

Iterative Target-Constrained Interference-Minimized Classifier for Hyperspectral Classification

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

Abstract—Despite the fact that many approaches to hyperspectral image classification are reported, specifically spectral–spatial based methods, this paper presents a rather different approach from a viewpoint of mixed pixel classification, referred to iterative target-constrained interference-minimization classifier (ITCIMC), which includes an iterative Gaussian filtered feedback process to capture the spatial contextual information so as to improve hyperspectral image classification for multiple classes at one-shot operation. In order to evaluate classification performance more effectively, new performance measures other than commonly used overall accuracy (OA) are introduced, particularly, precision rate (PR), misclassification (MC) rate which have been overlooked in hyperspectral image classification. To illustrate the differences among OA, MC rate, and PR, two concepts of *a priori* classification and *a posteriori* classification are also proposed from a statistical signal processing point of view. As shown by experiments, ITCIMC generally performs significantly better than the existing

Manuscript received May 8, 2017; revised September 2, 2017 and January 6, 2018; accepted January 31, 2018. Date of current version April 11, 2018. The work of C. Yu was supported by National Nature Science Foundation of Liaoning Province (20170540095). The work of M. Song was supported by National Nature Science Foundation of China under Grant 61601077 and Fundamental Research Funds for the Central Universities under Grant 3132017124. The work of Y. Wang was supported in part by the Fundamental Research Funds for the Central Universities under Grant 3132017080 and in part by the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences under Grant LSIT201707D. The work of C.-I Chang was supported by the Fundamental Research Funds for Central Universities under Grant 3132016331. (*Corresponding author: Bai Xue.*)

C. Yu and S. Li are with the Center of Hyperspectral Imaging in Remote Sensing, Information and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: yuchunyan1997@126.com; listen@dlmu.edu.cn).

B. Xue is with the Center of Hyperspectral Imaging in Remote Sensing, Information and Technology College, Dalian Maritime University, Dalian 116026, China, and also with Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore, MD 21250 USA (e-mail: baixuel@umbc.edu).

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

Y. Wang is with Information and Technology College, Dalian Maritime University, Dalian 116026, China and also with State Key Laboratory of Integrated Services Networks, Xi'an 710000, China (e-mail: wangyulei.wyl@163.com).

C.-I Chang is with the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliu 64002, Taiwan, Department of Computer Science and Information Management, Providence University, Taichung 09212, Taiwan, Center of hyperspectral Imaging in Remote Sensing, Information and Technology College, Dalian Maritime University, Dalian 116026, China, and also with Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, 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/JSTARS.2018.2802041

spectral–spatial hyperspectral image classification techniques in terms of PR and MC rate at the expense of slight loss of OA.

Index Terms—Hyperspectral image classification, iterative TCMIC (ITCIMC), misclassification (MC) rate, Otsu's method, overall accuracy (OA), precision rate (PR), target-constrained interference-minimization classifier (TCIMC), target-constrained interference-minimization filter (TCIMF).

I. INTRODUCTION

ONE of the major tasks in hyperspectral data exploitation is image classification [1]. However, when mixed pixel classification is used to perform hyperspectral image classification for class membership assignment it generally does not work effectively in [1, Ch. 16] and [2]. It is even worse when a hyperspectral image is heavily mixed due to low spatial resolution such as Purdue Indiana Indian Pine data [3], [4] in [1] and [2]. Since mixed pixel classification is generally pixel-based and estimates abundant fractions of material substances present in a single pixel vector, its classification is usually performed by thresholding its found abundance fractions to produce class-membership maps. When these abundance fractions are close or relatively low, two scenarios not encountered in classification will occur. One is that if the threshold is set too low, it may result in multiple-class assignment. On the other hand, if the threshold is set too high, it may be very likely that no class assignment can be made. Both cases are referred to as rejection class in pattern recognition such as optical character recognition and biometric recognition. Therefore, finding an appropriate threshold for mixed pixel classification is very challenging. As a consequence, on many occasions mixed pixel classification does not perform classification as well as pure-pixel classification such as maximum likelihood classification [5], [6] because the former does not take into spatial information into account, while the latter does. Accordingly, using mixed pixel classification to perform image classification is generally not preferable or dismissed if hyperspectral images contain important spatial contextual information attributed to classification. This leads to the development of many spectral–spatial techniques [7]–[28] which use joint spectral and spatial information to perform classification. Most notable are the spectral–spatial based methods which use support vector machine (SVM) for spectral classification and then follow up with spatial techniques to capture spatial contextual information to complete hyperspectral image classification. Unfortunately, SVM is a pure pixel-based

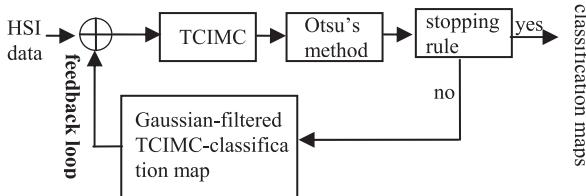


Fig. 1. Diagram of implementing ITCIMC.

spectral classifier, which directly produces classification maps and does not really take full advantage of abundance fractions of material substances present in a single pixel vector that may be crucial to actually determine class memberships. To address this issue, we reinvent a wheel by reviving mixed pixel classification in a novel way. Specifically, we look into a mixed pixel classifier, called target-constrained interference-minimized filter (TCIMF) [29] which is now available in the latest version of the most widely used popular ENVI software developed by analytical imaging and geophysics (AIG). It estimates the abundance fractions of multiple material substances simultaneously. If we consider each material substance to represent a spectral class, the material abundance fractions provided by TCIMF for data sample vectors can be used as the likelihood of data sample vectors to be assigned to their classes specified by particular material substances. With this interpretation, the resulting TCIMF can be further used as a classifier, referred to target-constrained interference-minimized classifier (TCIMC). However, this advantage is offset by the fact that TCIMC does not take into account spatial information in classification. To resolve this issue, a Gaussian filter is introduced into TCIMC in such a way that the TCIMC-mixed pixel classification map is further being processed by Gaussian filters to capture the spatial information. This Gaussian-filtered TCIMC-mixed pixel classification maps are then used as additional band images to be fed back to create a new hyperspectral image cube for TCIMC to be reprocessed again for next iteration. Such a process of repeatedly implementing TCIMC by feeding back Gaussian-filtered TCIMC-classification maps in an iterative manner is called iterative TCIMC (ITCIMC). At each iterative stage, ITCIMC also performs classification using a commonly used thresholding technique, Otsu's method [30] to produce classification maps for classes specified by each of material substances. These Otsu-thresholded TCIMC-classification a stopping rule maps can be further used to provide stopping rule to automatically terminate ITCIMC. Fig. 1 depicts a diagram of implementing ITCIMC.

In addition to the new development of ITCIMC, this paper also introduces several new performance measures to better evaluate the classification performance. Over the past years OA, AA or kappa coefficient are major classification criteria to measure the classification performance. This is mainly due to the fact that the classification is performed only based on classes of interest while discarding the effect of the background (BKG) class. However, such practice may only be valid when there is prior knowledge about the BKG class to allow users to remove the BKG from consideration. In order to address the BKG class issue more effectively, three specific new performance

classification measures are also introduced other than OA. One is misclassification (MC) rate which calculates the rate of misclassifying data samples in a particular class into other classes. Its role is very similar to false alarm probability P_F in signal detection theory [31]. Another is accuracy rate (AR) which extends OA to include the BKG class as a single separate class for classification. A third one is an introduction of new concept, called precision rate (PR) developed to evaluate the effectiveness of a classifier which cannot be measured by OA or AR. Like AR this new measure, PR also includes the BKG class for classification. Nevertheless, PR and AR are two separate concepts and one does not imply another. In order to illustrate the differences among OA, MC rate, AR, and PR for hyperspectral image classification, two concepts of *a priori* and *a posteriori* concepts from a statistical signal processing point of view are developed. When OA is used as a criterion for classification, the resulting classification is called *a priori* classification as opposed to *a posteriori* classification which uses PR as classification measure. As demonstrated by experiments such as *a priori* classification by OA and *a posteriori* classification by PR provide quite different insights into the classification, which have been overlooked in hyperspectral image classification in the past.

II. HYPERSPECTRAL MIXED PIXEL CLASSIFICATION

Assume that a hyperspectral image is represented by a collection of image pixel vectors, denoted by $\{\mathbf{r}_i\}_{i=1}^N$ where $\mathbf{r}_i = (r_{i1}, r_{i2}, \dots, r_{iL})^T$ for $1 \leq i \leq N$ is an L -dimensional vector, N is the total number of pixels in the image and L is the total number of spectral bands.

A. Linearly Constrained Minimum Variance

Suppose that $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_p$ are p specific target signatures of interest, each of which specifies a particular target class. We can now form a target class signature matrix, denoted by $\mathbf{T} = [\mathbf{t}_1 \mathbf{t}_2 \dots \mathbf{t}_p]$. The goal is to design a finite impulse response (FIR) linear filter specified by an L -dimensional weighting vector $\mathbf{w} = (w_1, w_2, \dots, w_L)^T$ that minimizes the filter output energy subject to the following constraint:

$$\mathbf{T}^T \mathbf{w} = \mathbf{c} \text{ where } \mathbf{t}_j^T \mathbf{w} = \sum_{l=1}^L w_l t_{jl} = c_j \text{ for } 1 \leq j \leq p \quad (1)$$

where $\mathbf{c} = (c_1, c_2, \dots, c_p)^T$ is a constraint vector.

Now, let y_i be the output of the designed FIR filter with the input \mathbf{r}_i . Then, y_i can be expressed by

$$y_i = \sum_{l=1}^L w_l r_{il} = \mathbf{w}^T \mathbf{r}_i = \mathbf{r}_i^T \mathbf{w}. \quad (2)$$

According to the LCMV beamformer [32], we can design an LCMV-based target detector using (1) which minimizes the

following average energy of the filter outputs $\{y_1, y_2, \dots, y_N\}$:

$$\begin{aligned} (1/N) \sum_{i=1}^N y_i^2 &= (1/N) \sum_{i=1}^N (\mathbf{r}_i^T \mathbf{w})^T (\mathbf{r}_i \mathbf{w}) \\ &= \mathbf{w}^T \left((1/N) \sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^T \right) \mathbf{w} = \mathbf{w}^T \mathbf{R}_{L \times L} \mathbf{w} \end{aligned} \quad (3)$$

subject to the constrained (1), specifically,

$$\min_{\mathbf{w}} \{\mathbf{w}^T \mathbf{R}_{L \times L} \mathbf{w}\} \text{ subject to } \mathbf{T}^T \mathbf{w} = \mathbf{c} \quad (4)$$

where $\mathbf{R}_{L \times L} = (1/N) \sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^T$ is the global autocorrelation sample matrix of the hyperspectral image. The solution to (4) is obtained in [33] by

$$\mathbf{w}^{\text{LCMV}} = \mathbf{R}_{L \times L}^{-1} \mathbf{T} (\mathbf{T}^T \mathbf{R}_{L \times L}^{-1} \mathbf{T})^{-1} \mathbf{c} \quad (5)$$

which gives rise to an LCMV target detector $\delta^{\text{LCMV}}(\mathbf{r})$ for each hyperspectral image pixel vector \mathbf{r} given by

$$\delta^{\text{LCMV}}(\mathbf{r}) = (\mathbf{w}^{\text{LCMV}})^T \mathbf{r}. \quad (6)$$

Apparently, the LCMV detector specified by (6) is not a classifier. In order to make an LCMV detector and LCMV classifier, we introduce a set of constraint vectors, denoted by $\{\mathbf{c}_j\}_{j=1}^p$, where each $\mathbf{c}_j = (0, \dots, c_j, \dots, 0)_{1 \times p}^T$ is a p -dimensional column vector with “ c_j ” in the j th component and “0”s in all other components to specify a particular class C_j . Substituting $\{\mathbf{c}_j\}_{j=1}^p$ for the constraint \mathbf{c} in (4) for $1 \leq j \leq p$ yields the following constrained p -class ($\{C_j\}_{j=1}^p$) classification problem:

$$\min_{\mathbf{w}_j} \mathbf{w}_j^T \mathbf{R}_{L \times L} \mathbf{w}_j \text{ subject to } \mathbf{T}^T \mathbf{w}_j = \mathbf{c}_j \text{ for } 1 \leq j \leq p \quad (7)$$

where \mathbf{w}_j is used to classify class C_j . The optimal solution to (7), $\mathbf{w}_j^{\text{LCMV}}$ can be obtained in [32] and [33] by

$$\mathbf{w}_j^{\text{LCMV}} = \mathbf{R}_{L \times L}^{-1} \mathbf{T} (\mathbf{T}^T \mathbf{R}_{L \times L}^{-1} \mathbf{T})^{-1} \mathbf{c}_j. \quad (8)$$

A classifier that uses the set of $\{\mathbf{w}_j^{\text{LCMV}}\}_{j=1}^p$ specified by (8) to classify p classes using p class signatures in \mathbf{T} is called an LCMV classifier.

B. Target-Constrained Interference-Minimized Classifier

One drawback of the LCMV classifier is that if there is prior knowledge about undesired target signal sources, it does not take advantage of such knowledge by annihilating these target signal sources as OSP does in [34]. To remedy this shortcoming, an approach, called TCIMF was developed in [29] to detect multiple target signal sources while also eliminating a set of undesired signal sources. Its idea combines CEM and OSP by expanding the unity vector $\mathbf{1} = (1, 1, \dots, 1, 0, 0, \dots, 0)^T$ such that one component “1” is used to constrain a particular desired target signal source and one component “0” is used to annihilate the undesired signal source. By virtue of such constraint vector $(\mathbf{1}, \mathbf{0})^T$, TCIMF can enhance desired target detectability by simultaneously eliminating effects caused by undesired signatures as well as minimizing the interfering effects resulting from unknown signal sources such as BKG.

Now, we can extend TCIMF to TCIMC in a similar manner that LCMV detector is extended to LCMV classifier. TCIMC assumes that an image scene has two types of signal classes, \mathbf{D} (desired target classes) and \mathbf{U} , (undesired target classes) plus BKG in the data. The TCIMC implements a constraint vector that can be used to simultaneously constrain \mathbf{D} and \mathbf{U} in such a way that it can classify target signal classes specified by the desired class signatures in \mathbf{D} while eliminating the undesired signal classes specified by undesired class signatures in \mathbf{U} . As a result, LCMV classifier described in Section II-A can be considered as a special case of TCIMC where the constraint vector used by LCMV is only used to class target signal sources, but not to annihilate undesired classes specified by undesired class signatures in \mathbf{U} as proposed in TCIMC so as to enhance the target classification. In order for TCIMF to perform mixed pixel classification as TCIMC, we can augment the constrained target signature vector to a constrained class signature matrix as follows.

Let $\mathbf{D} = [\mathbf{d}_1 \mathbf{d}_2 \dots \mathbf{d}_p]$ and $\mathbf{U} = [\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_q]$ denote the desired class signature matrix and the undesired class signature matrix, respectively. A constraint vector replaces the target signature matrix \mathbf{T} in (4) with the desired-undesired class signature matrix $[\mathbf{DU}]$. In addition, the constraint vector \mathbf{c} in (4) is also replaced with the desired-undesired class signature constraint vector $\mathbf{c} = (\mathbf{1}_{p \times 1}^T, \mathbf{0}_{q \times 1}^T)$ as follows:

$$[\mathbf{DU}]^T \mathbf{w} = \begin{bmatrix} \mathbf{1}_{p \times 1} \\ \mathbf{0}_{q \times 1} \end{bmatrix} \quad (9)$$

where $\mathbf{1}_{p \times 1}$ is a $p \times 1$ column vector with ones in all components and $\mathbf{0}_{q \times 1}$ is a $q \times 1$ column vector with all zeros in its components. Using (9) as a constraint imposed on (1) yields the following linearly constrained optimization problem:

$$\min_{\mathbf{w}} \{\mathbf{w}^T \mathbf{R}_{L \times L} \mathbf{w}\} \text{ subject to } [\mathbf{DU}]^T \mathbf{w} = \begin{bmatrix} \mathbf{1}_{p \times 1} \\ \mathbf{0}_{q \times 1} \end{bmatrix} \quad (10)$$

with the optimal weight vector $\mathbf{w}^{\text{TCIMC}}$ given by

$$\mathbf{w}^{\text{TCIMC}} = \mathbf{R}_{L \times L}^{-1} [\mathbf{DU}] \left([\mathbf{DU}]^T \mathbf{R}_{L \times L}^{-1} [\mathbf{DU}] \right)^{-1} \begin{bmatrix} \mathbf{1}_{p \times 1} \\ \mathbf{0}_{q \times 1} \end{bmatrix}. \quad (11)$$

The classifier using the weight vector $\mathbf{w}^{\text{TCIMC}}$ specified by (11) is called TCIMC and can be implemented as

$$\delta^{\text{TCIMC}}(\mathbf{r}) = \mathbf{r}^T \mathbf{R}_{L \times L}^{-1} [\mathbf{DU}] \left([\mathbf{DU}]^T \mathbf{R}_{L \times L}^{-1} [\mathbf{DU}] \right)^{-1} \begin{bmatrix} \mathbf{1}_{p \times 1} \\ \mathbf{0}_{q \times 1} \end{bmatrix}. \quad (12)$$

In particular, if there are p classes needed to be classified by p desired class signatures, $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_p$, the sizes of the class signature matrix $\mathbf{T} = [\mathbf{DU}] = [\mathbf{d}_1 \mathbf{d}_2 \dots \mathbf{d}_p \mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_q]$, weight matrix, $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_p]$ and constrained matrix $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_p]$ are $L \times (p+q)$, $L \times p$, $(p+q) \times p$, respectively, where $\mathbf{c}_j = (0, \dots, 0, \underbrace{1}_{j}, 0, \dots, 0)^T$ is a $(p+q)$ -dimensional vector for $1 \leq j \leq p$. In this case, the constrained class signature matrix can be specified by

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & 0 \end{bmatrix}_{(p+q) \times p} \quad (13)$$

Using (13), (12) becomes

$$\delta^{\text{TCIMC}}(\mathbf{r}) = \mathbf{r}^T \mathbf{R}_{L \times L}^{-1} [\mathbf{DU}] \left([\mathbf{DU}]^T \mathbf{R}_{L \times L}^{-1} [\mathbf{DU}] \right)^{-1} \mathbf{C}. \quad (14)$$

The most important advantage provided by TCIMC is its combination of both strengths of constrained energy minimization (CEM) [32], [35], [36], and OSP [34] by suppressing BKG as CEM does and also annihilating undesired signatures as OSP does.

As a final comment, it is important to note that TCIMC only requires the prior knowledge of the desired class signatures \mathbf{D} and undesired class signatures \mathbf{U} which can be obtained in two different sources, one from database or spectral library or the other from class information such as training sample vectors. Technically speaking, TCIMC deals with class signatures not training sample vectors. For example, class signatures can be obtained by the class sample means or averaging training sample vectors from individual classes as will be demonstrated in the following experiments.

III. ITERATIVE TCIMC

This section presents an iterative version of TCIMC, to be called ITCIMC, which iteratively feeds back its Gaussian-filtered TCIMC-mixed pixel classification maps to be added to create new expanded sets of band images after each iteration. The feedback loop is continued on until a stopping rule is satisfied. Fig. 2 describes a graphic diagram of Fig. 1 which implements ITCIMC with a feedback loop updating the desired target signature matrix \mathbf{D} , undesired target signature matrix \mathbf{U} , and the sample correlation matrix \mathbf{R} from the Gaussian-filtered TCIMC-mixed pixel classification maps iteratively. A detailed step-by-step implementation of ITCIMC in Fig. 2 is given as follows.

IV. STOPPING RULE FOR ITCIMC

In order to develop a stopping rule for ITCIMC, we need a criterion that can tell when ITCIMC must be terminated. The Tanimoto index (TI) defined in [37] can be used for this purpose as follows:

$$\text{TI}^{(k)} = \frac{|S_k \cap S_{k-1}|}{|S_k \cup S_{k-1}|} \quad (15)$$

ITCIMC

-
1. *Initial condition:* Let $\Omega^{(0)}$ be the original band set, $\mathbf{D}^{(0)}$ be the original desired class signatures, and $\mathbf{U}^{(0)}$ be the original undesired class signatures. Let $k = 1$.
 2. Implement δ^{TCIMC} on $\Omega^{(k)}$ using $\mathbf{D}^{(k)}$ and $\mathbf{R}^{(k)}$ to produce $\mathbf{B}_k^{\text{TCIMC}}$ which is the classification map.
 3. Use a Gaussian filter to blur $|\mathbf{B}_k^{\text{TCIMC}}|$, where $|\mathbf{B}_k^{\text{TCIMC}}|$ is the absolute value of $\mathbf{B}_k^{\text{TCIMC}}$. The resulting image is denoted by Gaussian-filtered TCIMC-classification map $|\mathbf{GB}_k^{\text{TCIMC}}|$.
 4. Form a new hyperspectral image cube $\Omega^{(k)}$ by augmenting the hyperspectral image cube, $\Omega^{(k-1)}$ obtained at $k - 1$ iteration by adding the new $|\mathbf{GB}_k^{\text{TCIMC}}|$ obtained at the k th iteration, i.e., $\Omega^{(k)} = \Omega^{(k-1)} \cup |\mathbf{GB}_k^{\text{TCIMC}}|$ where the desired signature matrix $\mathbf{D}^{(k)}$ is updated by adding p new class sample means from the TCIMC-classified map $|\mathbf{GB}_k^{\text{TCIMC}}|$ and $\mathbf{U}^{(k)}$ is also obtained similarly from $|\mathbf{GB}_k^{\text{TCIMC}}|$.
 5. Check if $\mathbf{B}_k^{\text{TCIMC}}$ satisfies a given stopping rule to be discussed in Section IV, go to step 7. Otherwise, continue.
 6. Form $\Omega^{(k+1)} = \Omega^{(k)} \cup \{ |\mathbf{GB}_k^{\text{TCIMC}}| \}$. Let $k \leftarrow k + 1$ and go to step 2.
 7. ITCIMC is terminated and Otsu's method is applied to threshold $|\mathbf{B}_k^{\text{TCIMC}}|$ into multiclass classification map.
-

where $|S|$ is size of a set S , S_k , and S_{k-1} are the k th thresholded binary image of the k th TCIMC detection map, $|\mathbf{B}_k^{\text{TCIMC}}|$ and $k - 1$ st thresholded binary image of the $k - 1$ st TCIMC classification map $|\mathbf{B}_{k-1}^{\text{TCIMC}}|$. Fig. 3 describes a flowchart of a stopping rule using TI as a measure.

V. PERFORMANCE MEASURES FOR CLASSIFICATION

How to evaluate classification performance is crucial to justify if a classifier is effective. In the past, OA, AA, or kappa coefficient have been widely used for hyperspectral image classification. Unfortunately, these criteria only tell half of a story. This section provides another half of a story to complete an entire picture of how a classifier can be evaluated from two view points of statistical signal processing, *a priori* and *a posteriori* concepts.

A. Accuracy and Precision Rates

In general, a classifier can be evaluated by two criteria. One is that for a given set of known data samples how well a classifier performs in sense of their class accuracy, known as AR. This can be considered as *a priori* classification performance since the classification is evaluated based on the dataset that is known *a priori* and provided by the known ground truth. The other is that for a given set of data samples that are already classified by a classifier how effectively this classifier performs in the sense of precision, known as PR. This type of classification is referred

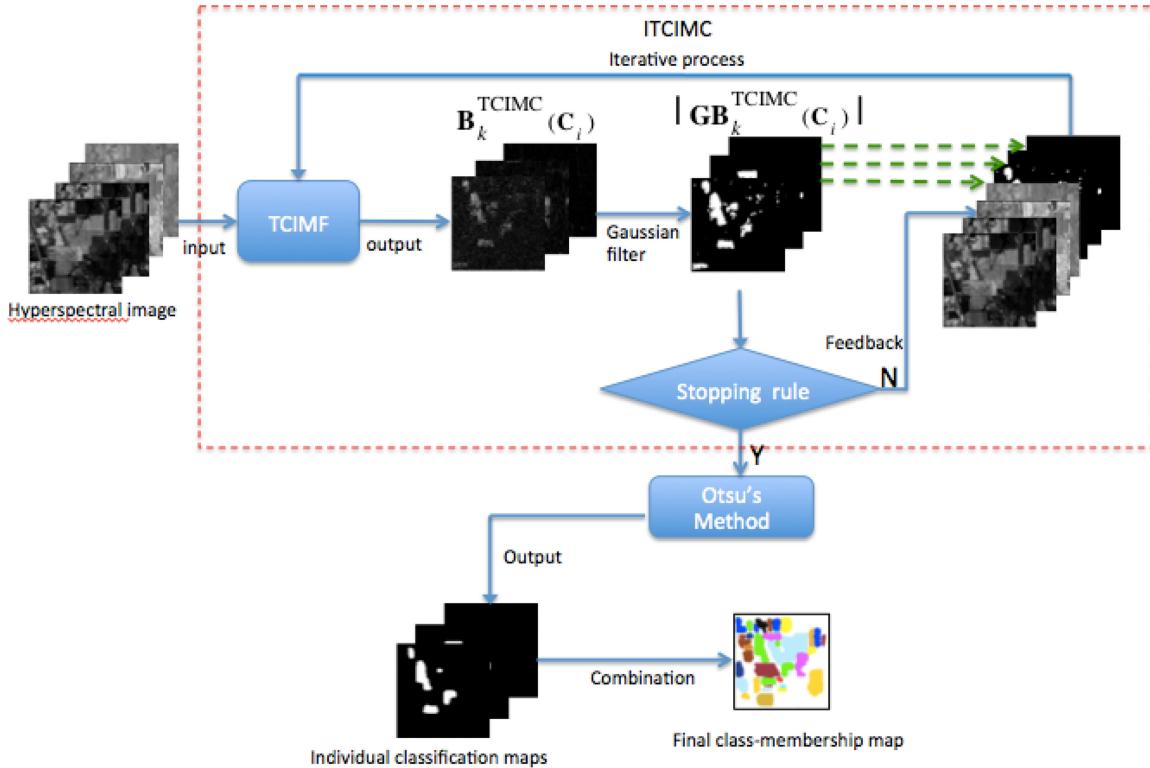


Fig. 2. Graphic implementation of ITCIMC in Fig. 1.

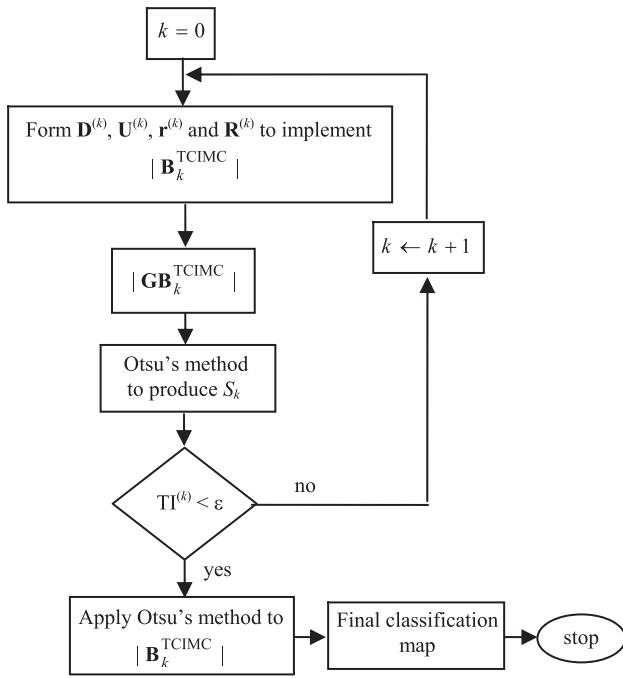


Fig. 3. Flowchart of a stopping rule.

to as *a posteriori* classification because the classification performance is evaluated based on classified data samples which are produced by a classifier *a posteriori*.

More specifically, let $\{C_j\}_{j=1}^p$ be p classes of data samples of interest provided by the ground truth, in which case $\{C_j\}_{j=1}^p$

can be considered as *a priori* classification information. Also, let $\{\hat{C}_j\}_{j=1}^p$ be the p classes of data samples that are being classified by a classifier, in which case $\{\hat{C}_j\}_{j=1}^p$ can be considered as *a posteriori* classification information produced by a classifier where the hat “ $\hat{\cdot}$ ” indicates “*a posteriori*.” By virtue of these two *a priori* and *a posteriori* classes, $\{C_j\}_{j=1}^p$ and $\{\hat{C}_j\}_{j=1}^p$, we can define the following measures that are derived from a classification point of view. Specifically, we use the subscript of n , “ ij ” to present that the “ i ” indicates for the i th classified class provided by *a posteriori* class information produced by a classifier and the “ j ” indicates the j th known class provided by the ground truth as *a priori* classification.

- 1) p = the number of classes.
- 2) n_j = the number of data samples in the j th class, C_j ;
- 3) \hat{n}_j = number of dample samples classified into \hat{C}_j .
- 4) n_{ij} = the number of data samples in the j th class, C_j to be classified into the i th class \hat{C}_i .
- 5) \hat{n}_{ij} = the number of data samples in the i th class, \hat{C}_i which are supposed to in the j th class, C_j .
- 6) n_{jj} = the number of data samples in the j th class correctly classified into the j th class.
- 7) $\hat{n}_i = \sum_{j=1}^p \hat{n}_{ij}$.
- 8) $n_j = \sum_{i=1}^p n_{ij}$.
- 9) N = total number of data samples, $N = \sum_{j=1}^p n_j = \sum_{i=1}^p \hat{n}_i$.
- 10) $p(C_j) =$ prior probability of $C_j = \frac{n_j}{N}$.
- 11) \tilde{C}_j = the set of data samples in the j th class but misclassified into other classes.

- 12) C_{ij} = the set of data samples in the j th class, C_j but misclassified into the i th class C_i .

$$\begin{aligned} P_C(C_j) &= \text{correct rate of classifying the } j\text{th class } C_j \\ &= p(C_{jj}|C_j) = \frac{n_{jj}}{n_j}. \end{aligned} \quad (16)$$

P_C = correct classification rate

$$= \sum_{j=1}^p \left(\frac{n_j}{N} \right) \frac{n_{jj}}{n_j} = \frac{1}{N} \sum_{j=1}^p n_{jj}. \quad (17)$$

$$P_{MC}(C_j) = p(\tilde{C}_j)$$

$$\begin{aligned} &= \text{rate of misclassifying the } j\text{th class, } C_j \\ &= \sum_{i=1, i \neq j}^p p(C_{ij}|C_j) = \frac{\sum_{i=1, i \neq j}^p n_{ij}}{\sum_{i=1}^p n_{ij}} \\ &= 1 - p(C_{jj}|C_j) = 1 - P_C(C_j). \end{aligned} \quad (18)$$

P_{MC} = misclassification rate

$$= \sum_{j=1}^p p(C_j) P_{MC}(C_j). \quad (19)$$

For a given i th classified class \hat{C}_i , we can define an *a posteriori* measure, PR denoted by $P_{\text{precision}}(\hat{C}_i)$ as

$$P_{\text{precision}}(\hat{C}_i) = p(\{C_j\}_{j=1}^p | \hat{C}_i) = \frac{\hat{n}_{ii}}{\sum_{j=1}^p \hat{n}_{ij}} \quad (20)$$

$$\begin{aligned} P_{\text{precision}}(\{C_j\}_{j=1}^p | \{\hat{C}_i\}_{i=1}^p) &= \sum_{i=1}^p p(\hat{C}_i) p_{\text{precision}}(\hat{C}_i) \\ &= \frac{1}{N} \sum_{i=1}^p \hat{n}_i p_{\text{precision}}(\hat{C}_i). \end{aligned} \quad (21)$$

On the other hand, for a given ground truth class C_j , we can define *a priori* measure, AR denoted by $P_A(C_j)$ as

$$\begin{aligned} P_A(C_j) &= \text{accuracy of classifying the } j\text{th class } C_j \\ &= p\left(\{\hat{C}_i\}_{i=1}^p | C_j\right) = \frac{n_{jj}}{\sum_{i=1}^p n_{ij}} = \frac{n_{jj}}{n_j} = P_C(C_j) \end{aligned} \quad (22)$$

$$\begin{aligned} P_{OA} &= P\left(\{\hat{C}_i\}_{i=1}^p | \{C_j\}_{j=1}^p\right) = \sum_{j=1}^p p(C_j) p\left(\{\hat{C}_i\}_{i=1}^p | C_j\right) \\ &= \frac{1}{N} \sum_{j=1}^p n_j \left(\frac{n_{jj}}{\sum_{i=1}^p n_{ij}} \right) = \sum_{j=1}^p p(C_j) P_A(C_j) = P_A. \end{aligned} \quad (23)$$

It should be noted that according to (23), P_{OA} and P_A are the same measure if all the classes $\{C_j\}_{j=1}^p$ are considered as classes of interest. Therefore, the classification is performed according to P_{MC} in (17)–(19), PA in (22), and P_{OA} (23) is called *a priori classification* because these criteria assume that the ground truth class knowledge is given, i.e., $\{C_j\}_{j=1}^p$. On the

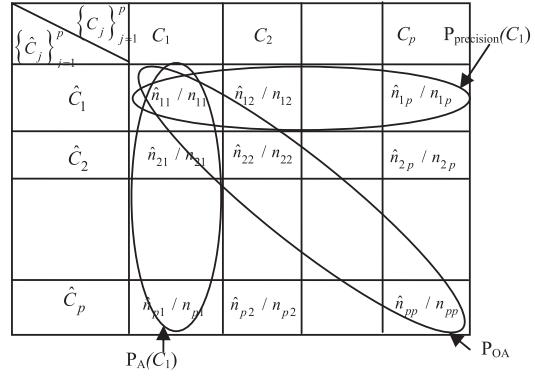


Fig. 4. AR, PR, and OA calculated from p -class confusion matrix.

other hand, the classification is performed based on $P_{\text{precision}}$ in (20), (21) is called *a posteriori* classification because $P_{\text{precision}}$ makes use of classes $\{\hat{C}_i\}_{i=1}^p$ that are already classified by a classifier.

B. Accuracy and Precision Rates Calculated From Confusion Matrix

A common criterion to measure the classification is to use a confusion matrix shown in Fig. 4 where the element n_{ij} in the i th row and the j th column represents the number of data samples in the j th class to be classified into the i th class.

For example, the first row represents $P_{\text{precision}}$ of a given classified class, \hat{C}_1 obtained by (20) using $\{n_{1i}\}_{i=1}^p$ to calculate $P_{\text{precision}}(\hat{C}_1) = p(\{C_i\}_{i=1}^p | \hat{C}_1) = \frac{\hat{n}_{11}}{\sum_{i=1}^p \hat{n}_{1i}}$. Similarly, the first column represents the AR of a given known class C_1 from the ground truth obtained by (22) using $\{n_{j1}\}_{j=1}^p$ to calculate $P_A(C_1) = p(\{\hat{C}_i\}_{i=1}^p | C_1) = \frac{n_{11}}{\sum_{j=1}^p n_{j1}}$. Finally, the diagonal elements represents P_{OA} calculated by (23) using $\{n_{ii}\}_{i=1}^p$.

It should be noted that *a priori* classification is a classification based on the given ground truth, that is, all labeled data samples are provided *a priori* and the classification is evaluated according to their memberships. By contrast, *a posteriori* classification is performed based on observations, which are classified data samples by a classifier and the classification performance is then evaluated in accordance with the ground truth. Therefore, there is a significant difference between *a priori* classification which yields OA and *a posteriori* classification which gives arise to PR.

C. Effect of BKG Class on Classification

In Sections V–A and V–B, the classes of interest make up the entire data samples. However, due to significant high spectral resolution many unknown data samples can be uncovered by a hyperspectral imaging sensor but cannot be labeled. In this case, when it comes to hyperspectral image classification, these unlabeled data samples are generally grouped into a BKG class which is considered as an uninteresting class. As a consequence, such BKG class is usually excluded from hyperspectral image classification. Unfortunately, on many occasions, the BKG class plays an integral part of the data, specifically in defense appli-

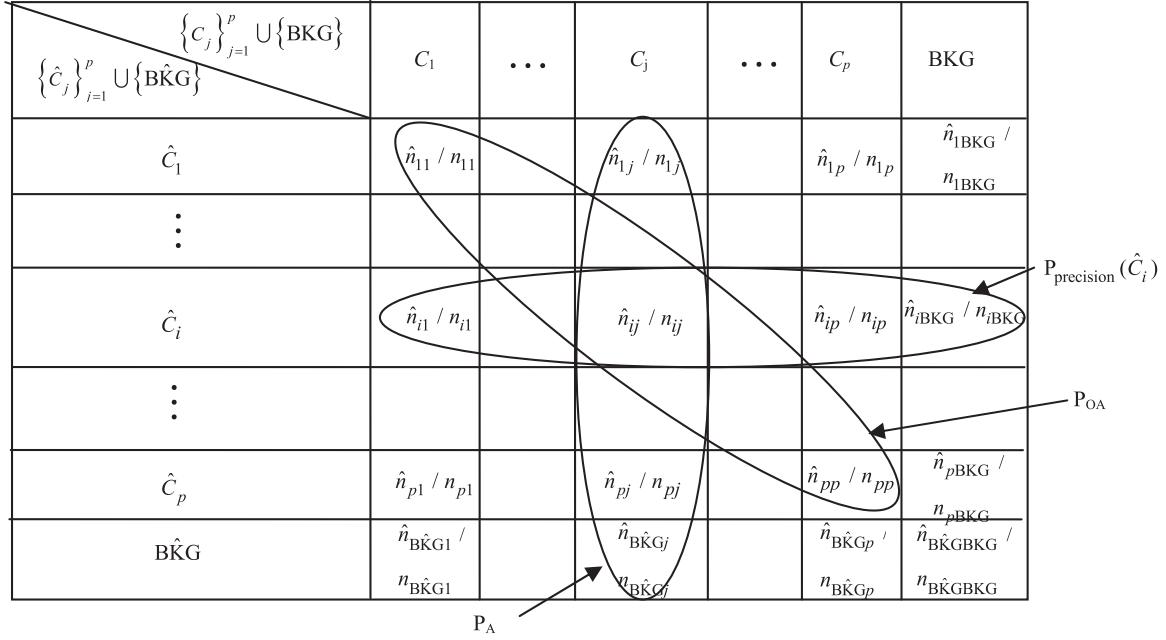


Fig. 5. P_A , $P_{\text{precision}}$, and P_{OA} calculated p -class confusion matrix along with BKG class to be considered as a single separate class.

cations where subtle BKG information provides crucial information. Discarding or ignoring the BKG class may result in a loss of significant information. Most interestingly, it is the BKG class that attributes to MC, i.e., P_{MC} in (18) and (19), an issue which has been overlooked in supervised hyperspectral image classification. Fig. 5 is a modified confusion matrix from Fig. 4 by separating the BKG class from classes of interest ($\{C_j\}_{j=1}^p$) to stand alone as a single class.

As should be noted in Fig. 5, P_{OA} is only defined on the classes of interest as many reports did in the literature where the BKG class was considered as a class of no interest and were not to be considered for classification. However, this is not true for P_A and $P_{\text{precision}}$ since they are designed to account for the effect of BKG class on other classes of interest including the BKG class. In other words, when the BKG class is removed from consideration in P_A , P_A is then reduced to P_{OA} . In particular, MC rate, P_{MC} in (18) and (19) and $P_{\text{precision}}$ in (20) and (21) were not considered in [18].

D. TI Calculated by Accuracy and PRs

As noted in ITCIMC, the stopping rule is determined by the TI defined by (15). By taking advantage of the concepts provided by $P_{\text{precision}}$ in (20) and P_A in (22) it turns out that TI can be actually calculated by

$$\text{TI} = \sum_{j=1}^p \frac{|C_j \cap \hat{C}_j|}{|C_j \cup \hat{C}_j|} p(C_j) \quad (24)$$

which makes a perfect sense. In other words, TI measures the discrepancy between *a priori* classes $\{C_j\}_{j=1}^p$, which are used to calculate P_A and *a posteriori* classes $\{\hat{C}_j\}_{j=1}^p$, which are used to calculate $P_{\text{precision}}$. As their overlapped class sample vectors $C_j \cap \hat{C}_j$ achieve a certain level, ITCIMC is terminated. This

implies that when the two rates P_A and $P_{\text{precision}}$ are sufficiently close, ITCIMC has completed its task and is then terminated.

VI. REAL HYPERSPECTRAL IMAGE EXPERIMENTS

Three real hyperspectral images were used for experiments, Purdue University's Indiana Indian Pines, Italy and Salinas, University of Pavia, each of which has its own unique feature characteristics worth being explored. According to the recent work [18], a comprehensive comparative analysis was conducted among most recently developed spectral-spatial techniques where the four EPF-based techniques, EPF-B-c, EPF-G-c, EPF-B-g, and EPF-G-g were shown to be best classification techniques with "B" and "G" used to specify bilateral filter and guided filter, respectively, and "g" and "c" indicate that the first principal component and color composite of three principal components are used as reference images [18]. Therefore, in the following experiments, the performance of ITCIMC will be evaluated in comparison with these four EPF-based techniques due to two main reasons. One is that these four techniques are available on website and we could reimplement them for comparison. Another is that these four techniques were compared to other existing spectral-spatial classification methods in [18] to show their superiority.

In the following experiments, ITCIMC was implemented in four different versions depending upon how to make use of the knowledge of undesired class signatures in \mathbf{U}

- 1) *ITCIMC-1*: \mathbf{U} is specified by the BKG mean;
- 2) *ITCIMC-2*: \mathbf{U} is specified by the local sample mean of an area extracted from the upper-right corner region with 30 sample vectors;
- 3) *ITCIMC-3*: \mathbf{U} is specified by a single target found in BKG by automatic target generation process (ATGP) [38], $\mathbf{U} = [\mathbf{t}^{\text{ATGP}}]$;

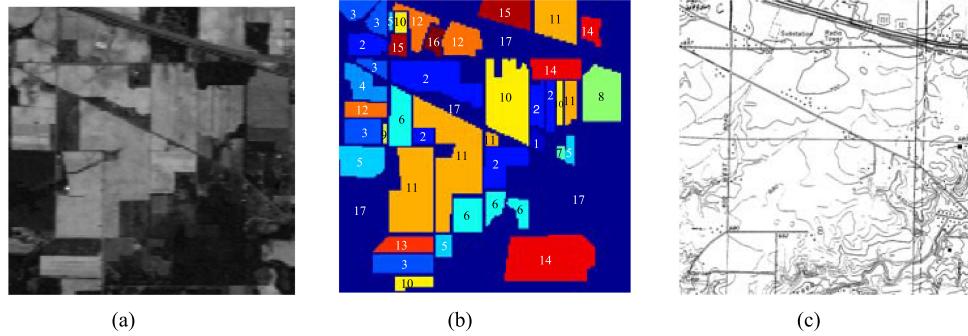


Fig. 6. AVIRIS image scene: Purdue Indiana Indian Pines test site. (a) Band 186 (2162.56 nm). (b) Ground truth. (c) USGS Quadrangle map of the test site.

TABLE I
CLASS LABELS OF 17 CLASSES

class 1 (54)	Alfalfa	class 7 (26)	grass/pasture-mowed	class 13 (212)	wheat
class 2 (1434)	corn-notill	class 8 (489)	hay-windrowed	class 14 (1294)	woods
class 3 (834)	corn-min	class 9 (20)	oats	class 15 (380)	bldg-grass green-drives
class 4 (234)	corn	class 10 (968)	soybeans-notill	class 16 (95)	stone-steel towers
class 5 (497)	grass/pasture	class 11 (2468)	soybeans-min	class 17 (10659)	BKG
class 6 (747)	grass/trees	class 12 (614)	soybeans-clean		

TABLE II
SPECIFICATIONS OF PARAMETERS USED BY ITCIMC FOR PURDUE INDIAN PINES SCENE

D	16 class sample means
Class number	1–16
σ used in Gaussian filter	0.5 with window size 5×5
Thresholding method	Otsu
Stopping threshold (TI)	0.99

- 4) *ITCIMC-4*: $\mathbf{U} = \emptyset$, i.e., no undesired class signature is used.

A. Purdue Indiana Indian Pines

A real image to be used for experiments is a well-known Airborne Visible Infrared Imaging Spectrometer (AVIRIS) image scene, Purdue Indiana Indian Pine test site is shown in Fig. 6(a). Table I also tabulates all the specific types of 16 classes. It has size of 145×145 pixel vectors taken from an area of mixed agriculture and forestry in Northwestern Indiana, USA, with details of band and wavelength is given in caption. The dataset used for experiments is available at website [3]. It was recorded in June 1992 with 220 bands including water absorption bands (bands 104–108 and 150–163, 220) which were removed in [20]. While the complete ground truth for this scene is unknown, it is believed that there are no endmembers present in the scene since the pixels in this scene are heavily mixed.

Table II tabulates specifications of parameters and various methods used by ITCIMC where \mathbf{D} in (9) consists of 16 class sample means. Fig. 7 shows 16 class maps produced by four versions of ITCIMC, ITCIMC-1 with \mathbf{U} specified by BKG mean, ITCIMC-2 with \mathbf{U} specified by local mean, ITCIMC-3 with \mathbf{U} specified by one ATGP-found target in BKG, ITCIMC-4 with $\mathbf{U} = \emptyset$ using Otsu's thresholding method. According to visual

inspection based on the ground truth map in Fig. 6(b), the best results in Fig. 7 seemed to be those produced by ITCIMC-3 and ITCIMC-4. As a matter of fact, the results in Fig. 7 suggested that the less BKG information was used, the better the classification was. Tables III and IV further tabulate respective results produced by EPF-B-g, EPF-B-c, EPF-G-g, and EPF-G-c and ITCIMC-1, ITCIMC-2, ITCIMC-3, ITCIMC-4 using five performance measures $P_C(C_i)$ and P_C in (16) and (17), P_{MC} in (19), $P_{precision}(\hat{C}_i)$ and $P_{precision}$ in (20) and (21), P_{OA} in (23), $P_A(C_j)$ in (22) and P_A in (23) in terms of percentage (%) where the $P_C\%$, $P_{MC}\%$, and $P_{precision}^{C_i}$ are shown in the three columns under each method and P_{OA} and P_A are shown in the last two rows with the bold-faced values indicating the best results where P_A (BKG) specifically emphasize the fact that BKG is considered as a single and separate class for classification.

Interestingly, the quantitative study in Table IV further showed that the best result was the one produced by ITCIMC-4 with $\mathbf{U} = \emptyset$ which did not use any BKG information in \mathbf{U} . This is due to the fact that some BKG information might contain needed class information which has been eliminated by \mathbf{U} . In order to see how the spatial information provided by Gaussian-filtered TCIMC improves classification ITCIMC-4 is selected for illustration since it was the best version of TCIMC. Fig. 8 shows progressive color classification maps produced by ITCIMC-4, where the number of iterations indicated underneath each figure. As we can see, when no spatial information is included in TCIMC the classification result was poor as shown at the first iteration in Fig. 8. As the Gaussian-filtered spatial information was gradually added to the processed hyperspectral image cubes via feedback loops during iterative processes, the classification results began to improve until it satisfied the stopping rule at the 12th iteration. In particular, initially there were many scattering classified data sample vectors in the first iteration and then the number of scattering data sample vectors

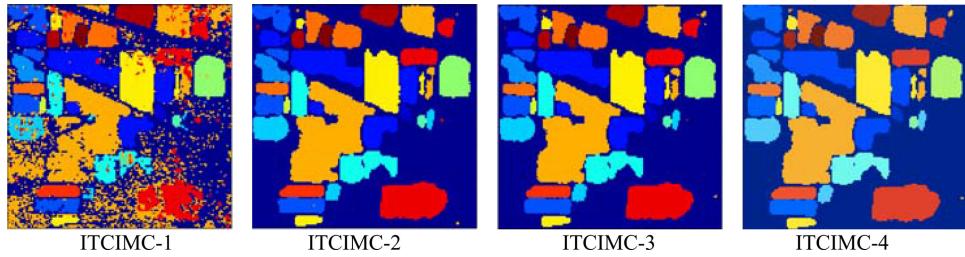


Fig. 7. Classification maps by ITCIMC of Purdue data.

TABLE III
POA AND PA (BKG) PRODUCED BY EPF-B-c, EPF-B-g, EPF-G-c, AND EPF-G-G FOR PURDUE'S INDIAN PINES

Class	EPF-B-g			EPF-B-c			EPF-G-g			EPF-G-c		
	PA%	PMC%	Pprecision%	PA%	PMC%	Pprecision%	PA%	PMC%	Pprecision%	PA%	PMC%	Pprecision%
1	100.00	0.16	57.50	97.83	0.16	57.69	97.83	0.15	58.44	100.00	0.14	60.53
2	85.01	3.11	67.03	84.94	3.12	66.91	85.22	3.11	67.09	84.45	3.27	65.76
3	93.13	1.33	74.40	94.10	1.41	73.54	92.41	1.41	73.12	92.41	1.34	74.11
4	99.16	0.62	64.74	99.16	0.62	64.56	99.16	0.75	60.26	99.16	0.76	59.95
5	93.58	3.46	39.65	93.37	3.20	41.45	94.00	3.11	42.23	93.58	2.93	43.55
6	100.00	3.88	49.06	99.73	3.89	48.92	99.73	4.25	46.82	100.00	3.94	48.70
7	96.43	0.08	62.79	96.43	0.07	65.85	96.43	0.10	56.25	96.43	0.10	56.25
8	100.00	1.01	69.88	100.00	0.99	70.40	100.00	0.92	71.77	100.00	0.94	71.34
9	95.00	0.14	38.78	100.00	0.13	42.55	100.00	0.08	55.56	95.00	0.05	63.33
10	82.30	2.41	62.70	82.82	2.50	61.97	81.79	2.45	62.21	82.51	2.44	62.51
11	95.23	3.95	76.71	95.64	3.99	76.63	94.46	3.84	77.02	94.70	3.76	77.45
12	98.82	1.75	62.47	98.65	1.68	63.38	98.48	1.81	61.67	98.65	1.85	61.13
13	99.02	0.25	79.61	99.02	0.29	77.19	99.51	0.30	76.69	99.51	0.36	73.12
14	98.26	15.48	31.91	98.50	15.80	31.58	98.10	15.46	31.90	98.50	15.48	31.97
15	96.63	26.91	7.85	96.89	26.59	7.94	94.30	26.91	7.67	99.48	27.28	7.99
16	96.77	0.35	54.88	98.92	0.32	57.86	100.00	0.41	52.25	100.00	0.40	52.54
POA		94.83		95.33				94.99			94.60	
PA (BKG)		46.23		46.47				46.31			46.12	

was significantly reduced from first to second and then third iterations. When it reached the sixth iteration and afterward, only a few scattering data sample vectors were remained. This phenomenon demonstrated why mixed pixel classification generally did not work effectively because those scattering data sample vectors were result of mixed pixel classification on a single pixel basis. If we included in spatial information provided by Gaussian filters these scattering data sample vectors would disappear and correctly classified into classes to which they were supposed to belong.

According to the ground truth in Table I, each of corn, soybean, and grass has three different types. In this case, 16 classes are divided into four subclasses, corn-like classes 2–4, grass-like classes 5–7, and soybean-like classes 10–12, and other remaining classes plus BKG class 17. Fig. 9 shows the spectral profiles of these four classes plotted by 16 class means where the spectral signatures of three corn classes are very close to each other and so are the three soybean classes. As expected, classifying these classes will be very challenging.

According to Tables III and IV, the best POA was the one produced by EPF-B-c, while the best accuracy was produced by ITCIMC using Otsu's thresholding. As for the other three performance measures, P_C , P_{MC} , and $P_{precision}$, EPF-G-c generally produced the best P_C values for 8 out of 16 classes,

and ITCIMC produced best values of P_{MC} and $P_{precision}$ almost across board except class 9 whose best values of P_{MC} and $P_{precision}$ were produced by EPF-G-c. The most interesting finding is P_A defined in (23) which includes BKG classification into OA. As shown in the last row of Table III, ITCIMC produced nearly twice better $P_{accuracy}$ than four EPF-based methods which did not include BKG classification in their results in [18]. It is known that OA has been the major performance measure to be used to evaluate hyperspectral image classification [18], [20]. Unfortunately, OA only tells half of a story which does not account for BKG classification but rather calculates the correct classification of test data samples where both training samples and testing samples come from the same class. That is, OA does not include MC of data samples from other classes. To be more specific, if we consider multiclass confusion matrix, OA only calculates the correct classification rates for all classes along the diagonal line without BKG, while discarding all MC rates off diagonal line. Such MC rates are exactly the other half of the story must be told in accuracy calculation. However, how to address MC rates is challenging because there are $\binom{c}{2} = \frac{c(c-1)}{2}$ combinations of MC rates, where c is the total number of classes, which can be referred to as one against one strategy. An alternative strategy is one against rest which considers a class of interest as the desired class and all the rest of classes including

TABLE IV
 P_{OA} AND P_A (BKG) PRODUCED BY FOUR DIFFERENT VERSIONS OF ITCIMC FOR PURDUE'S INDIAN PINES

Class	ITCIMC-1			ITCIMC-2			ITCIMC-3			ITCIMC-4		
	$P_A\%$	$P_{MC}\%$	$P_{precision}\%$	$P_A\%$	$P_{MC}\%$	$P_{precision}\%$	$P_A\%$	$P_{MC}\%$	$P_{precision}\%$	$P_A\%$	$P_{MC}\%$	$P_{precision}\%$
1	95.65	0.07	74.58	95.65	0.07	75.86	95.65	0.06	77.19	95.65	0.06	77.19
2	90.39	0.70	90.20	94.75	0.55	92.67	95.73	0.63	91.47	96.01	0.61	91.96
3	86.47	0.55	86.47	95.43	0.39	91.04	96.51	0.47	89.40	96.99	0.46	89.74
4	91.87	0.02	98.26	99.16	0.08	93.25	98.73	0.10	91.76	98.73	0.09	92.86
5	60.59	0.19	88.92	89.23	0.15	93.29	89.44	0.17	92.70	89.44	0.16	92.90
6	77.30	0.78	78.59	96.99	0.39	89.96	96.99	0.40	89.85	97.12	0.39	89.97
7	100	0.04	77.78	100	0.05	73.68	100	0.04	75.68	100	0.04	75.68
8	98.14	0.10	95.56	98.95	0.09	96.14	98.74	0.10	95.93	98.74	0.10	95.74
9	95.24	0.08	52.63	100	0.08	54.05	100	0.08	54.05	100	0.09	52.63
10	85.10	0.33	92.77	90.23	0.39	91.74	93.21	0.39	92.07	93.93	0.40	91.85
11	85.67	21.23	38.63	93.65	0.79	93.99	94.62	0.92	93.18	94.70	0.78	94.17
12	80.66	0.16	94.15	95.28	0.30	90.11	94.95	0.31	89.81	95.45	0.33	89.43
13	93.78	0.04	95.61	98.54	0.07	93.52	98.54	0.07	93.52	98.54	0.07	93.52
14	59.47	2.55	62.52	92.96	0.24	96.16	92.96	0.21	96.63	93.52	0.19	96.89
15	83.26	0.05	97.10	90.67	0.08	95.37	90.67	0.06	96.42	90.67	0.06	96.69
16	80.70	0.07	85.98	98.92	0.09	83.64	98.92	0.08	85.19	98.92	0.07	85.98
P_{OA}	81.43			94.41			94.82			95.09		
P_A (BKG)	74.53			93.59			93.68			93.96		

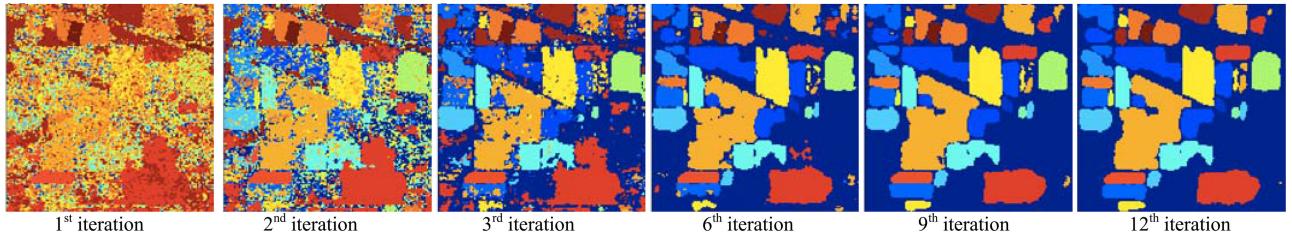


Fig. 8. Progressive color classification maps by ITCIMC-4 for Purdue's Indian Pines scene.

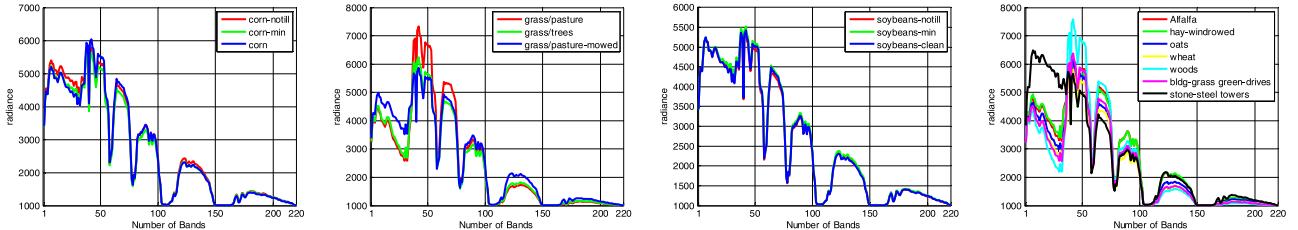


Fig. 9. Spectral profiles of 16 class sample means in Purdue Indian Pines.

BKG class as the background (BKG). By doing so, a multiclass confusion matrix is further simplified to a binary confusion matrix in which case a standard binary hypothesis testing problem can be applied. With this interpretation, we can include the BKG class, class 17 as a new class, a case not considered in [18]. In other words, we can consider the class of interest as signal class specified by the alternative hypothesis H_1 and the BKG as the null hypothesis H_0 . Then, P_C is calculated by the correct classification rate of samples from the signal class whereas P_{MC} is the false alarm rate defined as the rate of misclassifying the samples from BKG into the signal class. Such misclassified data samples are referred to falsely alarmed or falsely classified data samples. The measures of $P_C(C_i)$ and P_C in (16) and (17), P_{MC} in (19), $P_{precision}(\hat{C}_i)$ and $P_{precision}$ in (20) and (21), P_{OA} in (23), $P_A(C_j)$ in (22), and P_A in (23)

are particularly designed to address this issue. Based on these five performance measures calculated in Tables III and IV, ITCIMC-4 was the best to produce best values of P_{MC} , $P_{precision}$, and P_A , while EPF-G-c and EPF-B-c were the best to produce best values of P_C and P_{OA} , respectively. Therefore, generally speaking, ITCIMC-4 is the best classification technique to deal with correct classification and MC issues.

B. Salinas Scene

The Salinas image shown in Fig. 10(a) was captured by the AVIRIS sensor over Salinas Valley, CA, USA, and with a spatial resolution of 3.7-m per pixel with spectral resolution of 10 nm.

The image has size of $512 \times 217 \times 224$. Fig. 10(b), (c) shows the color composite of the Salinas image and the corresponding

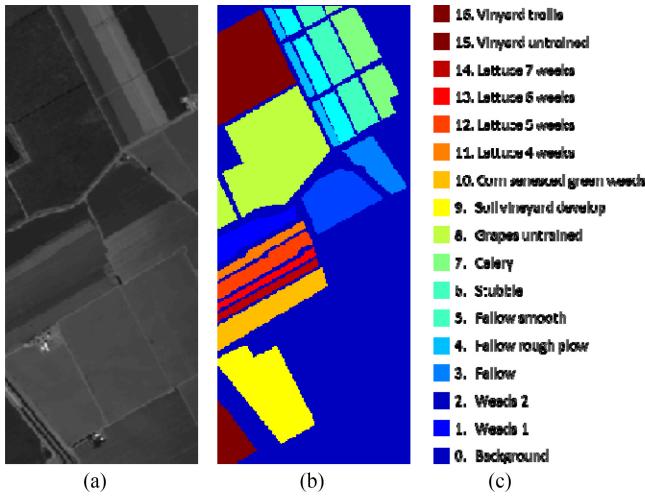


Fig. 10. Ground-truth of Salinas scene with 16 classes. (a) Salinas scene. (b) Color ground-truth image. (c) Ground truth class labels.

TABLE V
SPECIFICATIONS OF PARAMETERS USED BY ITCIMC FOR SALINAS SCENE

D	16 class sample means
Class number	1–16
σ used in Gaussian filter	0.5 with window size 5×5
Thresholding method	Otsu
Stopping threshold (TI)	0.99

ground truth class labels. The dataset used for experiments includes 20 water absorption bands, are 108–112, 154–167, and 224 which were also removed in [20]. Table V tabulates the specifications of parameters and various methods used by ITCIMC, where \mathbf{D} in (9) consists of 16 class sample means.

According to the ground truth in Fig. 10(c), 16 classes are divided into 5 subclasses, weeds classes 1–2, fallow classes 3–5, and lettuce classes 11–14, vineyard classes 15–16, and other remaining classes plus BKG class 0. Fig. 11 shows the spectral profiles of these 5 classes plotted by 16 class means.

Fig. 12 shows 16 class maps produced by 4 versions of ITCIMC, ITCIMC-1 with \mathbf{U} specified by BKG mean, ITCIMC-2 with \mathbf{U} specified by local mean, ITCIMC-3 with \mathbf{U} specified by one ATGP-found target, ITCIMC-4 with $\mathbf{U} = \emptyset$ using Otsu's thresholding method. Like Purdue data experiments, the results in Fig. 12 suggested that the less BKG information was used, the better the classification was. As discussed in the Purdue data experiments, we also used one against rest to calculate the five performance measures, $P_C(C_i)$ and P_C in (16) and (17), P_{MC} in (19), $P_{precision}(\hat{C}_i)$, and $P_{precision}$ in (20) and (21), P_{OA} in (23), $P_A(C_j)$ in (22) and P_A in (23) where the BKG class, class 0 is also included for classification.

Tables VI and VII further tabulate respective results produced by EPF-B-g, EPF-B-c, EPF-G-g, and EPF-G-c and ITCIMC-1, ITCIMC-2, ITCIMC-3, ITCIMC-4 using five performance measures, $P_C(C_i)$ and P_C in (16) and (17), P_{MC} in (19), $P_{precision}(\hat{C}_i)$ and $P_{precision}$ in (20) and (21), P_{OA} in (23), $P_A(C_j)$ in (22), and P_A in (23) in terms of percent-

age (%) where the $P_C\%$, $P_{MC}\%$, and $P_{precision}^{C_i}$ are shown in the three columns under each method and P_{OA} and P_A are shown in the last two rows with the bold faced values indicating the best results, where P_A (BKG) specifically emphasize the fact that BKG is considered as a single and separate class for classification.

In analogy with Table IV, the quantitative study in Table VII also showed that the best result was the one produced by ITCIMC-4 with $\mathbf{U} = \emptyset$ which did not use any BKG information in \mathbf{U} . This indicated that some BKG information might contain crucial class information which was eliminated by \mathbf{U} . In order to see how the spatial information provided by Gaussian-filtered TCIMC improves classification, ITCIMC-4 is selected for illustration since it was the best version of TCIMC. Fig. 13 shows the progressive color classification maps produced by ITCIMC-4 where the number of iterations indicated underneath each figure. As we can see, when no spatial information is included in TCIMC the classification result was poor as shown at the first iteration in Fig. 13. As the Gaussian-filtered spatial information was gradually added to the processed hyperspectral image cubes via feedback loops during iterative processes, the classification results began to improve until it satisfied the stopping rule at the 19th iteration. In particular, initially there were many scattering classified data sample vectors in the first iteration and then the number of scattering data sample vectors was significantly reduced from first iteration to second and third iterations. When it reached the ninth iteration and afterward, only a few scattering data sample vectors were remained. This also explained the dilemma of mixed pixel classification in classification. However, this issue can be resolved by including spatial information provided by iteratively applying Gaussian filters to TCIMC-classification maps.

From Tables VI and VII, the best OA was the one produced by EPF-G-c, while the best P_A was produced by ITCIMC. As for the other three performance measures, P_C , P_{MC} , and $P_{precision}$, EPF-G-c generally produced the best P_C values for 14 out of 16 classes, and ITCIMC produced best values of P_{MC} and $P_{precision}$ almost across board except class 8 whose best values of P_{MC} and $P_{precision}$ were produced by EPF-G-c. In analogy with the Purdue Indian Pines scene, ITCIMC-4 was the best to produce best values of P_{MC} , $P_{precision}$, and P_A . This also concludes that ITCIMC-4 is generally the best classification technique to deal with correct classification and MC issues.

C. University of Pavia

The University of Pavia image capturing an urban area surrounding the University of Pavia, Italy, 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 per pixel and a spectral coverage ranging from 0.43 to 0.86 μm with spectral resolution of 4 nm (12 most noisy channels were removed before experiments). Nine classes of interest plus BKG class, class 0 are considered for this image. Fig. 14 shows the University of Pavia image in (a), three-band color composite in (b), and ground truth class labels in (c). Table VIII tabulates specifications of parameters

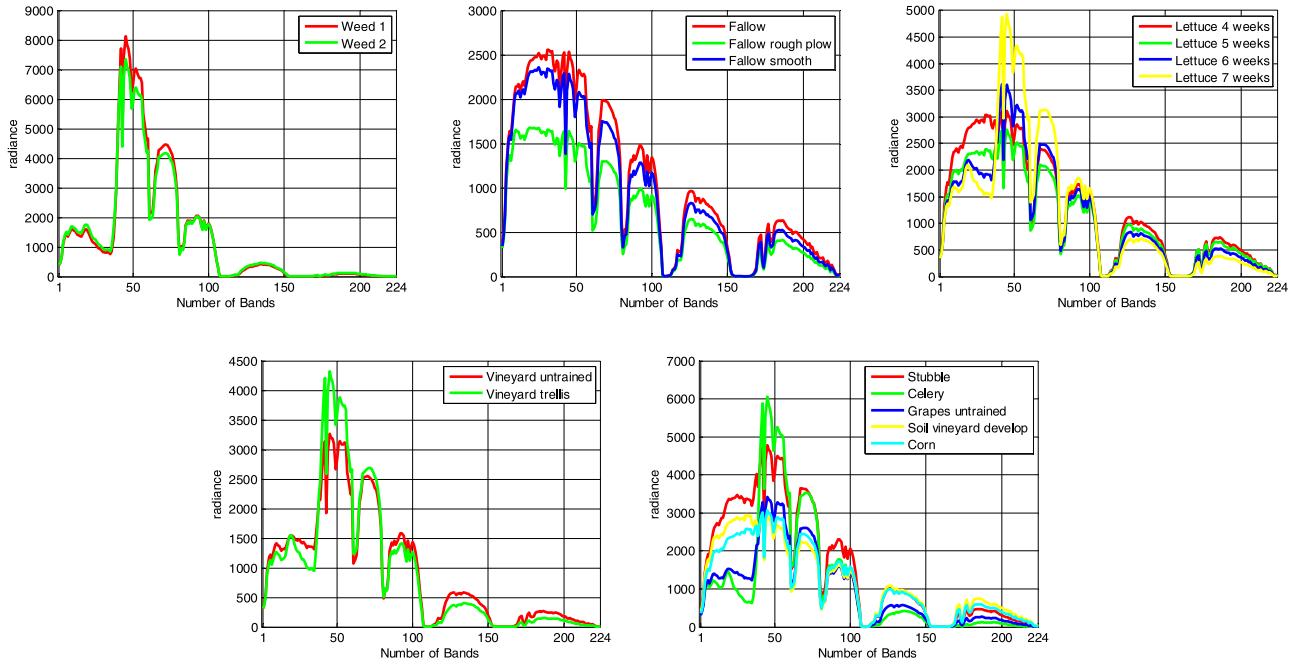


Fig. 11. Spectral profiles of 16 class sample means in Salinas scene.

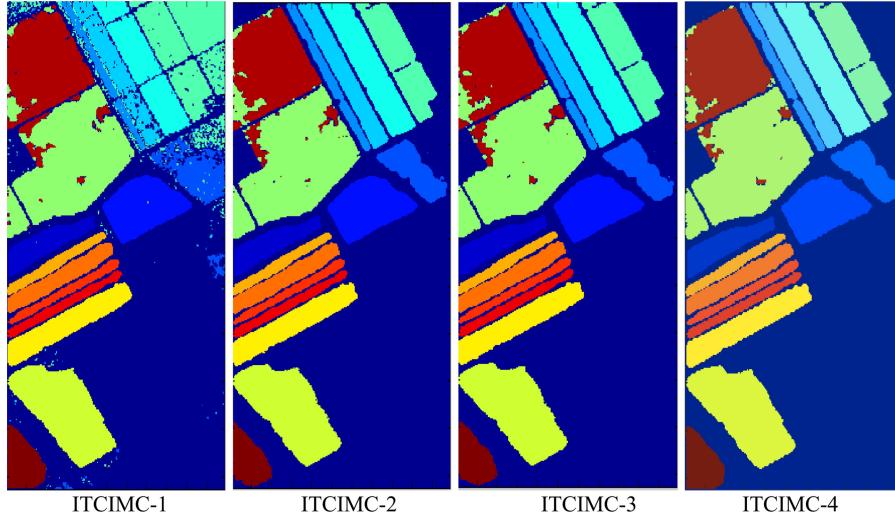


Fig. 12. Classification maps by four different versions of ITCIMC for Salinas data.

and various methods used by ITCIMC, where \mathbf{D} in (9) consists of nine class sample means.

Fig. 15 shows nine class maps produced by four versions of ITCIMC, ITCIMC-1 with \mathbf{U} specified by BKG mean, ITCIMC-2 with \mathbf{U} specified by local mean, ITCIMC-3 with \mathbf{U} specified by one ATGP-found target, ITCIMC-4 with $\mathbf{U} = \emptyset$ using Otsu's thresholding method. Unlike Purdue and Salina data experiments, differences in the results in Fig. 15 are not appreciable by visual inspection.

As discussed in previous experiments, Tables IX and X further tabulate respective results produced by EPF-B-g, EPF-B-c, EPF-G-g, and EPF-G-c and ITCIMC-1, ITCIMC-2, ITCIMC-3, ITCIMC-4 using five performance measures $P_C(C_i)$ and P_C in (16) and (17), P_{MC} in (19), $P_{precision}(\hat{C}_i)$ and $P_{precision}$ in

(20) and (21), P_{OA} in (23), $P_A(C_j)$ in (22), and P_A in (23) in terms of percentage (%) where the $P_C\%$, $P_{MC}\%$, and $P_{precision}^{C_i}$ are shown in the three columns under each method and P_{OA} and P_A are shown in the last two rows with the bold faced values indicating the best results where P_A (BKG) specifically emphasize the fact that BKG is considered as a single and separate class for classification. Despite that ITCIMC-4 produced lower $P_{OA} = 77.63$ compared to 98–99% produced by the four EPF-based methods, ITCIMC did produce the best $P_A = 80.21$ which was about four times better than around 20% produced by the four EPF-based methods. As for the other three performance measures, P_C , P_{MC} , and $P_{precision}$, EPF-G-c generally produced the best P_C values for eight out of nine classes, and ITCIMC produced best values of P_{MC} and $P_{precision}$ across board. Like

TABLE VI
POA AND PA (BKG) PRODUCED BY EPF-B-C, EPF-B-G, EPF-G-C, AND EPF-G-G FOR SALINAS SCENE

Class	EPF-B-g			EPF-B-c			EPF-G-g			EPF-G-c		
	PA%	P _{MC} %	P _{precision} %	PA%	P _{MC} %	P _{precision} %	PA%	P _{MC} %	P _{precision} %	PA%	P _{MC} %	P _{precision} %
1	100.00	0.68	73.19	100.00	0.65	74.05	100.00	0.65	74.00	100.00	0.75	71.22
2	100.00	2.78	56.22	99.97	2.78	56.21	100.00	2.72	56.72	100.00	2.96	54.67
3	100.00	14.99	12.20	100.00	14.95	12.22	100.00	15.09	12.13	100.00	15.14	12.10
4	100.00	4.41	23.13	100.00	4.38	23.25	100.00	4.43	23.04	100.00	4.54	22.64
5	98.51	1.95	55.95	98.36	1.97	55.66	98.47	1.92	56.29	98.84	1.83	57.61
6	100.00	0.83	81.75	100.00	0.86	81.18	100.00	0.84	81.58	100.00	0.89	80.65
7	100.00	0.87	79.34	100.00	0.83	80.14	100.00	0.83	80.21	99.97	0.87	79.48
8	81.52	1.35	87.11	81.47	1.38	86.84	82.37	1.22	88.33	83.52	1.18	88.82
9	99.85	12.16	35.25	99.84	12.11	35.33	99.87	12.20	35.19	99.87	12.38	34.90
10	96.19	8.10	28.06	96.06	8.21	27.77	96.49	8.17	27.96	97.86	7.40	30.16
11	100.00	3.25	23.57	99.91	3.24	23.58	100.00	3.22	23.71	100.00	3.15	24.09
12	100.00	4.90	27.42	100.00	4.92	27.36	100.00	4.96	27.20	100.00	4.95	27.23
13	99.13	0.40	67.16	99.45	0.40	67.38	99.56	0.41	66.91	99.78	0.42	66.52
14	100.00	0.50	65.97	100.00	0.51	65.64	100.00	0.52	65.40	100.00	0.46	67.85
15	93.82	2.09	76.17	93.26	2.11	75.86	94.32	2.02	76.86	96.19	1.94	77.94
16	99.61	0.73	69.47	99.56	0.74	69.19	99.39	0.70	70.35	99.67	0.70	70.38
POA	95.87			95.70			96.01			96.55		
PA (BKG)	46.71			46.63			46.77			47.04		

TABLE VII
POA AND PA (BKG) PRODUCED BY TCIMC IN COMPARISON WITH DIFFERENT U

Class	ITCIMC-1			ITCIMC-2			ITCIMC-3			ITCIMC-4		
	PA%	P _{MC} %	P _{precision} %	PA%	P _{MC} %	P _{precision} %	PA%	P _{MC} %	P _{precision} %	PA%	P _{MC} %	P _{precision} %
1	95.62	0.09	94.73	95.42	0.09	94.95	95.52	0.10	94.86	95.57	0.09	94.91
2	98.15	0.08	97.47	98.09	0.10	97.03	98.17	0.10	97.29	98.17	0.09	97.55
3	79.61	3.23	30.57	93.78	0.17	90.97	93.98	0.14	94.28	94.13	0.13	92.91
4	71.96	0.12	88.74	95.84	0.13	90.27	95.98	0.14	89.68	95.48	0.13	90.48
5	82.97	3.24	37.91	95.11	0.31	88.31	95.74	0.32	88.20	96.53	0.33	87.81
6	97.96	4.86	41.57	98.28	0.29	92.71	98.36	0.28	92.78	98.31	0.29	92.71
7	97.23	4.42	41.02	98.04	0.25	92.83	98.07	0.25	92.86	98.04	0.25	92.78
8	90.90	0.65	93.40	92.57	0.16	90.84	92.78	1.00	91.31	94.47	0.73	93.64
9	95.97	0.10	98.07	95.55	0.12	97.87	95.71	0.12	97.89	96.37	0.12	98.02
10	95.31	0.07	97.41	94.87	0.09	97.04	94.94	0.08	97.28	95.55	0.06	97.97
11	95.79	0.11	88.65	95.69	0.11	89.02	95.69	0.11	89.18	96.07	0.11	89.53
12	97.15	0.11	93.18	97.56	0.14	92.25	97.41	0.14	92.42	97.04	0.13	93.08
13	94.76	0.15	82.90	92.69	0.19	79.87	93.67	0.20	79.74	95.31	0.17	82.44
14	94.69	0.11	88.28	93.27	0.15	86.03	93.55	0.14	86.44	94.58	0.14	86.64
15	90.68	0.75	88.57	88.61	0.66	90.34	90.17	0.62	91.11	92.50	0.45	93.51
16	94.33	0.10	93.61	87.66	0.06	95.88	94.63	0.11	93.34	94.13	0.11	93.36
POA	92.51			94.13			94.71			95.54		
PA (BKG)	81.01			94.54			94.78			95.33		

TABLE VIII
SPECIFICATIONS OF PARAMETERS USED BY ITCIMC

D	9 class sample means
Class number	1–9
σ used in Gaussian filter	0.5 with window size 5×5
Thresholding method	Otsu
Stopping threshold (TI)	0.95

Purdue, Indian scene and Salinas scene ITCIMC are also the best classification techniques to deal with correct classification and MC issues.

If we further plot the spectral profiles of the nine classes by their class means, all spectral signatures are very similar as shown in Fig. 16. This fact was evidenced in the last row in Table IX, where the ARs produced by four EPF-based methods were only around 20.40% even through their P_C and P_{OA} can be very high. This indicates that P_{OA} did not reflect the difficulty of classifying these nine classes. Instead, P_{accuracy} should be the one to be used to measure the classification difficulty.

In analogy with Tables IV and VII, Table X also showed that the best results were those produced by ITCIMC-3 with $U = [t^{ATGP}]$ and ITCIMC-4 with $U = \emptyset$. In order to see how the spatial information provided by Gaussian-filtered TCIMC improves classification ITCIMC-4 is selected for illustration. Fig. 17 shows progressive color classification maps produced

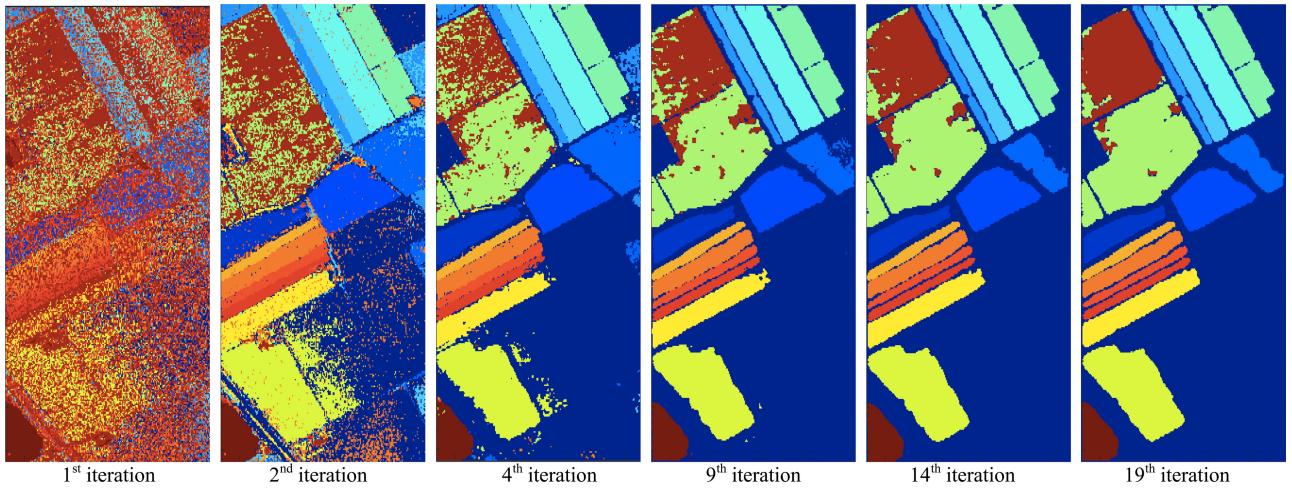


Fig. 13. Progressive color classification maps by ITCIMC-4 of Salinas scene.

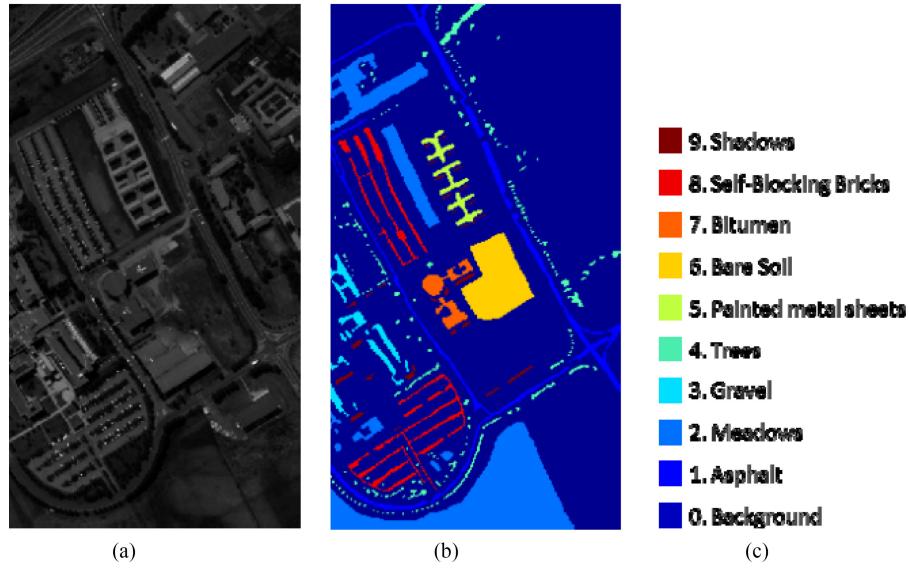


Fig. 14. Ground-truth of University of Pavia scene with 9 classes. (a) University of Pavia scene. (b) Color ground-truth image. (c) Ground truth class labels.

by ITCIMC-4, where the number of iterations indicated underneath each figure. When no spatial information is included in TCIMC, the classification result was very poor as shown in the first iteration in Fig. 17, where nothing was visible. As ITCIMC completed, the second iteration with the Gaussian-filtered spatial information was first fed back and added to the processed hyperspectral image cubes, the classification results was improved significantly but still far from satisfaction. As the iterative process was continued, the classification results were gradually improved until it reached the 13th iteration. This example demonstrated that the spatial information for the University of Pavia is much more crucial than the Purdue Indian Pine scene and the Salina scene in classification.

Five comments on ITCIMC are worth mentioning.

- 1) There are two major advantages of ITCIMC over spectral-spatial hyperspectral image classification techniques including EPF-based methods in [18]. One is that ITCIMC classifies all classes simultaneously using the constraint

matrix \mathbf{C} in (13) in one-shot operation as opposed to SVM used in [18] which extends a binary SVM classifier to a multiclass classifier via one against one or against all strategy. The other is that ITCIMC evaluates all data sample vectors compared to SVM which evaluates only both training and test sample vectors from the same class.

- 2) Theoretically speaking, ITCIMC-1, ITCIMC-2, and ITCIMC-3 should perform better than ITCIMC-4 by taking advantage of removing the effects of the undesired signatures in \mathbf{U} on the desired signatures in \mathbf{D} . Our experiments showed otherwise. The reason for this is caused by complicated BKG. Since BKG samples vary so much, particularly for the University of Pavia it requires as many BKG samples to be removed by \mathbf{U} . As a result, using limited number of BKG samples does not help much in classification performance. This evidence also explains why most hyperspectral image classification techniques reported in the literature did not address BKG issue.

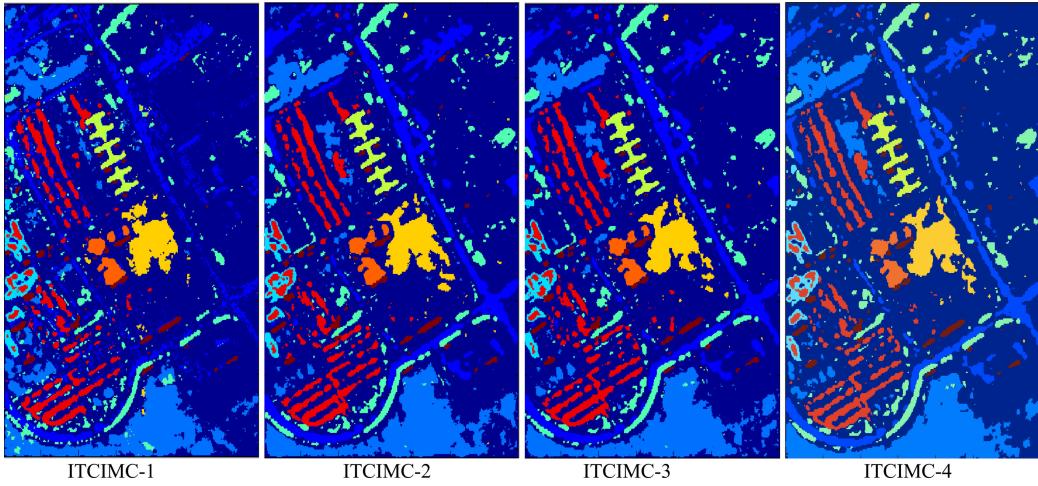


Fig. 15. Classification maps by four different version of ITCIMC for University Pavia.

- 3) The Otsu method used in ITCIMC is an effective threshold method, but not necessarily optimal. ITCIMC can be further improved by any better thresholding technique if there exists one.
- 4) The value of the parameter σ used by Gaussian filters is also selected empirically and not optimized. How to find an appropriate value of σ is a challenging issue since it is determined by the data to be used for processing. Nevertheless, according to our experiments, the used values for $\sigma = 0.5$ were relatively robust.
- 5) The classification performance of ITCIMC can be further improved if there is a better filter to replace Gaussian filters used in ITCIMC. However, the experimental results conducted in this paper suggested that Gaussian filters were sufficiently effective and also rather simple.
- 6) Our experimental results demonstrate that the ITCIMC generally performed better than SVM-based hyperspectral image classification techniques using several quantitative objective classification measures.

D. Comparison of EPF-Based Methods With Composite Kernel-SVM-Based Methods

In addition to comparing the EPF-based methods in Section VI-C, this section also compare spectral-spatial composite kernel (CK)-SVM-based methods developed in [24] which have received interest in hyperspectral image classification. The CK makes use of a spectral kernel K_ω and a spatial kernel K_s to develop four different CK-SVM-based methods, referred to as $K_{\{s,\omega\}}$, $K_s + K_\omega$, $\mu K_s + (1 - \mu)K_\omega$, and $K_s + K_\omega + K_{sw} + K_{ws}$. The first one is $K_{\{s,\omega\}}$ which is a stacked feature approach taking into account the spectral and textual information. The second one is $K_s + K_\omega$ which is a weighted summation kernel taking direct summation kernel. The third one is $\mu K_s + (1 - \mu)K_\omega$ which uses μ as a summation weight. The fourth one is $K_s + K_\omega + K_{sw} + K_{ws}$ which is a cross-information kernel. Parameters set for SVM and the window sizes of spatial kernels were set to be the same as [24]; how-

ever, the number of training samples was set to be the same as EPF-based methods for comparison. The four CK-SVM methods, $K_{\{s,\omega\}}$, $K_s + K_\omega$, $\mu K_s + (1 - \mu)K_\omega$, and $K_s + K_\omega + K_{sw} + K_{ws}$ methods were implemented in two ways, using only the mean, μ of neighborhood pixels in a window, and using the mean μ and standard deviation, σ of neighborhood pixels in a window. Tables XI, XIII, and XV tabulate P_{OA} , P_A (BKG) P_A and P_{MC} produced by $K_{\{s,\omega\}}$, $K_s + K_\omega$, $\mu K_s + (1 - \mu)K_\omega$, and $K_s + K_\omega + K_{sw} + K_{ws}$ methods using only μ of neighborhood pixels in a window for Purdue's data, Salinas, and University of Pavia, respectively, while Tables XII, XIV, and XVI tabulate P_{OA} , $P_{precision}$, P_D , and P_F produced by $K_{\{s,\omega\}}$, $K_s + K_\omega$, $\mu K_s + (1 - \mu)K_\omega$, and $K_s + K_\omega + K_{sw} + K_{ws}$ methods using μ and σ of neighborhood pixels in a window for Purdue's data, Salinas, and University of Pavia, respectively. The best results boldfaced in these tables suggest that a better CK-SVM method was $K_s + K_\omega$ since except the OA of Purdue's data $K_s + K_\omega$ performed generally better than other three CK-SVM methods.

Since not each method can produce best P_{OA} and P_A (BKG) for all the three image scenes, Table XVII summarizes the results in Tables III-IV, VI-VII, IX-X, and XI-XVI by selecting best possible methods from each of three categories, 4 ITCIMC-based methods, 4 EPF-based methods, 4 CK-SVM based methods in terms of P_{OA} and P_A (BKG) for each of three hyperspectral image scenes, Purdue Indian Pines, Salinas, and University of Pavia where an N/A indicates "not applicable." As we can see from Table XVII, ITCIMC-4 was the only one produced both best P_{OA} and P_A (BKG) among the four versions of ITCIMC. Also, from Table XVII, the best EPF method clearly outperformed the best CK-SVM methods with P_{OA} at least 5–6% better for Purdue's data and Salians and, at least 2% better for the University of Pavia. In addition, the best EPF-based method and CK-SVM-based method produced P_A (BKG) in the range of 39–47% for Purdue's data, the range of 47–50% for Salinas and the range of 20–21% for University of Pavia. Also, except Salinas scene, the four EPF methods produced better P_A (BKG) than the CK-SVM-based methods. Most importantly, ITCIMC

TABLE IX
POA AND PA (BKG) PRODUCED BY EPF-B-C, EPF-B-G, EPF-G-C, AND EPF-G-G FOR UNIVERSITY OF PAVIA SCENE

Class	EPF-B-g			EPF-B-c			EPF-G-g			EPF-G-c			
	PA %	P _{MC} %	P _{precision} %	PA %	P _{MC} %	P _{precision} %	PA %	P _{MC} %	P _{precision} %	PA %	P _{MC} %	P _{precision} %	
1	97.07	16.71	18.28	97.10	16.82	18.20	96.53	16.25	18.56	96.95	16.13	18.72	
2	98.10	20.22	36.51	98.09	20.26	36.48	98.13	19.63	37.10	98.16	19.32	37.42	
3	91.47	1.83	34.19	91.71	1.81	34.55	91.76	1.76	35.11	91.81	1.77	35.02	
4	95.04	14.64	10.03	93.99	14.64	9.93	94.35	14.20	10.21	98.14	14.02	10.69	
5	100.00	0.88	42.92	100.00	0.82	44.42	100.00	0.92	41.81	100.00	0.82	44.63	
6	100.00	30.14	9.69	100.00	29.92	9.74	100.00	29.96	9.73	100.00	30.02	9.72	
7	100.00	0.99	39.82	100.00	1.01	39.19	100.00	0.94	40.82	100.00	0.92	41.34	
8	99.02	8.15	19.19	98.72	8.10	19.23	98.91	8.09	19.28	99.51	8.13	19.31	
9	100.00	4.62	9.41	100.00	4.77	9.15	100.00	6.05	7.44	100.00	6.49	7.00	
P _{OA}		98.97			98.95			98.84			99.17		
P _A (BKG)		20.41			20.41			20.39				20.45	

TABLE X
POA AND PA (BKG) PRODUCED BY TCIMC IN COMPARISON WITH DIFFERENT U

Class	ITCIMC-1			ITCIMC-2			ITCIMC-3			ITCIMC-4		
	PA %	P _{MC} %	P _{precision} %	PA %	P _{MC} %	P _{precision} %	PA %	P _{MC} %	P _{precision} %	PA %	P _{MC} %	P _{precision} %
1	65.94	9.68	19.33	76.52	5.55	32.15	76.46	5.47	32.48	76.56	5.35	32.93
2	68.64	2.99	70.44	75.55	2.59	74.16	81.53	2.21	78.45	81.68	2.19	78.65
3	51.97	0.94	41.11	46.28	0.91	40.42	48.79	0.87	41.83	45.74	0.88	41.26
4	61.60	3.14	24.60	81.24	3.53	26.18	80.27	3.46	26.36	79.97	3.34	26.93
5	92.57	0.28	69.13	98.59	0.29	68.56	98.44	0.30	68.32	98.44	0.28	69.25
6	74.65	0.37	81.91	71.75	0.60	74.67	70.73	0.67	72.44	71.12	0.61	74.32
7	79.60	0.16	77.21	84.51	0.18	75.49	87.22	0.17	77.13	85.64	0.18	76.24
8	74.20	2.79	32.50	80.20	3.28	31.15	77.16	3.34	29.35	79.17	3.39	30.19
9	74.22	0.70	34.51	78.46	0.72	33.39	79.73	0.76	32.46	78.56	0.67	34.83
P _{OA}		67.94		75.25			77.71				77.63	
P _A (BKG)		75.59		78.97			79.73				80.21	

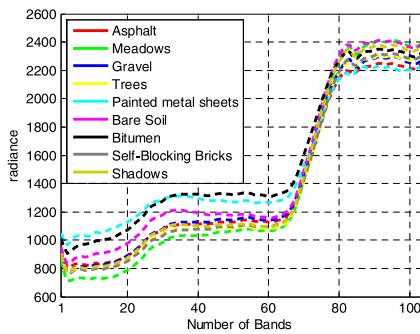


Fig. 16. Spectral profiles of 9 class sample means in University of Pavia scene.

produced the best P_A (BKG) among all the test methods across aboard for all the three datasets. It is also interesting to note in Tables III, VI, IX, XV, and XVI that the BKG issue has a significant impact on P_{OA} and P_A (BKG) for the University of Pavia when four EPF-based and four CK-SVM methods were implemented. By excluding BKG from classification, these methods produced very high P_{OA} around 97–99% rates but produced very poor P_A (BKG) around 20–21% rates. By contrast, ITCIMC-4 produced P_{OA} = 77.63% with 20% lower than EPF-based CK-SVM methods but far better P_A (BKG) = 80.21% almost four times better than 20–21% produced by EPF-based and CK-SVM based methods. This particular image scene demon-

strated that removing BKG data samples from classification may be misleading. For example, by only looking at P_A for Purdue's Indian Pines, Salinas, and University of Pavia, we may conclude that University of Pavia image scene is less complicated than the other two image scenes. However, the P_A (BKG) values showed otherwise.

Finally, this section also shows that the four CK-SVM-based methods are not as competitive as EPF-based methods are in classification of all three image scenes. This fact provides evidence that when it comes to spectral–spatial approaches to hyperspectral image classification, EPF-based methods are better options than CK-SVM methods. This is mainly because using windows conjunction with composite kernels as *a priori* information to obtain spatial information is not as effective as EPF-based methods using the followed-up spatial filters right after SVM-classified maps to capture *a posteriori* spatial contextual information from the entire classification maps.

VII. DISCUSSIONS

This section discusses five major issues encountered in hyperspectral image classification, BKG, spatial filter selection, class signature selection, test sample vector selection, and performance evaluation.

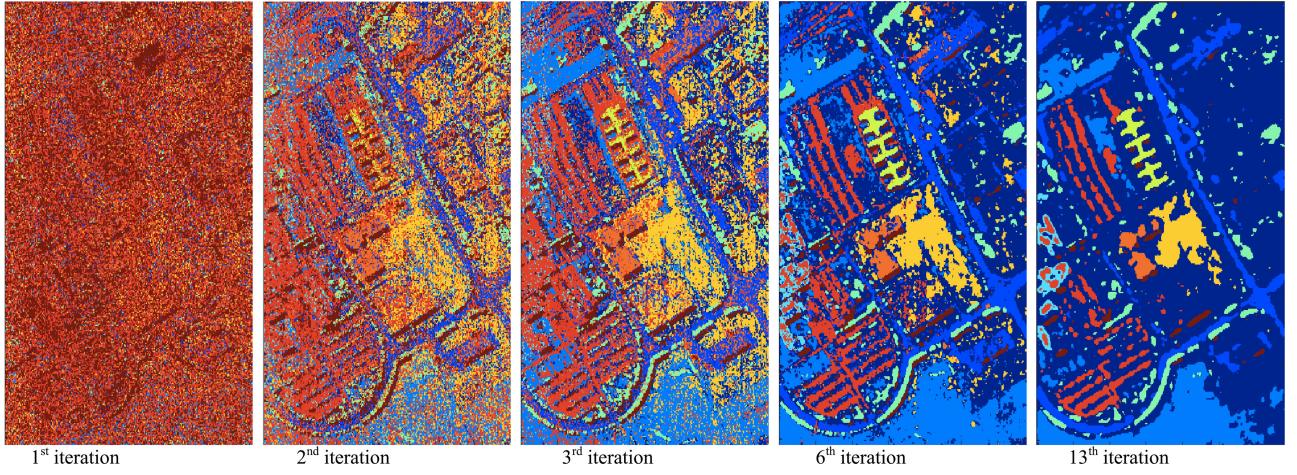


Fig. 17. Progressive color classification maps by ITCIMC for University of Pavia data.

TABLE XI
POA AND PA (BKG) PRODUCED BY CK-SVM PERFORMANCE USING ONLY μ OF NEIGHBORHOOD PIXELS IN A WINDOW AMONG $K_{\{s, \omega\}}$, $K_s + K_\omega$, $\mu K_s + (1 - \mu)K_\omega$ AND $K_s + K_\omega + K_{sw} + K_{ws}$ METHODS FOR PURDUE'S INDIAN PINES SCENE

Class	$K_{\{s, \omega\}}$			$K_s + K_\omega$			$\mu K_s + (1 - \mu)K_\omega$			$K_s + K_\omega + K_{sw} + K_{ws}$		
	P _A %	P _{MC} %	P _{precision} %	P _A %	P _{MC} %	P _{precision} %	P _A %	P _{MC} %	P _{precision} %	P _A %	P _{MC} %	P _{precision} %
1	91.30	0.39	33.87	95.65	0.40	34.65	95.65	0.40	34.65	97.83	0.68	23.94
2	76.47	3.92	58.68	77.73	3.21	63.83	78.43	3.30	63.42	92.86	3.28	67.38
3	75.90	4.43	41.31	76.75	3.01	51.16	77.47	3.33	48.86	90.60	3.23	53.52
4	89.45	2.82	26.53	92.83	4.08	20.60	92.41	3.94	21.10	94.94	2.35	31.56
5	93.58	8.71	20.16	93.17	10.02	17.94	93.17	10.17	17.72	88.61	6.09	25.51
6	94.79	4.25	44.53	91.10	4.07	44.57	91.64	4.18	44.07	92.60	4.04	45.22
7	100.00	0.39	25.69	96.43	0.24	35.06	96.43	0.28	31.40	100.00	1.31	9.21
8	97.91	0.26	89.83	97.91	0.55	80.55	97.91	0.53	81.25	98.74	0.29	88.72
9	100.00	0.90	9.62	100.00	0.17	36.36	100.00	0.20	31.75	100.00	1.02	8.51
10	76.75	3.12	54.37	87.04	2.30	64.73	85.19	2.29	64.34	89.92	2.47	63.84
11	67.90	3.72	70.73	72.91	2.88	77.02	71.41	2.95	76.22	88.19	2.92	79.98
12	86.00	1.70	59.51	79.43	3.65	38.70	79.60	3.47	39.97	95.28	1.86	59.73
13	100.00	2.40	29.12	100.00	0.97	50.49	100.00	1.06	48.12	100.00	3.41	22.43
14	88.85	8.73	39.45	93.52	11.05	35.15	93.20	11.14	34.87	85.53	10.88	33.48
15	87.05	16.28	9.09	81.09	14.51	9.46	79.53	14.14	9.52	78.76	13.86	9.61
16	100.00	0.74	37.50	96.77	0.54	44.12	96.77	0.51	45.92	100.00	0.39	53.45
POA			81.15		83.32			82.88			90.35	
PA (BKG)			38.23		41.44			40.77			39.77	

A. BKG Issue

BKG is generally an overlooked issue in hyperspectral image classification. According to Tables III–IV, VI–VII, and IX–X, the correct classification rate P_C and OA rate POA of four EPF-based methods in [18] were higher and better than those produced by our proposed four versions of ITCIMC. This is because these rates only calculate classes of interest and did not include BKG data sample vectors as test sample vectors. However, if we include BKG as part of test data sample vectors and also performed EPF-based methods using the program available at website provided by one of the authors in [18] by including BKG as test sample vectors which were not considered in [18], their PRs, MC rates, and ARs were all low, particularly, ARs which included the BKG class for classification and were indeed very poor. In contrast, all the four versions of ITCIMC performed significantly better than EPF-based methods in these

three rates. This indicates that ITIMC can deal with the BKG issue more effectively than EPF-based methods.

B. Class Signature Issue

As noted in the last comment in Section II, the class signatures $\mathbf{T} = [\mathbf{DU}]$ used in (9) by TCIMC can be obtained by any type of prior knowledge, such as database or spectral library. They generally do not have to be obtained from real data sample vector or training sample vectors. In our experiment, such class signatures were obtained by class means without training sample vectors. However, we can also use training sample vectors to gather the knowledge of \mathbf{T} . A simple way to do so is the common practice used by supervised classification techniques by randomly sampling a certain percentage of entire class sample vectors such as 5%, 10%, etc., and further find the sample

TABLE XII
**P_{OA} AND P_A (BKG) PRODUCED BY CK-SVM PERFORMANCE USING μ AND σ OF NEIGHBORHOOD PIXELS IN A WINDOW AMONG
 $K_{\{s, \omega\}}$, $K_s + K_\omega$, $\mu K_s + (1 - \mu)K_\omega$ AND $K_s + K_\omega + K_{s\omega} + K_{\omega s}$ METHODS FOR PURDUE'S INDIAN PINES SCENE**

Class	$K_{\{s, \omega\}}$			$K_s + K_\omega$			$\mu K_s + (1 - \mu)K_\omega$		
	P _A %	P _{MC} %	P _{precision} %	P _A %	P _{MC} %	P _{precision} %	P _A %	P _{MC} %	P _{precision} %
1	95.65	0.41	34.11	95.65	0.42	33.08	95.65	0.43	32.84
2	81.44	3.98	59.89	81.44	3.43	63.34	80.53	3.52	62.50
3	81.57	5.05	39.89	79.04	3.17	50.58	78.19	3.25	49.69
4	95.36	2.39	31.30	94.51	3.34	24.37	93.25	3.41	23.79
5	94.20	6.97	24.11	94.20	9.23	19.35	94.20	9.46	18.97
6	96.71	3.15	52.49	94.79	3.38	50.25	93.97	3.37	50.11
7	100.00	0.40	25.23	100.00	0.24	35.90	96.43	0.27	32.53
8	96.86	0.23	90.61	97.70	0.53	81.08	97.70	0.53	81.08
9	100.00	0.75	11.30	100.00	0.09	51.28	100.00	0.11	46.51
10	77.88	2.00	65.31	85.49	2.31	64.17	83.02	2.35	63.10
11	71.20	4.37	68.31	74.79	2.79	77.96	73.48	2.91	76.93
12	90.22	1.86	58.47	87.02	3.08	45.03	86.17	3.09	44.71
13	100.00	0.96	50.62	100.00	0.73	57.58	100.00	0.79	55.41
14	91.54	9.26	38.76	92.65	10.04	37.14	92.09	10.19	36.65
15	87.31	18.93	7.94	85.75	17.33	8.47	84.46	16.99	8.51
16	100.00	0.67	39.74	98.92	0.53	45.54	97.85	0.51	45.96
P _{OA}		84.06		85.20				84.18	
P _A (BKG)		41.06		43.83				42.90	

TABLE XIII
**P_{OA} AND P_A (BKG) PRODUCED BY CK-SVM PERFORMANCE USING ONLY μ OF NEIGHBORHOOD PIXELS IN A WINDOW AMONG
 $K_{\{s, \omega\}}$, $K_s + K_\omega$, $\mu K_s + (1 - \mu)K_\omega$ AND $K_s + K_\omega + K_{s\omega} + K_{\omega s}$ METHODS FOR SALINAS SCENE**

Class	$K_{\{s, \omega\}}$			$K_s + K_\omega$			$\mu K_s + (1 - \mu)K_\omega$			$K_s + K_\omega + K_{s\omega} + K_{\omega s}$		
	P _A %	P _{MC} %	P _{precision} %	P _A %	P _{mc} %	P _{precision} %	P _A %	P _{MC} %	P _{precision} %	P _A %	P _{MC} %	P _{precision} %
1	98.36	0.18	90.81	99.55	0.40	82.17	99.35	0.41	81.87	98.61	0.53	77.38
2	99.89	2.79	55.40	99.68	2.99	53.62	99.68	3.00	53.56	99.76	1.34	72.09
3	99.70	15.09	10.69	98.89	13.77	11.51	98.43	13.90	11.36	99.54	15.29	10.54
4	99.57	4.27	22.87	99.64	4.45	22.15	99.64	4.48	22.02	99.71	4.92	20.47
5	97.16	2.23	51.79	97.91	2.82	46.13	97.76	2.70	47.18	96.45	2.02	54.13
6	99.44	0.95	79.50	99.85	0.86	81.14	99.85	0.86	81.05	99.72	0.77	82.77
7	98.55	0.43	88.51	99.41	0.49	87.14	99.41	0.49	87.18	98.99	0.99	76.97
8	75.06	2.64	76.25	74.33	2.54	76.74	75.74	2.76	75.57	74.98	3.14	72.92
9	99.44	11.00	34.84	99.74	12.43	32.18	99.74	12.35	32.33	99.66	11.51	33.86
10	90.09	1.80	60.38	90.88	1.69	62.02	90.48	1.57	63.73	90.63	1.50	64.70
11	95.88	3.78	19.77	97.94	4.32	18.02	97.66	4.33	17.95	96.25	3.76	19.90
12	98.44	4.56	27.60	99.79	5.35	24.76	99.69	5.42	24.51	98.91	4.76	26.81
13	98.47	0.84	49.48	98.58	0.44	65.10	98.47	0.45	64.57	98.58	0.57	58.98
14	98.97	2.29	29.56	96.26	0.72	56.38	95.70	0.72	56.48	98.88	1.94	33.09
15	77.19	3.47	60.86	78.52	3.35	62.11	75.83	3.29	61.75	75.56	3.29	61.67
16	98.23	1.47	52.47	99.23	1.04	61.17	99.23	1.04	61.19	98.45	1.53	51.61
P _{OA}		90.47		90.78				90.63			90.34	
P _A (BKG)		47.69		49.55				49.55			48.11	

mean of sampled data for each of classes to represent class signatures for \mathbf{T} . Since TCIMC-4 is the best among all the four versions, Fig. 18(a)–(b) shows the respective results of AR, P_A, and OA produced by TCIMC-4 for three scenes, Purdue Indian Pines, Salinas, and University of Pavia, where the class signatures were calculated by randomly sampling data according to initial 10% up 100% with a step size 10% and finding their class sample means for \mathbf{T} . As we can see from these figures, P_A and P_{OA} of Salinas scene were nearly the same after 20% randomly sampling. On the other hand, P_A and P_{OA} of University of Pavia were nearly stable in the beginning and then gradually increased at the end. As for Purdue Indian Pines, its P_A and P_{OA} gradually

increased until 40% of data sample vectors randomly sampled and both rates became flat after that.

C. Spatial Filter Issue

In order to capture spatial information, a general approach is to use a spatial filter such as bilateral filter or guided filter in [18]. These two filters are designed to capture the morphological features such as edges. Our proposed ITCIMC is a mixed pixel-based method and has little to do with spatial features. Therefore, using Gaussian filters in this paper to capture spatial information instead of spatial features is more effective and bet-

TABLE XIV
POA AND PA (BKG) PRODUCED BY CK-SVM PERFORMANCE USING μ AND σ OF NEIGHBORHOOD PIXELS IN A WINDOW AMONG STACKED $K_{\{s, \omega\}}$, $K_s + K_\omega$, $\mu K_s + (1 - \mu)K_\omega$ AND $K_s + K_\omega + K_{sw} + K_{ws}$ METHODS FOR SALINAS SCENE

Class	$K_{\{s, \omega\}}$			$K_s + K_\omega$			$\mu K_s + (1 - \mu)K_\omega$		
	PA %	PMC %	Pprecision %	PA %	PMC %	Pprecision %	PA %	PMC %	Pprecision %
1	99.25	0.31	85.69	99.05	0.40	81.86	99.10	0.41	81.50
2	99.89	2.46	58.48	99.76	3.01	53.47	99.57	3.01	53.44
3	98.18	12.23	12.69	99.14	13.26	11.92	98.99	13.74	11.54
4	99.64	2.60	32.72	99.57	4.35	22.54	99.57	4.39	22.38
5	97.09	3.97	37.63	97.65	3.38	41.67	97.05	3.08	43.75
6	99.77	0.70	83.97	99.85	0.87	80.85	99.85	0.87	80.85
7	98.38	0.61	84.38	99.41	0.48	87.33	99.41	0.48	87.27
8	70.22	4.13	65.76	73.69	2.66	75.75	73.45	2.64	75.82
9	98.65	10.20	36.39	99.77	12.63	31.83	99.71	12.42	32.19
10	92.71	2.45	53.51	91.95	1.72	61.97	91.40	1.55	64.22
11	96.25	2.90	24.39	99.06	3.73	20.50	98.60	3.75	20.35
12	97.98	4.89	26.14	99.64	5.76	23.38	99.69	5.77	23.38
13	98.80	5.36	13.28	98.36	0.44	65.10	98.36	0.44	65.20
14	97.20	0.51	65.00	96.54	0.68	58.16	96.36	0.69	57.50
15	77.26	4.09	56.93	74.78	3.57	59.47	75.08	3.74	58.40
16	97.73	0.99	61.99	99.28	1.06	60.67	99.23	1.06	60.76
POA	89.47			90.21			90.10		
PA (BKG)	47.00			49.21			49.33		

TABLE XV
POA AND PA (BKG) PRODUCED BY CK-SVM PERFORMANCE USING ONLY μ OF NEIGHBORHOOD PIXELS IN A WINDOW AMONG $K_{\{s, \omega\}}$, $K_s + K_\omega$, $\mu K_s + (1 - \mu)K_\omega$ AND $K_s + K_\omega + K_{sw} + K_{ws}$ METHODS FOR UNIVERSITY OF PAVIA SCENE

Class	$K_{\{s, \omega\}}$			$K_s + K_\omega$			$\mu K_s + (1 - \mu)K_\omega$			$K_s + K_\omega + K_{sw} + K_{ws}$		
	PA %	PMC %	Pprecision %	PA %	PMC %	Pprecision %	PA %	PMC %	Pprecision %	PA %	PMC %	Pprecision %
1	97.15	13.26	19.49	96.50	12.06	20.91	95.82	12.17	20.64	94.93	14.08	18.22
2	98.11	18.08	34.90	97.75	20.22	32.32	97.46	19.66	32.88	96.81	20.32	32.01
3	93.47	4.03	19.18	93.95	3.58	21.17	93.23	3.24	22.76	89.42	2.74	25.00
4	98.34	12.13	10.84	97.98	11.88	11.00	97.91	11.80	11.06	98.66	13.24	10.05
5	100.00	1.33	32.88	99.93	0.76	46.19	99.93	0.65	50.09	99.93	1.38	32.08
6	97.06	21.79	9.97	97.85	20.64	10.54	98.35	21.23	10.33	96.86	19.13	11.18
7	98.42	1.58	28.73	97.67	2.18	22.40	97.89	2.23	22.07	96.47	1.37	31.23
8	92.80	6.19	21.32	91.80	6.52	20.28	92.56	6.84	19.66	92.88	6.69	20.06
9	100.00	4.50	9.25	100.00	5.22	8.08	100.00	5.26	8.03	100.00	4.35	9.54
POA	97.28			97.00			96.86			96.12		
PA (BKG)	18.65			19.29			19.75			18.94		

TABLE XVI
POA AND PA (BKG) PRODUCED BY CK-SVM PERFORMANCE USING μ AND σ OF NEIGHBORHOOD PIXELS IN A WINDOW AMONG STACKED $K_{\{s, \omega\}}$, $K_s + K_\omega$, $\mu K_s + (1 - \mu)K_\omega$ AND $K_s + K_\omega + K_{sw} + K_{ws}$ METHODS FOR UNIVERSITY OF PAVIA SCENE

Class	$K_{\{s, \omega\}}$			$K_s + K_\omega$			$\mu K_s + (1 - \mu)K_\omega$		
	PA %	PMC %	Pprecision %	PA %	PMC %	Pprecision %	PA %	PMC %	Pprecision %
1	96.74	14.15	18.42	96.61	14.67	17.87	95.96	13.55	18.96
2	98.06	14.83	39.52	97.92	14.91	39.35	97.86	15.22	38.84
3	93.81	2.93	24.63	93.43	3.31	22.39	93.47	3.53	21.32
4	98.96	14.88	9.07	99.05	14.97	9.03	98.76	14.23	9.43
5	100.00	1.50	30.37	99.93	0.65	50.13	99.93	0.66	49.83
6	98.95	20.36	10.78	98.95	20.44	10.74	98.87	22.15	9.98
7	99.02	1.77	26.50	98.42	1.24	33.82	97.37	1.31	32.38
8	92.72	7.45	18.36	92.83	7.29	18.71	92.91	6.72	19.98
9	100.00	4.73	8.85	100.00	5.16	8.16	100.00	5.32	7.94
POA	97.49			97.38			97.20		
PA (BKG)	18.65			21.02			20.87		

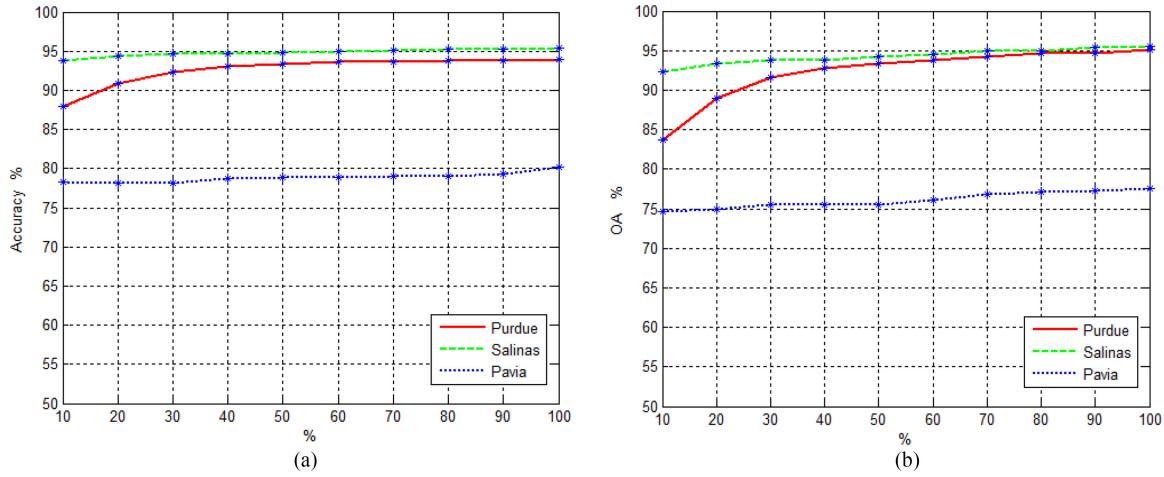


Fig. 18. AR and OA produced by ITCIMC-4 using different percentage of training sample vectors to obtain class signatures \mathbf{T} . (a) AR. (b) OA rate.

TABLE XVII
POA AND PA (BKG) PRODUCED BY BEST POSSIBLE METHODS FROM EACH OF THREE CATEGORIES FOR THREE IMAGE SCENES

Best methods	Purdue Indian Pines		Salinas		University of Pavia	
Classification accuracy	POA	PA (BKG)	POA	PA (BKG)	POA	PA (BKG)
ITCIMC-4	95.09	93.96	N/A	N/A	N/A	N/A
EPF-B-c	95.33	46.47	N/A	N/A	N/A	N/A
$K_s + K_\omega + K_{sw} + K_{ws}$ using μ and σ	90.35	39.77	N/A	N/A	N/A	N/A
ITCIMC-4	N/A	N/A	95.54	95.33	N/A	N/A
EPF-G-c	N/A	N/A	96.55	47.04	N/A	N/A
$K_s + K_\omega$ using μ	N/A	N/A	90.78	49.55	N/A	N/A
ITCIMC-4	N/A	N/A	N/A	N/A	77.63	80.21
EPF-G-c	N/A	N/A	N/A	N/A	99.17	20.45
$K_s + K_\omega$ using μ and σ	N/A	N/A	N/A	N/A	97.38	21.02

TABLE XVIII
ACCURACY AND OA WITH $\sigma = 0.5$

	Purdue		Salinas		Pavia	
W	P _A (%)	P _{OA} (%)	P _A (%)	P _{OA} (%)	P _A (%)	P _{OA} (%)
3 × 3	93.94	95.07	95.20	95.36	80.24	77.62
5 × 5	93.96	95.07	95.33	95.54	80.21	77.63
7 × 7	93.96	95.09	95.33	95.54	80.21	77.63
9 × 9	93.96	95.09	95.33	95.54	80.21	77.63
11 × 11	93.96	95.09	95.33	95.54	80.21	77.63

ter than using the bilateral and guided filters. An issue in using a Gaussian spatial filter is the selection of parameters used to specify the filter. In our experiments, these parameters are the standard deviation σ and the window size, w . Tables XVIII–XX classification results produced by ITCIMC-4 with various values of σ and w , respectively, where TCIMC-4 was selected to represent all the four versions of ITCIMC. As we can see from these tables, ITCIMC-4 is very robust to its used window size ranging from 3×3 to 11×11 and the value of σ around $0.4 \leq \sigma \leq 1$.

D. Test Sample Vector Issue

One major issue in hyperspectral image classification is the use of test sample vectors which also come from the same

TABLE XIX
ACCURACY AND OA WITH WINDOWS SIZE IS 5*5

	Purdue		Salinas		Pavia	
σ	P _A (%)	P _{OA} (%)	P _A (%)	P _{OA} (%)	P _A (%)	P _{OA} (%)
0.1	71.80	78.07	81.47	79.99	25.56	63.02
0.2	71.81	78.08	81.47	79.99	25.27	63.03
0.3	84.31	85.45	84.33	81.45	43.76	68.78
0.4	94.55	94.43	91.28	87.85	80.94	75.71
0.5	93.94	95.07	95.33	95.54	80.21	77.63
0.6	93.52	95.50	95.27	95.82	78.33	77.85
0.7	92.98	95.81	95.25	96.04	77.17	78.99
0.8	92.32	95.66	95.09	96.07	75.55	79.23
0.9	91.81	95.66	95.12	96.35	74.38	79.71
1.0	91.31	95.64	95.05	96.47	74.52	81.32

class trained by the training sample vectors. As a result, it is expected that P_{OA} is very high. However, if test sample vectors include BKG sample vectors are not necessarily coming from the same training class the classification performance would be dropped drastically as shown in Tables III–IV, VI–VII, and IX–X by MC rate P_{MC} and PR P_{precision}. This indicates that P_{OA} may miss some crucial information about classification which is in fact very important in other pattern recognition areas, such as optical character recognition, biometric recognition, and medical diagnosis.

TABLE XX
COMPUTING TIME IN SECONDS FOR FOUR ITCIMC VERSIONS AND FOUR SVM-EPF METHODS

Classification methods	ITCIMF1	ITCIMF2	ITCIMF3	ITCIMF4	EPF-B-g	EPF-B-c	EPF-G-g	EPF-G-c
Indian Pines	21.55	10.92	10.98	10.71	128.16	130.43	129.60	131.37
Salinas	149.88	100.32	125.85	73.56	168.02	202.79	169.54	183.44
Univ. of Pavia	88.28	63.14	63.82	59.71	220.49	229.53	240.65	215.93

E. Performance Evaluation Issue

In hyperspectral image classification, commonly used performance measures are POA, AA, or kappa coefficient in [10]–[28]. As already noted above, these measures may not effectively characterize images from various aspects. There are other useful measures such as correct classification rate, PC, MC rate, P_{MC} , and PR, $P_{precision}$ and P_A (BKG) have been used in pattern classification but not considered in [10]–[28]. Tables III–IV, VI–VII, IX–X, and XI–VII provide hard evidence that OA is not sufficiently effective to evaluate the classification performance.

It should be noted that the TCIMF developed in [29] was primarily designed for hyperspectral target detection not for hyperspectral image classification. In order for TCIMF to perform classification as ITCIMC, it must make hard decisions not soft decisions. This process is carried out by Otsu's thresholding method, a well-known technique in image processing. Accordingly, comparing ITCIMC to soft classifiers seems not appropriate. However, it is also noted that TCIMF is a pixel-based abundance fraction estimator which produces abundance fractional maps that generally do not correspond to soft decisions produced by classifiers such as fuzzy classifiers. Many reports on comparison of TCIMF with other abundance estimators have been reported in the literature such as [1], [2], [29], [33]. Those who are interested in such comparisons can consult these references.

F. Computing Time

In order to evaluate the efficiency of TCIMC compared to four SVM-EPF methods, Table XIII tabulates their respective computing times in seconds where a computer environment was specified by an Intel i7-6500U 2.5 GHz base frequency CPU and 12 GB 2133 MHz memory. All experiments were implemented using MATLAB and the SVM available in LIBSVM [39]. In particular, the Gaussian kernel parameter σ and slack variables used by SVM were selected by fivefold cross validation. As shown in the table, all the four versions of ITCIMC required less computing time than four SVM-EPF methods did. In particular, ITCIMC4 was the best and was nearly ten times faster than four SVM-EPF methods for Purdue data, 2.5 times for Salinas, and 4 times for University of Pavia.

VIII. CONCLUSION

Since TCIMC is a pixel-based technique, it does not take care of spatial information. Therefore, it works well if targets of interest are relatively small and do not exhibit spatial patterns such as HYDICE data in [1]. However, TCIMC may not work effectively if a hyperspectral image contains crucial spatial

information such as urban scenes. Therefore, for TCIMC to capture such spatial information an iterative version of ITCIMC, called ITCIMC, is derived, which makes use of Gaussian filters to obtain *a posteriori* neighboring spatial information of each classified data sample vector and then iteratively feeds back these Gaussian-filtered ITCIMC-classified spatial image to form a new set of images cubes for reprocessing ITCIMC. Many contributions are noteworthy

- 1) First and foremost is the design rationale of ITCIMC from a mixed pixel classification perspective which is completely different from commonly used spectral-spatial approaches reported in the literature.
- 2) ITCIMC classifies all classes simultaneously at the same time via (13), (14) which resolve issues arising from extending binary classification to multiclass classification such as SVM using one against one or one against all strategy.
- 3) ITCIMC introduces new evaluation measures, correct classification rate P_C , MC rate, P_{MC} , PR, $P_{precision}$ and AR, P_A which reflect real scenarios.
- 4) ITCIMC uses P_A , P_{MC} , and $P_{precision}$ to evaluate all data sample vectors including BKG data sample vectors for classification not only those in the same class.
- 5) ITCIMC repeatedly makes use of feedback loops by incorporating *a posteriori* spatial contextual information captured by Gaussian spatial filters into ITCIMC to improve classification performance iteratively.
- 6) ITCIMC can perform pure-pixel classification by Otsu's method as well as mixed pixel classification by finding class abundance fractions, while SVM-based spectral-spatial techniques perform pure-pixel spectral classification by SVM.
- 7) ITCIMC is easily to be implemented using only two robust parameters σ and Gaussian window size w .
- 8) ITCIMC requires computing time less than that required by SVM-EPF methods.
- 9) ITCIMC can be used to update TCIMF in ENVI to become a major classification technique for hyperspectral image classification.
- 10) Last but not least, the extensive experiments are conducted in this paper to compare two well-known SVM-based techniques, EPF and CK-based methods. These quantitative analyses suggest that CK-SVM-based methods are generally not competitive as the EPF-based methods are. To authors' best knowledge, no such detailed and comprehensive studies were available in the literature. Furthermore, both performed very poorly in P_A (BKG) and P_{MC} rates compared to ITCIMC which performed significantly better in both rates.

ACKNOWLEDGMENT

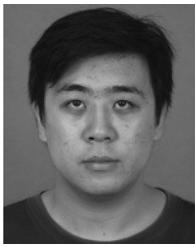
The authors would like to thank Dr. X. Kang for sharing his codes available at website <http://xudongkang.weebly.com>.

REFERENCES

- [1] C.-I Chang, *Hyperspectral Data Processing: Algorithm Design and Analysis*. Hoboken, NJ, USA: Wiley, 2013.
- [2] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York, NY, USA: Kluwer Academic/Plenum Publishers, 2003.
- [3] [Online]. Available: <https://purr.purdue.edu/publications/1947/serve/1?el=1>
- [4] W. W. Liu, X. Jiao, B. Ji, S. Yang, and C.-I Chang, "A study on Purdue's Indian Pine test site," *Recent Advances in Hyperspectral Signal and Image Processing*, C.-I Chang Ed. Trivandrum, Kerala, India: Research Signpost, 2006, ch. 4, pp. 93–139.
- [5] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag, 1999.
- [6] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. New York, NY, USA: Wiley, 2003.
- [7] 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.
- [8] 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.
- [9] G. Vamps-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.
- [10] 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.
- [11] 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. 9, no. 4, pp. 736–740, Oct. 2010.
- [12] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression model with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
- [13] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptive hyperspectral Markov random field approach for classification of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 11, pp. 4085–4098, Nov. 2011.
- [14] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.
- [15] Y. Chen, N. Nasabardi, and T. D. Tran, "Hyperspectral image segmentation using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
- [16] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected makers," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1267–1279, Oct. 2011.
- [17] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.
- [18] 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.
- [19] W. Fu, S. Li, L. Fang, X. Kang, and J. A. Benediktsson, "Hyperspectral image classification via shape-adaptive joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 556–567, Feb. 2016.
- [20] X. Kang, S. Li, L. Fang, M. Li, and J. A. Benediktsson, "Extended random walker-based classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 144–153, Jan. 2015.
- [21] T. Lu, S. Li, L. Fang, L. Bruzzone, and J. A. Benediktsson, "Set-to-set distance-based spectral spatial classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7122–7134, Dec. 2016.
- [22] 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.
- [23] X. Ma, H. Wang, and J. Wang, "Semisupervised classification of hyperspectral image based on multi-level labeling and deep feature learning," *JSPRS J. Photogramm. Remote Sens.*, vol. 120, pp. 99–107, 2016.
- [24] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "CKs for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
- [25] P. Gurram and H. Kwon, "Contextual SVM using Hilbert space embedding for hyperspectral classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1031–1035, Sep. 2013.
- [26] J. Peng, Y. Zhou, and C. Chen, "Region-kernel-based support vector machines for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4810–4824, Sep. 2015.
- [27] S. Sun, P. Zhong, H. Xiao, and R. Wang, "Active learning with Gaussian process classifier for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1746–1760, Apr. 2015.
- [28] R. Pullanagari, G. Kereszturi, I. J. Yule, and P. Ghamisi, "Assessing the performance of multiple spectral-spatial features of a hyperspectral image for classification of urban land cover classes using support vector machines and artificial neural network," *J. Appl. Remote Sens.*, vol. 11, no. 2, pp. 026009-1–026009-21, Apr.–Jun. 2017.
- [29] H. Ren and C.-I Chang, "Target-constrained interference-minimized approach to subpixel target detection for hyperspectral imagery," *Opt. Eng.*, vol. 39, no. 12, pp. 3138–3145, Dec. 2000.
- [30] 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.
- [31] H. V. Poor, *An Introduction to Signal Detection and Estimation Theory*. New York, NY, USA: Springer, 1994.
- [32] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- [33] C.-I Chang, "Target signature-constrained mixed pixel classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 2, pp. 1065–1081, May 2002.
- [34] J. C. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.
- [35] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Dept. Elect. Eng., Univ. Maryland, Baltimore, MD, USA, Aug. 1993.
- [36] W. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailing in the coeur d'Alene river valley, Idaho, through the use of constrained energy minimization technique," *Remote Sens. Environ.*, vol. 59, pp. 64–76, 1997.
- [37] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. New York, NY, USA: Academic, 1999, p. 366.
- [38] H. Ren and C.-I Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, Oct. 2003.
- [39] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, Apr. 2011. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

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. From June 2013 to June 2016, she was a Postdoctoral Fellow with the Information Science and Technology College, Dalian Maritime University. From September 2014 to September 2015, she was a Visiting Scholar with the College of Physicians and Surgeons, Columbia University, New York, 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.



Bai Xue (S'15) received the B.E. degree in automation from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2015. He is currently working toward the Ph.D. degree in electrical engineering at the University of Maryland Baltimore County, Baltimore, MD, USA.

From 2014 to 2015, he was a program exchange undergraduate student visiting the Department of Electrical and Computer Engineering, University of Detroit Mercy (UDM), Detroit, MI, USA. He is also a Visiting Research Assistant in the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Dalian Maritime University, Dalian, China. His research interest includes multispectral/hyperspectral imaging, pattern recognition, medical imaging and medical data analysis.



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

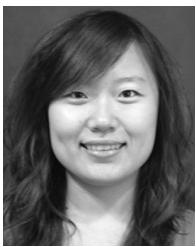
He has been with the University of Maryland, Baltimore County (UMBC), since 1987 and is currently a Professor in the Department of Computer Science and Electrical Engineering. He is currently holding Chang Jiang Scholar Chair Professorship and the Director of Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Dalian Maritime University, Dalian, China, since 2016, Hua Shan Scholar Chair Professorship with Xidian University, Xian, China, since 2016 and the Feng-Tai Chair Professor with National Yunlin University of Science and Technology since 2017. In addition, he is also the Chair Professor of Providence University, Taichung, Taiwan, since 2012. He was a Visiting Research Specialist in the Institute of Information Engineering, National Cheng Kung University, Tainan, Taiwan, from 1994 to 1995 and also a Distinguished Visiting Fellow/Fellow Professor, both of which were sponsored by National Science Council in Taiwan, from 2009 to 2010. He was also a Distinguished Lecturer Chair at the National Chung Hsing University, Taichung, Taiwan, sponsored by the Ministry of Education in Taiwan, from 2005 to 2006 and a Distinguished Chair Professor at National Chung Hsing University from 2014 to 2017. He was a plenary speaker for *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*, December 30, 2010, Industrial Technology Research Institute, Hsinchu; *2009 Annual Meeting of the Radiological Society of the Republic of China*, 2009, Taichung; *2008 International Symposium on Spectral Sensing Research* in 2008; and the *Conference on Computer Vision, Graphics, and Image Processing* 2003, Kimen and 2013, Nan-Tou, Taiwan. He has seven patents on hyperspectral image processing. He authored four books, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Kluwer Academic Publishers, 2003) and *Hyperspectral Data Processing: Algorithm Design and Analysis* (John Wiley & Sons, 2013), *Real Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection* (Springer, 2016) and *Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation* (Springer, 2017). In addition, he also edited two books, *Recent Advances in Hyperspectral Signal and Image Processing*, 2006 and *Hyperspectral Data Exploitation: Theory and Applications*, (John Wiley & Sons, 2007) and co-edited with A. Plaza a book on *High Performance Computing in Remote Sensing*, (CRC Press, 2007). His research interests include multispectral/hyperspectral image processing, automatic target recognition, and medical imaging.

Dr. Chang was a Guest Editor of a special issue of the *Journal of High Speed Networks on Telemedicine and Applications* (April 2000) and a Co-Guest Editor of another special issue of the same journal on *Broadband Multimedia Sensor Networks in Healthcare Applications* in April 2007. He is also a Co-Guest Editor of special issues on *High Performance Computing of Hyperspectral Imaging* for the *International Journal of High Performance Computing Applications* in December 2007, *Signal Processing and System Design in Health Care Applications* for the *EURASIP Journal on Advances in Signal Processing* in 2009, *Multispectral, Hyperspectral, and Polarimetric Imaging Technology* for the *Journal of Sensors and Hyperspectral Imaging and Applications for Remote Sensing*, 2017. He was an Associate Editor in the area of hyperspectral signal processing for *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* in 2001–2007. 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*, the *Journal of Robotics and Artificial Intelligence*, and *Open Remote Sensing Journal*. He received a National Research Council Senior Research Associateship Award from 2002 to 2003 sponsored by the U.S. Army Soldier and Biological Chemical Command, Edgewood Chemical and Biological Center, Aberdeen Proving Ground, Maryland. He is a Fellow of SPIE.



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



Yulei Wang received the B.S. and Ph.D. degrees in signal and information processing from Harbin Engineering University, Harbin, China, 2009 and 2015, respectively. She is awarded by China Scholarship Council as a joint Ph.D. student to study in Remote Sensing Signal and Image Processing Laboratory, University of Maryland, Baltimore County, College Park, MD, USA, in 2011.

She is currently a Lecturer with the College of Information Science and Technology, Dalian Maritime University.

Her current research interests include hyperspectral image processing and vital signs signal processing.



Sen Li received the B.S. degree in microelectronics from Liao Ning University, Shenyang, China, in 1996, the M.S. degree in information and communication engineering from Dalian Maritime University, Dalian, China, in 1999, and the Ph.D. degree in information and signal processing from Dalian University of Technology, Dalian, China, in 2011.

She was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada, in 2013. She is currently an Associate Professor with the Department of Information Science and Technology, Dalian Maritime University. Her research interests include array signal processing, communication signal processing, non-Gaussian signal processing, and hyperspectral image processing.