62	88.07	87.95	88.89
3	73.34	84.38	82.29	83.33	82.14	84.63	85.37	84.06	87.53	83.75
4	79.85	78.90	78.72	76.66	76.17	79.22	62.66	74.81	67.43	77.01
5	98.81	97.84	93.38	98.88	96.95	97.99	98.51	98.14	98.81	98.81
6	91.49	89.93	91.79	91.33	93.32	87.11	94.84	92.90	93.64	90.02
7	89.10	91.35	83.93	82.78	87.53	87.44	94.36	90.45	93.91	86.92
8	81.10	83.32	74.15	76.26	75.64	76.34	70.89	76.37	81.78	78.54
9	78.46	79.09	77.72	79.51	79.83	79.30	75.50	75.50	75.53	80.25
P_{OA}	84.32	85.19	83.45	83.85	84.33	84.25	82.5	85.39	85.03	85.32
P_{PR}	98.18	99.33	96.58	98.81	98.95	98.88	99.20	99.22	99.27	99.11

and selected, respectively. Specifically, ICSCBSC using bands selected by two CSCBS-BS methods, CSCBS-SFBS and CSCBS-SBBS, generally outperformed ICSCBSC using full bands in terms of OA and precision in most of cases. Most interestingly, Tables III–V showed that ICSCBSC using bands selected by UBS may not do better than that using bands selected by CSCBS-BS methods but clearly performed better than that using full bands.

Two concluding remarks are noteworthy.

- 1) It is important to note that the proposed ICSCBSC in Section V is designed based on LCMV specified by (4), which makes use of the inverse of the sample correlation matrix \mathbf{R} in (3), i.e., \mathbf{R}^{-1} , to suppress the BKG, as demonstrated in Figs. 7(g)–(j), 8(g)–(j), and 9(g)–(j). In addition, Tables III–V also show that the better classification results were indeed produced by ICSCBSC using bands selected by CSCBS-based methods in terms of OA and PR. Moreover, the effectiveness of BKG suppression using full bands has been also demonstrated by many experiments using the same three image scenes conducted in [30] and [31]. This paper further demonstrates that the same BKG-suppression

effectiveness is also applied when the bands are selected by CSCBS.

- 2) Defining the number of bands is different from determining n_{BS} . In other words, before determining n_{BS} , we must define what the number of bands means. In this paper, we define the number of bands based on applications in classification, where the selected bands are found and ranked by objection functions specified by classification criteria. Accordingly, all the compared BS methods are selected based on their classification performance and, so are the proposed CSCBS-based methods. Furthermore, since for each class, its class sample mean signature can be considered as a spectrally distinct signature, the concept of VD was applied to determine how many number of bands, n_{BS} , required for BS to select. Nevertheless, a recent work [82] develops an interesting approach from an information theory's point of view using the channel capacity to define so-called band capacity that can be used to define the number of bands solely based on data itself without referring to applications (see [82, Sec. VII]). More details can be found in [82].

IX. CONCLUSION

This paper develops a new approach to BS, called CSCBS-based BS, for HSIC. It is particularly designed to address BKG issue encountered in HSIC. In doing so, CBCBS borrows the idea of LCMV in adaptive beamforming by interpreting the desired and undesired signal arriving directions as classes of interest and BKG signatures, respectively. Accordingly, the variance resulting from the arrival directions of undesired signals impinging upon an array of sensors in adaptive beamforming can be interpreted exactly the same as the misclassification error caused by CSCBS. Furthermore, CSCBS can be also implemented as CSCBS-BS using a partial set of bands selected by a BS method instead of using full bands. In order to find an appropriate partial set of desired bands that minimize the variances, two types of CSCBS-based BS can be also derived, which are CSCBS-BP and CSCBS-BS, each of which also gives rise to its own corresponding two versions, FCSCBS-BP and BCSCBS-BP, and CSCBS-SFBS and CSCBS-SBBS, respectively. Experimental results demonstrate that judiciously selecting a partial band subset can actually improve classification accuracy over that produced by using full-band set for HSIC.

As a conclusion, we summarize several major contributions derived from this paper. First of all, we take advantage of the MV in (4), which can be considered as minimum least squares error resulting from the LCMV formulation by (2) as a criterion for CBCBS to select bands. Second and most importantly, we have proved that the MV of (4) is monotonically decreasing as the number of bands is increased in Theorem 1. As a result, it can be used for BP and BS. Third, we developed two types of CBCBS-based BP criteria as well as their corresponding CBCBS-based BS methods. Fourth, these CBCBS-based BP criteria and CBCBS-based BS methods can be further fused to two new versions of BS, CSCBS-SFBS and CSCBS-SBBS. Finally, in order to evaluate the performance of CBCBS-based selected bands, a comparative study is conducted by comparing two most recent algorithms, DSEBS, developed in [49], and MDPP, developed in [59], both of which were derived from CBS in [37], where their used criterion is CEM, which is actually a special case of LCMV, as shown in [73] and [74]. The experimental results demonstrated that CBCBS-based selected bands can actually improve the classification performance over using full bands and also performed better than DSEBS and MDPP.

APPENDIX

Proof of Theorem 1: For simplicity, let $\mathbf{D}_{\Omega_l}^M = \mathbf{D}_l$, $\mathbf{D}_{\Omega_{l+1}}^M = \mathbf{D}_{l+1}$, $\mathbf{c}^M = \mathbf{c}$, and represent a band image \mathbf{B}_l as a band vector \mathbf{b}_l , i.e., $\mathbf{B}_l = \mathbf{b}_l$. Also, let

$$\mathbf{X}_l = [\mathbf{r}_l^1 \mathbf{r}_l^2 \dots \mathbf{r}_l^N] = \begin{bmatrix} r_1^1 & r_1^2 & \dots & r_1^N \\ r_2^1 & r_2^2 & \dots & r_2^N \\ \vdots & \vdots & \ddots & \vdots \\ r_l^1 & r_l^2 & \dots & r_l^N \end{bmatrix} \quad (\text{A1})$$

denote the data matrix using the first l bands, which is an $l \times N$ dimensional matrix, where $\mathbf{r}_l^i = (r_1^i, r_2^i, \dots, r_l^i)^T$. Then, from (A1), the data matrix \mathbf{X}_{l+1} using $(l+1)$ bands is given by

$$\mathbf{X}_{l+1} = \begin{bmatrix} \mathbf{X}_l \\ \mathbf{r}_{l+1}^T \end{bmatrix} \quad (\text{A2})$$

where $\mathbf{r}_{l+1} = (r_{l+1}^1, r_{l+1}^2, \dots, r_{l+1}^N)^T$. Let $\mathbf{R}_{\Omega_l} = (1/N) \sum_{i=1}^N (\mathbf{r}_l^i (\mathbf{r}_l^i)^T) = (1/N) \mathbf{X}_l \mathbf{X}_l^T$. We have

$$\begin{aligned} \mathbf{R}_{\Omega_l \cup \{\mathbf{b}_{l+1}\}} &= \frac{1}{N} \sum_{i=1}^N (\mathbf{r}_{l+1}^i (\mathbf{r}_{l+1}^i)^T) = \frac{1}{N} \mathbf{X}_{l+1} \mathbf{X}_{l+1}^T \\ &= \frac{1}{N} \begin{bmatrix} \mathbf{X}_l \\ \mathbf{r}_{l+1}^T \end{bmatrix} \begin{bmatrix} \mathbf{X}_l^T & \mathbf{r}_{l+1} \end{bmatrix} \\ &= \frac{1}{N} \begin{bmatrix} \mathbf{X}_l \mathbf{X}_l^T & \mathbf{X}_l \mathbf{r}_{l+1} \\ \mathbf{r}_{l+1}^T \mathbf{X}_l^T & \mathbf{r}_{l+1}^T \mathbf{r}_{l+1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R}_{\Omega_l} & (1/N) \mathbf{X}_l \mathbf{r}_{l+1} \\ (1/N) \mathbf{r}_{l+1}^T \mathbf{X}_l^T & (1/N) \mathbf{r}_{l+1}^T \mathbf{r}_{l+1} \end{bmatrix}. \end{aligned} \quad (\text{A3})$$

Now, in order to take care of the scalar $1/N$ in (A3) to simplify the following mathematical derivations, we define $\tilde{\mathbf{X}}_l = (1/\sqrt{N}) \mathbf{X}_l$ and $\tilde{\mathbf{r}}_{l+1} = (1/\sqrt{N}) \mathbf{r}_{l+1}$, both of which absorb $1/\sqrt{N}$ into their notations to yield:

$$\mathbf{R}_{\Omega_l} = \frac{1}{N} \sum_{i=1}^N (\mathbf{r}_l^i (\mathbf{r}_l^i)^T) = \sum_{i=1}^N (\tilde{\mathbf{r}}_l^i (\tilde{\mathbf{r}}_l^i)^T) = \tilde{\mathbf{X}}_l \tilde{\mathbf{X}}_l^T. \quad (\text{A4})$$

Then, (A3) becomes

$$\begin{aligned} \mathbf{R}_{\Omega_l \cup \{\mathbf{b}_{l+1}\}} &= \frac{1}{N} \begin{bmatrix} \mathbf{X}_l \mathbf{X}_l^T & \mathbf{X}_l \mathbf{r}_{l+1} \\ \mathbf{r}_{l+1}^T \mathbf{X}_l^T & \mathbf{r}_{l+1}^T \mathbf{r}_{l+1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R}_{\Omega_l} & \tilde{\mathbf{X}}_l \tilde{\mathbf{r}}_{l+1} \\ \tilde{\mathbf{r}}_{l+1}^T \tilde{\mathbf{X}}_l^T & \tilde{\mathbf{r}}_{l+1}^T \tilde{\mathbf{r}}_{l+1} \end{bmatrix}. \end{aligned} \quad (\text{A5})$$

Using (A4) and (A5) and the following matrix identity:

$$\begin{bmatrix} \mathbf{U}^T \mathbf{U} & \mathbf{U}^T \mathbf{d} \\ \mathbf{d}^T \mathbf{U} & \mathbf{d}^T \mathbf{d} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{U}^T \mathbf{U})^{-1} + \beta \mathbf{U}^{\#} \mathbf{d} \mathbf{d}^T (\mathbf{U}^{\#})^T & -\beta \mathbf{U}^{\#} \mathbf{d} \\ -\beta \mathbf{d}^T (\mathbf{U}^{\#})^T & \beta \end{bmatrix} \quad (\text{A6})$$

with $\mathbf{U} = \tilde{\mathbf{X}}_l^T$ and $\mathbf{d} = \tilde{\mathbf{r}}_{l+1}$, we can further derive

$$\begin{aligned} \mathbf{R}_{\Omega_l \cup \{\mathbf{b}_{l+1}\}}^{-1} &= \begin{bmatrix} \mathbf{R}_{\Omega_l} & \tilde{\mathbf{X}}_l \tilde{\mathbf{r}}_{l+1} \\ \tilde{\mathbf{r}}_{l+1}^T \tilde{\mathbf{X}}_l^T & \tilde{\mathbf{r}}_{l+1}^T \tilde{\mathbf{r}}_{l+1} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{R}_{\Omega_l}^{-1} + \beta (\tilde{\mathbf{X}}_l^T)^{\#} \tilde{\mathbf{r}}_{l+1} \tilde{\mathbf{r}}_{l+1}^T \tilde{\mathbf{X}}_l^{\#} & (\tilde{\mathbf{X}}_l^T)^{\#} \tilde{\mathbf{r}}_{l+1} \\ \tilde{\mathbf{r}}_{l+1}^T \tilde{\mathbf{X}}_l^{\#} & \beta \end{bmatrix} \end{aligned} \quad (\text{A7})$$

where $(\tilde{\mathbf{X}}_l^T)^{\#} = (\tilde{\mathbf{X}}_l \tilde{\mathbf{X}}_l^T)^{-1} \tilde{\mathbf{X}}_l^T$ and $\beta = (\tilde{\mathbf{r}}_{l+1}^T (\mathbf{I} - \tilde{\mathbf{X}}_l^T (\tilde{\mathbf{X}}_l^T)^{\#}) \tilde{\mathbf{r}}_{l+1})^{-1} = (\tilde{\mathbf{r}}_{l+1}^T P_{\tilde{\mathbf{X}}_l^T}^{\perp} \tilde{\mathbf{r}}_{l+1})^{-1}$ due to the fact that $P_{\tilde{\mathbf{X}}_l^T}^{\perp} = \mathbf{I} - \tilde{\mathbf{X}}_l^T (\tilde{\mathbf{X}}_l^T)^{\#}$ is idempotent, i.e., $(P_{\tilde{\mathbf{X}}_l^T}^{\perp})^2 = P_{\tilde{\mathbf{X}}_l^T}^{\perp}$.

$$\begin{aligned} (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} &= (\mathbf{D}_l^T \mathbf{R}_{\Omega_l}^{-1} \mathbf{D}_l)^{-1} \\ &\quad - \frac{\beta (\mathbf{D}_l^T \mathbf{R}_{\Omega_l}^{-1} \mathbf{D}_l)^{-1} [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}] [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]^T N^{-1} (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1}}{1 + \beta [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]^T (\mathbf{D}_l^T \mathbf{R}_{\Omega_l}^{-1} \mathbf{D}_l)^{-1} [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]} \end{aligned} \quad (\text{A10})$$

$$\begin{aligned} \mathbf{c}^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} \mathbf{c} &= \mathbf{c}^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} \mathbf{c} \\ &\quad - \frac{\beta \mathbf{c}^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}] [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} \mathbf{c}}{1 + \beta [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]} \end{aligned} \quad (\text{A11})$$

$$\mathbf{c}^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} \mathbf{c} - \mathbf{c}^T (\mathbf{D}_l^T \mathbf{R}_{\Omega_l}^{-1} \mathbf{D}_l)^{-1} \mathbf{c} = -\frac{\beta \|\mathbf{c}^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]\|^2}{1 + \beta \|[\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1}\|^{1/2}} < 0 \quad (\text{A14})$$

Using (A7) and letting $\mathbf{D}_{\Omega_l \cup \{\mathbf{b}_{l+1}\}}^T \mathbf{R}_{\Omega_l \cup \{\mathbf{b}_{l+1}\}}^{-1} \mathbf{D}_{\Omega_l \cup \{\mathbf{b}_{l+1}\}} = \begin{bmatrix} \mathbf{D}_l \\ \mathbf{d}_{l+1}^T \end{bmatrix}$ and $\mathbf{d}_{l+1} = (d_{l+1}^1, d_{l+1}^2, \dots, d_{l+1}^p)^T$, we can derive

$$\begin{aligned} &\mathbf{D}_{\Omega_l \cup \{\mathbf{b}_{l+1}\}}^T \mathbf{R}_{\Omega_l \cup \{\mathbf{b}_{l+1}\}}^{-1} \mathbf{D}_{\Omega_l \cup \{\mathbf{b}_{l+1}\}} \\ &= [\mathbf{D}_l^T \quad \mathbf{d}_{l+1}] \\ &\quad \times \begin{bmatrix} \mathbf{R}_{\Omega_l}^{-1} + \beta (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} \tilde{\mathbf{X}}_l^T \tilde{\mathbf{X}}_l^\# & -\beta (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} \\ -\beta \tilde{\mathbf{r}}_{l+1}^T \tilde{\mathbf{X}}_l^\# & \beta \end{bmatrix} \begin{bmatrix} \mathbf{D}_l \\ \mathbf{d}_{l+1}^T \end{bmatrix} \\ &= \mathbf{D}_l^T (\mathbf{R}_{\Omega_l}^{-1} + \beta (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} \tilde{\mathbf{r}}_{l+1}^T \tilde{\mathbf{X}}_l^\#) \mathbf{D}_l - \beta \mathbf{d}_{l+1} \tilde{\mathbf{r}}_{l+1}^T \tilde{\mathbf{X}}_l^\# \mathbf{D}_l \\ &\quad - \beta \mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} \mathbf{d}_{l+1}^T + \mathbf{d}_{l+1} \mathbf{d}_{l+1}^T \beta \\ &= \mathbf{D}_l^T \mathbf{R}_{\Omega_l}^{-1} \mathbf{D}_l + \beta \mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} \tilde{\mathbf{r}}_{l+1}^T \mathbf{X}^\# \mathbf{D}_l \\ &\quad - 2\beta \mathbf{d}_{l+1} \tilde{\mathbf{r}}_{l+1}^T \tilde{\mathbf{X}}_l^\# \mathbf{D}_l + \mathbf{d}_{l+1} \mathbf{d}_{l+1}^T \beta \\ &= \mathbf{D}_l^T \mathbf{R}_{\Omega_l}^{-1} \mathbf{D}_l + \beta [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}] \\ &\quad \times [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]^T \end{aligned} \quad (\text{A8})$$

where $\beta = (\mathbf{d}_{l+1}^T P_{\tilde{\mathbf{X}}_l^T}^\perp \mathbf{d}_{l+1})^{-1}$.

Using Woodbury's identity

$$[\mathbf{A} + \mathbf{u}\mathbf{v}^T]^{-1} = \mathbf{A}^{-1} - \frac{[\mathbf{A}^{-1}\mathbf{u}][\mathbf{v}^T \mathbf{A}^{-1}]}{1 + \mathbf{v}^T \mathbf{A}^{-1} \mathbf{u}} \quad (\text{A9})$$

with $\mathbf{A} = \mathbf{D}_l^T \mathbf{R}_{\Omega_l}^{-1} \mathbf{D}_l$, $\mathbf{u} = \mathbf{v} = \sqrt{\beta} [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]$, (A10) and (A11), as shown at the top of this page.

Since $\beta > 0$,

$$\begin{aligned} &\mathbf{c}^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}] \\ &\quad \times [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]^T N (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} \mathbf{c} \\ &= \|\mathbf{c}^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]\|^2 > 0 \end{aligned} \quad (\text{A12})$$

and

$$\begin{aligned} &[\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1} \\ &\quad \times [\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}] \\ &= \|[\mathbf{D}_l^T (\tilde{\mathbf{X}}_l^T)^\# \tilde{\mathbf{r}}_{l+1} - \mathbf{d}_{l+1}]^T (\mathbf{D}_{l+1}^T \mathbf{R}_{\Omega_{l+1}}^{-1} \mathbf{D}_{l+1})^{-1/2}\|^2 > 0 \end{aligned} \quad (\text{A13})$$

substituting (A13) and (A12) into (A11) yields (A14), as shown at the top of this page, which shows that (8) is monotonically decreasing. Theorem 1 is proved.

REFERENCES

- [1] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [2] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [3] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
- [4] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.
- [5] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.
- [6] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.
- [7] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.
- [8] 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.
- [9] X. Guo, X. Huang, L. Zhang, L. Zhang, A. Plaza, and J. A. Benediktsson, "Support tensor machines for classification of hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3248–3264, Jun. 2016.
- [10] X. Ma, J. Geng, and H. Wang, "Hyperspectral image classification via contextual deep learning," *EURASIP J. Image Video Process.*, vol. 2015, p. 20, Dec. 2015.
- [11] X. Ma, H. Wang, and J. Wang, "Semisupervised classification for hyperspectral image based on multi-decision labeling and deep feature learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 120, pp. 99–107, Oct. 2016.
- [12] T. Lu, S. Li, L. Fang, X. Jia, and J. A. Benediktsson, "From subpixel to superpixel: A novel fusion framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4398–4411, Aug. 2017.
- [13] L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. Cao, "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5585–5599, Oct. 2017.

- [14] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140–7151, Dec. 2017.
- [15] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.
- [16] H. Li, Y. Song, and C. L. P. Chen, "Hyperspectral image classification based on multiscale spatial information fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5302–5312, Sep. 2017.
- [17] M. A. Hossain, X. Jia, and J. A. Benediktsson, "One-class oriented feature selection and classification of heterogeneous remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, pp. 1606–1612, Apr. 2016.
- [18] B. Du, Z. Wang, L. Zhang, L. Zhang, and D. Tao, "Robust and discriminative labeling for multi-label active learning based on maximum correntropy criterion," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1694–1707, 2017.
- [19] B. Du *et al.*, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1017–1027, Apr. 2017.
- [20] X. Lu, X. Zheng, and Y. Yuan, "Remote sensing scene classification by unsupervised representation learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5148–5157, 2017.
- [21] X. Lu, Y. Chen, and X. Li, "Hierarchical recurrent neural hashing for image retrieval with hierarchical convolutional features," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 106–120, 2018.
- [22] X. Lu, W. Zhang, and X. Li, "A hybrid sparsity and distance-based discrimination detector for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1704–1717, 2018.
- [23] X. Lu, B. Wang, X. Li, and X. Zheng, "Exploring models and data for remote sensing image caption generation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5148–5157, 2017.
- [24] W. Zhao, S. Du, and W. J. Emery, "Object-based convolutional neural network for high-resolution imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3386–3396, Jul. 2017.
- [25] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.
- [26] Y. Chen, L. Zhu, P. Ghamisi, X. Jia, G. Li, and L. Tang, "Hyperspectral images classification with Gabor filtering and convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2355–2359, Dec. 2017.
- [27] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.
- [28] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.
- [29] J. Li, X. Zhao, Y. Li, Q. Du, B. Xi, and J. Hu, "Classification of hyperspectral imagery using a new fully convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 292–296, Feb. 2018.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. JPROC-60, no. 8, pp. 926–935, Aug. 1972.
- [34] 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.
- [35] C.-I Chang, Q. Du, T.-L. 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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, 1990.
- [40] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection," *Photogramm. Eng. Remote Sens.*, vol. 10, no. 7, pp. 793–802, Jul. 2004.
- [41] 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.
- [42] S. De 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.
- [43] S. De Backer, P. Kempeneers, W. Debruyn, and P. Scheunders, "Band selection for hyperspectral remote sensing band selection for hyperspectral remote sensing," *Pattern Recognit.*, vol. 2, no. 3, pp. 319–323, 2005.
- [44] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.
- [45] H. Yang, Q. Du, H. 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.
- [46] 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.
- [47] 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.
- [48] W. Wei, Q. Du, and N. H. Younan, "Fast supervised hyperspectral band selection using graphics processing unit," *J. Appl. Remote Sens.*, vol. 6, no. 1, pp. 061504-1–061504-12, 2012.
- [49] K. Sun, X. Geng, L. Ji, and Y. Lu, "A new band selection method for hyperspectral image based on data quality," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2697–2703, Jun. 2014.
- [50] X. Geng, K. Sun, and L. Ji, "Band selection for target detection in hyperspectral imagery using sparse CEM," *Remote Sens. Lett.*, vol. 5, no. 3, pp. 1022–1031, 2014.
- [51] K. Sun, X. Geng, and L. Ji, "A new sparsity-based band selection method for target detection of hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 329–333, Feb. 2015.
- [52] K. Sun, X. Geng, and L. Ji, "Exemplar component analysis: A fast band selection method for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 998–1002, May 2015.
- [53] A. Zare and P. Gader, "Hyperspectral band selection and endmember detection using sparsity promoting priors," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 256–260, Apr. 2008.
- [54] 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.
- [55] X. Lu, X. Li, and L. Mou, "Semi-supervised multitask learning for scene recognition," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1967–1976, Sep. 2015.
- [56] 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.
- [57] Y. Yuan, J. Lin, and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1431–1445, Mar. 2016.
- [58] 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.

- [59] 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.
- [60] X. Lu, Y. Yuan, and X. Zheng, "Joint dictionary learning for multispectral change detection," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 884–897, Apr. 2017.
- [61] X. Lu, X. Zheng, and Y. Yuan, "Remote sensing scene classification by unsupervised representation learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5148–5157, Sep. 2017.
- [62] 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.
- [63] J. Feng, L. C. Jiao, X. Zhang, and T. Sun, "Hyperspectral band selection based on trivariate mutual information and clonal selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4092–4105, Jul. 2014.
- [64] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, "Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2956–2969, May 2015.
- [65] C. Wang, M. Gong, M. Zhang, and Y. Chan, "Unsupervised hyperspectral image band selection via column subset selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1411–1415, Jul. 2015.
- [66] 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.
- [67] 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.
- [68] J. E. Ball, L. M. Bruce, and N. H. Younan, "Hyperspectral pixel unmixing via spectral band selection and DC-insensitive singular value decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 382–386, Jul. 2007.
- [69] K. Koonsanit, C. Jaruskulchai, and A. Eiumnoh, "Band selection for dimension reduction in hyper spectral image using integrated information gain and principal components analysis technique," *Int. J. Mach. Learn. Comput.*, vol. 2, no. 3, pp. 248–251, Jun. 2012.
- [70] W. Xia, B. Wang, and L. Zhang, "Band selection for hyperspectral imagery: A new approach based on complex networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1229–1233, Sep. 2013.
- [71] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994.
- [72] Accessed: Apr. 27, 2018. [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes
- [73] 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.
- [74] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Norwell, MA, USA: Kluwer, 2003.
- [75] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. New York, NY, USA: Academic, 1999, p. 366.
- [76] N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
- [77] 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.
- [78] C.-I Chang, *Hyperspectral Data Processing: Algorithm Design and Analysis*. Hoboken, NJ, USA: Wiley, 2013.
- [79] 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.
- [80] J. C. Harsanyi, W. Farrand, and C.-I Chang, "Detection of subpixel spectral signatures in hyperspectral image sequences," in *Proc. Annu. Meeting, Amer. Soc. Photogramm. Remote Sens.*, Reno, NV, USA, 1994, pp. 236–247.
- [81] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
- [82] C.-I Chang, "Spectral inter-band discrimination capacity of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1749–1766, Mar. 2018.



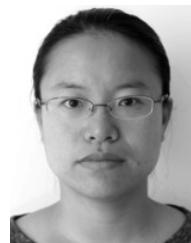
Chunyan Yu (M'17) received the B.Sc. and Ph.D. degrees in environment engineering from Dalian Maritime University, Dalian, China, in 2004 and 2012, respectively.

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 a Lecturer with the Information Science and Technology College, Dalian Maritime University. Her research interests include image segmentation, hyperspectral image classification, and pattern recognition.



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, and the joint Ph.D. degree from the Remote Sensing Signal and Image Processing Laboratory, University of Maryland at Baltimore, Baltimore, MD, USA, funded by the China Scholarship Council in 2011.

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 signs signal processing.



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 University of Maryland at Baltimore, Baltimore, MD, USA. She is currently an Associate Professor with the College of Information Science and Technology, Dalian Maritime University, Dalian, China. Her research includes remote sensing and hyperspectral image processing.



College Park, MD, USA.

He has been with the University of Maryland at Baltimore, 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. Also, he was the Distinguished Lecturer Chair with National Chung Hsing University, Taichung, sponsored by the Ministry of Education, Taiwan, from 2005 to 2006, and a Distinguished Chair Professor with National Chung Hsing University from 2014 to 2017. He was also a Distinguished Visiting Fellow/Fellow Professor, which were sponsored by the National Science Council, Taiwan, from 2009 to 2010. He has also been the Chair Professor of Providence University, Taichung, Taiwan, since 2012. He has also held the Chang Jiang Scholar Chair Professorship and the Director of the Center for Hyperspectral Imaging in Remote Sensing, Dalian Maritime University, Dalian, China, since 2016, the Hua Shan Scholar Chair Professorship with Xidian University, Xi'an, China, since 2016, and the Feng-Tai Chair Professor with the National Yunlin University of Science and Technology since 2017. He has authored four books: the *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Kluwer Academic Publishers, 2003), the *Hyperspectral Data Processing: Algorithm Design and Analysis* (John Wiley & Sons, 2013), the *Real Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection* (Springer, 2016), and the *Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation* (Springer, 2017). In addition, he also edited two books on the *Recent Advances*

in Hyperspectral Signal and Image Processing (2006) and the *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). He holds seven patents on hyperspectral image processing. His research interests include multispectral/hyperspectral image processing, automatic target recognition, and medical imaging.

Dr. Chang is a fellow of SPIE. He was a recipient of the 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. He was the Guest Editor of a special issue of the *Journal of High Speed Networks on Telemedicine and Applications* (2000) and Co-Guest Editor of another special issue of the same journal on Broadband Multimedia Sensor Networks in Healthcare Applications (2007). 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 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* (2007), Signal Processing and System Design in Health Care Applications for the *EURASIP Journal on Advances in Signal Processing* (2009), Multippectral, Hyperspectral, and Polarimetric Imaging Technology for the *Journal of Sensors and Hyperspectral Imaging and Applications for Remote Sensing*, (2017). He is currently an Associate Editor of *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*, *Journal of Robotics and Artificial Intelligence*, and *Open Remote Sensing Journal*. He was a plenary speaker of the Society for Photo-Optical Instrumentation Engineers (SPIE) Optics+Applications, Remote Sensing Symposium, in 2009. He was also a keynote speaker at the User Conference of Hyperspectral Imaging 2010, the 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 2003, Kimen, and 2013, Nan-Tou, Taiwan.