

© 2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

C. -I. Chang, Y. -M. Kuo and P. F. Hu, "Unsupervised Rate Distortion Function-Based Band Subset Selection for Hyperspectral Image Classification," in IEEE Transactions on Geoscience and Remote Sensing, doi: 10.1109/TGRS.2023.3296728.

<https://doi.org/10.1109/TGRS.2023.3296728>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Unsupervised Rate Distortion Function-Based Band Subset Selection for Hyperspectral Image Classification

Chein-I Chang^{1b}, *Life Fellow, IEEE*, Yi-Mei Kuo, and Peter Fuming Hu^{2b}

Abstract—Due to significant interband correlation resulting from the use of hundreds of contiguous spectral bands, band selection (BS) is one of the most widely used methods to reduce data dimensionality for band redundancy removal. A challenge for BS is how to design an effective criterion that can select bands with preserving crucial spectral information, while also avoiding selecting highly correlated bands. Information theory turns out to be one of the best means to address such issues in terms of information redundancy, specifically, the rate distortion function (RDF) of Shannon's third noisy source coding (or joint source and channel coding) theorem, which has been widely used in image compression/coding. This article presents a novel unsupervised RDF-based band subset selection (RDFBSS) for hyperspectral image classification (HSIC). To accomplish this goal, a new concept of the area under an RDF curve, A_{RDF} similar to the area under a receiver operating characteristic (ROC), A_z defined in hyperspectral target detection is coined and defined as a criterion for BSS. Since BSS generally requires an exhaustive search for an optimal band subset, two iterative algorithms similar to sequential (SQ) N finder (N-FINDR) and successive (SC) N-FINDR for finding endmembers, called sequential (SQ) RDFBSS and successive (SC) RDFBSS, can be also derived and coupled with A_{RDF} as a criterion to find optimal band subsets. The experimental results demonstrate that RDFBSS is indeed a very effective BS method to find the best possible band subsets and also performs better than most recent BS methods.

Index Terms—Band selection (BS), band subset selection (BSS), mutual information (MI), rate distortion function (RDF), sequential RDF BSS (SQ-RDFBSS), successive RDF BSS (SC-RDFBSS), virtual dimensionality (VD).

Manuscript received 16 February 2023; revised 22 June 2023; accepted 14 July 2023. Date of publication 19 July 2023; date of current version 2 August 2023. This work was supported in part by the National Science and Technology Council (NSTC) under Grant 111-2634-F-006-012. (Corresponding author: Chein-I Chang.)

Chein-I Chang is with the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information and Technology College, Dalian Maritime University, Dalian 116026, China, also with the Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250 USA, also with the Department of Electrical Engineering, National Cheng Kung University, Tainan 70101, Taiwan, and also with the Department of Computer Science and Information Management, Providence University, Taichung 02912, Taiwan (e-mail: cchang@umbc.edu).

Yi-Mei Kuo is with the Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250 USA (e-mail: ccgsh10923@gmail.com).

Peter Fuming Hu is with the Department of Anesthesia, School of Medicine, R. A. Cowley Shock Trauma Center, Shock Trauma Anesthesia Organized Research Center, University of Maryland Baltimore County, Baltimore, MD 21201 USA (e-mail: phu@som.umaryland.edu).

Digital Object Identifier 10.1109/TGRS.2023.3296728

NOMENCLATURE

A_{RDF}	Area under rate distortion function (RDF) curve.
A_z	Area under the receiver operating characteristic (ROC) curve.
BS	Band selection.
BSS	Band subset selection.
CC	Channel capacity.
EPF	Edge preserving filter.
HSIC	Hyperspectral image classification.
IEPF	Iterative EPF.
MI	Mutual information.
RDF	Rate distortion function.
ROC	Receiver operating characteristic.
SC-RDFBSS	Successive RDF BSS.
SCR	Spectral compression ratio.
SMI	Self-MI.
SQ-RDFBSS	Sequential RDF BSS.
VD	Virtual dimensionality.

I. INTRODUCTION

DUE to the use of hundreds of contiguous spectral bands, a hyperspectral imaging sensor can acquire a wealth of spectral information that can help resolve many issues arising in multispectral imaging [1] such as subpixel detection, mixed pixel classification, endmember finding, spectral unmixing, etc. However, such benefits are also traded for abundant spectral information which generally provides more information than what it needs, specifically, redundant information resulting from interband correlation. To deal with this problem, data compression is generally used as a preprocessing step to reduce spectral dimensionality. BS is one of the most effective means of accomplishing this goal [2]. For example, Fu et al. [3] proposed an approach coupling a neighborhood grouping normalized matched filter (NGNMF) with an enhanced 2-D singular spectrum analysis (E2DSSA) for BS so that the spatial context and structural information from each selected band, which can reduce the data dimension while preserving the corresponding spectral information.

In order to address the issue of information redundancy, Sun et al. [4] proposed a concrete autoencoder-based uniform BS (UBS) framework to search for an optimal band subset from the concrete encoder based on an information entropy

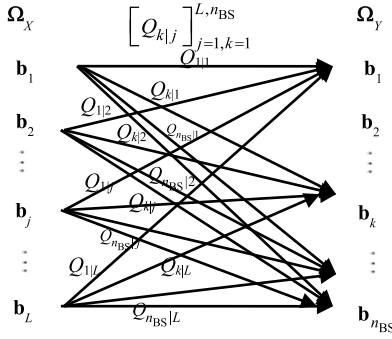


Fig. 1. Hyperspectral BS channel specified by $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ and $\Omega_Y = \{\mathbf{b}_k\}_{k=1}^{n_{BS}}$.

criterion. Also shown in [5], since entropy is an infinite-order statistics criterion, it is generally a better measure than second-order statistics-based criteria such as variance [5]. In particular, information theory is probably the most effective means to provide theoretical support [6], [7], [8]. It is considered as the fundamental theory of communication, which provides theoretical limits on information transmission. Surprisingly, information theory-based BS has not received much attention over the past years. In general, we can formulate hyperspectral BS as an information transmission system through a channel that has full bands, $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ as input space Ω_X and selected bands $\Omega_Y = \{\mathbf{b}_k\}_{k=1}^{n_{BS}}$ as output space Ω_Y , where L is the total number of spectral bands and n_{BS} is the number of selected bands. Specifically, a hyperspectral BS channel is generally characterized by three components, full band input space Ω_X , selected band output space Ω_Y , and a channel transition matrix $\mathbf{Q} = [Q_{y|x}]_{x \in \Omega_X, y \in \Omega_Y}$ as a triplet $(\Omega_X, \Omega_Y, \mathbf{Q})$ where $\mathbf{Q} = [Q_{k|l}]_{l=1, k=1}^{L, n_{BS}}$ is a matrix formed by a set of conditional probabilities, say $Q_{k|l} = p(\mathbf{b}_k | \mathbf{b}_l)$ that describes a transition from the l th band, \mathbf{b}_l in Ω_X to the k th selected band, and \mathbf{b}_k in Ω_Y as shown in Fig. 1.

By virtue of the hyperspectral band channel depicted in Fig. 1, its MI can be defined as

$$I(\Omega_X; \Omega_Y) = H(\Omega_X) - H(\Omega_X | \Omega_Y) \quad (1)$$

where $H(\Omega_X)$ is entropy of Ω_X and $H(\Omega_X | \Omega_Y)$ is the conditional entropy of Ω_X given Ω_Y . For a given channel transition matrix $Q_{k|l} = p(\mathbf{b}_k | \mathbf{b}_l)$, the CC, C is then defined as

$$C = \max_{\mathbf{p}} I(\Omega_X; \Omega_Y) \quad (2)$$

where the maximum is taken over all possible input probability vectors, \mathbf{p} is defined on Ω_X to find the most favorable probability vector \mathbf{p}^* to yield C . On the other hand, for a given input probability \mathbf{p} and a certain amount of distortion D , the RDF is defined as

$$R(D) = \min_{Q_{k|l}} I(\Omega_X; \Omega_Y) \text{ subject to } E[d] = \sum_{l=1}^L \sum_{k=1}^{n_{BS}} p_l Q_{k|l} d(\mathbf{b}_l, \mathbf{b}_k) \leq D \quad (3)$$

where $d(\mathbf{b}_l, \mathbf{b}_k)$ is a distortion measure between input band \mathbf{b}_l and output selected band \mathbf{b}_k and the minimum is taken over all possible channel transition matrices, $Q_{k|l} = [p(\mathbf{b}_k | \mathbf{b}_l)]_{l,k}^{L, n_{BS}}$.

Now, let R be the information transmission rate defined as

$$R = \frac{n_{BS}}{L}. \quad (4)$$

Then the well-known three Shannon's coding theorems can be stated by the following two inequalities:

$$R(D) \leq R \leq C. \quad (5)$$

More specifically, the right-hand side inequality of (5)

$$R \leq C \quad (6)$$

is Shannon's channel coding theorem specified as the second coding theorem, while the left-hand side inequality of (5)

$$R(D) \leq R \quad (7)$$

is Shannon's third theorem, referred to as Shannon's RDF theorem despite that it is also known under various names, RDF theorem, joint source and channel coding theorem, source coding subject to a fidelity criterion theorem, or simply noisy source coding theorem. In particular, as a special situation of (7) with zero-distortion $D = 0$ and distortion matrix specified by the Hamming matrix (i.e., probability error matrix), $R(D)$ is then reduced to $R(0) = H(\Omega_X)$. In this case, (7) becomes

$$R(0) = H(\Omega_X) \leq R \quad (8)$$

which is precisely the Shannon noiseless source coding theorem as the first coding theorem. On the other hand, the essence of Shannon's third coding theorem has a very important and crucial implication for data transmission. It indeed offers a feasible solution to what we should do when the required information transmission rate R is greater than C , that is, the capacity of a given channel, $\mathbf{Q} = [Q_{k|l}]_{l=1, k=1}^{L, n_{BS}}$. Under this circumstance, the best solution is to reduce R at the expense of distortion D by making R satisfy (6) to yield

$$R(D) = C. \quad (9)$$

In other words, the best scenario is to make the two ways of inequalities equal

$$R(D) = R = C \quad (10)$$

which culminates in one of the most celebrated results in data communication. This is exactly the same idea of why a hyperspectral image generally requires data spectral dimensionality reduction to satisfy constraints on the identification of material substances by a certain amount of distortion which is incurred by data compression or BS with SCR to be achieved by taking reciprocal of R in (1) to yield $SCR = R^{-1} = (L/n_{BS})$. This issue can be interpreted as the pigeon-hole principle in [2]. For those who are interested in this principle, we refer to the details in [9].

Despite the fact that there are reviews and surveys on BS such as [10] and [11], many most recently developed approaches were not included and reviewed, particularly, information theory-based BS and BSS methods. As noted

above, information theory has its core in three of Shannon's coding theorems. For example, band prioritization (BP) methods developed by entropy-based criteria, spectral information divergence (SID) [12], [13], Kullback-Leibler (KL) divergence [14], sample entropy and class entropy in [15], and self-information-based class information used to select bands for HSIC [16] can be traced back to Shannon's first coding theorem, while CC-based BSS [17], interband spectral capacity [18] can be considered as applications of Shannon's second coding theorem. Most recently, a new information theoretic approach to BS, to be called SMI-based BS also took advantage of Shannon's second coding theorem, combined with finding prominent peaks of SMI-generated curves to select optimal bands [19]. It first defined a new concept of SMI as the MI between one single band, \mathbf{b} and the full band set, Ω_Y as the band channel output space Y , denoted by $I(\mathbf{b}; \Omega_Y)$, and then used SMI as a BP criterion to rank bands for BS. The SMI-based BS is quite different from the CC-based BSS developed in [17], which uses CC as a BS criterion to select band subsets. Specifically, SMI-BS considers the MI between a particular band of interest and the full band set Ω rather than the MI between the channel input space and channel output space. The role of SMI plays in BS is similar to self-information that defines entropy [20]. Interestingly, to the authors' best knowledge, by far there exists no work reported in the literature on applying Shannon's third coding theorem to BS, even though RDF introduced by Shannon's third RDF theorem has been widely used to measure the performance of image compression/coding [21], [22].

This article takes an approach from an RDF perspective of Shannon's third coding theorem which is completely opposite to those in [17], [18], and [19] using Shannon's second channel coding theorem from a CC aspect. In other words, instead of maximizing MI in (2) over input probability vectors \mathbf{p} to find C as was done in [17], [18], and [19], the proposed RDF-based BSS is to minimize (3) to find a least favorable channel transition matrix $Q_{k|l}^* = [p(\mathbf{b}_k|\mathbf{b}_l)]_{l,k}^{L,n_{BS}}$ over all possible $Q_{k|l}$ for a given probability vector of a full band set Ω_X , $\{p(\mathbf{b}_l)\}_{l=1}^L$ and a prescribed distortion $D \geq 0$. More specifically, (2) finds a most favorable distribution \mathbf{p}^* over full band set Ω_X for a given channel transition matrix $Q_{k|l} = [p(\mathbf{b}_k|\mathbf{b}_l)]_{l,k}^{L,n_{BS}}$ so as to achieve the least upper bound of (5), C , while (3) finds a least favorable channel transition matrix $Q_{k|l}^* = [p(\mathbf{b}_k|\mathbf{b}_l)]_{l,k}^{L,n_{BS}}$ for a given probability distribution \mathbf{p} over full band set Ω_X and distortion D to achieve the greatest lower bound of (5), $R(D)$. It seems that one is against the other in terms of solving optimization problems. The truth of the matter is that (2) and (3) are indeed compatible because finding optimal solutions for (2) and (3) has different interpretations. One is the most favorable distribution \mathbf{p}^* by (2) compared to $Q_{k|l}^* = [p(\mathbf{b}_k|\mathbf{b}_l)]_{l,k}^{L,n_{BS}}$ by (3) which is the least favorable channel transition matrix $Q_{k|l}^* = [p(\mathbf{b}_k|\mathbf{b}_l)]_{l,k}^{L,n_{BS}}$. The key issue is how to make it work since CC-based BSS approaches are not applicable to RDFBSS. To avoid being trapped in CC-based approaches, we need to reinvent the wheel by looking into a different problem formulation.

First of all, we reinterpret a distortion matrix between inputs and outputs as a band discrimination (BD) matrix. This is quite

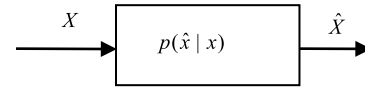


Fig. 2. Estimation model used to describe Shannon's third RDF theorem.

different from the CC-based BSS which characterizes channel transition matrix by a BD measure. In other words, we use BD as a distortion metric to measure band redundancy among bands. Specifically, in CC, the channel is given and CC is to find optimal channel input probabilities. Therefore, in CC-based methods, spectral angle mapper (SAM), SID, and joint SID and SAM (SIDAM) are used to characterize the channel. Conversely, in calculating RDF, the source input probabilities are given and the used distortion measure must be specified, SID [12], SAM [1], and SIDAM [13] in this article to calculate distortion D . In this case, RDF is to find the least favorable channel (i.e., worst channel) for given source probabilities and distortion D . Therefore, RDF is completely different from CC. As a result, their definitions are different even though they use SID, SAM, and SIDAM for different purposes. Second, in order to evaluate the effectiveness of RDFBSS performance, we specifically introduce a new concept, called the area under an RDF curve, $A_{R(D)}$, as a criterion that is similar to the area under a ROC curve widely used in statistical detection theory [23]. Third, the input probability of each band, \mathbf{b}_l to be used for calculating RDF is obtained by its band entropy, $H(\mathbf{b}_l)$. Fourth, we select a band subset as the channel output space Ω_Y and then implement a numerical algorithm developed by Blahut [24] to find RDF for the channel specified by the full bands as input channel space Ω_X and the selected band subset as Ω_Y . Fifth, to exhaust possible band subsets, we appeal for sequential and successive processes developed for N finder algorithm (N-FINDER) in [2], [25], [26], and [27] to find an optimal RDFBSS. Last but not least, RDFBSS is completely unsupervised without using any training dataset.

Several main contributions presented in this article are summarized as follows:

1. It formulates RDF as a BSS problem which is believed to be the first work ever reported in the literature.
2. It constructs a distortion matrix as a BD matrix by replacing the distortion measure with the BD measure.
3. It introduces a new concept of the area under an RDF curve, $A_{R(D)}$ as a RDFBSS criterion, which has never been explored before.
4. Finally, it combines Blahut's RDF algorithm with the idea of SQ/SC N-FINDER algorithms to develop unsupervised SQ/SC RDFBSS algorithms.

The remainder of this article is organized as follows. Section II describes how to construct a distortion matrix using BD measures. Section III reviews the basic properties of RDF and derives some important results. Section IV defines a new concept of the area under an RDF curve criterion that will be used to design RDF-based BSS algorithms. Section V develops two versions of RDF-based BSS algorithms, SQ-RDFBSS, and SC-RDFBSS. Section VI describes the experiment setup. Section VII presents experimental results and discussions. Section VIII draws some conclusions.

II. CONSTRUCTION OF A DISTORTION MATRIX

Shannon's third RDF theorem has shown its potential in many applications such as detection, classification, coding, etc. It can be generally described by an estimation/reconstruction model

$$\hat{x} = x + \varepsilon \quad (11)$$

where x in Ω and $\hat{x} \in \hat{\Omega}$ are considered as target signal and estimated/reconstructed signal, respectively, and ε can be interpreted as error result from an estimator such as maximum likelihood estimator (MLE)

$$\max_x p(\hat{x}|x) = \hat{x}_{\text{MLE}}(x) \quad (12)$$

or reconstruction error resulting from a decoder by $p(\hat{x}|x)$, which is the conditional probability of \hat{x} given x as shown in Fig. 2.

Specifically, when ε is interpreted as noise n , then (11) becomes a signal detection model. To evaluate the effectiveness of using model (11) is measured by a criterion, to be called distortion measure denoted by $d(x, \hat{x})$. For example, when (11) is applied for estimation

$$d(x, \hat{x}) = (x - \hat{x})^2 \quad (13)$$

is generally used as the mean squared error distortion measure. On the other hand, if (11) is applied for image coding

$$d(x, \hat{x}) = \begin{cases} 1, & \text{if } x \neq \hat{x} \\ 0, & \text{if } x = \hat{x} \end{cases} \quad (14)$$

is the Hamming distortion measure or probability error matrix.

Once a distortion measure is determined, then we can define the expected distortion by

$$E[d] = \sum_{j=1}^J \sum_{k=1}^K p_j Q_{k|j} d(x_j, \hat{x}_k) \quad (15)$$

where $Q_{k|j} = Q(\hat{x}_k|x_j)$ and $p_j Q_{k|j} = p(x_j, \hat{x}_k)$ is a joint probability. $E[d] = \sum_{j=1}^J \sum_{k=1}^K p(x_j, \hat{x}_k) d(x_j, \hat{x}_k)$.

However, when we consider RDF for BS, the distortion measures specified by (13) and (14) are no longer applicable. To address this issue, the distortion measure $d(x, \hat{x})$ in (13), which is the mean squared error should be interpreted by a BD measure, $d(\mathbf{b}_j, \mathbf{b}_k)$ which can better capture and calculate spectral variations of the interband correlation between any two bands, \mathbf{b}_j and \mathbf{b}_k . There are three BD measures, SID, SAM, and SIDAM can be used as a distortion measure to construct a distortion matrix, each of which is described in detail as follows.

A. SID

SID was originally developed to measure the discrepancy between two spectral bands based on their probability distributions. To calculate SID, we first define the probability vector $\mathbf{p}_j = (p_{j1}, p_{j2}, \dots, p_{jN})^T$ for band \mathbf{b}_j and $\mathbf{p}_k = (p_{k1}, p_{k2}, \dots, p_{kN})^T$ for \mathbf{b}_k as follows:

$$p_{jl} = b_{jl} / \sum_{n=1}^N b_{jn} \text{ and } p_{kl} = b_{kl} / \sum_{n=1}^N b_{kn}. \quad (16)$$

Then, SID between two bands $\mathbf{b}_j, \mathbf{b}_k$ is defined by

$$\text{SID}(\mathbf{b}_j, \mathbf{b}_k) = D(\mathbf{p}_j||\mathbf{p}_k) + D(\mathbf{p}_k||\mathbf{p}_j) \quad (17)$$

where

$$D(\mathbf{p}_j||\mathbf{p}_k) = \sum_{n=1}^N p_{jn} \log \frac{p_{jn}}{p_{kn}}. \quad (18)$$

B. SAM

SAM is another popular spectral measure defined by

$$\text{SAM}(\mathbf{b}_j, \mathbf{b}_k) = \cos^{-1} \left(\frac{\mathbf{b}_j^T \mathbf{b}_k}{\|\mathbf{b}_j\| \|\mathbf{b}_k\|} \right). \quad (19)$$

C. SIDAM

A third measure is SIDAM developed by Du et al. [13], which combines SID and SAM as follows:

$$\text{SIDAM}(\mathbf{b}_j, \mathbf{b}_k) = \text{SID}(\mathbf{b}_j, \mathbf{b}_k) \tan(\text{SAM}(\mathbf{b}_j, \mathbf{b}_k)). \quad (20)$$

III. RATE DISTORTION FUNCTION

This section presents the RDF theory derived by Blahut [24] and also proves some important and useful results that will be needed to develop RDF-based BSS algorithms in Section V.

A. Blahut's RDF Theory

Now, we can express MI in (1) in terms of the hyperspectral BS channel in Fig. 1 as

$$I(\Omega_X; \Omega_Y) = \sum_{l=1}^L p_l \sum_{k=1}^{n_{\text{BS}}} \log \frac{Q_{k|l}}{\sum_{l=1}^L p_l Q_{k|l}} \quad (21)$$

where the probability vector $\mathbf{p} = (p_1, p_2, \dots, p_L)^T$ is probability distribution defined on the full band input space $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ and p_l is the probability of the l th band input \mathbf{b}_l .

In order to make our notations more general and consistent with the notations used in [24], we replace L and n_{BS} in (21) with J and K in the following discussions.

It is very difficult to solve for (3) directly, if not impossible. In [24], Blahut reformulated the optimization problem of (3) as a double minimum optimization by introducing a new objective function specified by

$$J(\mathbf{p}, \mathbf{Q}, \mathbf{q}) = \sum_{j=1}^J \sum_{k=1}^K p_j Q_{k|j} \log \frac{Q_{k|j}}{q_k} - s \left(\sum_{j=1}^J \sum_{k=1}^K p_j Q_{k|j} \rho_{jk} - D \right) \quad (22)$$

where s is a Lagrange multiplier. Then Blahut proved the following theorems:

Theorem 1:
1)

$$R(D) = sD + \min_{\mathbf{q}} \min_{\mathbf{Q}} \{J(\mathbf{p}, \mathbf{Q}, \mathbf{q})\} \quad (23)$$

and

$$D = \sum_{k=1}^K p_j Q_{k|j}^{*} sd(x_j, x_k) \quad (24)$$

where \mathbf{Q}^* is the point that achieves the minimum of (23).

2) For a fixed \mathbf{Q} , $J(\mathbf{p}, \mathbf{Q}, \mathbf{q})$ is minimized by

$$q_k = \sum_{j=1}^J p_j Q_{k|j}. \quad (25)$$

3) For a fixed \mathbf{q} , $J(\mathbf{p}, \mathbf{Q}, \mathbf{q})$ is minimized by

$$Q_{k|j} = \frac{q_k \exp(s\rho_{jk})}{\sum_{k=1}^K q_k \exp(s\rho_{jk})} \quad (26)$$

where $\exp(s\rho_{jk}) = e^{sd(x_j, \hat{x}_k)}$

Theorem 2:

A necessary and sufficient condition on an output probability vector \mathbf{q} via a channel transition matrix given by (26) to yield a point on the RDF curve is

$$c_k = \sum_{j=1}^J p_j \sum_{k=1}^K q_k \exp(s\rho_{jk}) = 1 \text{ for } q_k \neq 0 \quad (27)$$

$$c_k = \sum_{j=1}^J p_j \sum_{k=1}^K q_k \exp(s\rho_{jk}) \leq 1 \text{ for } q_k = 0. \quad (28)$$

For those who are interested in Theorems 1 and 2, more details can be found in [8] and [24].

B. Domain Range of RDF

In order to generate an RDF curve, we first need to identify the range of distortion D . Let

$$D_{\min} = \sum_x p(x) \min_{\hat{x}} d(x, \hat{x}) \quad (29)$$

$$D_{\max} = \min_{\hat{x}} \sum_x p(x) d(x, \hat{x}). \quad (30)$$

Then we can prove the following theorem [6]:

Theorem 3:

1) $R(D)$ is achievable only if $D \geq D_{\min}$.

2) $R(D) = 0$ if and only if $D \geq D_{\max}$.

Proof: See Appendix.

C. Blahut's RDF Algorithm

As noted above, directly finding $R(D)$ is not feasible. In this case, we must rely on a numerical algorithm to accomplish this task. Theorem 1 shows how an iterative algorithm can be designed to find RDF, $R(D)$, while Theorem 2 provides a stopping rule for the designed algorithm. A detailed step-by-step implementation of Blahut's RDF algorithm is described in the following Algorithm 1 with its flowchart which can be found in [8] and [24].

It is worth noting that although we are given the input probability \mathbf{p} , the input to the algorithm is actually the output probability \mathbf{q} , not \mathbf{p} as shown in (25) and also in Theorem 2. In addition, it should be also noted that the Lagrange multiplier s is an independent parameter in (23), while the D in (23) becomes a parameter depending upon s . In other words,

Algorithm 1 Calculating RDF

1. Initial conditions:

(i) Assume that $\{p_j\}_{j=1}^J$ be the channel input probabilities where p_j is the probability of the j^{th} input x_j and $\mathbf{D} = [\rho_{jk}]_{j,k}^{J,K}$ be a distortion matrix with $\rho_{jk} = d(x_j, \hat{x}_k)$.

(ii) Also let D be a distortion to be tolerated.

(iii) Let $\mathbf{q}^{(0)} = (q_1^{(0)}, q_2^{(0)}, \dots, q_K^{(0)})^T$ be an initial output probability vector.

(iv) Let $s < 0$ be specified by a prescribed value and ε be a prescribed error tolerance.

(v) Let $A_{jk} = e^{s\rho_{jk}}$ and $l = 0$.

2. Let $l \leftrightarrow l + 1$ and calculate

$$q_k^{(l)} = q_k^{(l-1)} \sum_{j=1}^J p_j \frac{A_{jk}}{\sum_{k=1}^K q_k^{(l-1)} A_{jk}} \quad (31)$$

$$Q_{k|j}^{(l)} = \frac{q_k^{(l-1)} e^{s\rho_{jk}}}{\sum_{k=1}^K q_k^{(l-1)} e^{s\rho_{jk}}} \quad (32)$$

3. Find

$$c_j^{(l)} = \sum_{k=1}^K p_j^{(l)} \frac{A_{jk}}{\sum_{k=1}^K q_k^{(l)}} \quad (33)$$

4. Calculate

$$T_U^{(l)} = \max_k \log c_k^{(l)} \quad (34)$$

$$T_L^{(l)} = \sum_{k=1}^K q_k^{(l)} c_k^{(l)} \quad (35)$$

5. If $T_U^{(l)} - T_L^{(l)} < \varepsilon$ Then stop. Otherwise, go to step 2.

to calculate (23), we specify $s < 0$ rather than D where both D and $R(D)$ are generated for the point on the $R(D)$ curve that has slope s . This is very similar to the ROC curve which is generated by an independent threshold τ instead of the false alarm probability, β shown in [28].

IV. RDF-BASED CRITERION

As Shannon's RDF theorem is applied to BS, the channel input space is fixed at the full band set, $\mathbf{\Omega}_X = \{\mathbf{b}_l\}_{l=1}^L$ and the channel output space is designated by a selected band subset, $\mathbf{\Omega}_Y = \{\mathbf{b}_k\}_{k=1}^{n_{\text{BS}}}$. As a result, RDF is solely determined by 1) the selected band subset $\mathbf{\Omega}_Y = \{\mathbf{b}_k\}_{k=1}^{n_{\text{BS}}}$ and 2) the distortion matrix discussed in Section II because a distortion matrix is constructed between $\mathbf{\Omega}_X = \{\mathbf{b}_l\}_{l=1}^L$ and $\mathbf{\Omega}_Y = \{\mathbf{b}_k\}_{k=1}^{n_{\text{BS}}}$ by a BD measure such as SAM($\mathbf{b}_j, \mathbf{b}_k$) in (19), SID($\mathbf{b}_j, \mathbf{b}_k$) in (17) and SIDAM($\mathbf{b}_j, \mathbf{b}_k$) in (20) between $\mathbf{\Omega}_X = \{\mathbf{b}_l\}_{l=1}^L$ and $\mathbf{\Omega}_Y = \{\mathbf{b}_k\}_{k=1}^{n_{\text{BS}}}$. This implies that different selected band subsets produce different distortion matrices which also result in different domain ranges for distortion D . Consequently, it may be often the case that two different RDF curves may actually yield the same performance. This dilemma is also encountered in detection theory when the effectiveness of a detector is evaluated by its ROC curve. It may also lead to the fact that two detectors that generate different ROC curves actually have the same effectiveness of detection [29, Fig. 3]. To resolve this issue, the area under an ROC

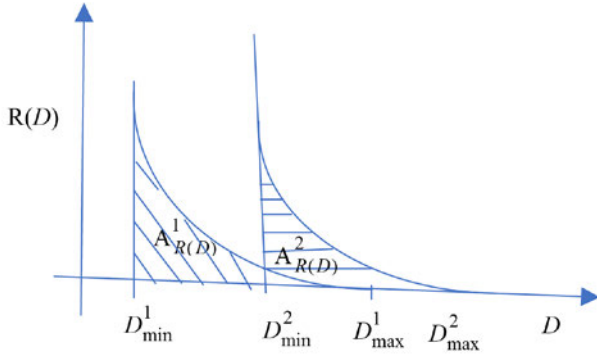


Fig. 3. Two $R(D)$ curves along with their own $A^1_{R(D)}$ and $A^2_{R(D)}$.

curve, A_z , is calculated and also normalized to unity so that all detectors producing different ROC curves can be fairly compared on the same basis [29]. Inspired by the use of A_z in ROC analysis, this section introduces a new and similar concept, called the area under an RDF curve, $A_{R(D)}$ which can be used to evaluate the effectiveness of BS performance produced by different selected band subsets. As an illustrative example, Fig. 3 shows two $R(D)$ curves produced by two different selected band subsets $\Omega_Y^1 = \{\mathbf{b}_k\}_{k=1}^{n_{BS}}$ and $\Omega_Y^2 = \{\mathbf{b}_k\}_{k=1}^{n_{BS}}$ with two different domain ranges $[D^1_{min}, D^1_{max}]$ and $[D^2_{min}, D^2_{max}]$ with their own $A^1_{R(D)}$ and $A^2_{R(D)}$ with different shades but with the same area value, i.e., $A^1_{R(D)} = A^2_{R(D)}$.

Following a similar approach proposed in [29], the distortion D is first normalized to \tilde{D} in the unit interval $[1, 0]$ according to Theorem 3 by

$$\tilde{D} = \frac{D}{D_{max} - D_{min}}. \quad (36)$$

Then $R(\tilde{D})$ is further normalized to $\tilde{R}(\tilde{D})$ by

$$\tilde{R}(\tilde{D}) = \frac{R(\tilde{D}) - R(\tilde{D}_{max})}{R(\tilde{D}_{min}) - R(\tilde{D}_{max})} = \frac{R(\tilde{D})}{R(\tilde{D}_{min})} \quad (37)$$

since $R(\tilde{D}_{max}) = 0$ according to Theorem 3.

By virtue of (37), $\tilde{R}(\tilde{D})$ can be used as a criterion to measure the effectiveness of various BSS methods in performance.

V. RDFBSS ALGORITHMS

In analogy with the CC computation problem formulated for BSS in [17] and [18], a penalty function specified by a Lagrange multiplier s can be also defined by

$$F(\mathbf{p}, \{Q_{k|l}\}, \mathbf{q}) = \sum_{l=1}^L p_l \sum_{k=1}^{n_{BS}} Q_{k|l} \log \frac{Q_{l|k}}{q_k} - s \sum_{l=1, k=1}^{L, n_{BS}} p_j Q_{k|l} d(\mathbf{b}_l, \mathbf{b}_k). \quad (38)$$

By virtual of (23) and (38), we can reformulate (3) as a double minima problem for each selected band subset $\Omega_Y = \{\mathbf{b}_k\}_{k=1}^{n_{BS}}$ given by

$$R(D) = sD + \min_{\mathbf{q}} \min_{\{Q_{k|l}\}} F(\mathbf{p}, \{Q_{k|l}\}, \mathbf{q}) \quad (39)$$

with s used as the Lagrange multiplier to constrain $\sum_{l=1, k=1}^{L, n_{BS}} p_l Q_{k|l} d(\mathbf{b}_l, \mathbf{b}_k) = D$ where the probability output

vector $\mathbf{q} = (q_1, q_2, \dots, q_{n_{BS}})^T$ is used as a dummy variable ξ considered in the dummy variable trick [30] and will be determined and vanished by the optimal solution $q_k^* = \sum_{l=1}^L p_l Q_{k|l}$.

Apparently, solving (39) for an optimal band subset $\Omega_Y^* = \{\mathbf{b}_k^*\}_{k=1}^{n_{BS}}$ directly by an exhausting search is generally not possible. To mitigate this dilemma, we take advantage of these two algorithms, SQ N-FINDR and SC N-FINDR developed to avoid an exhaust search required by N-FINDR [25], [26], [27] to further develop SQ-RDFBSS and SC-RDFBSS to do exactly what SQ N-FINDR and SC N-FINDR do for N-FINDR.

To design an iterative algorithm for RDFBSS, three components are required, initial conditions, learning process, and replacement rule as summarized as follows:

1) Initial conditions:

a) The number of bands to be selected, n_{BS} , which can be estimated by VD [31], [32], [33], denoted by n_{VD} .

b) The probability of each band in $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$, say p_l , can be calculated by its entropy

$$p_l = \frac{H(\mathbf{b}_l)}{\sum_{j=1}^L H(\mathbf{b}_j)} \quad (40)$$

where $H(\mathbf{p}_l) = -\sum_{j=1}^L p_{lj} \log p_{lj}$ is the entropy of the l th band \mathbf{b}_l .

2) Learning process:

SQ/SC iterative process can be derived from SQ N-FINDR and SC N-FINDR and also SQ-CCBSS and SC-CCBSS [18].

3) Replacement rule:

$\tilde{R}(\tilde{D})$ is a measure to determine if the current band subset needs to be replaced. There are two criteria that can be designed for this purpose. One is to select a band subset with $\min \tilde{R}(\tilde{D})$. The other selects a band subset with $\max \tilde{R}(\tilde{D})$. Both versions will be used for experiments.

It should be noted that there are no training samples required for initial conditions.

A. SQ-RDFBSS-MIN

The SQ-RDFBSS-MIN presented in this section implements two loops, the outer loop indexed by l and the inner loop indexed by p . It starts with an initial band subset with $n_{BS} = n_{VD}$ bands uniformly selected from the full band set $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$, $\Omega_Y^{(0)} = \{\mathbf{b}_k^{(0)}\}_{k=1}^{n_{BS}}$. It then begins with its outer loop by selecting the first band, \mathbf{b}_1 from $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ with setting index $l = 1$ to run through bands in $\Omega_Y^{(0)} = \{\mathbf{b}_k^{(0)}\}_{k=1}^{n_{BS}}$ one at a time via its inner loop index by p to check if the visited bands need to be replaced according to $\min \tilde{R}(\tilde{D})$. The resulting band subset is denoted by $\Omega_Y^{(1)} = \{\mathbf{b}_k^{(1)}\}_{k=1}^{n_{BS}}$. It then goes back to the outer loop by setting $l = 2$ to select the next band, \mathbf{b}_2 from $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$. The \mathbf{b}_2 will then once again go into the inner loop to run through $\Omega_Y^{(1)} = \{\mathbf{b}_k^{(1)}\}_{k=1}^{n_{BS}}$ one at a time via index p for band replacement according to $\min \tilde{R}(\tilde{D})$. The entire procedure is repeated over and over again by processing the outer loop and then the inner loop iteratively via interchanging two indices, l and p , denoted by (l, p) since it begins with the outer loop at index l followed by the inner loop by p . The details of carrying out SQ-RDFBSS-MIN in step-by-step implementation are summarized in the following Algorithm 2.

Algorithm 2 SQ-RDFBSS-MIN

Step 1: Initial condition:

- (i) If $T_U^{(l)} - T_L^{(l)} < \varepsilon$. Then stop. Otherwise, go to step 2.
- (ii) Let n_{BS} be the number of bands needed to be selected determined by VD.
- (iii) Let the channel input space $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ and the channel output space $\Omega_{BS} = \Omega_X^{(0)} = \{\hat{\mathbf{b}}_k^{(0)}\}_{k=1}^{n_{VD}}$ uniformly selected from the full band set. Then calculate the distortion matrix $\mathbf{D}^{(0)} = [\rho_{jk}^{(0)}]_{j=1,k=1}^{L,n_{VD}}$ by SIDAM, SID or SAM.
- (iv) Calculate the input distribution $\{p_l\}_{l=1}^L$ over $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ with p_l given by (40).
- (v) Implement the Blahut's rate-distortion algorithm to find and normalize RDF to obtain $\tilde{R}(\tilde{D})$ according to (36)-(37).
- (vi) Calculate the area under $\tilde{R}^{(0)}(\tilde{D})$ curve denoted by $A_{\tilde{R}^{(0)}(\tilde{D})}$.
- (vii) Let $A_{RDF} = \tilde{R}^{(0)}(\tilde{D})$.

Step 2: a. Outer loop: for $l = 1, \dots, L$ If $\mathbf{b}_l \in \Omega_{BS}$, go to step 3. Otherwise, let

$$\Omega_{\hat{X}} = \Omega_{BS}.$$

b. Inner loop: for $p = 1, \dots, n_{VD}$, do

- (i) Let $\Omega_{\hat{X}}^{(l,p)} = \{\hat{\mathbf{b}}_k^{(l,p)}\}_{k=1}^{n_{VD}} = \Omega_{\hat{X}}$ and then replace the p th band of $\Omega_{\hat{X}}^{(l,p)}$ with \mathbf{b}_l .
- (ii) Construct a new distortion matrix $\mathbf{D}^{(l,p)} = [\rho_{jk}^{(l,p)}]_{j=1,k=1}^{L,n_{BS}}$ by SIDAM, SID or SAM with input space Ω_X and output space $\Omega_{\hat{X}}^{(l,p)}$.
- (iii) Implement the Blahut's rate-distortion algorithm to find and normalize RDF to obtain $\tilde{R}^{(l,p)}(\tilde{D})$.
- (iv) Calculate the $A_{\tilde{R}^{(l,p)}(\tilde{D})}$ of $\tilde{R}^{(l,p)}(\tilde{D})$.
- (v) If $A_{\tilde{R}^{(l,p)}(\tilde{D})} < A_{RDF}$, let $A_{RDF} = A_{\tilde{R}^{(l,p)}(\tilde{D})}$, $\Omega_{BS} = \Omega_{\hat{X}}^{(l,p)}$. Otherwise, continue.
- (vi) If $p < n_{VD}$, let $p \leftarrow p+1$ and go to step 2b. Otherwise, go to step 3.

Step 3: If $l < L$, let $l \leftarrow l+1$ and go to step 2a. Otherwise, continue.Step 4: Output final band subset, $\Omega_{\hat{X}}^{(l,p)}$.

Another version of SQ-RDFBSS is called SQ-RDFBSS-MAX, which uses $\max \tilde{R}(\tilde{D})$ instead of $\min \tilde{R}(\tilde{D})$ as a criterion to produce an optimal band subset. All steps of its algorithm implementation are identical to SQ-RDFBSS-MIN except step 2b(v), where $A_{\tilde{R}^{(l,p)}(\tilde{D})} < A_{RDF}$ is replaced by $A_{\tilde{R}^{(l,p)}(\tilde{D})} > A_{RDF}$.

B. SC-RDFBSS-MIN

The SC-RDFBSS-MIN presented in this section swaps the two loops implemented in SQ-RDFBSS-MIN by making the inner loop indexed by p an outer loop and an outer loop indexed by l an inner loop. Like SQ-RDFBSS-MIN, it also starts with an initial band subset with $p = n_{BS} = n_{VD}$ bands uniformly selected from the full band set $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$, $\Omega_Y = \{\mathbf{b}_k^{(0)}\}_{k=1}^{n_{BS}}$. It first begins with its outer loop by selecting the first band, $\mathbf{b}_1^{(0)}$ from $\Omega_Y = \{\mathbf{b}_k^{(0)}\}_{k=1}^{n_{BS}}$ with setting index $p = 1$, and then goes into its inner loop index by l to run through all bands in $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ one at a time to check if

the visited bands need to be replaced according to $\min \tilde{R}(\tilde{D})$. The resulting band subset is denoted by $\Omega_Y^{(p)} = \{\mathbf{b}_k^{(p)}\}_{k=1}^{n_{BS}}$. It then goes back to the outer loop by setting $p = 2$ to select the next band, $\mathbf{b}_2^{(1)}$ from $\Omega_Y^{(1)} = \{\mathbf{b}_k^{(1)}\}_{k=1}^p$. This $\mathbf{b}_2^{(1)}$ will then once again go into the inner loop to run through $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ one at a time via index l for band replacement according to $\min \tilde{R}(\tilde{D})$. The entire procedure is repeated over and over again by processing the outer loop and then the inner loop iteratively via interchanging two indices, p and l , denoted by (p, l) since it begins with the outer loop at index p followed by the inner loop by l . The details of carrying out SC-RDFBSS-MIN in step-by-step implementation are summarized in the following Algorithm 3.

Algorithm 3 SC-RDFBSS-MIN

Step 1: Initial condition:

- (i) Let n_{BS} be the number of bands needed to be selected determined by VD.
- (ii) Let the channel input space $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ and the channel output space $\Omega_{BS} = \Omega_X^{(0)} = \{\hat{\mathbf{b}}_k^{(0)}\}_{k=1}^{n_{VD}}$ uniformly selected from the full band set. Then calculate the distortion matrix $\mathbf{D}^{(0)} = [\rho_{jk}^{(0)}]_{j=1,k=1}^{L,n_{VD}}$ by SIDAM, SID or SAM.
- (iii) Calculate the input distribution $\{p_l\}_{l=1}^L$ over $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ with p_l given by (40).
- (iv) Implement the Blahut's rate-distortion algorithm to find and normalize RDF to obtain $\tilde{R}(\tilde{D})$ according to (36)-(37).
- (v) calculate the area under $\tilde{R}^{(0)}(\tilde{D})$ curve denoted by $A_{\tilde{R}^{(0)}(\tilde{D})}$.
- (vi) Let $A_{RDF} = \tilde{R}^{(0)}(\tilde{D})$.

Step 2: a. Outer loop: for $p = 1, \dots, n_{VD}$ Let $\Omega_{\hat{X}}^{(p)} = \{\mathbf{b}_k^{(p)}\}_{k=1}^p = \Omega_{BS}^p$.b. Inner loop: for $l = 1, \dots, L$, do

- (i) If $\mathbf{b}_l \in \Omega_{BS}^p$, go to step 2b(v). Otherwise, replace the p th band of $\Omega_{\hat{X}}^{(p,l)}$ with \mathbf{b}_l .
- (ii) Construct a new distortion matrix $\mathbf{D}^{(p,l)} = [\rho_{jk}^{(p,l)}]_{j=1,k=1}^{L,p}$ by SIDAM, SID or SAM with input space Ω_X and output space $\Omega_{\hat{X}}^{(p,l)}$.
- (iii) Implement the Blahut's rate-distortion algorithm to find and normalize RDF to obtain $\tilde{R}^{(p,l)}(\tilde{D})$ by
- (iv) Calculate the $A_{\tilde{R}^{(p,l)}(\tilde{D})}$ of $\tilde{R}^{(p,l)}(\tilde{D})$.
- (v) If $A_{\tilde{R}^{(p,l)}(\tilde{D})} < A_{RDF}$, let $A_{RDF} = A_{\tilde{R}^{(p,l)}(\tilde{D})}$, $\Omega_{BS} = \Omega_{\hat{X}}^{(p,l)}$. Otherwise, continue.
- (vi) If $l < L$, let $l \leftarrow l+1$ and go to step 2b. Otherwise, go to step 3.

Step 3: If $p < n_{VD}$, let $p \leftarrow p+1$ and go to step 2a. Otherwise, continue.Step 4: Output final band subset, $\Omega_{\hat{X}}^{(p,l)}$.

Like SQ-RDFBSS-MAX described in Section V-A, a version of SC-RDFBSS, called SC-RDFBSS-MAX, can be also derived by only replacing $A_{\tilde{R}^{(p,l)}(\tilde{D})} < A_{RDF}$ in step 2(v) of SC-RDFBSS-MIN with $A_{\tilde{R}^{(p,l)}(\tilde{D})} > A_{RDF}$. Figs. 4 and 5 depict the block diagrams of SQ-RDFBSS-MIN and SC-RDFBSS-MIN.

It is worth noting that the area under an RDF curve $\tilde{R}(\tilde{D})$ represents the region of rates that cannot be achieved

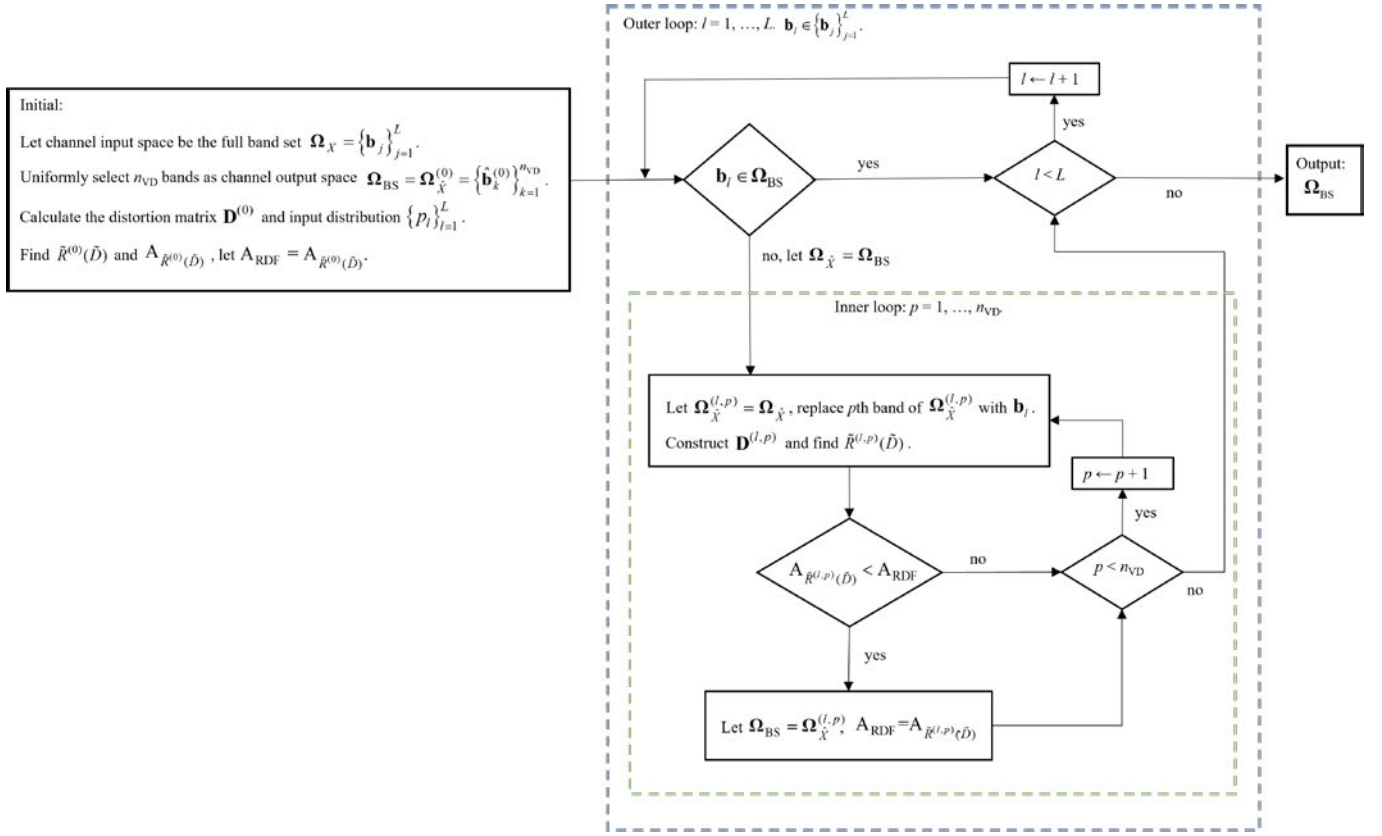


Fig. 4. Block diagram of SQ-RDFBSS-MIN.

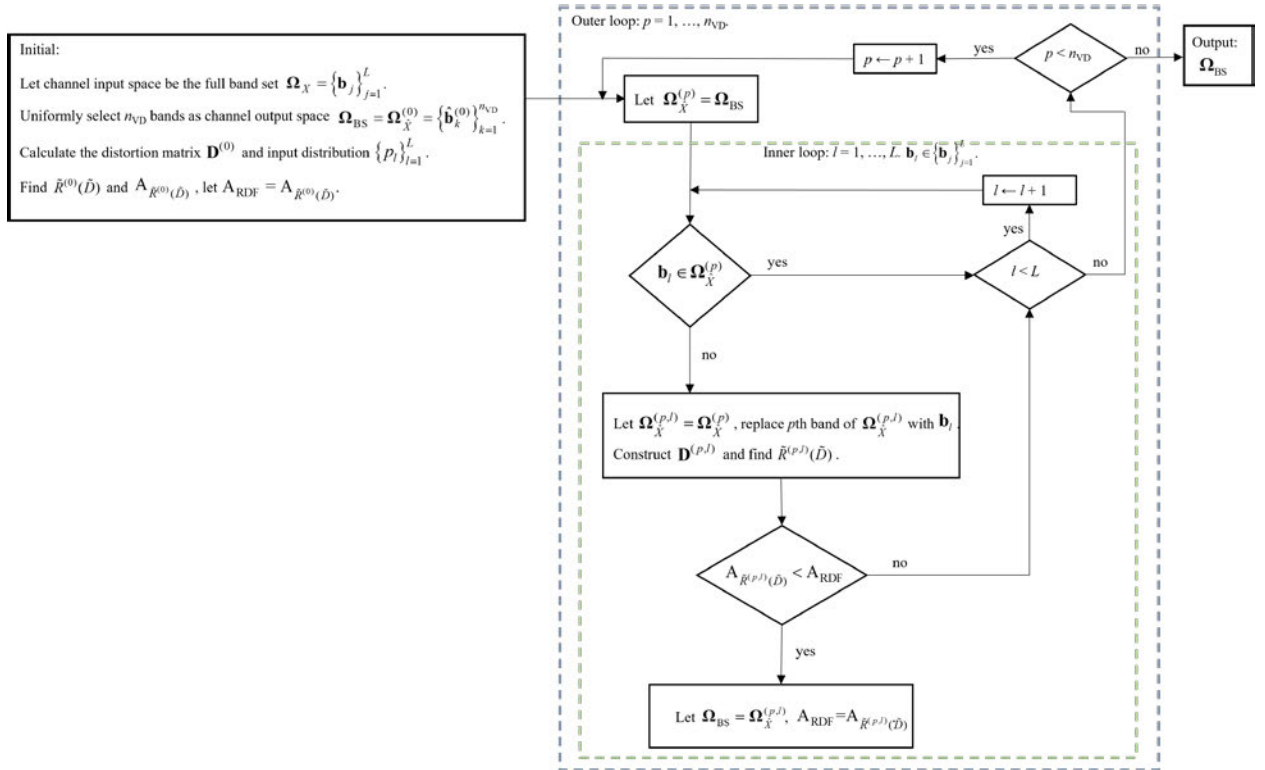


Fig. 5. Block diagram of SC-RDFBSS-MIN.

by a given source subject to the distortion D . Therefore, minimization of $\tilde{R}(\tilde{D})$, $\min \tilde{R}(\tilde{D})$, indicates the minimal possible unachievable rate region. On the other hand, maximization of $\tilde{R}(\tilde{D})$, $\max \tilde{R}(\tilde{D})$ indicates the maximal possible

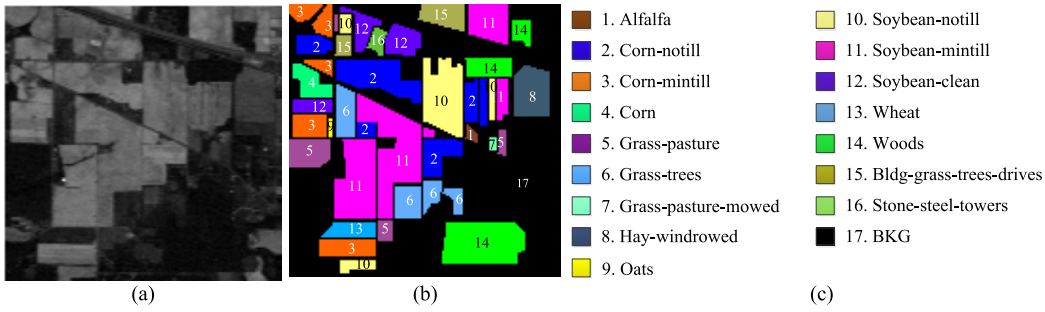


Fig. 6. Purdue's data and their ground truth with 16 classes. (a) Band 186 (2162.56 nm). (b) Ground truth map. (c) Classes by colors.

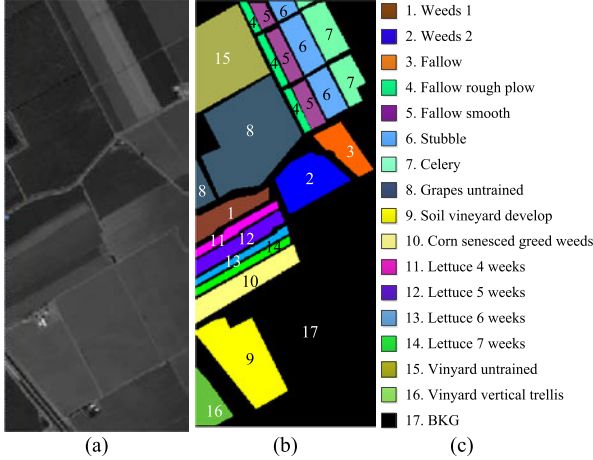


Fig. 7. Salinas and its ground truth with 16 classes. (a) Salinas scene. (b) Ground-truth image. (c) Classes by colors.

unachievable rate region. Since each dataset has different image characteristics, neither $\min \tilde{R}(\tilde{D})$ nor $\max \tilde{R}(\tilde{D})$ can be used as a universal criterion for BS. This is the reason why we develop SQ/SC RDBSS-MIN and SQ/SC RDBSS-MAX for BS.

As a final comment, the most crucial difference between SQ-RDBSS-MIN and SC-RDBSS-MIN is that SQ-RDBSS processes band subsets, $\Omega_Y^{(l)} = \{\mathbf{b}_k^{(l)}\}_{k=1}^p$ sequentially in its inner loop indexed by p for each selected band indexed by l , while SC-RDBSS processes bands in $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ one at a time in the inner loop index by l successively for each selected band subset $\Omega_Y^{(p)} = \{\mathbf{b}_k^{(p)}\}_{k=1}^{n_{\text{vd}}}$ indexed by p .

VI. EXPERIMENT SETUP

This section describes the experimental setup including images to be used for the experiment, and compared BS methods and classification measures used to measure classification performance.

A. Images Used for Experiments

To evaluate the performance of RDBSS, three popular hyperspectral images were used for experiments and studied extensively for HSIC.

1) *Purdue Indiana Indian Pines*: The first one is the Purdue Indiana Indian Pines test site with an aerial view shown in Fig. 6(a) along with its ground truth of 17 class maps in Fig. 6(b) and their class labels in Fig. 6(c). It is an airborne

visible/infrared imaging spectrometer (AVIRIS) image scene and has a size of $145 \times 145 \times 220$ pixel vectors with water absorption bands (bands 104–108 and 150–163, 220). Therefore, a total of 220 bands were used for experiments. It should be noted that in many reports, 200 bands were used by excluding water absorption bands. However, for BS, it is believed that a full set of 220 bands should be used. Table I tabulates the number of data samples in each class where there are four small classes with less than 100 samples, classes 7, 9, 1, 16, and 3 three classes with more than 1000 samples, classes, 2, 11, and 14. Therefore, this scene clearly has an imbalanced class issue in classification.

2) *Salinas*: A second image dataset is Salinas in Fig. 7(a) which is also an AVIRIS scene. It was collected over Salinas Valley, CA, USA, with a spatial resolution of 3.7-m/pixel with a spectral resolution of 10 nm. It has a size of $512 \times 217 \times 224$ with 20 water absorption bands, 108–112, 154–167, 224. Therefore, a total of 224 bands were used for experiments. Fig. 7(b) and (c) shows the color composite of the Salinas image along with the corresponding ground truth class labels. Unlike the Purdue data, the Salinas scene does not have an issue in imbalanced classes as tabulated in Table II where the smallest class is class 13 with 916 data samples.

3) *University of Pavia*: The third and last hyperspectral image data used for experiments were the University of Pavia image shown in Fig. 8 which is an urban area surrounding the University of Pavia, Pavia, Italy. It was recorded by the ROSIS-03 satellite sensor over an urban area surrounding the University of Pavia, Pavia, Italy. 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 μm with a spectral resolution of 4 nm (12 most noisy channels were removed before experiments). Fig. 8(b) provides its ground-truth map of nine classes along with color class labels in Fig. 8(c). Table III also tabulates the number of data samples in parentheses collected for each class. Like the Salinas scene, this scene also has very large classes with only one small class with less than 1000 data samples, class 9 with 947 samples. However, this scene has a more complicated background (BKG) than the other two studied scenes as already shown in [34] where the precision rate (PR) of this scene was much lower than the other two scenes.

According to recent work [35], [36], the Harsany-Farrand-Chang (HFC) method [37] along with its noise-whitened HFC method developed in [1] and [25] was used to estimate VD for the three hyperspectral images and their estimated VD

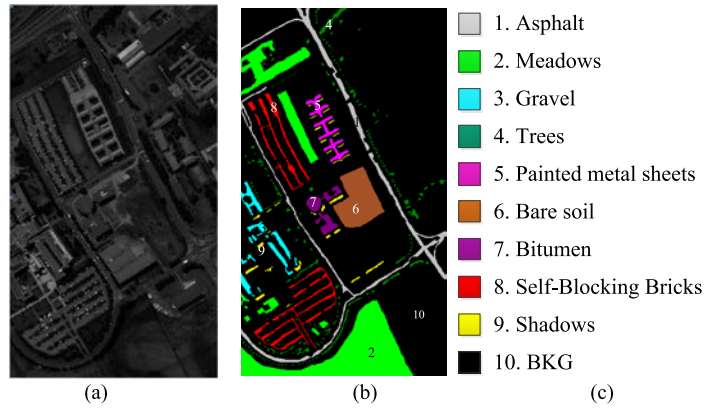


Fig. 8. University of Pavia and its ground truth with nine classes. (a) University of Pavia scene. (b) Ground-truth map. (c) Classes by colors.

TABLE I
CLASS LABELS OF PURDUE'S DATA

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		

TABLE II
CLASS LABELS OF CLASSES IN 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		

TABLE III
CLASS LABELS OF UNIVERSITY 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		

value, n_{VD} with various false alarm probabilities, and P_F are tabulated in Table IV.

In order to make a fair comparison with the results in [35] and [36], n_{VD} was set to $n_{VD} = 18$ for the Purdue data, $n_{VD} = 21$ for Salinas, and $n_{VD} = 14$ for the University of Pavia in which case, n_{VD} was determined by the false alarm probability $P_F = 10^{-4}$.

B. Classification Methods and Test BS Methods to Be Used for Comparison

To demonstrate the advantages of RDFBSS for HSIC, three important components are crucial to justify its utility. One is the classification methods selected to perform classification. Another is to test BS methods selected to conduct comparative studies and analyses. A third one is classification measures used to evaluate classification performance.

TABLE IV

n_{VD} ESTIMATED BY HFC/NWHFC					
	$P_F=10^{-1}$	$P_F=10^{-2}$	$P_F=10^{-3}$	$P_F=10^{-4}$	$P_F=10^{-5}$
Purdue	73/21	49/19	35/18	27/18	25/17
Salinas	32/33	28/24	25/21	21/21	20/20
U. of Pavia	25/34	21/27	16/17	14/14	13/12

1) *Classification Methods*: There are many HSIC methods have been reported in the literature. Among them, spectral-spatial classification is one of the most widely used and promising methods for HSIC. Of particular interest is an EPF-based approach developed in [38] which used a support vector machine (SVM) as a spectral classifier followed by an EPF to perform the final classification. It was later extended to IEPF methods in [39] which significantly improved EPF-based methods. Since four IEPF methods proposed in [36] performed similarly, IEPF-G-g was selected to be used for experiments.

TABLE V
18 SELECTED BANDS FOR PURDUE DATA USING DIFFERENT BS METHODS

	Selected bands ($n_{VD} = 18$)																	
SC-RDFBSS-SID-MIN	11	80	117	43	60	146	4	99	16	115	131	6	38	153	145	130	183	210
SC-RDFBSS-SAM-MIN	8	148	189	38	61	69	80	102	149	57	4	145	35	158	79	185	167	5
SC-RDFBSS-SIDAM-MIN	19	29	80	146	58	45	77	98	81	79	13	121	28	149	174	4	99	210
SQ-RDFBSS-SID-MIN	91	1	35	177	97	6	80	152	187	116	36	153	184	9	174	145	78	2
SQ-RDFBSS-SAM-MIN	57	73	35	168	58	75	174	74	62	116	71	145	61	38	7	37	64	186
SQ-RDFBSS-SIDAM-MIN	93	86	84	87	85	70	81	71	14	59	77	145	51	108	133	186	2	46
SC-RDFBSS-SID-MAX	101	30	111	57	197	92	82	213	196	170	215	184	110	84	166	148	112	210
SC-RDFBSS-SAM-MAX	147	33	74	87	71	70	196	175	219	138	197	72	86	177	169	73	150	176
SC-RDFBSS-SIDAM-MAX	101	102	57	147	127	125	181	115	182	116	113	212	176	214	175	173	195	209
SQ-RDFBSS-SID-MAX	49	12	48	50	47	210	41	218	106	8	45	7	104	219	44	42	43	208
SQ-RDFBSS-SAM-MAX	159	43	42	41	106	160	104	152	162	150	130	108	151	163	157	158	30	101
SQ-RDFBSS-SIDAM-MAX	5	6	167	47	42	44	43	41	32	51	49	28	48	29	30	45	46	39
SMI-BS [19]	59	117	125	38	40	180	173	20	168	31	34	12	114	192	160	108	95	92
UBS	1	14	27	40	53	66	79	92	105	118	131	144	157	170	183	196	209	220

TABLE VI
CLASSIFICATION RESULTS FOR THE PURDUE DATA PRODUCED BY IEPF-G-g USING VARIOUS VERSIONS OF SQ/SC-RDFBSS-MIN, SMI-BS, UBS, AND FULL BANDS

IEPF-G-g									
class	SC-RDFBSS-SID-MIN	SC-RDFBSS-SAM-MIN	SC-RDFBSS-SIDAM-MIN	SQ-RDFBSS-SID-MIN	SQ-RDFBSS-SAM-MIN	SQ-RDFBSS-SIDAM-MIN	SMI-BS [19]	UBS	Full bands
1	1.0000	0.9978	0.9957	0.9978	0.9935	0.9957	1.0000	1.0000	0.9957
2	0.9644	0.9584	0.9587	0.9675	0.9745	0.9604	0.9575	0.9266	0.9336
3	0.9842	0.9861	0.9805	0.9782	0.9882	0.9811	0.9851	0.9873	0.9660
4	0.9996	0.9975	0.9970	1.0000	0.9975	0.9970	0.9987	0.9962	0.9886
5	0.9807	0.9810	0.9778	0.9791	0.9758	0.9754	0.9797	0.9743	0.9629
6	0.9982	0.9990	0.9985	0.9990	0.9979	0.9988	0.9993	0.9985	0.9967
7	0.9893	0.9929	0.9893	0.9929	0.9857	0.9857	0.9929	0.9893	0.9893
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	0.9950	1.0000	1.0000	1.0000	1.0000	1.0000
10	0.9663	0.9790	0.9746	0.9608	0.9819	0.9608	0.9679	0.9504	0.9418
11	0.9608	0.9521	0.9600	0.9474	0.9699	0.9473	0.9505	0.9415	0.9266
12	0.9892	0.9943	0.9943	0.9902	0.9933	0.9926	0.9922	0.9894	0.9860
13	0.9956	0.9971	0.9971	0.9971	0.9956	0.9961	0.9951	0.9971	0.9966
14	0.9934	0.9936	0.9955	0.9957	0.9914	0.9906	0.9963	0.9966	0.9869
15	0.9979	0.9966	0.9922	0.9961	0.9922	0.9922	0.9953	0.9959	0.9883
16	0.9957	0.9968	0.9978	0.9968	0.9978	0.9978	0.9989	0.9989	0.9957
P _{OA}	0.9784	0.9772	0.9781	0.9750	0.9833	0.9732	0.9757	0.9673	0.9596
P _{APR}	0.9745	0.9780	0.9739	0.9689	0.9793	0.9752	0.9736	0.9691	0.9504
P _{AA}	0.9885	0.9889	0.9881	0.9871	0.9897	0.9857	0.9881	0.9839	0.9784

Correspondingly, IEPF-G-g was also selected for this reason. When IEPF-G-g was implemented, the training samples were set up in exactly the same way as it was used in [36]. It should be noted that IEPF-B-g was also used for experiments as the classifier in [19]. As shown in [39], both IEPF-B-g and IEPF-G-g performed similarly. The IEPF-G-g was chosen as its alternative for comparison in [19] because IEPF-G-g used guided filters as opposed to IEPF-B-g which used bilateral filters.

2) *Compared BS Methods*: Many BS methods are proposed in the past. As shown in [20], SMI-BS already performed at least comparably or better than most recent BS techniques such as minimum estimated abundance covariance (MEAC) [40], forward class signature constrained band selection (CSCBS)-BP (FCSCBS-BP) [35], backward CSCBS-BP (FCSCBS-BP) [35], linearly constrained minimum variance-sequential feed-forward BS (LCMV-SFBS) [36], LCMV-sequential backward BS (LCMV-SBBS) [36], SQ-CCBSS [18], SC-CCBSS [18], multigraph determinantal point process (MDPP) [41], and dominant set extraction BS (DSEBS) in [42]. In this case, we only need to compare RDFBSS to SMI-BS using three BD measures, SAM, SID,

SIDAM, and UBS which do not need any criterion or prior knowledge, and it is considered to be an efficient BS method.

C. Classification Measures

Three types of classification measures defined in [43] will be used for experiments. Let, n_{mm} = the number of signal samples in the m th class correctly classified into the m th class \hat{C}_m , n_{jm} = the number of data samples in the m th class C_m but actually classified into the j th class, \hat{C}_j , M = the number of classes, C_m = the set of data samples in the m th class, by ground truth, $n_m = \sum_{j=1}^M n_{jm}$ = the number of data samples in C_m , and N = total number of data samples, $N = \sum_{m=1}^M n_m$.

In traditional HSIC, the popular performance measurements are class accuracy (A) and overall accuracy (OA), given by

$$p_A(C_m) = \text{accuracy of the } m\text{th class} = \frac{n_{mm}}{\sum_{j=1}^M n_{jm}} = \frac{n_{mm}}{n_m} \quad (41)$$

$$P_{OA} = \frac{1}{N} \sum_{m=1}^M n_{mm} = \sum_{m=1}^M \frac{n_m}{N} p_A(C_m) \quad (42)$$

TABLE VII
CLASSIFICATION RESULTS FOR THE PURDUE DATA PRODUCED BY IEPF-G-G USING VARIOUS VERSIONS
OF SQ/SC-RDFBSS-MAX, SMI-BS, UBS, AND FULL BANDS

class	IEPF-G-g					
	SC-RDFBSS-SID-MAX	SC-RDFBSS-SAM-MAX	SC-RDFBSS-SIDAM-MAX	SQ-RDFBSS-SID-MAX	SQ-RDFBSS-SAM-MAX	SQ-RDFBSS-SIDAM-MAX
1	1.0000	0.9957	0.9891	0.9957	0.9978	0.9957
2	0.9652	0.9550	0.9722	0.9404	0.9227	0.9578
3	0.9811	0.9827	0.9793	0.9851	0.9605	0.9851
4	0.9983	0.9970	0.9937	0.9996	0.9987	0.9970
5	0.9810	0.9772	0.9710	0.9737	0.9723	0.9689
6	0.9968	0.9973	0.9944	0.9947	0.9951	0.9959
7	0.9964	0.9893	0.9893	0.9857	0.9750	0.9857
8	1.0000	1.0000	0.9992	1.0000	1.0000	1.0000
9	1.0000	1.0000	0.9900	1.0000	1.0000	1.0000
10	0.9741	0.9615	0.9628	0.9502	0.9176	0.9698
11	0.9526	0.9375	0.9529	0.9440	0.8908	0.9606
12	0.9917	0.9941	0.9821	0.9870	0.9755	0.9895
13	0.9956	0.9961	0.9922	0.9976	0.9956	0.9956
14	0.9968	0.9933	0.9914	0.9940	0.9880	0.9915
15	0.9964	0.9982	0.9953	0.9904	0.9984	0.9813
16	1.0000	0.9978	0.9968	1.0000	0.9957	0.9892
P_{OA}	0.9775	0.9710	0.9751	0.9688	0.9472	0.9762
P_{APR}	0.9767	0.9723	0.9802	0.9727	0.9595	0.9763
P_{AA}	0.9891	0.9858	0.9845	0.9836	0.9740	0.9852

TABLE VIII
21 SELECTED BANDS FOR SALINAS USING DIFFERENT BS METHODS

Selected bands ($n_{VD} = 21$)																					
SC-RDFBSS-SID-MIN	151	8	28	105	166	50	18	81	224	99	222	154	6	116	112	106	128	172	189	169	111
SC-RDFBSS-SAM-MIN	4	37	148	34	119	61	71	138	86	104	221	124	41	114	65	176	174	190	192	204	216
SC-RDFBSS-SIDAM-MIN	106	149	11	224	167	7	41	59	151	104	124	193	131	143	154	139	158	19	150	195	173
SQ-RDFBSS-SID-MIN	112	201	31	111	44	166	40	82	100	102	154	123	222	101	192	83	224	1	194	171	117
SQ-RDFBSS-SAM-MIN	169	217	177	41	175	61	72	115	218	220	202	223	186	171	117	40	219	210	192	216	221
SQ-RDFBSS-SIDAM-MIN	3	221	30	38	44	7	4	2	19	223	113	224	220	145	168	1	105	12	5	104	222
SC-RDFBSS-SID-MAX	158	155	159	160	156	140	157	82	187	188	192	189	185	134	138	164	142	196	186	184	136
SC-RDFBSS-SAM-MAX	111	166	158	154	162	159	61	157	142	160	155	110	108	134	165	164	109	185	156	212	204
SC-RDFBSS-SIDAM-MAX	107	110	165	125	134	93	86	130	92	85	122	123	121	120	124	175	174	184	194	205	215
SQ-RDFBSS-SID-MAX	10	19	2	17	13	61	4	3	11	12	5	18	138	155	1	164	156	160	137	159	157
SQ-RDFBSS-SAM-MAX	95	89	45	96	44	160	90	108	92	98	159	43	132	143	97	162	91	99	158	157	109
SQ-RDFBSS-SIDAM-MAX	75	16	60	41	51	32	47	78	70	22	71	125	74	77	72	73	76	34	59	30	58
SMI-BS [19]	15	32	34	40	63	117	150	157	184	206	223	22	41	64	71	94	118	151	176	185	202
UBS	1	12	23	34	45	56	67	78	89	100	111	122	133	144	155	166	177	188	199	210	224

which shows that P_{OA} utilizes sample ratios as weights to average the accuracy of each class. In addition, we can consider all classes to be equally likely by letting $(n_m/N) = (1/M)$ to further define P_{AA} as average accuracy (AA) by the number of classes, M , expressed as

$$P_{AA} = \frac{1}{M} \sum_{m=1}^M \frac{n_{mm}}{n_m} = \frac{1}{M} \sum_{m=1}^M P_A(C_m). \quad (43)$$

Another measure is PR given by

$$p_{PR}(C_m) = \text{precision rate of } C_m = \frac{\hat{n}_{mm}}{\hat{n}_m} \quad (44)$$

where $\hat{n}_m = \sum_{j=1}^M \hat{n}_{mj}$ is the total number of data samples that be classified into the m th class and \hat{n}_{mj} is the number of data samples classified into the m th class but supposed to be in the j th class. Therefore, the average PR (APR), P_{APR} , in correspondence to P_{AA} in (46) is defined by

$$P_{APR} = \frac{1}{M} \sum_{m=1}^M P_{PR}(\hat{C}_m). \quad (45)$$

Since n_{mm} is identical to \hat{n}_{mm} and $\hat{N} = \sum_{i=1}^M \hat{n}_i$ equal to $N = \sum_{i=1}^M n_i$. As a result, an overall PR (OPR), P_{OPR} can be

also defined as the counterpart of P_{OA} in (42) by

$$\begin{aligned} P_{OPR} &= \sum_{i=1}^M p(\hat{C}_i) P_{PR}(\hat{C}_i) = \sum_{i=1}^M \left(\frac{\hat{n}_i}{\hat{N}} \right) P_{PR}(\hat{C}_i) \\ &= \sum_{i=1}^M \left(\frac{n_i}{N} \right) P_A(C_i) = P_{OA} \end{aligned} \quad (46)$$

which implies that P_{OPR} is essentially the same as P_{OA} .

It should be noted that PR, also known as user's accuracy, is proposed to address the BKG issue, while many existing classification methods have removed all unlabeled data samples as BKG from classification. PR is the one that can include all data samples for evaluation.

VII. EXPERIMENTAL RESULTS AND DISCUSSIONS

According to Table IV [35], [36], n_{BS} was selected to be 18 for the Purdue data, 21 for Salinas, and 14 for the University of Pavia with $P_F = 10^{-4}$.

A. Purdue Data

Table V tabulates 18 bands selected by various versions of SQ/SC-RDFBSS-MIN and SQ/SC-RDFBSS-MAX along

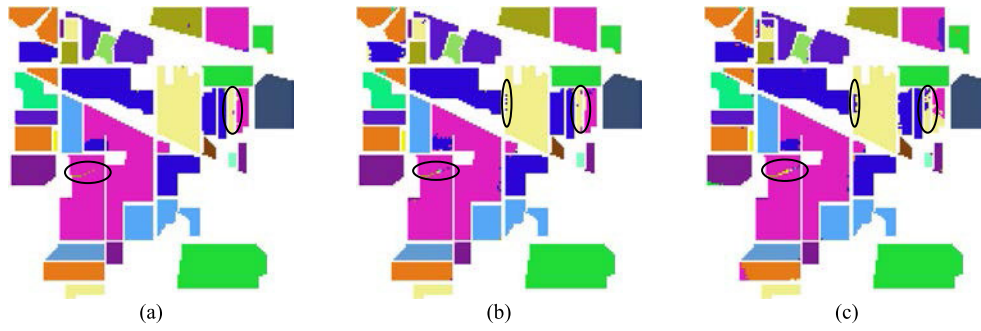


Fig. 9. Classification maps by IEPF-G-g using the best result, (a) SQ-RDFBSS-SAM-MIN, (b) SMI-BS, and (c) full bands.

TABLE IX
CLASSIFICATION RESULTS FOR SALINAS PRODUCED BY IEPF-G-G USING VARIOUS VERSIONS OF
R SQ/SC-RDFBSS-MIN, SMI-BS, UBS, AND FULL BANDS

class	IEPF-G-g						SMI-BS [19]	UBS	Full bands
	SC-RDFBSS- SID-MIN	SC-RDFBSS- SAM-MIN	SC-RDFBSS- SIDAM-MIN	SQ-RDFBSS- SID-MIN	SQ-RDFBSS- SAM-MIN	SQ-RDFBSS- SIDAM-MIN			
1	0.9961	0.9999	1.0000	1.0000	0.9990	1.0000	0.9998	0.9998	1.0000
2	0.9953	0.9961	0.9981	0.9981	0.9957	0.9993	0.9947	0.9964	0.9967
3	0.9966	0.9995	0.9993	0.9999	0.9987	1.0000	0.9996	0.9997	0.9986
4	0.9958	0.9976	0.9972	0.9987	0.9979	0.9973	0.9979	0.9986	0.9973
5	0.9827	0.9904	0.9904	0.9911	0.9861	0.9877	0.9909	0.9916	0.9899
6	0.9947	0.9980	0.9989	0.9989	0.9972	0.9996	0.9982	0.9988	0.9978
7	0.9939	0.9962	0.9969	0.9972	0.9957	0.9976	0.9955	0.9963	0.9958
8	0.9267	0.9054	0.9034	0.8835	0.8808	0.8841	0.8611	0.8886	0.8944
9	0.9951	0.9967	0.9982	0.9964	0.9932	0.9982	0.9945	0.9957	0.9964
10	0.9772	0.9892	0.9869	0.9863	0.9829	0.9867	0.9822	0.9889	0.9820
11	0.9940	0.9962	0.9967	0.9983	0.9946	0.9984	0.9961	0.9982	0.9960
12	0.9985	0.9999	1.0000	1.0000	0.9996	1.0000	1.0000	1.0000	1.0000
13	0.9900	0.9931	0.9942	0.9941	0.9928	0.9977	0.9936	0.9927	0.9936
14	0.9868	0.9905	0.9903	0.9913	0.9903	0.9939	0.9915	0.9932	0.9894
15	0.9305	0.8990	0.9113	0.8969	0.9172	0.9072	0.8946	0.8728	0.9098
16	0.9857	0.9915	0.9897	0.9911	0.9892	0.9908	0.9912	0.9903	0.9901
P_{OA}	0.9700	0.9638	0.9653	0.9592	0.9598	0.9610	0.9532	0.9569	0.9624
P_{APR}	0.9845	0.9806	0.9803	0.9789	0.9785	0.9797	0.9761	0.9782	0.9779
P_{AA}	0.9837	0.9837	0.9845	0.9826	0.9819	0.9837	0.9801	0.9814	0.9830

SMI-BS along with UBS. Tables VI and VII also tabulate their classification results produced by IEPF-G-g in terms of P_{OA} , P_{AA} , and P_{APR} , respectively, with the best results boldfaced in red. Also included in Table VI are results produced by full bands in the last column for comparison. As shown in Table VI, all RDF-BBS versions outperformed full bands and the best result was produced by SQ-RDFBSS-SAM-MIN. Compared Table VI to Table VII, SQ/SC-RDFBSS-MIN performed better than SQ/SC-RDFBSS-MAX. Nevertheless, except SQ-RDFBSS-SAM-MAX, all other SQ/SC-RDFBSS-MAX still outperformed full bands. Interestingly, despite the fact that SMI-BS was not the best one but performed very comparably to all SQ/SC RDFBSS versions.

Fig. 9 shows the classification maps produced by IEPF using the best result, SQ-RDFBSS-SAM-MIN, SMI-BS, and full bands for visual inspection. For example, the regions of classes 10 and 11 circled by black showed the smallest classification errors produced by using SQ-RDFBSS-SAM-MIN in Fig. 9(a) compared to the largest classification errors produced by using full bands in Fig. 9(c), while SMI-BS performed right in between.

B. Salinas

Table VIII tabulates 21 bands selected by various versions of SQ/SC-RDFBSS-MIN and SQ/SC-RDFBSS-MAX along

SMI-BS along with UBS. Tables IX and X also tabulate their classification results produced by IEPF-G-g in terms of P_{OA} , P_{AA} , and P_{APR} , respectively with the best results boldfaced in red. Also included in Table IX are results produced by full bands in the last column for comparison. As shown in Table IX, all SC-RDFBSS versions outperformed full bands and the best result was produced by SC-RDFBSS-SID-MIN. Unlike the Purdue data, none of the SQ-RDFBSS-MIN versions performed better than full bands. Compared Table IX to Table X, SQ/SC-RDFBSS-MIN also performed better than SQ/SC-RDFBSS-MAX and none of all SQ/SC-RDFBSS-MAX versions could perform better than full bands. Once again, although SMI-BS did perform the best but did perform very comparably to all SQ/SC-RDFBSS versions. Fig. 10 shows the classification maps produced by IEPF using the best result, SC-RDFBSS-SID-MIN, SMI-BS, and full bands for visual inspection where the classification map of class 8 circled by black produced by using SC-RDFBSS-SID-MIN was clearly better than those produced by SMI-BS and full bands.

C. University of Pavia

Table XI tabulates 14 bands selected by various versions of SQ/SC-RDFBSS-MIN and SQ/SC-RDFBSS-MAX along SMI-BS along with UBS. Tables XII and XIII also tabulate

TABLE X
CLASSIFICATION RESULTS FOR SALINAS PRODUCED BY IEPF-G-g USING VARIOUS VERSIONS OF
SQ/SC-RDFBSS-MAX, SMI-BS, UBS, AND FULL BANDS

class	IEPF-G-g					
	SC-RDFBSS-SID-MAX	SC-RDFBSS-SAM-MAX	SC-RDFBSS-SIDAM-MAX	SQ-RDFBSS-SID-MAX	SQ-RDFBSS-SAM-MAX	SQ-RDFBSS-SIDAM-MAX
1	0.9997	0.9992	0.9995	0.9999	0.9998	0.9996
2	0.9972	0.9930	0.9973	0.9992	0.9971	0.9954
3	0.9964	0.9904	0.9955	0.9999	0.9978	0.9988
4	0.9960	0.9974	0.9953	0.9978	0.9996	0.9964
5	0.9872	0.9836	0.9881	0.9862	0.9848	0.9855
6	0.9963	0.9961	0.9976	0.9990	0.9970	0.9979
7	0.9953	0.9935	0.9958	0.9972	0.9950	0.9953
8	0.8699	0.9072	0.8885	0.8465	0.9042	0.9072
9	0.9927	0.9924	0.9937	0.9977	0.9940	0.9957
10	0.9809	0.9823	0.9838	0.9900	0.9880	0.9832
11	0.9933	0.9934	0.9970	0.9993	0.9939	0.9952
12	0.9954	0.9916	0.9980	0.9995	0.9995	1.0000
13	0.9939	0.9939	0.9972	0.9989	0.9954	0.9933
14	0.9859	0.9898	0.9921	0.9954	0.9926	0.9943
15	0.9049	0.9018	0.8846	0.8897	0.8888	0.9138
16	0.9842	0.9828	0.9851	0.9887	0.9898	0.9928
P_{OA}	0.9553	0.9618	0.9572	0.9508	0.9614	0.9654
P_{APR}	0.9755	0.9796	0.9788	0.9765	0.9794	0.9814
P_{AA}	0.9793	0.9805	0.9806	0.9803	0.9823	0.9840

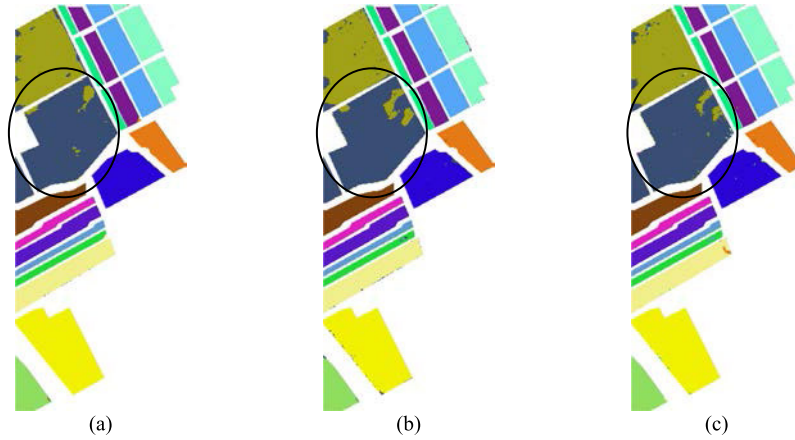


Fig. 10. Classification maps by IEPF-G-g using the best result, (a) SC-RDFBSS-SID-MIN, (b) SMI-BS, and (c) full bands.

TABLE XI
14 SELECTED BANDS FOR UNIVERSITY OF PAVIA USING DIFFERENT BS METHODS

	Selected bands ($n_{VD} = 14$)													
	5	11	3	18	35	41	71	26	65	74	77	38	32	101
SC-RDFBSS-SID-MIN	5	24	8	30	3	46	80	51	65	68	77	57	84	102
SC-RDFBSS-SAM-MIN	72	2	11	1	35	3	20	68	62	26	76	74	55	97
SC-RDFBSS-SIDAM-MIN	7	16	5	28	72	74	37	77	61	68	76	83	3	96
SQ-RDFBSS-SID-MIN	7	46	86	27	33	42	50	55	63	67	78	4	84	102
SQ-RDFBSS-SAM-MIN	48	64	13	60	3	63	49	61	62	1	73	58	65	74
SQ-RDFBSS-SIDAM-MIN	24	48	88	45	78	80	81	59	61	69	46	84	63	58
SC-RDFBSS-SID-MAX	71	91	18	54	94	50	53	51	70	85	52	96	93	97
SC-RDFBSS-SAM-MAX	38	78	77	68	70	39	45	55	66	69	67	65	51	50
SC-RDFBSS-SIDAM-MAX	62	59	100	101	60	61	99	63	95	70	58	98	102	103
SQ-RDFBSS-SID-MAX	7	10	9	62	8	14	13	61	17	15	16	39	12	11
SQ-RDFBSS-SAM-MAX	84	83	80	68	52	51	49	85	82	69	53	81	87	86
SQ-RDFBSS-SIDAM-MAX	22	41	91	9	18	21	37	40	48	57	66	82	92	94
SMI-BS [19]	1	9	17	25	33	41	49	57	65	73	81	89	97	103
UBS														

their classification results produced by IEPF-G-g in terms of P_{OA} , P_{AA} , and P_{APR} , respectively with the best results were boldfaced in red. Also included in Table XI are results produced by full bands in the last column for comparison. Interestingly, as shown in Table XII, none of the

SQ/SC-RDFBSS-MIN and SQ/SC-RDFBSS-MAX versions could perform better than full bands, while the best result among all SQ/SC-RDFBSS versions was produced by SC-RDFBSS-SAM-MAX with P_{AA} and P_{OA} and SQ-RDFBSS-SID-MIN with P_{AA} . Nevertheless, their performances were

TABLE XII
CLASSIFICATION RESULTS FOR UNIVERSITY OF PAVIA PRODUCED BY IEPF-G-G USING VARIOUS VERSIONS OF
SQ/SC-RDFBSS-MIN, SMI-BS, UBS, AND FULL BANDS

IEPF-G-g									
class	SC-RDFBSS-SID-MIN	SC-RDFBSS-SAM-MIN	SC-RDFBSS-SIDAM-MIN	SQ-RDFBSS-SID-MIN	SQ-RDFBSS-SAM-MIN	SQ-RDFBSS-SIDAM-MIN	SMI-BS [19]	UBS	Full bands
1	0.9813	0.9802	0.9805	0.9833	0.9832	0.9854	0.9782	0.9868	0.9858
2	0.9546	0.9660	0.9644	0.9671	0.9697	0.9203	0.9613	0.9534	0.9761
3	0.9832	0.9862	0.9878	0.9852	0.9840	0.9791	0.9879	0.9884	0.9911
4	0.9924	0.9903	0.9899	0.9858	0.9904	0.9818	0.9933	0.9946	0.9879
5	0.9967	0.9983	0.9981	0.9992	0.9987	0.9978	0.9967	0.9961	0.9978
6	0.9827	0.9901	0.9938	0.9967	0.9898	0.9481	0.9873	0.9854	0.9972
7	0.9961	0.9980	0.9984	0.9993	0.9983	0.9932	0.9972	0.9969	0.9997
8	0.9748	0.9766	0.9805	0.9842	0.9751	0.9764	0.9733	0.9903	0.9819
9	0.9962	0.9971	0.9969	0.9992	0.9967	0.9936	0.9957	0.9958	0.9997
P _{OA}	0.9714	0.9774	0.9776	0.9795	0.9792	0.9521	0.9746	0.9738	0.9841
P _{APR}	0.9684	0.9725	0.9730	0.9748	0.9746	0.9579	0.9688	0.9727	0.9786
P _{AA}	0.9842	0.9870	0.9878	0.9889	0.9873	0.9751	0.9857	0.9875	0.9908

TABLE XIII
CLASSIFICATION RESULTS FOR UNIVERSITY OF PAVIA PRODUCED BY IEPF-G-G USING VARIOUS VERSIONS
OF SQ/SC-RDFBSS-MAX, SMI-BS, UBS, AND FULL BANDS

IEPF-G-g						
class	SC-RDFBSS-SID-MAX	SC-RDFBSS-SAM-MAX	SC-RDFBSS-SIDAM-MAX	SQ-RDFBSS-SID-MAX	SQ-RDFBSS-SAM-MAX	SQ-RDFBSS-SIDAM-MAX
1	0.9773	0.9808	0.9748	0.9789	0.9799	0.9739
2	0.9541	0.9697	0.9529	0.9635	0.9235	0.9646
3	0.9865	0.9824	0.9779	0.9807	0.9746	0.9808
4	0.9895	0.9902	0.9899	0.9868	0.9772	0.9874
5	0.9984	0.9997	0.9980	0.9993	0.9953	0.9990
6	0.9765	0.9939	0.9740	0.9831	0.9562	0.9876
7	0.9953	0.9986	0.9974	0.9989	0.9950	0.9986
8	0.9733	0.9851	0.9554	0.9639	0.9672	0.9776
9	0.9968	0.9985	0.9914	0.9996	0.9893	0.9994
P _{OA}	0.9697	0.9802	0.9665	0.9737	0.9521	0.9752
P _{APR}	0.9653	0.9757	0.9607	0.9692	0.9550	0.9685
P _{AA}	0.9831	0.9888	0.9791	0.9839	0.9731	0.9854

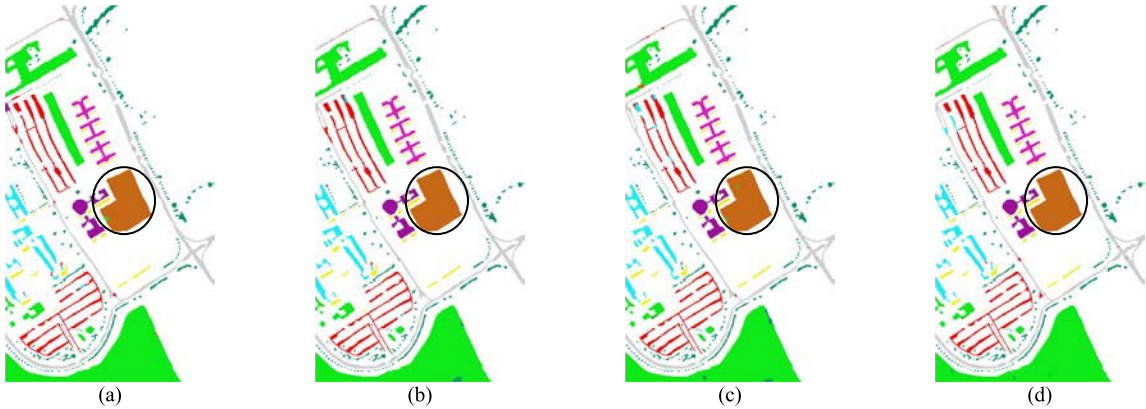


Fig. 11. Classification maps by IEPF-G-g for the University of Pavia using two best results, (a) SC-RDFBSS-SAM-MAX, (b) SQ-RDFBSS-SID-MIN, (c) SMI-BS, and (d) full bands.

very close to that produced by full bands. The reason for this is that the University of Pavia has a very complicated BKG as noted in [34] with its data description in Section VI-A3. In this case, IEPF requires full bands to perform well. Unlike the Purdue data and Salinas, Tables XII and XIII, showed that both SQ/SC-RDFBSS-MIN and SQ/SC-RDFBSS-MAX performed very comparably. Once again, SMI-BS also performed very comparably to all SQ/SC RDFBSS versions. Fig. 11 shows the classification maps produced by IEPF using the two best results, SC-RDFBSS-SAM-MAX and

SQ-RDFBSS-SID-MIN, SMI-BS, and full bands for visual inspection. By inspecting class 6 circled by black in Fig. 11, it clearly showed that full bands and SQ-RDFBSS-SID-MIN produced the best results.

D. Discussions

According to the above experimental results in Sections VII-A–VII-C, we summarize several interesting observations and findings in the following:

TABLE XIV
COMPARISON BETWEEN CCBSS [17] AND RDFBSS

	CCBSS [17]	RDFBSS
Motivation	Shannon's 2 nd coding theorem	Shannon's 3 rd coding theorem
Design rationale	<ul style="list-style-type: none"> Mutual information (MI) between channel input space $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$ and channel output space $\Omega_Y = \{\mathbf{b}_k\}_{k=1}^L$, $I(\Omega_X; \Omega_Y)$ $= \sum_{j=1}^L p_j \sum_{k=1}^L Q_{k j} \log \frac{Q_{k j}}{\sum_{j=1}^L p_j Q_{k j}}$ To find bands which maximum SMI 	<ul style="list-style-type: none"> Mutual information (MI) between channel input space $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$, channel output space $\Omega_{BS} = \{\mathbf{b}_k\}_{k=1}^{n_{BS}}$, $I(\Omega_X; \Omega_{BS})$ $= \sum_{l=1}^L p_l \sum_{k=1}^{n_{BS}} \log \frac{Q_{k l}}{\sum_{l=1}^L p_l Q_{k l}}$ To find a least favorable channel transition matrix that minimizes mutual information over all possible channels subject to a distortion D
input information	n_{BS} , $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$, $\Omega_Y = \{\mathbf{b}_k\}_{k=1}^L$, channel transition matrix $\{Q_{k j}\}_{j=1, k=1}^{L, n_{BS}}$	n_{BS} , $\Omega_X = \{\mathbf{b}_l\}_{l=1}^L$, $\Omega_{BS} = \{\mathbf{b}_k\}_{k=1}^{n_{BS}}$, distortion measure $d(x, \hat{x})$ and distortion D
Objective function	Self-mutual information (SMI) $I(\mathbf{b}_j; \Omega_Y) = \sum_{k=1}^L Q_{k j} \log \frac{Q_{k j}}{\sum_{j=1}^L p_j Q_{k j}}$	Rate distortion function (RDF) $R(D) = \min_{Q_{kl}} I(\Omega_X; \Omega_{BS})$ subject to $E[d] = \sum_{l=1}^L \sum_{k=1}^{n_{BS}} p_l Q_{k l} d(\mathbf{b}_l, \mathbf{b}_k) \leq D$ and normalize $R(D)$, denoted by $\tilde{R}(\tilde{D})$
Finding band subsets	Calculate n_{BS} -prominent peaks of the curve plotted by $\{I(\mathbf{b}_j; \Omega_Y)\}_{j=1}^L$	Calculate the area under $\tilde{R}(\tilde{D})$, $A_{\tilde{R}(\tilde{D})}$
output	n_{BS} -band subset which yields the n_{BS} prominent peaks	n_{BS} -band subset which yields A_{RDF} either by SQ/SC RDFBSS-MIN or SQ/SC RDFBSS-MAX

- 1) SQ/SC-RDFBSS-MIN generally performed better than SQ/SC-RDFBSS-MAX.
- 2) SMI-BS may perform slightly worse than the best versions of SQ/SC-RDFBSS-MIN/MAX but it did perform comparably to their best versions. These experiments demonstrated that Shannon's second and third coding theorems can have benefits on the design and development of BS methods.
- 3) It has been shown in [19] that SMI-BS performed better than most recently developed BS methods. Since SQ/SC-RDFBSS-MIN also performed better than SMI-BS based on our conducted experiments, this indicates that SQ/SC-RDFBSS-MIN also performed better than these BS methods as well.
- 4) Experiments also demonstrate that RDFBSS and SMI-BS indeed performed very comparably. This evidence shows that information theory-based BS methods can be more effective in addressing spectral redundancy.
- 5) There are several reasons for tabulating the detailed results in Tables V, VIII, and XI. One is to show different BS methods and select different BSS. Another is to show the interband correlation exhibited by selected bands. A third one is for comparison among the bands selected by other BS methods. In many references published in the literature, the selected bands were not reported for comparison. In this case, it is very hard to reproduce the BS results.

- 6) Tables VI, IX, and XII provided evidence that three datasets exhibit different image characteristics, which implies that no BS method can claim to perform universally better than other BS methods for all three datasets. In other words, one BS method could perform better than another on one dataset but may also perform worse on another dataset. This is indeed the case shown by our experiments.
- 7) Table XIV tabulates the running times required by each of the test BS methods including full bands.
- 8) Finally, since RDFBSS is unsupervised, one of its potential applications is anomaly detection which is currently under investigation.

VIII. CONCLUSION

The three Shannon's coding theorem in information theory have found their roles in many applications in communications and signal processing such as KL divergence [13] for the generative adversarial network (GAN), cross entropy for objection detection, etc. Of particular interest is their applications to BS. Two of the most widely used concepts are self-information [15] and entropy [5] for BP, both of which arise from Shannon's first coding theorem. Most recently, Shannon's second channel coding theorem has also found its applications in BSS [16], [17]. Interestingly, it seems that by far no work was reported in the past in terms of applying Shannon's third RDF theorem to BS. This article is believed

to be the first work to do so by developing an RDF approach to BSS. The main issue is how to interpret RDF to formulate a BSS problem. The approach based on CC developed in [16] and [17] is not applicable to RDF. In this case, we need to reinvent the wheel by looking into alternative means. One of the key ideas is to convert a channel transition matrix to a distortion matrix which can be constructed by a BD measure. Another which is crucial to success is to design an RDF-based criterion that can be used to measure the BSS performance. Finally, to find an optimal band subset the idea of SQ/SC-N-FINDR [24], [25], [26] is further incorporated to derive SQ/SC-RDFBSS algorithms.

In what follows, we summarize several novelties of this article as follows:

- 1) It formulates RDF as a BSS problem which is believed to be the first work ever reported in the literature.
- 2) It reinterprets the RDF theory from a BSS perspective as discussed in Section I. Specifically, it interprets three of Shannon's coding theorems for BSS which have not been done in the past.
- 3) It reinterprets channel transition matrices, $Q_{k|j} = [p(\mathbf{b}_k|\mathbf{b}_j)]_{j,k}^{L,n_{BS}}$ as a distortion matrix in Section II.
- 4) It designs a new RDF curve-based criterion for BSS which is very similar to the ROC curve-based criterion discussed in Section IV. This entire section is completely new and considered to be the very first work ever reported in the literature.
- 5) Last but not least, Table XIV compares CCBSS in [17] and RDFBSS in terms of their motivations, design rationales, objection functions, and BSS.

APPENDIX

Proof of Theorem 3:

1)

Since

$$\begin{aligned} E[d] &= \sum_{x,\hat{x}} p(x,\hat{x})d(x,\hat{x}) \\ &\geq \sum_x p(x) \sum_{\hat{x}} p(\hat{x}|x) \min_{\hat{x}} d(x,\hat{x}) \\ &\geq \sum_{x,\hat{x}} p(x,\hat{x}) \min_{\hat{x}} d(x,\hat{x}) = D_{\min}. \end{aligned} \quad (\text{A.1})$$

2) (\Leftarrow) If $D \geq D_{\max}$, then $R(D) = 0$.

Find a test channel that maps all x into a $\hat{x}_{\max} = \arg\{\min_{\hat{x}} \sum_x p(x)d(x,\hat{x})\}$. In this case,

$$p(\hat{x}|x) = \begin{cases} 1, & \text{if } \hat{x} = \hat{x}_{\max} \\ 0, & \text{if } \hat{x} \neq \hat{x}_{\max}. \end{cases}$$

This implies that

$$p(x,\hat{x}) = \begin{cases} p(\hat{x}|x)p(x) = p(x), & \text{if } \hat{x} = \hat{x}_{\max} \\ p(\hat{x}|x)p(x) = 0, & \text{if } \hat{x} \neq \hat{x}_{\max} \end{cases}$$

and

$$p(x)p(\hat{x}) = \begin{cases} p(x), & \text{if } \hat{x} = \hat{x}_{\max} \\ 0, & \text{if } \hat{x} \neq \hat{x}_{\max}. \end{cases}$$

Therefore, $p(x,\hat{x}) = p(x)p(\hat{x}) \Rightarrow I(X;\hat{X}) = 0$ and

$$\begin{aligned} E[d] &= \sum_{x,\hat{x}} p(x,\hat{x})d(x,\hat{x}) = \sum_{\hat{x}} p(\hat{x}) \sum_x p(x|\hat{x})d(x,\hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \left(\min_{\hat{x}} \sum_x p(x|\hat{x})d(x,\hat{x}) \right) = \sum_{\hat{x}} p(\hat{x}) D_{\min}. \end{aligned} \quad (\text{A.2})$$

(\Rightarrow) If $R(D) = 0$, then $D \geq D_{\max}$

If $R(D) = 0$, then $I(X;\hat{X}) = 0$ which implies that X and \hat{X} are independent

$$\begin{aligned} E[d] &= \sum_{x,\hat{x}} p(x,\hat{x})d(x,\hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x|\hat{x})d(x,\hat{x}) \\ &\geq \sum_{\hat{x}} p(\hat{x}) \left[\min_{\hat{x}} \left(\sum_x p(x)d(x,\hat{x}) \right) \right] \end{aligned} \quad (\text{A.3})$$

because $R(D) = 0$

$\Rightarrow I(X;\hat{X}) = 0$

$\Rightarrow \sum_x p(\hat{x}) D_{\max} = D_{\max}.$

REFERENCES

- [1] C.-I. Chang, *Hyperspectral Imaging: Spectral Techniques for Detection and Classification*. Norwell, MA, USA: Kluwer, 2003.
- [2] C.-I. Chang, *Hyperspectral Data Processing: Signal Processing Algorithm Design and Analysis*. Hoboken, NJ, USA: Wiley, 2013.
- [3] H. Fu et al., "A novel band selection and spatial noise reduction method for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5535713.
- [4] H. Sun, J. Ren, H. Zhao, P. Yuen, and J. Tschannerl, "Novel Gumbel-Softmax trick enabled concrete autoencoder with entropy constraints for unsupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5506413.
- [5] 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.
- [6] R. J. McEliece, *Theory of Information and Coding*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [7] R. G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ, USA: Wiley, 1968.
- [8] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA, USA: Addison-Wesley, 1987.
- [9] C.-I. Chang, "Introduction: Two fundamental principles behind hyperspectral imaging," in *Advances in Hyperspectral Image Processing Techniques*, C.-I. Chang, Ed. Hoboken, NJ, USA: Wiley, 2022, ch. 1.
- [10] S. Sawant and M. Prabukumar, "A survey of band selection techniques for hyperspectral image classification," *J. Spectral Imag.*, vol. 9, pp. 1–18, Jun. 2020.
- [11] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.
- [12] C.-I. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.
- [13] Y. Du, C.-I. Chang, H. Ren, C.-C. Chang, and J. O. Jensen, "New hyperspectral discrimination measure for spectral characterisation," *Opt. Eng.*, vol. 43, no. 8, pp. 1777–1786, Aug. 2004.
- [14] S. Kullback, *Information Theory and Statistics*. New York, NY, USA: Wiley, 1968.
- [15] C.-I. Chang, K. Y. Ma, C.-C. Liang, Y.-M. Kuo, S. Chen, and S. Zhong, "Iterative random training sampling spectral spatial classification for hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3986–4007, 2020.
- [16] M. Song, X. Shang, Y. Wang, C. Yu, and C.-I. Chang, "Class information-based band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8394–8416, Nov. 2019.

- [17] C.-I Chang, L.-C. Lee, B. Xue, M. Song, and J. Chen, "Channel capacity approach to hyperspectral band subset selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4630–4644, Oct. 2017.
- [18] 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.
- [19] C.-I Chang, Y.-M. Kuo, S. Chen, C.-C. Liang, K. Y. Ma, and P. F. Hu, "Self-mutual information-based band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5979–5997, Jul. 2021.
- [20] R. M. Fano, *Transmission of Information*. Cambridge, MA, USA: MIT Press, 1961.
- [21] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Upper Saddle River, NJ, USA: Prentice-Hall, 1984.
- [22] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. New York, NY, USA: Kluwer Academics, 1992.
- [23] H. V. Poor, *An Introduction to Detection and Estimation Theory*, 2nd ed. New York, NY, USA: Springer, 1994.
- [24] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 460–473, Jul. 1972.
- [25] C. C. Wu, S. Chu, and C.-I Chang, "Sequential N-FINDR algorithm," *Proc. SPIE*, vol. 7086, pp. 106–117, Aug. 2008.
- [26] W. Xiong, C.-I Chang, C.-C. Wu, K. Kalpakakis, and H. M. Chen, "Fast algorithms to implement N-FINDR for hyperspectral endmember extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 545–564, Sep. 2011.
- [27] C.-I Chang, "Maximum simplex volume-based endmember extraction algorithms," U.S. Patent 8417 748 B2, Apr. 9, 2013.
- [28] C.-I Chang, "Comprehensive analysis of receiver operating characteristic (ROC) curves for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5541124.
- [29] C.-I Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5131–5153, Jun. 2021.
- [30] C.-I Chang, "Constrained energy minimization anomaly detection for hyperspectral imagery via dummy variable trick," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517119.
- [31] 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.
- [32] 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.
- [33] C.-I Chang, W. Xiong, and C.-H. Wen, "A theory of high-order statistics-based virtual dimensionality for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 188–208, Jan. 2014.
- [34] 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.
- [35] 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.
- [36] C. Yu, M. Song, and C.-I Chang, "Band subset selection for hyperspectral image classification," *Remote Sens.*, vol. 10, no. 1, p. 113, Jan. 2018, doi: 10.3390/rs10010113.
- [37] J. C. Harsanyi, W. Farrand, and C.-I Chang, "Detection of subpixel spectral signatures in hyperspectral image sequences," in *Proc. Amer. Soc. Photogram. Remote Sens. (ASPRS)*, vol. 1, 1994, pp. 236–247.
- [38] 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.
- [39] S. Zhong, C.-I Chang, and Y. Zhang, "Iterative edge preserving filtering approach to hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 90–94, Jan. 2019.
- [40] 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.
- [41] 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.

- [42] 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.
- [43] C.-I Chang, "Statistical detection theory approach to hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2057–2074, Apr. 2019.



Chein-I Chang (Life Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 1987.

He has been a Chang Jiang Scholar Chair Professor and the Director of Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Dalian Maritime University, Dalian, China, since 2016, and a Yushan Scholar Chair Professor with National Cheng Kung University, Tainan, Taiwan. In addition, he has been a Chair Professor with National Yang Ming Chiao Tung University, Hsinchu, Taiwan, since 2019. He is currently a Professor with the University of Maryland Baltimore County (UMBC), Baltimore, MD, USA. 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), *Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation* (Springer, 2017), and also edited three books, *Recent Advances in Hyperspectral Signal and Image Processing* (2006), *Hyperspectral Data Exploitation: Theory and Applications* (John Wiley & Sons, 2007), and *Advances in Hyperspectral Image Processing Techniques* (Wiley, 2023). In addition, he also 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.

Dr. Chang is a fellow of Society for Photo-Optical Instrumentation Engineers (SPIE). He is currently serving as an Associate Editor for *Remote Sensing* and *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*.



Yi-Mei Kuo received the B.S. degree in computer science and information engineering from the National University of Kaohsiung, Kaohsiung, Taiwan, in 2014, and the M.S. degree in electrical engineering from National Chung Hsing University, Taichung, Taiwan, in 2017. She is currently pursuing the Ph.D. degree in electrical engineering with the University of Maryland Baltimore County, Baltimore, MD, USA.

Her research interests include hyperspectral imaging, remote sensing, pattern recognition, and medical data analysis.



Peter Fuming Hu received the B.S. degree in electrical engineering from Shanghai University, Shanghai, China, in 1984, and the M.S. and Ph.D. degrees in computer science from the University of Maryland Baltimore County (UMBC), Baltimore, MD, USA, in 1992 and 2013, respectively.

He has been with the School of Medicine, UMBC, since 1992, where he is currently a Professor with the Department of Anesthesiology, Surgery, and Epidemiology. He also holds the Adjunct Professor with UMBC. In addition, he is the Chief

Technologist with the Shock Trauma Center, UMBC. He has more than 25 years' experience in clinical research with the School of Medicine and Shock Trauma Center, UMBC. During that time, he worked as a Principal Investigator (PI), co-PI, or co-investigator on over 50 successfully completed peer-reviewed grants funded by various U.S. government agencies. His research interests include applying big data machine learning (AI) methods for the prediction of trauma patient outcomes.

Dr. Hu is also a member of the Editorial Board of *Critical Care Medicine* which is the leading clinical journal dedicated to Intensive Care Unit (ICU) patient care.