Least Squares Subspace Projection Approach to Mixed Pixel Classification for Hyperspectral Images

Chein-I Chang, *Senior Member, IEEE*, Xiao-Li Zhao, Mark L. G. Althouse, *Member, IEEE*, and Jeng Jong Pan

Abstract— An orthogonal subspace projection (OSP) method using linear mixture modeling was recently explored in hyperspectral image classification and has shown promise in signature detection, discrimination, and classification. In this paper, the OSP is revisited and extended by three unconstrained least squares subspace projection approaches, called signature space OSP, target signature space OSP, and oblique subspace projection, where the abundances of spectral signatures are not known *a priori* but need to be estimated, a situation to which the OSP cannot be directly applied. The proposed three subspace projection methods can be used not only to estimate signature abundance, but also to classify a target signature at subpixel scale so as to achieve subpixel detection. As a result, they can be viewed as *a posteriori* OSP as opposed to OSP, which can be thought of as *a priori* OSP. In order to evaluate these three approaches, their associated least squares estimation errors are cast as a signal detection problem in the framework of the Neyman–Pearson detection theory so that the effectiveness of their generated classifiers can be measured by characteristics (ROC) analysis. All results are demonstrated by computer simulations and Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data.

Index Terms— Classification, detection, hyperspectral image, oblique subspace projection classifier (OBC), orthogonal subspace projection (OSP), receiver operating characteristics (ROC), signature space orthogonal projection classifier (SSC), target signature space orthogonal projection classifier (TSC).

I. INTRODUCTION

THE ADVENT of high spatial resolution airborne and satellite sensors improves the capability of ground-based data collection in the fields of geology, geography, and agriculture. One major advantage of hyperspectral imagery over multispectral imagery is that the former images a scene using as many as 224 contiguous bands, such as Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [1], as opposed to the latter using only 4–7 discrete bands, such as SPOT and LANDSAT images. As a result, hyperspectral image data permit the expansion of detection and classification activities to targets previously unresolved in multispectral images. In addition, hyperspectral imagery provides more information with which it can differentiate very similar reflectance spectra, a task that multispectral imagery generally has difficulty with.

In multispectral/hyperspectral imagery, a scene pixel is generally mixed by a number of spectral signatures (or endmembers) due to improved spectral resolution with large spatial coverage from 10 to 20 m. Two models have been proposed in the past to describe such activities of mixed pixels. One is the marcospectral mixture [2] that models a mixed pixel as a linear combination of signatures resident in the pixel with relative concentrations. A second model suggested by Hapke in [3], called the intimate spectral mixture, is a nonlinear mixing of signatures present within the pixel. Nevertheless, Hapke's model can be linearized by a method proposed by Johnson *et al.* [4]. In this paper, only the linear spectral mixture model will be considered. By taking advantage of linear modeling, many image processing techniques can be applied. Of most interest is the principal components analysis (PCA), also known as Karhunen–Loeve transformation, which is widely used for data projection, so as to achieve data dimensionality reduction as well as feature extraction. As a result of PCA, the data coordinates will be rotated along with the direction of the maximum variance of the data matrix so that the significant information of the data can be prioritized in accordance with the magnitude of the eigenvalues of the data covariance matrix. Two disadvantages arise from the PCA approach. One is that the pixels in the PCA-transformed data are still a mixing of spectral signatures with unknown abundances. So, the determination and identification of individual spectral signatures are not mitigated. Malinowski [5] and Heute [6] proposed a solution. They first reconstructed the original data using the largest PCA-generated eigenvalue and measured the error between the raw data and the reconstructed data to see if the error falls within the prescribed tolerance. If not, they gradually added to data reconstruction the eigenvalues in decreasing magnitude until the error resulted within the desired level. A second disadvantage resulting from PCA is that PCA is only optimal in the sense of minimum mean square error, but not necessarily optimal in terms of class discrimination and separability [7]–[9].

Recently, an orthogonal subspace projection (OSP) method was proposed in [10] for hyperspectral image classification. It formulated an image classification problem as a generalized eigenvalue problem, thereby, Fisher's linear discriminant analysis can be used to classify mixed pixels. The classifier

resulting from OSP is an operator composed of two linear filters, one derived from the simultaneous diagonalization filter developed in [11] and the second, called the matched filter, derived from communication systems. However, the model on which the OSP classifier was based assumed the complete knowledge of signature abundance *a priori*, which is generally difficult to obtain in practice. In order for the OSP classifier to be applied to real scene experiments, this knowledge was obtained by the ground truth and directly extracted from scene pixels, as done in [10], where the signatures were estimated from the AVIRIS data themselves. Despite the success of OSP in classification of AVIRIS data [10], there is a lack of theory to support this experiment. In this paper, we will revisit the OSP approach and offer a theoretical background for the OSP from an estimation theory point of view to explain why the OSP can be applied to real hyperspectral data, as demonstrated in [10]. The theory is derived based on unconstrained least squares estimation and can be viewed as *a posteriori* approach. More precisely, all required information for data analysis is obtained from observed data not prior information assumed in the model used in [10]. As a result, the approaches presented in this paper can be referred to as *a posteriori* OSP, while the OSP in [10] can be regarded as *a priori* OSP. A recent work in [12] also derived an unconstrained maximum likelihood estimation (UMLE) that generated the same classification feature vector as the OSP but with an extra constant. However, it is this constant associated with abundance estimation. Moreover, there is a difference in their approaches. The UMLE in [12] maximized the conditional probability distribution of an unknown constant specified by signature abundance, while the OSP maximized the SNR based on Fisher's discriminant criterion, which only depends upon the noise second-order statistics. Accordingly, the UMLE and the OSP are indeed different methods.

Three approaches will be presented to extend the OSP classifier in [10] to the case in which the signature abundances are not necessarily known *a priori*, but can be estimated from the images of interest on the basis of the least squares error criterion. They are unconstrained least squares estimation methods and are derived from the subspace projection principle. The first approach was proposed in [13], which projected observed pixels (in this case, the observed data pixels are pixels in real data) into a signature space generated by an entire set of spectral signatures to reduce unwanted interference. The projector, named LSOSP in [13], will be referred to in this paper as the signature space orthogonal projection classifier (SSC) to reflect the projection space. Since the target signature is the one needed to be classified, a second approach is to directly project the observed pixels into the space spanned by only the target signature rather than the entire signatures. The resulting classifier will be referred to as the target signature space orthogonal projection classifier (TSC). As we might expect, TSC should perform better than SSC in the sense that the target signature space contains no other signatures but the target signature itself. Unfortunately, as shown in this paper, this is generally not the case. The drawback of TSC is that, since the desired signatures are not necessarily orthogonal to the target signature, the undesired signatures may be scrambled into the target signature space rather than mapping their own respective spanned spaces, as does SSC. As a result of such mixing, a signature bias becomes indispensable. In order to cope with this problem, a third approach based on oblique subspace projection [14] is suggested to eliminate such a bias. It projects the target signature and undesired signatures into two separate spaces, its range space and null space, respectively. Since these two spaces are disjoint, no mixing will occur and, thus, no signature bias will be generated. The resulting classifier will be referred to as an oblique subspace projection classifier (OBC). The OBC comes at a price, however. It is no longer orthogonal like SSC and TSC. Nevertheless, it is still a projection. What it is interesting about the OBC is that SSC can actually be decomposed into two oblique projections, one of which is exactly the OBC. In this paper, we will show a surprising result that SSC and OBC are essentially equivalent in the sense of classification, regardless of the fact that one is orthogonal and the other is not. In addition, we will further show that the OBC is exactly identical to the UMLE derived in [12]. It should also be noted that the techniques presented in this paper can be used for subpixel target detection and classification, even though they are primarily developed for signature abundance estimation here.

In order to evaluate the performance of these three classifiers, we model their associated least squares errors as a signal detection problem in which the true target signature abundance is the desired signal and the estimation error is treated as noise. By means of this detection model, the effectiveness of each classifier can be measured using the receiver operating characteristics (ROC) analysis via the Neyman–Pearson detection theory and both SSC and OBC generate identical ROC curves. Thus, they are essentially the same classifier in terms of detection power for target abundance. Most importantly, the OSP classifier derived in [10] produced the same classification feature vector as that produced by SSC and OBC with an extra constant, a result also noted in [12]. Since this extra constant can only affect the magnitude of the classification feature vector that only determines the amount of signature abundance contained in classified pixels, it does not alter the classification results. Consequently, UMLE, OSP, SSC, and OBC can be viewed as the same classifier. This validates the AVIRIS experiments conducted by the OSP classifier in [10]; in which case, the OSP is essentially equivalent to UMLE, SSC, and OBC in classification. As mentioned previously, TSC produces a signature bias that deteriorates its performance. However, it will outperform SSC and OBC, as shown in computer simulations, if the signature bias is known and removed. However, it is very rare in real applications to obtain such signature bias for TSC. So, the TSC proposed here only serves as a theoretic approach and a transition model from SSC to OBC.

The remainder of this paper is organized as follows. Section II briefly reviews the OSP approach. Section III describes three least squares error-based classifiers: SSC, TSC, and OBC. Section IV evaluates the estimation errors using ROC analysis. Section V conducts computer simulations and

experimental results using AVIRIS data. Section VI includes a brief conclusion.

II. OSP APPROACH

A. Linear Spectral Mixture Model

Linear spectral unmixing is a widely used approach in multispectral/hyperspectral imagery to determine and quantify individual spectral signatures in a mixed pixel. Let r_i be an $l \times 1$ column vector, and denote the ith pixel in a multispectral/hyperspectral image, where the bold face is used for vectors. In this case, each pixel is viewed as a pixel vector with dimension l. Assume that M is an $l \times p$ signature matrix denoted by $(\mathbf{m}_1 \ \mathbf{m}_2 \ \cdots \ \mathbf{m}_p)$, where $\mathbf{m}_j$ is an $l \times 1$ column vector represented by the jth signature resident in the pixel $\mathbf{r}_i$ and p is the number of signatures of interest. Let $\alpha_i = (\alpha_{i1} \ \alpha_{i2} \ \cdots \ \alpha_{ip})^T$ be a $p \times 1$ abundance column vector associated with $\mathbf{r}_i$, where α_{ij} denotes the fraction of the jth signature in the pixel $\mathbf{r}_i$. A linear spectral mixture model is described by

$$\mathbf{r}_i = M\alpha_i + \mathbf{n}_i \tag{1}$$

where $\mathbf{n}_i$ is an $l \times 1$ column vector representing an additive white Gaussian noise with zero mean and variance $\sigma^2 I$ and I is the $l \times l$ identity matrix.

B. OSP [10]

In the following, we briefly review the OSP approach given in [10]. First of all, we rewrite model (1) as

$$\mathbf{r} = \mathbf{d}\alpha_p + U\gamma + \mathbf{n} \tag{2}$$

where the subscript i is suppressed and M is made up of a desired signature $\mathbf{d} = \mathbf{m}_p$ and the undesired spectral signature matrix $U = (\mathbf{m}_1 \ \mathbf{m}_2 \ \cdots \ \mathbf{m}_{p-1})$. Model (2) can be extended straightforwardly to more than one desired signature. Here, we assume without loss of generality that the last signature is the desired signature $\mathbf{d}$. We also assume that $\langle \mathbf{d} \rangle$, $\langle M \rangle$ and $\langle U \rangle$ are the spaces linearly spanned by $\mathbf{d}$, U, and M, respectively. The reason for separating U from M is to allow us to design an OSP to annihilate U from an observed pixel prior to classification. One of such desired OSP's was derived in [10], given by $P_U^\perp = I - UU^\#$, where $U^\# = (U^T U)^{-1} U^T$ is the pseudoinverse of U and the notation $\perp$ in $P_U^\perp$ indicates that the projector $P_U^\perp$ maps the observed pixel $\mathbf{r}$ into the range space $\langle U \rangle^\perp$, the orthogonal complement of $\langle U \rangle$.

Now, applying $P_U^\perp$ to model (2) results in a new spectral signature model

$$P_U^\perp \mathbf{r} = P_U^\perp \mathbf{d}\alpha_p + P_U^\perp \mathbf{n} \tag{3}$$

where the undesired signatures in U have been eliminated and the original noise has been suppressed to $P_U^\perp \mathbf{n}$.

Equation (3) represents a standard signal detection problem. If the optimal criterion for the signal detection problem specified by (3) is to maximize the SNR, given by

$$\text{SNR} = \frac{[\mathbf{x}^T P_U^\perp \mathbf{d}]\alpha_p^2[\mathbf{d}^T P_U^\perp \mathbf{x}]}{\mathbf{x}^T P_U^\perp E[\mathbf{n}\mathbf{n}^T]P_U^\perp \mathbf{x}} \quad \text{over} \quad \mathbf{x} \tag{4}$$

the maximum SNR of (4) can be obtained by a matched filter, denoted by $M_\mathbf{d}$, with the designed matched signal given by $\mathbf{d}$. In this case, the maximum SNR is obtained by letting $\mathbf{x} = \kappa \mathbf{d}$.

It is easily shown in [15] that maximizing (4) is equivalent to finding the maximum eigenvalue of the following generalized eigenvalue problem:

$$[P_U^\perp]^{-1}[P_U^\perp \mathbf{d}\alpha_p^2 \mathbf{d}^T P_U^\perp]\mathbf{x} = \lambda \mathbf{x}. \tag{5}$$

Since (4) and (5) present a two-class classification problem, the rank of the matrix on the left of (5) is one. This implies that the only nonzero eigenvalue is the maximum eigenvalue, which also solves (5) by letting $\mathbf{x} = \kappa \mathbf{d}$. Accordingly, this eigenvalue can be obtained as

$$
\begin{aligned}
\lambda_{\max} &= \text{SNR}_{\max} \\
&= \frac{\alpha_p^2}{\sigma^2} \frac{[\mathbf{d}^T P_U^\perp \mathbf{d}][\mathbf{d}^T P_U^\perp \mathbf{d}]}{\mathbf{d}^T P_U^\perp \mathbf{d}} = \frac{\alpha_p^2}{\sigma^2} \mathbf{d}^T P_U^\perp \mathbf{d}. \tag{6}
\end{aligned}
$$

Based on the approach outlined by (3)–(6), a mixed pixel classification can be carried out by a two-stage process, an undesired signature rejecter $P_U^\perp$ followed by a matched filter $M_\mathbf{d}$. More precisely, if we want to classify a target signature in a mixed pixel at the subpixel scale, say $\mathbf{d}$, based on model (1), we first apply $P_U^\perp$ to model (2) to eliminate U, then use the matched filter $M_\mathbf{d}$ to extract the $\mathbf{d}$ from (3). The operator coupling $P_U^\perp$ with $M_\mathbf{d}$ is called an orthogonal subspace classifier, the one derived in [10] and denoted by

$$q_{\text{OSP}}^T = M_\mathbf{d} P_U^\perp = \mathbf{d}^T P_U^\perp. \tag{7}$$

In the OSP approach, an assumption made about model (1) was that the complete knowledge of signatures M along with their fractions $\alpha_i = (\alpha_{i1} \ \alpha_{i2} \ \cdots \ \alpha_{ip})^T$ must be known *a priori*. Unfortunately, in the real image data experiments conducted in [10], it is not possible to know the abundance of spectral signatures in advance. Under this circumstance, we must estimate the signatures from the data themselves. So, model (1) may not be adequate in real-world applications and needs to be amended. One way to circumvent this problem was proposed in [13], where model (1) was reformulated based on *a posteriori* information obtained from images and given by

$$
\begin{aligned}
\mathbf{r} &= M\hat{\alpha}(\mathbf{r}) + \hat{\mathbf{n}}(\mathbf{r}) \\
&= \mathbf{d}\hat{\alpha}_p(\mathbf{r}) + U\hat{\gamma}(\mathbf{r}) + \hat{\mathbf{n}}(\mathbf{r}) \tag{8}
\end{aligned}
$$

where $\hat{\alpha}(\mathbf{r})$, $\hat{\alpha}_p(\mathbf{r})$, and $\hat{\gamma}(\mathbf{r})$ are estimates of α, α_p, and γ, respectively, based on the observed pixel itself $\mathbf{r}$. Because of this, the model depicted by (8) is called *a posteriori* model as opposed to model (1), which can be viewed as a Bayes or *a priori* model. It should be noted that for the purpose of simplicity, we will drop the dependency of $\mathbf{r}$ from all the notations of estimates throughout the rest of the paper.

III. LEAST SQUARES SUBSPACE PROJECTION CLASSIFIERS

In order to convert the Bayes model (1) to *a posteriori* model (8), we need to develop techniques that can estimate the abundance α for model (8). Three least squares subspace projection-based classifiers are proposed as follows. They are all unconstrained least squares error estimation techniques [16].
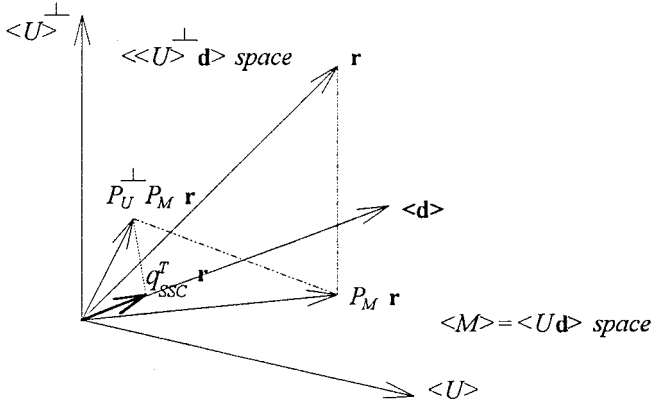
Fig. 1. Signature space orthogonal projection classifier (SSC).

A. SSC [13]

Using the least squares error as an optimal criterion yields the optimal least squares estimate of α, $\hat{\alpha}_{LS}(\mathbf{r})$ given by

$$\hat{\alpha}_{LS}(\mathbf{r}) = (M^T M)^{-1} M^T \mathbf{r}. \qquad (9)$$

Substituting (9) for the estimate of α in (8) results in

$$\mathbf{r} = M\hat{\alpha}_{LS} + \hat{\mathbf{n}}_{LS} \qquad (10)$$

where $\mathbf{r}$ is dropped from the notation of $\hat{\alpha}_{LS}(\mathbf{r})$ and

$$\hat{\mathbf{n}}_{LS} = \mathbf{r} - M\hat{\alpha}_{LS} = M(\alpha - \hat{\alpha}_{LS}) + \mathbf{n}. \qquad (11)$$

It should be noted that, if the estimate $\hat{\alpha}_{LS}(\mathbf{r})$ in (9) is replaced by the unconstrained ML estimate $\hat{\alpha}_{UMLE}$, (10) was the model used in [12].

Let $P_M = M(M^T M)^{-1} M^T$ denote the signature space orthogonal projector that projects $\mathbf{r}$ into the signature space $\langle M \rangle$. Then $P_M M = M$.

Applying P_M to model (10) and using (11) gives rise to

$$P_M = P_M M\hat{\alpha}_{LS} + P_M \hat{\mathbf{n}}_{LS} \qquad (12)$$
$$= M\hat{\alpha}_{LS} \qquad (13)$$

where the term $P_M \hat{\mathbf{n}}_{LS}$ vanishes in (13) since P_M annihilates $\hat{\mathbf{n}}_{LS}$.

By coupling P_M with the OSP classifier q_{OSP}^T given by (7), a new classifier q_{SSC}^T can be derived, called SSC and obtained as follows.

$$q_{SSC}^T = q_{OSP}^T P_M = \mathbf{d}^T P_U^\perp P_M. \qquad (14)$$

Fig. 1 shows the projection $q_{SSC}^T \mathbf{r}$.

Now, if q_{SSC}^T is applied to both *a priori* model (1) and *a posteriori* model (8), we obtain

$$q_{SSC}^T \mathbf{r} = \mathbf{d}^T P_U^\perp P_M M\alpha + q_{SSC}^T \mathbf{n}$$
$$= \mathbf{d}^T P_U^\perp P_M \mathbf{d}\alpha_p + q_{SSC}^T \mathbf{n} \qquad (15)$$

and

$$q_{SSC}^T \mathbf{r} = q_{SSC}^T (M\hat{\alpha}_{LS} + \mathbf{n}_{LS})$$
$$= \mathbf{d}^T P_U^\perp P_M \mathbf{d}\hat{\alpha}_p. \qquad (16)$$

Equating (15) and (16) yields

$$\mathbf{d}^T P_U^\perp P_M \mathbf{d}\hat{\alpha}_p = \mathbf{d}^T P_U^\perp P_M \mathbf{d}\alpha_p + q_{SSC}^T \mathbf{n}. \qquad (17)$$

Dividing (17) by $\mathbf{d}^T P_U^\perp P_M \mathbf{d}$, we obtain the estimate of α_p, denoted by $\hat{\alpha}_{SSC,p}$

$$\hat{\alpha}_{SSC,p} = \alpha_p + \frac{q_{SSC}^T \mathbf{n}}{\mathbf{d}^T P_U^\perp P_M \mathbf{d}} = \alpha_p + \frac{q_{SSC}^T \mathbf{n}}{\mathbf{d}^T P_U^\perp \mathbf{d}} \qquad (18)$$

where the last equality holds because of $\mathbf{d}^T P_U^\perp P_M \mathbf{d} = \mathbf{d}^T P_U^\perp \mathbf{d}$. It is worth noting that (18) is identical to (31) in [13].

Based on (18), a normalized SSC classifier $\tilde{q}_{SSC}^T$ can be defined by

$$\tilde{q}_{SSC}^T = \left(\mathbf{d}^T P_U^\perp P_M \mathbf{d}\right)^{-1} q_{SSC}^T = \left(\mathbf{d}^T P_U^\perp \mathbf{d}\right)^{-1} \mathbf{d}^T P_U^\perp P_M. \qquad (19)$$

Using (18) and (19), the estimation error is given by

$$\varepsilon_{SSC,p} = \hat{\alpha}_{SSC,p} - \alpha_p \qquad (20)$$
$$= \tilde{q}_{SSC}^T \mathbf{n} \qquad (21)$$
$$= \left(\mathbf{d}^T P_U^\perp \mathbf{d}\right)^{-1} \mathbf{d}^T P_U^\perp P_M \mathbf{n} \qquad (22)$$

with the corresponding $SNR_{SSC,max}$ given by the maximum eigenvalue

$$\lambda_{SSC,max} = \frac{\alpha_p^2}{\sigma^2} \frac{\left[\mathbf{d}^T P_U^\perp P_M \mathbf{d}\right]^2}{\mathbf{d}^T P_U^\perp \mathbf{d}} = \frac{\alpha_p^2}{\sigma^2} \mathbf{d}^T P_U^\perp \mathbf{d}. \qquad (23)$$

The last equality holds because of $\mathbf{d}^T P_U^\perp P_M \mathbf{d} = \mathbf{d}^T P_U^\perp \mathbf{d}$.

The following two comments are noteworthy.

1) Comparing (23) to (6), the maximum SNR (or eigenvalue) generated by q_{OSP}^T is exactly the same as that produced by $\tilde{q}_{SSC}^T$, due to $\mathbf{d}^T P_U^\perp P_M \mathbf{d} = \mathbf{d}^T P_U^\perp \mathbf{d}$. This shows that the SNR will not be decreased or increased by SSC. In other words, the maximum eigenvalue λ_{max} obtained by (6) remains unchanged after applying P_M. This observation can be explained by using Malinowski's error theory [17], in which the eigenvalues are divided into two classes, a primary set containing larger eigenvalues and a secondary set of smaller eigenvalues. The former corresponds to signatures with unremovable imbedded error, and the latter represents experimental errors that can be eliminated by designed techniques. According to this theory, the maximum eigenvalue belongs to the primary set, thus, it cannot be reduced by any means. Equation (23), equal to (6), illustrates this phenomenon. A detailed error analysis resulting from least squares subspace approaches was studied in [18].

2) The quantity $\mathbf{d}^T P_U^\perp \mathbf{d}$ in (23) determines $\lambda_{SSC,max}$ and provides very important information because the magnitude of SNR is determined by the degree of target signature $\mathbf{d}$ correlated with the undesired signatures in U. If $\mathbf{d}$ is very similar to one or more signatures in U, $\mathbf{d}^T P_U^\perp \mathbf{d}$ (i.e., the projection of $\mathbf{d}$ onto $\langle U \rangle^\perp$) will be small. This implies that SNR will be low. As a result, it will make the target signature discrimination very
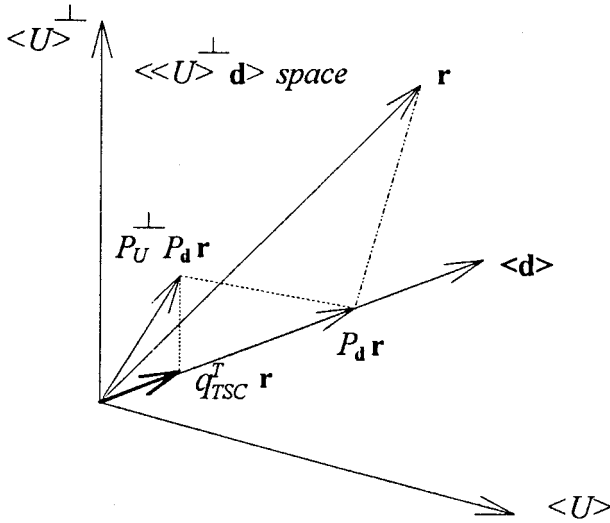
Fig. 2. Target signature space orthogonal projection classifier (TSC).

difficult. Therefore, the magnitude of $\mathbf{d}^T P_U^\perp \mathbf{d}$ can be used as a measure of the discrimination power of SSC. The larger the $\mathbf{d}^T P_U^\perp \mathbf{d}$, the better the discrimination.

B. TSC [19]

In SSC, P_M projects $\mathbf{r}$ onto the entire signature space $\langle M \rangle$. However, since we are only interested in classifying the target signature $\mathbf{d}$, a natural approach is to project the observed pixel $\mathbf{r}$ onto the target signature space $\langle \mathbf{d} \rangle$ rather than $\langle M \rangle$. This results in a second classifier, called TSC, denoted by q_{TSC}^T and given by

$$q_{\text{TSC}}^T = \mathbf{d}^T P_U^\perp P_{\mathbf{d}} \tag{24}$$

where $P_{\mathbf{d}}$ is defined in the same fashion that P_M was defined, i.e., $P_{\mathbf{d}} = \mathbf{d}(\mathbf{d}^T \mathbf{d})^{-1} \mathbf{d}^T$. Fig. 2 shows the operation of $q_{\text{TSC}}^T \mathbf{r}$.

Following the same argument for deriving (14) and (15), the counterparts of (17)–(20) for TSC can be obtained as follows:

$$(\mathbf{d}^T P_U^\perp \mathbf{d}) \hat{\alpha}_{\text{TSC},p} = (\mathbf{d}^T P_U^\perp \mathbf{d}) \alpha_p + q_{\text{TSC}}^T U\gamma + q_{\text{TSC}}^T \mathbf{n} \tag{25}$$

$$\tilde{q}_{\text{TSC}}^T = (\mathbf{d}^T P_U^\perp \mathbf{d})^{-1} q_{\text{TSC}}^T$$
$$= (\mathbf{d}^T P_U^\perp \mathbf{d})^{-1} \mathbf{d}^T P_U^\perp P_{\mathbf{d}} \tag{26}$$

$$\hat{\alpha}_{\text{TSC},p} = \alpha_p + \tilde{q}_{\text{TSC}}^T U\gamma + \tilde{q}_{\text{TSC}}^T \mathbf{n} \tag{27}$$

$$\varepsilon_{\text{TSC},p} = \hat{\alpha}_{\text{TSC},p} - \alpha_p = \tilde{q}_{\text{TSC}}^T U\gamma + \tilde{q}_{\text{TSC}}^T \mathbf{n}. \tag{28}$$

Comparing (27) to (18), there is an extra term $\tilde{q}_{\text{TSC}}^T U\gamma$ in (27) that does not appear in (18). The quantity $\tilde{q}_{\text{TSC}}^T U\gamma$ will be referred to as a signature bias due to the fact that the projector $P_{\mathbf{d}}$ used in $\tilde{q}_{\text{TSC}}^T$ cannot effectively eliminate U, as did $\tilde{q}_{\text{SSC}}^T$ in (20), and the signatures in U may be mixed into $\langle \mathbf{d} \rangle$ through $P_{\mathbf{d}}$. It is this bias to deteriorate the performance of TSC. However, if such a signature bias can be accurately estimated and removed prior to target classification, TSC will outperform SSC in terms of target classification power. This will be further studied in Section IV-B.
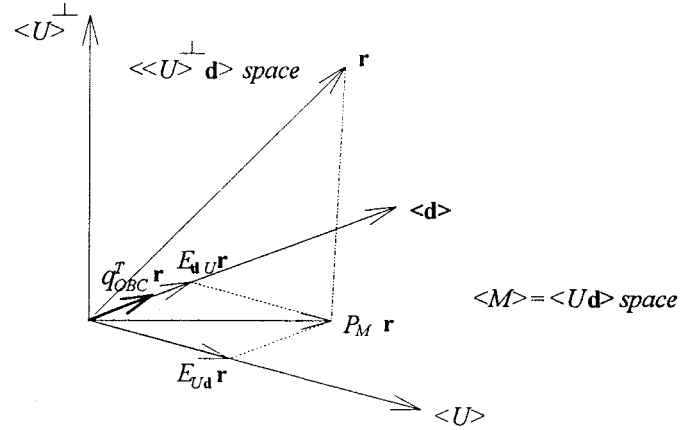

Fig. 3. Oblique subspace projection classifier (OBC).

The maximum SNR corresponding to (26) is given by the maximum eigenvalue

$$\lambda_{\text{TSC,max}} = \frac{\alpha_p^2}{\sigma^2} \frac{\left[\mathbf{d}^T P_U^\perp \mathbf{d} \right]^2}{\mathbf{d}^T P_{\mathbf{d}} P_U^\perp P_{\mathbf{d}} \mathbf{d}}. \tag{29}$$

Unlike (23), which is solely determined by $\mathbf{d}^T P_U^\perp \mathbf{d}$, the SNR of TSC is determined by the ratio of $[\mathbf{d}^T P_U^\perp \mathbf{d}]^2$ to $\mathbf{d}^T P_{\mathbf{d}} P_U^\perp P_{\mathbf{d}} \mathbf{d}$. If two operators $P_{\mathbf{d}}$ and $P_U^\perp$ commute, $\mathbf{d}^T P_{\mathbf{d}} P_U^\perp P_{\mathbf{d}} \mathbf{d} = \mathbf{d}^T P_U^\perp \mathbf{d}$ and (29) is reduced to (23). In this situation, TSC is the same as SSC.

C. OBC [19]

When TSC was compared to SSC, there was a signature bias resulting from TSC. The attribute of the bias is mainly due to the fact that $\langle \mathbf{d} \rangle$ is not orthogonal to $\langle U \rangle$, i.e., $\langle \mathbf{d} \rangle \cap \langle U \rangle \neq \emptyset$. In order to resolve this dilemma, we need to find a projector that can separate $\mathbf{d}$ from U. Fortunately, there exists such a projection in [14], called an oblique subspace projection, which designates $\langle \mathbf{d} \rangle$ as its range space and $\langle U \rangle$ as its null space. In other words, the oblique subspace projection maps an observed pixel into the target signature space, as did $P_{\mathbf{d}}$ in TSC, while annihilating all undesired signatures by mapping U into the null space, as did $P_U^\perp$ in SSC. The former operation projects the target signature onto the target signature space, and the latter projection forces the bias to zero. Of course, there is a tradeoff for doing so. That is, the oblique subspace projection is no longer orthogonal, although it is still an idempotent projection. More interestingly, it was shown in [14] that the signature space orthogonal projector P_M used in SSC can be decomposed as a sum of two oblique projectors, of which the proposed oblique subspace projection turns out to one of them.

Let E_{XY} be a projector with range space X and null space Y. The P_M can be decomposed as described in Fig. 3 and expressed by

$$P_M = E_{\mathbf{d}U} + E_{U\mathbf{d}} \tag{30}$$

with

$$E_{\mathbf{d}U} = \mathbf{d}(\mathbf{d}^T P_U^\perp \mathbf{d})^{-1} \mathbf{d}^T P_U^\perp \tag{31}$$

$$E_{U\mathbf{d}} = U(U^T P_{\mathbf{d}}^\perp U)^{-1} U^T P_{\mathbf{d}}^\perp \tag{32}$$

particularly, $E_{\mathbf{d}U}\mathbf{d} = \mathbf{d}$ and $E_{\mathbf{d}U}U = 0$.

In analogy with (14), an OBC denoted by q_{OBC}^T can be constructed via (31) by

$$q_{\text{OBC}}^T = \mathbf{d}^T E_{\text{d}U}. \tag{33}$$

Applying (33) to model (1) and model (8) results in

$$q_{\text{OBC}}^T \mathbf{r} = \mathbf{d}^T E_{\text{d}U} \mathbf{r} = \mathbf{d}^T \mathbf{d} \alpha_p + \mathbf{d}^T E_{\text{d}U} \mathbf{n} \tag{34}$$

and

$$\begin{aligned} q_{\text{OBC}}^T \mathbf{r} &= \mathbf{d}^T E_{\text{d}U} \mathbf{r} \\ &= \mathbf{d}^T \mathbf{d} \hat{\alpha}_{\text{OBC},p} + \mathbf{d}^T E_{\text{d}U} \hat{\mathbf{n}} \\ &= \mathbf{d}^T \mathbf{d} \hat{\alpha}_{\text{OBC},p} \end{aligned} \tag{35}$$

where $\mathbf{d}^T E_{\text{d}U} \hat{\mathbf{n}} = 0$.

Equating (34) and (35) yields

$$\mathbf{d}^T \mathbf{d} \hat{\alpha}_{\text{OBC},p} = \mathbf{d}^T \mathbf{d} \alpha_p + \mathbf{d}^T E_{\text{d}U} \mathbf{n} \tag{36}$$

and

$$\hat{\alpha}_{\text{OBC},p} = \alpha_p + (\mathbf{d}^T \mathbf{d})^{-1} \mathbf{d}^T E_{\text{d}U} \mathbf{n}. \tag{37}$$

So, from (37), a normalized OBC denoted by $\tilde{q}_{\text{OBC}}^T$ can be defined by

$$\tilde{q}_{\text{OBC}}^T = (\mathbf{d}^T \mathbf{d})^{-1} \mathbf{d}^T E_{\text{d}U} \mathbf{n} \tag{38}$$

which produces the maximum SNR given by the maximum eigenvalue

$$\lambda_{\text{OBC,max}} = \frac{\alpha_p^2}{\sigma^2} \left[\mathbf{d}^T P_U^\perp \mathbf{d} \right]^{-1}. \tag{39}$$

Equation (39) turns out to be equal to (23), i.e., $\lambda_{\text{OBC,max}} = \lambda_{\text{SSC,max}}$.

From (37), the estimation error $\varepsilon_{\text{OBC},p}$ can be obtained as

$$\varepsilon_{\text{OBC},p} = \hat{\alpha}_{\text{OBC},p} - \alpha_p = \left[\mathbf{d}^T P_U^\perp \mathbf{d} \right]^{-1} \mathbf{d}^T P_U^\perp \mathbf{n}. \tag{40}$$

Two interesting observations can be made from (40). If we compare (40) to the $\hat{\alpha}_p$ obtained in [12], they are identical. The estimate $\hat{\alpha}_p$ can also be derived by partitioning the signature matrix M into $\mathbf{d}$ and U, as was done in [12]. This implies that, when the UMLE and OBC are applied to linear spectral mixing problems specified by (1), they both arrive at identical classifiers. If we further compare (40) to (7), there is an extra constant $\left[\mathbf{d}^T P_U^\perp \mathbf{d} \right]^{-1}$ appearing in (40). This constant actually results from abundance estimation error, as will be shown in the next section. Since model (1) assumes the complete knowledge of the abundance α, there is no need for estimating abundance in model (1). Thus, no constant was included in (7). In other words, the constant $\left[\mathbf{d}^T P_U^\perp \mathbf{d} \right]^{-1}$ accounts for the distinction between the *a priori* model (1) and *a posteriori* model (8).

IV. ESTIMATION ERROR EVALUATED BY ROC ANALYSIS

In the previous section, three estimation errors were derived for SSC, TSC, and OBC. These errors are the penalties resulting from inaccurate estimation of unknown signature abundances. In order to evaluate the error performance of these classifiers, we cast their associated estimation errors as a standard signal detection problem [20], where α_p is viewed as a true target signature abundance corrupted by the noise represented by their estimation errors ε_p. By virtue of this formulation, these three classifiers can also be interpreted as subpixel target detectors that can be used to detect the presence of a target signature in a mixed pixel. The effectiveness of these detectors depends upon the accuracy of the abundance estimate $\hat{\alpha}_p$ and can be evaluated by the ROC analysis via the Neyman–Pearson detection theory. An ROC curve is a graph plotted by the detection power versus the false alarm probability. Instead of using ROC curves as a performance criterion in [13], we define a measure called detection rate (DR), which calculates the area under an ROC curve for the effectiveness of the detector. Obviously, DR always lies between 1/2 and one. The worst case occurs when DR = 1/2, i.e., the detection power is equal to the false alarm probability, which implies that the classifier is worthless. On the other hand, the best case occurs only when DR = 1, namely, the detection power is always one, regardless of the false alarm probability. This ROC analysis has been widely used in diagnostic imaging [21], [22] for evaluation of computer-aided diagnostic methods, in which the detection power is measured by the true-positive probability and the false alarm probability is represented by the false-positive probability.

Assume that z is the projection resulting from a classifier q^T applied to the observed pixel $\mathbf{r}$. A signal detection model based on z can be described by a test of two hypotheses as follows:

$$\begin{aligned} &H_0: z = q^T \mathbf{n} \cong p_0(z) \\ &\textit{versus} \\ &H_1: z = \alpha_p + q^T \mathbf{n} \cong p_1(z) \end{aligned} \tag{41}$$

where the null hypothesis H_0 and the alternative hypothesis H_1 represent the case of noise alone $z = q^T \mathbf{n}$ (i.e., estimation error) and the case of the true target signature α_p present in the z, respectively.

The Neyman–Pearson detector associated with (41) is given by

$$\delta_{\text{NP}}(z) = \begin{cases} 1, & z = q^T \mathbf{n} \geq \tau \\ 0, & z = q^T \mathbf{n} < \tau \end{cases} \tag{42}$$

where $\mathbf{r}$ is the observed pixel.

From (42), we can also define the false alarm probability and detection power (detection probability) [20] as follows:

$$P_F = \int_\tau^\infty p_0(z) \, dz \tag{43}$$

$$P_D = \int_\tau^\infty p_1(z) \, dz. \tag{44}$$

Using (44), the detection power specifies the capability of the detector δ_{NP} in detecting the true target signature α_p.

Therefore, the higher the detection power, the smaller the estimation error, the better classification the classifier.

A. SSC

Substituting $\tilde{q}_{SSC}^T$ specified by (17) for q^T in (41) results in a subpixel target detection model given by

$$
\begin{aligned}
&H_0: z = \tilde{q}_{SSC}^T \mathbf{n} = \hat{n}_{SSC} \cong p_0(z) \\
&versus \\
&H_1: z = \alpha_p + \hat{n}_{SSC} \cong p_1(z)
\end{aligned}
\tag{45}
$$

where the noise $\hat{n}_{SSC}$ is generated by the estimation error produced by SSC. The hypothesis H_0 in (45) represents the case in which the mixed pixel does not contain the target signature $\mathbf{d}$, while H_1 indicates the presence of $\mathbf{d}$ in the mixed pixel.

Based on (20)–(22) and (45), we obtain the error covariance matrix Σ_{SSC} for the estimation error $\hat{n}_{SSC}$ resulting from (20), as follows:

$$
\begin{aligned}
\Sigma_{SSC} &= E\left[\varepsilon_{SSC} \varepsilon_{SSC}^T\right] = E\left[\hat{n}_{SSC} \hat{n}_{SSC}^T\right] \\
&= \tilde{q}_{SSC}^T E[\mathbf{n}\mathbf{n}^T] \tilde{q}_{SSC} \\
&= \sigma^2 \frac{\mathbf{d}^T P_U^\perp P_M P_M P_U^\perp \mathbf{d}}{\mathbf{d}^T P_U^\perp P_M \mathbf{d}\mathbf{d}^T P_U^\perp P_M \mathbf{d}} \\
&= \frac{\sigma^2}{\mathbf{d}^T P_U^\perp P_M \mathbf{d}} = \frac{\sigma^2}{\mathbf{d}^T P_U^\perp \mathbf{d}}.
\end{aligned}
\tag{46}
$$

Assume that $\mathbf{n}$ is a white Gaussian noise with zero mean and covariance matrix $\sigma^2 I$. Substituting (46) into (45) yields

$$
\begin{aligned}
p_0(z) &\approx N\left(0, \sigma^2 (\mathbf{d}^T P_U^\perp \mathbf{d})^{-1}\right) \\
p_1(z) &\approx N\left(\alpha_p, \sigma^2 (\mathbf{d}^T P_U^\perp \mathbf{d})^{-1}\right)
\end{aligned}
\tag{47}
$$

and the threshold τ_{SSC} given by

$$
\tau_{SSC} = \sigma (\mathbf{d}^T P_U^\perp \mathbf{d})^{-1/2} \Phi^{-1}(1 - P_F)
\tag{48}
$$

where $\Phi(x)$ is the cumulative distribution of the standard Gaussian random variable given by

$$
\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt.
\tag{49}
$$

The desired detection power $P_{SSC,D}$ can be derived using (49)

$$
P_{SSC,D} = 1 - \Phi\left(\Phi^{-1}(1 - P_F) - \frac{\alpha_p}{\sigma (\mathbf{d}^T P_U^\perp \mathbf{d})^{-1/2}}\right)
\tag{50}
$$

$$
= 1 - \Phi\left(\Phi^{-1}(1 - P_F) - \sqrt{\lambda_{SSC,max}}\right).
\tag{51}
$$

In (50), $P_{SSC,D}$ is expressed in terms of $\mathbf{d}^T P_U^\perp \mathbf{d}$, which indicates the degree of the correlation between $\mathbf{d}$ and the projection $P_U^\perp \mathbf{d}$. On the other hand, (51) illustrates that the detection power is measured by the magnitude of SNR or, equivalently, the maximum eigenvalue. So, both (50) and (51) can be used to evaluate the performance of SSC $\tilde{q}_{SSC}^T$.

It is important to note that, for a fixed false alarm probability P_F, (50) shows that the discrimination power of $\tilde{q}_{SSC}^T$ is proportional to the value of $\mathbf{d}^T P_U^\perp \mathbf{d}$, whereas (51) suggests

that the discrimination power of $\tilde{q}_{SSC}^T$ is proportional to the magnitude of the maximum SNR or $\lambda_{SSC,max}$. This makes perfect sense since the former indicates the degree of similarity between the target signature $\mathbf{d}$ and the undesired signatures in U, and the latter demonstrates that SNR determines the performance. More interestingly, if we interpret $\mathbf{d}^T P_U^\perp \mathbf{d}$ as an inner product of $\mathbf{d}$ and the projection $P_U^\perp \mathbf{d}$, $\mathbf{d}^T P_U^\perp \mathbf{d}$ measures how much projection of $\mathbf{d}$ is projected onto the orthogonal complement space $\langle U \rangle^\perp$. The more the projection, the less similarity between $\mathbf{d}$ and U, thus, the better the discrimination. Additionally, the detection power equation (50) is also a function of α_p. The higher the α_p, the better the detection of $\mathbf{d}$. Despite no presence of α_p in (51), it is implicitly included in $\lambda_{SSC,max}$, given by (23). So, both (50) and (51) are also determined by the abundance strength of α_p. All of these relationships can be well explained by the ROC analysis.

B. TSC

The classifier $\tilde{q}_{TSC}^T$, specified by (26), produces the following subpixel detection problem associated with the estimation error $\hat{\alpha}_{TSC,p}$ modeled by the noise $\hat{n}_{TSC}$:

$$
\begin{aligned}
&H_0: z = \tilde{q}_{TSC}^T \mathbf{n} = \hat{n}_{TSC} \cong p_0(z) \\
&versus \\
&H_1: z = \alpha_p + \hat{n}_{TSC} \cong p_1(z)
\end{aligned}
\tag{52}
$$

where

$$
b = \tilde{q}_{TSC}^T U \gamma = (\mathbf{d}^T P_U^\perp \mathbf{d})^{-1} \mathbf{d}^T P_U^\perp P_{\mathbf{d}} U \gamma
\tag{53}
$$

is a signature bias resulting from γ, i.e., the unknown abundance of the undesired signature in U. The covariance matrix of $\hat{n}_{TSC}$, Σ_{TSC} is then given by

$$
\begin{aligned}
\Sigma_{TSC} &= E\left[\varepsilon_{TSC} \varepsilon_{TSC}^T\right] = E\left[\hat{n}_{TSC} \hat{n}_{TSC}^T\right] \\
&= \tilde{q}_{TSC}^T E[\mathbf{n}\mathbf{n}^T] \tilde{q}_{TSC} \\
&= \sigma^2 \frac{\mathbf{d}^T P_U^\perp P_{\mathbf{d}} P_U^\perp \mathbf{d}}{\mathbf{d}^T P_U^\perp P_{\mathbf{d}} \mathbf{d}\mathbf{d}^T P_U^\perp P_{\mathbf{d}} \mathbf{d}} \\
&= \sigma^2 \frac{\mathbf{d}^T P_U^\perp P_{\mathbf{d}} P_U^\perp \mathbf{d}}{(\mathbf{d}^T P_U^\perp \mathbf{d})^2}.
\end{aligned}
\tag{54}
$$

The probability density functions (pdf's) $p_0(z)$ and $p_1(z)$ under H_0 and H_1 hypotheses and the threshold τ_{TSC} can derived as follows:

$$
\begin{aligned}
p_0(z) &\approx N\left(0, \sigma^2 (\mathbf{d}^T P_U^\perp \mathbf{d})^{-2} \mathbf{d}^T P_U^\perp P_{\mathbf{d}} P_U^\perp \mathbf{d}\right) \\
p_1(z) &\approx N\left(\alpha_p + b, \sigma^2 (\mathbf{d}^T P_U^\perp \mathbf{d})^{-2} \mathbf{d}^T P_U^\perp P_{\mathbf{d}} P_U^\perp \mathbf{d}\right)
\end{aligned}
\tag{55}
$$

and

$$
\tau_{TSC} = \left[\sigma (\mathbf{d}^T P_U^\perp \mathbf{d})^{-1} \sqrt{\mathbf{d}^T P_U^\perp P_{\mathbf{d}} P_U^\perp \mathbf{d}}\right] \Phi^{-1}(1 - P_F)
\tag{56}
$$

TABLE I
TARGET SIGNATURES AND UNDESIRED SIGNATURES CHOSEN FOR DATA SETS 1 AND 2

	undesired signatures	target signature
Data set 1	dry grass, red soil	creosote leaves
Data set 2	sage brush, black brush	creosote leaves

TABLE II
50 SIMULATED PIXELS WITH ASSIGNED RELATIVE ABUNDANCES OF SIGNATURES FOR DATA SETS 1 AND 2

	pixel numbers 1-10	pixel numbers 11-20	pixel numbers 21-30	pixel numbers 31-40	pixel numbers 41-50
target signature	1 %	5 %	10 %	15 %	20 %
undesired signature 1	49.5 %	47.5 %	45 %	42.5 %	40 %
undesired signature 2	49.5 %	47.5 %	45 %	42.5 %	40 %
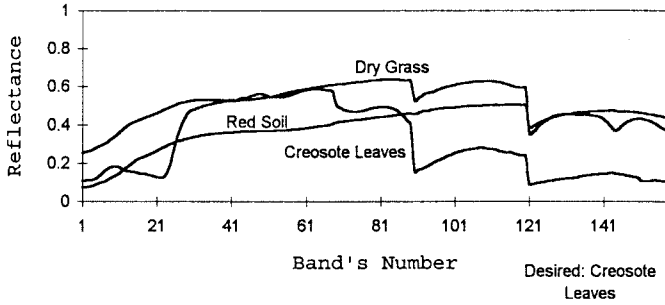
Fig. 4. Data set 1.
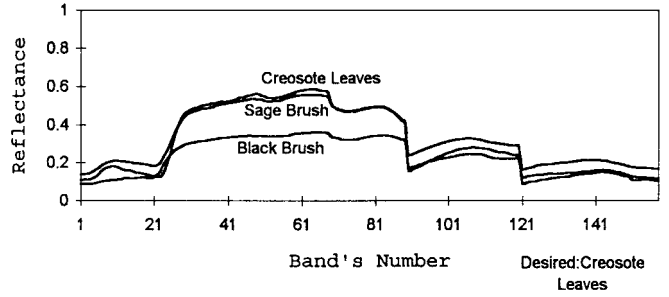
Fig. 5. Data set 2.

where $\Phi(x)$ is given by (49). From (55) and (56), the desired detection power $P_{\text{TSC},D}$ is

$$P_{\text{TSC},D} = 1 - \Phi$$
$$\left(\Phi^{-1}(1 - P_F) - \frac{\alpha_p + b}{\sigma \left(\mathbf{d}^T P_U^\perp \mathbf{d}\right)^{-1} \sqrt{\mathbf{d}^T P_U^\perp P_{\mathbf{d}} P_U^\perp \mathbf{d}}} \right)$$
$$= 1 - \Phi\left(\Phi^{-1}(1 - P_F)\right.$$
$$\left. - \left(\sqrt{\lambda_{\text{TSC,max}}} + \frac{\alpha_p + b}{\sigma \left(\mathbf{d}^T P_U^\perp \mathbf{d}\right)^{-1} \sqrt{\mathbf{d}^T P_U^\perp P_{\mathbf{d}} P_U^\perp \mathbf{d}}} \right) \right)$$
$$\tag{57}$$

where b is given by (53). In analogy with $\tilde{q}_{\text{SSC}}^T$, the detection power of $\tilde{q}_{\text{TSC}}^T$ can be evaluated by (57).

It is worth noting that, unlike $\tilde{q}_{\text{SSC}}^T$, the detection power of $\tilde{q}_{\text{TSC}}^T$ is also a function of the signature bias b. As mentioned previously, this bias is a result of the effect of the projector $P_{\mathbf{d}}$ mixing U into $\langle \mathbf{d} \rangle$, so that the bias b cannot be removed. This is, however, not the case of SSC, where U can be properly

projected by P_M into the signature space $\langle M \rangle$ and eliminated in a subsequent projection carried out by $P_U^\perp$. Nonetheless, if $\mathbf{d}$ is orthogonal to $\langle U \rangle$, then $P_{\mathbf{d}} U = 0$. This implies that no projection $\langle U \rangle$ can be projected by $P_{\mathbf{d}}$ into $\langle \mathbf{d} \rangle$ and, thus, the bias b will be zero.

1) Signature Bias Analysis for TSC Let V be a subset of M and $\langle V \rangle$ be the space linearly spanned by V. Assume that v is the dimensionality of $\langle V \rangle$, i.e., $v = \dim(\langle V \rangle)$. We can define the signature bias projection (SBP) in the following to measure the magnitude of the projection of U into $\langle V \rangle$:

$$\text{SBP} = \frac{p - \dim(\langle V \rangle)}{p} = \frac{p - v}{p}. \tag{58}$$

From (58), SBP describes the degree of the signature bias projection by which $\mathbf{d}$ is garbled and mixed by U. In the case of TSC, $V = \mathbf{d}$ with $v = 1$ and SBP is $\frac{p-1}{p}$. Then, SBP is gradually reduced by increasing v and finally reduced to zero when it reaches $V = M$ and $v = p$, in which case, TSC becomes SSC. This phenomenon was justified by computer simulations in [19].
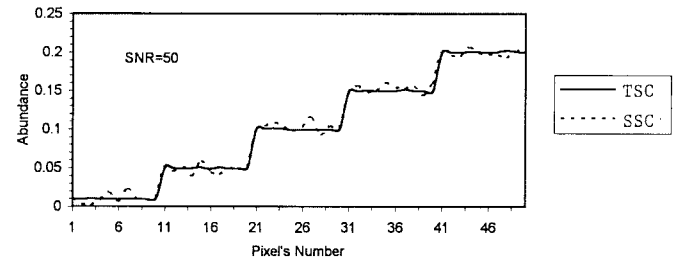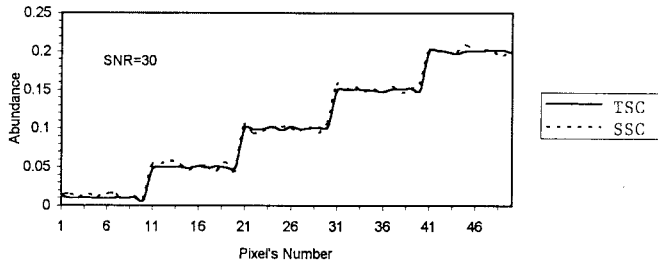
TABLE III
DR FOR DATA SET 1

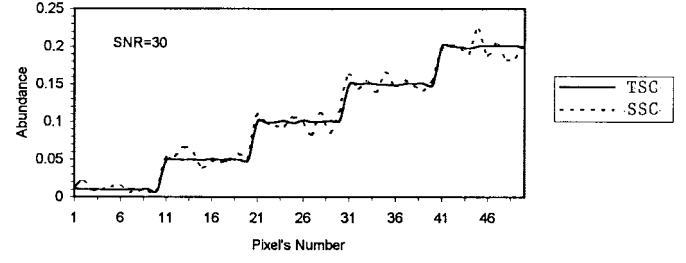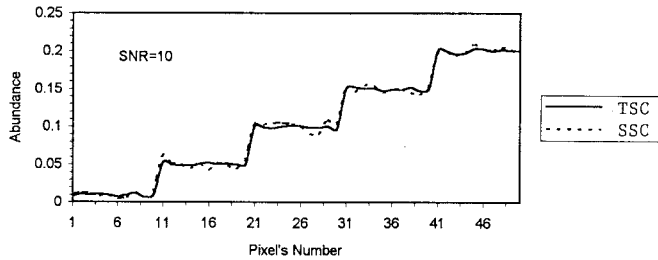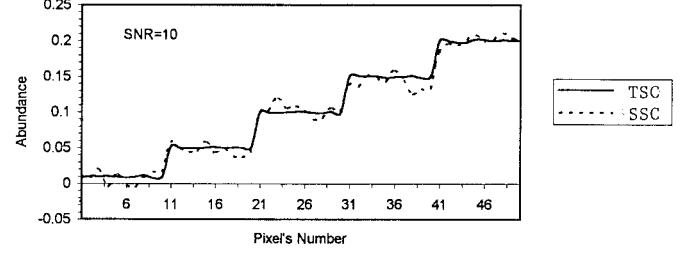| data set 1	$\alpha_p / \sigma = 0.5$	$\alpha_p / \sigma = 0.3$	$\alpha_p / \sigma = 0.1$
SSC = OBC	DR = 0.71	DR = 0.63	DR = 0.55
TSC without bias	DR = 0.95	DR = 0.83	DR = 0.63
TSC with bias	DR = 0.95	DR = 0.83	DR = 0.63
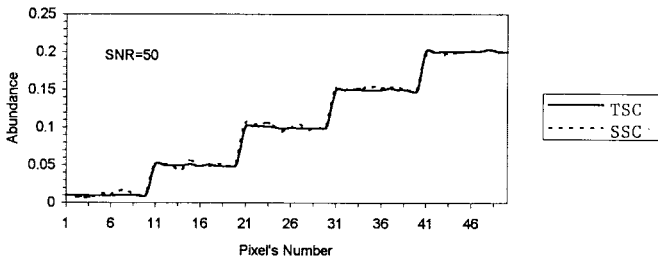
Fig. 6. Performance curves produced by SSC and TSC without bias, respectively, based on data set 1 with (a) SNR $= 50:1$, (b) SNR $= 30:1$, and (c) SNR $= 10:1$, where the solid curve is TSC and the dotted curve is SSC.

Fig. 7. Performance curves produced by SSC and TSC without bias, respectively, based on data set 1 with (a) SNR $= 50:1$, (b) SNR $= 30:1$, and (c) SNR $= 10:1$, where the solid curve is TSC and the dotted curve is SSC.

C. OBC

Analogous to (45) and (52), the OBC $\tilde{q}_{\text{TSC}}^T$ given by (38) generates the following subpixel target detection problem:
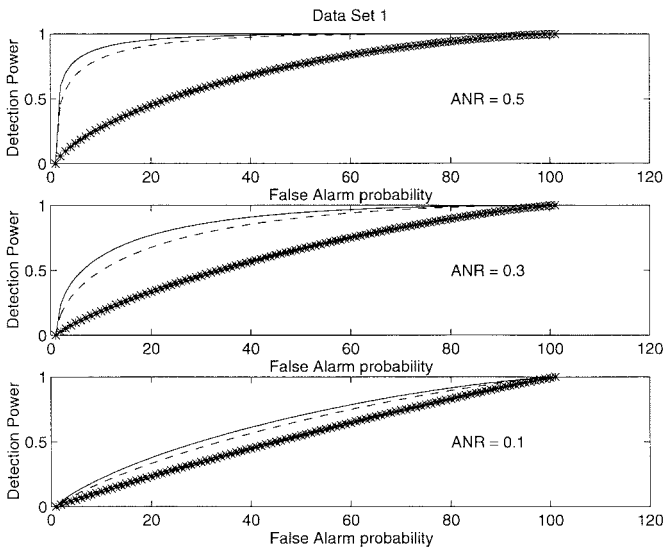
$$
\begin{aligned}
H_0: & \ z = \tilde{q}_{\text{OBC}}^T \mathbf{n} = \hat{n}_{\text{OBC}} \cong p_0(z) \\
& versus \\
H_1: & \ z = \alpha_p + \hat{n}_{\text{OBC}} \cong p_1(z)
\end{aligned}
\tag{59}
$$

with the noise covariance matrix Σ_{OBC} given by

$$
\begin{aligned}
\Sigma_{\text{OBC}} &= E\left[\varepsilon_{\text{OBC}}\varepsilon_{\text{OBC}}^T\right] = E\left[\hat{n}_{\text{OBC}}\hat{n}_{\text{OBC}}^T\right] \\
&= \tilde{q}_{\text{OBC}}^T E[\mathbf{nn}^T]\tilde{q}_{\text{OBC}} \\
&= (\mathbf{d}^T P_U^\perp \mathbf{d})^{-1}\mathbf{d}^T P_U^\perp E[\mathbf{nn}^T](\mathbf{d}^T P_U^\perp)^T(\mathbf{d}^T P_U^\perp \mathbf{d})^{-1} \\
&= \frac{\sigma^2}{\mathbf{d}^T P_U^\perp \mathbf{d}}.
\end{aligned}
\tag{60}
$$

TABLE IV
DR FOR DATA SET 2

Data set 2	$\alpha_p/\sigma = 0.5$	$\alpha_p/\sigma = 0.3$	$\alpha_p/\sigma = 0.1$
SSC = OBC	DR = 0.57	DR = 0.54	DR = 0.52
TSC without bias	DR = 0.95	DR = 0.83	DR = 0.63
TSC with bias	DR = 0.95	DR = 0.83	DR = 0.63

Fig. 8. ROC curves produced by SSC, TSC without bias, and TSC with bias, respectively, based on data set 1 with (a) $\alpha_3/\sigma = 0.5$, (b) $\alpha_3/\sigma = 0.3$, and (c) $\alpha_3/\sigma = 0.1$, where the asterisked, dotted, and solid curves are generated by SSC, TSC without bias, and TSC with bias, respectively.
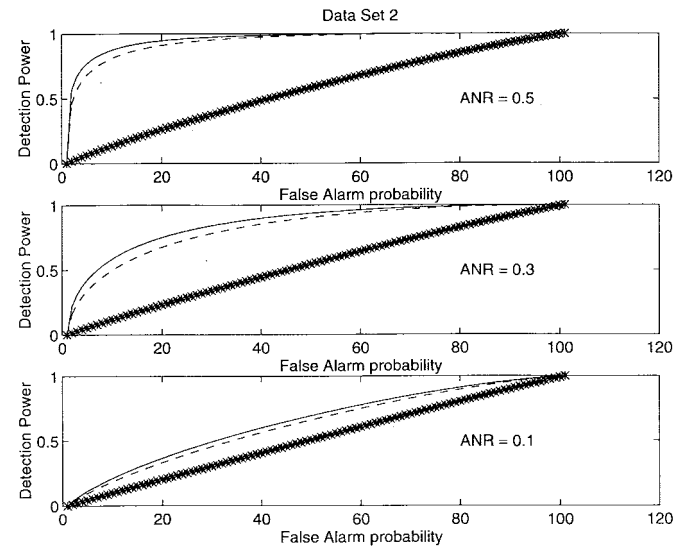
Fig. 9. ROC curves produced by SSC, TSC without bias, and TSC with bias, respectively, based on data set 2 with (a) $\alpha_3/\sigma = 0.5$, (b) $\alpha_3/\sigma = 0.3$, and (c) $\alpha_3/\sigma = 0.1$, where the asterisked, dotted, and solid curves are generated by SSC, TSC without bias, and TSC with bias, respectively.

Substituting (60) into (59), we obtain the same pdf's $p_0(z)$ and $p_1(z)$ given by (47) for the detection problem equation (59). As a result, the threshold τ_{OBC} is equal to τ_{SSC} given by (48) and the desired detection power $P_{OBC,D}$

$$P_{OBC,D} = 1 - \Phi\left(\Phi^{-1}(1 - P_F) - \frac{\alpha_p}{\sigma(\mathbf{d}^T P_U^\perp \mathbf{d})^{-1/2}} \right) \quad (61)$$

$$= 1 - \Phi\left(\Phi^{-1}(1 - P_F) - \sqrt{\lambda_{SSC,\max}} \right) \quad (62)$$

is identical to the detection power $P_{SSC,D}$. This implies that $\tilde{q}_{OBC}^T$ is essentially equivalent to $\tilde{q}_{SSC}^T$, in the sense that they both generate identical ROC curves.

V. COMPUTER SIMULATIONS AND EXPERIMENTAL RESULTS

In this section, computer simulations and a scene of AVIRIS data will be used to evaluate the relative performance of three proposed classifiers: SSC, TSC, and OBC.

A. Computer Simulations

In the following simulations, two laboratory data sets in [10] were used and each data set contains three field spectral reflectances with spectral range from 0.4 to 2.5 μm. In this case, the signature matrix is $M = (\mathbf{m}_1 \ \mathbf{m}_2 \ \mathbf{m}_3)$, consisting of three spectral signatures with abundance given by $\alpha = (\alpha_1 \ \alpha_2 \ \alpha_3)^T$. We also let $\mathbf{d} = \mathbf{m}_3$ be the target signature specified by abundance α_3 and $U = (\mathbf{m}_1, \mathbf{m}_2)$ be the matrix made up of undesired signatures with abundances given by $(\alpha_1 \ \alpha_2)^T$. Data set 1 shown in Fig. 4 contains dry grass, red soil, and creosote leaves, with creosote leaves designated as $\mathbf{d}$ (see Table I). Data set 2 is shown in Fig. 5 with U made up of sage brush, black brush, and $\mathbf{d}$ = creosote leaves (see Table I). The difference between data set 1 and data set 2 is that the spectrum of the target signature in data set 2 is very similar to that of sage brush, while all three spectra in data set 1 are distinguishable. Fifty mixed pixels are simulated with abundances in accordance with Table II. In this paper, we
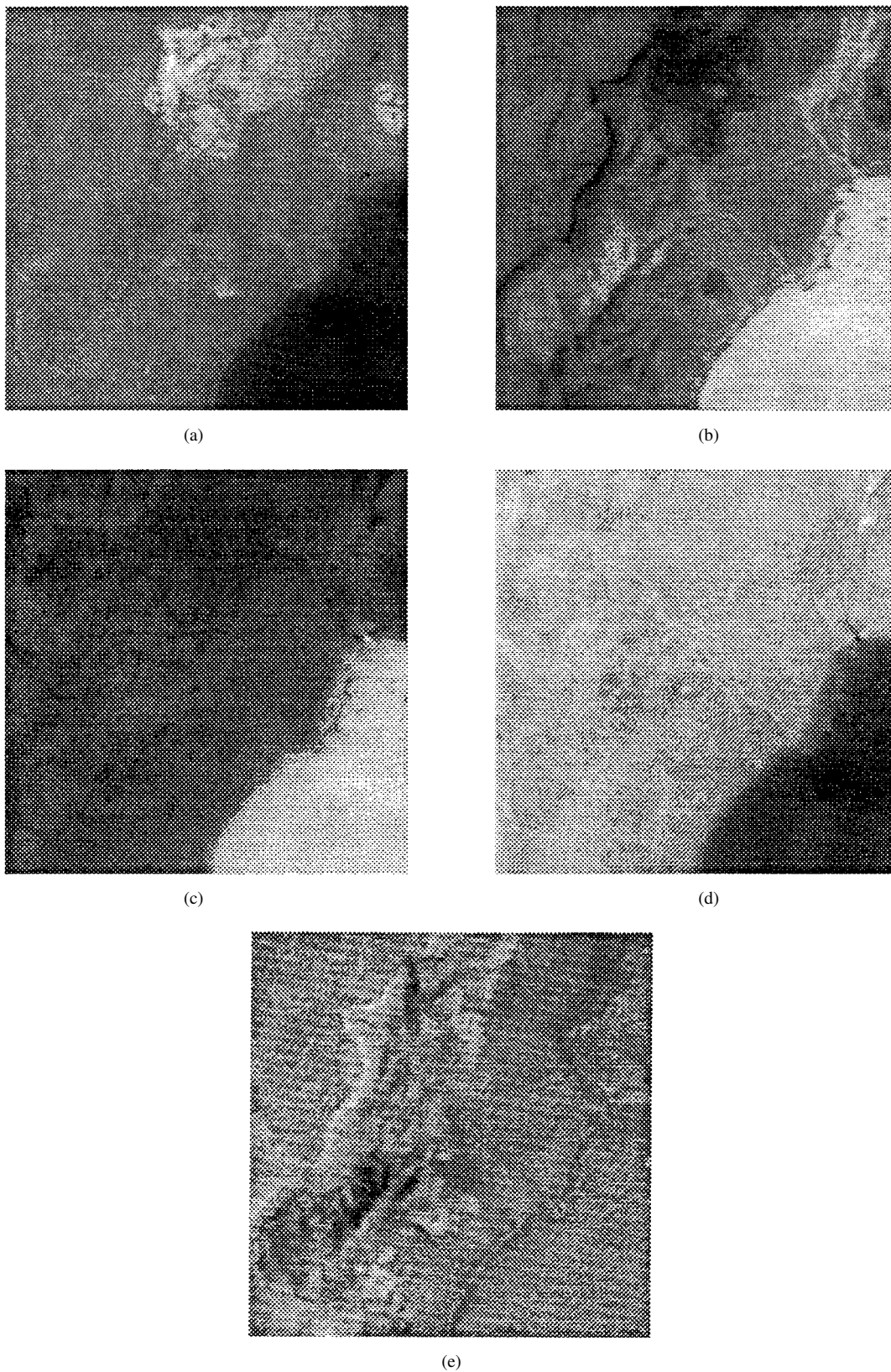
Fig. 10. LCVF scene of AVIRIS data classified by OSP with (a) cinders as the target signature, (b) rhyolite as the target signature, (c) playa as the target signature, (d) vegetation as the target signature, and (e) shade.

only consider the case in which the undesired signature vectors share their abundances evenly for illustrative purpose. For the case of uneven abundances in U, refer to [19], which shows no appreciable difference in the experiments, particularly for the desired signature with high abundance. In addition to signature vectors, different Gaussian noise levels are also simulated and
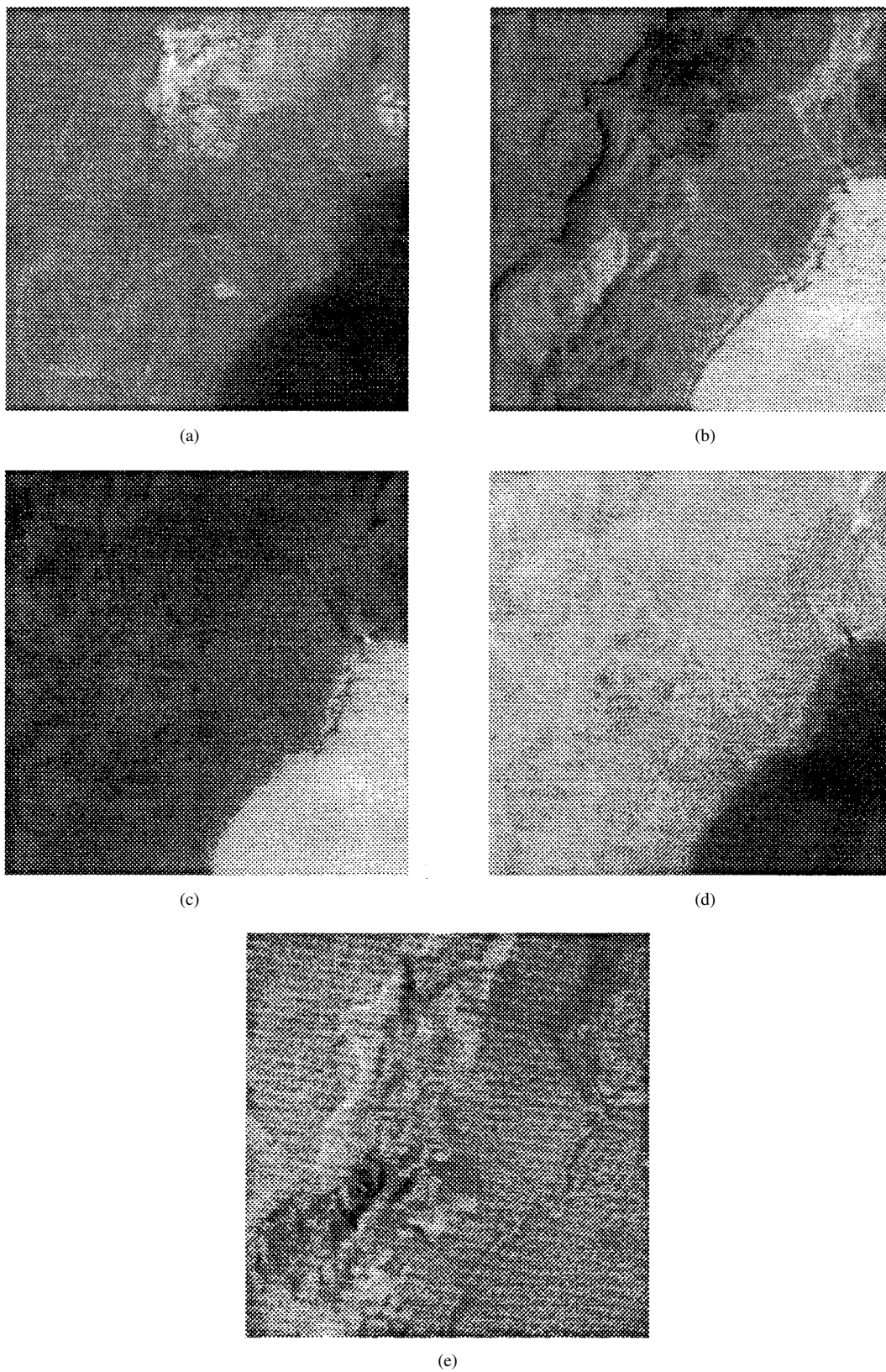
Fig. 11. LCVF scene of AVIRIS data classified by SSC with (a) cinders as the target signature, (b) rhyolite as the target signature, (c) playa as the target signature, (d) vegetation as the target signature, and (e) shade.

added to generate SNR 50:1, 30:1, and 10:1, with the SNR defined in [10] as 50% reflectance divided by the standard deviation of the noise. Since OBC is essentially equivalent to SSC, only results generated by SSC are plotted in all figures. Figs. 6 and 7 show performance comparisons between SSC and TSC for two data sets (Fig. 6 for data set 1 and Fig. 7 for
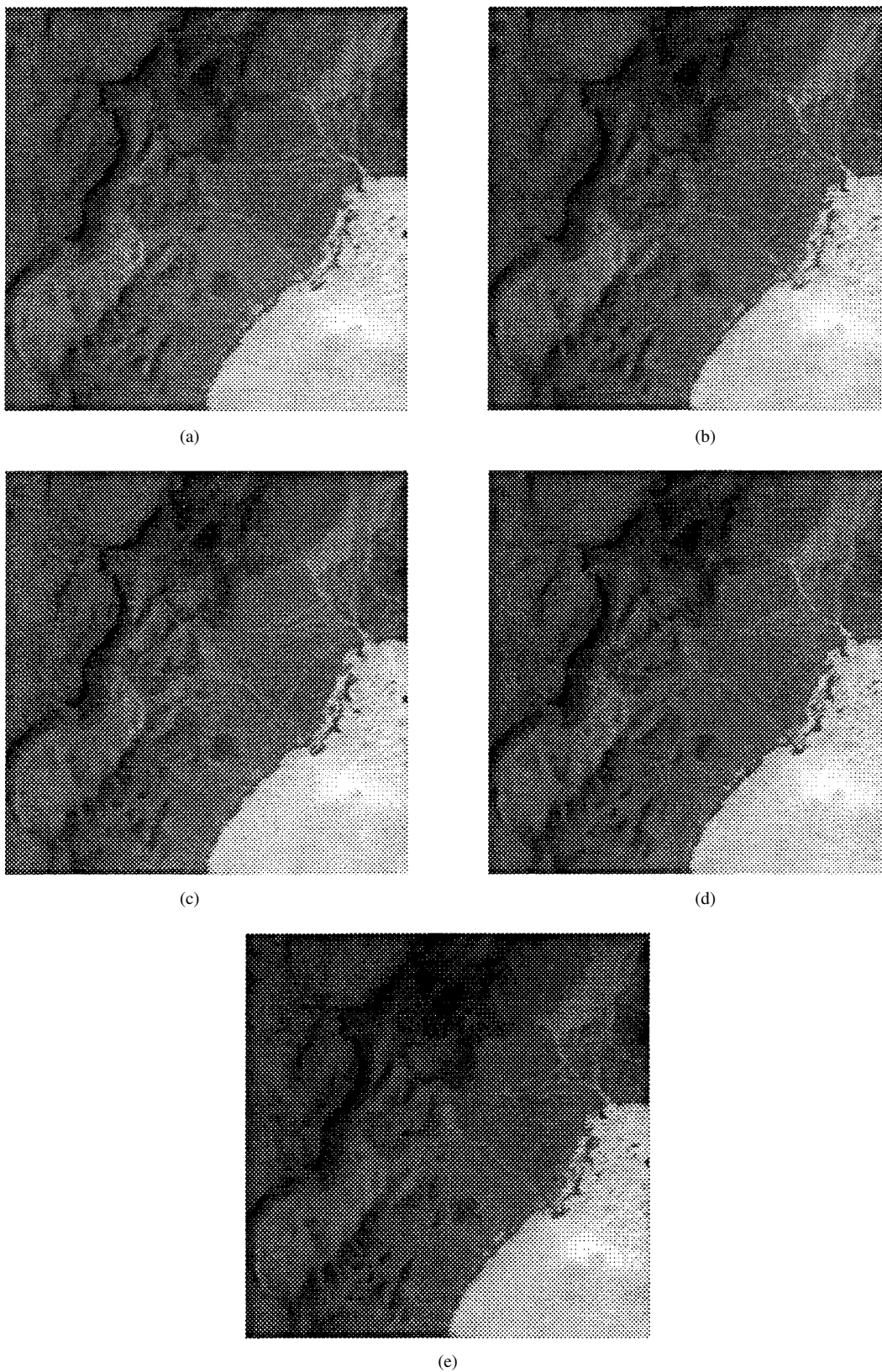
Fig. 12.   LCVF scene of AVIRIS data classified by TSC with (a) cinders as the target signature, (b) rhyolite as the target signature, (c) playa as the target signature, (d) vegetation as the target signature, and (e) shade.

data set 2), where figures labeled by (a), (b), and (c) are results obtained based on SNR 50:1, 30:1, and 10:1, respectively. The dotted and solid curves are produced by SSC and TSC, respectively, and the TSC used here has already eliminated the signature bias. As shown in these figures, TSC with bias removed performed better than SSC. However, it is not true if

TSC is used with bias. It was shown in [19] that the abundance curves produced by TSC with the bias were nearly flat across all 50 pixels (i.e., 1.175 for data set 1 and 0.85 for data set 2). This implies that all pixels contain almost the same abundance. Obviously, it is not the case that the pixels were simulated. This phenomenon can also be explained in terms of ROC analysis and will be further justified in Fig. 12 for AVIRIS, where the classification results using TSC with bias are erroneous because in real data it is impossible to estimate and eliminate the bias.

In order to evaluate the estimation error performance of SSC and TSC, their ROC curves are plotted in Figs. 8 and 9 for the two data sets with the same three signatures used for Figs. 6 and 7. Figures labeled by (a), (b), and (c) are results generated by $\alpha_3/\sigma = 0.5$, 0.3, and 0.1, respectively. The asterisked curve is the ROC curve generated by SSC. The dotted and solid curves are the ROC curves produced by TSC without bias and with bias, respectively. It should be noted that each ROC curve is generated by one mixed pixel with a given desired signature abundance-to-noise ratio (ANR), α_3/σ. For instance, if $\alpha_3/\sigma = 0.5$ and $\alpha_3 = 0.3$, the noise level will be $\sigma = 0.6$ and the other two undesired signatures will evenly split the remaining abundance 0.7. Their detection rates (DR's) are tabulated in Tables III and IV, which measure effects of different ANR's and spectral similarity. As expected, detection rates for TSC with bias removed are higher than that for SSC. However, what is unexpected is that TSC with bias even produces a little bit higher DR values than those without bias. This is due to the fact that when the bias is not known in real data, it must be considered to be a part of the target signature in the detection power equation (57). As a result, the target signal is strengthened by the bias in generating the ROC curves for TSC. From Figs. 6–9, it is also shown that SNR and spectral similarity play a role in performance.

B. AVIRIS Data

The AVIRIS data used in this experiment were the same data in [10], which is a scene of the Lunar Crater Volcanic Field, Northern Nye County, NV. As described in [10], the AVIRIS experiments were based on radiance spectra extracted directly from the image itself, not really based on prior knowledge of model (1). Nevertheless, OSP was proved to be effective for radiance spectra, not necessarily to be calibrated to reflectance spectra, as assumed in model (1). Here we apply SSC and TSC to the same AVIRIS data and compare their results to that produced by OSP in [10]. Figs. 10 (OSP), 11 (SSC), and 12 (TSC) show the experiments, where four signatures of interest in these images are "red oxidized," "basaltic cinders," "rhyolite," "playa (dry lakebed)," and "vegetation." Figures labeled (a)–(d) show cinders, rhyolite, playa, and vegetation as targets, respectively, and figures labeled by (e) are results of the shades, where we refer the original single-band image to Figs. 6 and 7(a) in [10]. As indicated in Section V-A of computer simulations, Fig. 12 produced the worst performance and cannot detect the targets correctly, due to unknown bias that cannot be eliminated in data processing. On the other hand, the images generated by OSP and SSC show no visible

difference. This implies that both SSC and OSP yield the same performance when no prior knowledge about data is available. As a conclusion, SSC can be viewed as an *a posteriori* OSP and a practical version of OSP.

VI. CONCLUSION

An OSP classifier was recently developed for hyperspectral image classification. However, the model on which it is based requires *a priori* knowledge about the abundance of signatures, which is generally not case in real data. In this paper, three least squares subspace projection-based classifiers, SSC, TSC, and OBC, were introduced to estimate signature abundance prior to classification. SSC estimates the target signature by projecting a mixed pixel into the signature space from which the target signature can be extracted by a matched filter. Rather than mapping a mixed pixel into the entire signature space, TSC projects the pixel directly into the target signature space. Unfortunately, it does not produce satisfactory performance, due to the fact that the undesired signatures are also projected and mixed into the target signature space. As a consequence, an unknown signature bias is created. In order to eliminate this bias, OBC is further proposed to project the target signature into its range space while projecting the undesired signatures into its null space. The paid price is that OBC is no longer an orthogonal projector, but still an idempotent projection. Despite such a difference in projection principle between SSC and OBC, it was shown in this paper that SSC and OBC are essentially equivalent, in the sense that they generate identical ROC curves. Their comparative performances are evaluated by ROC analysis through computer simulations and AVIRIS experiments. Since all three classifiers, SSC, TSC, and OBC, are designed on the basis of observed pixels, they work as *a posteriori* OSP, as opposed to *a priori* OSP in [10]. More importantly, in addition to mixed pixel classification, the *a posteriori* OSP can also be used for estimating the abundance of a desired target signature as well as for subpixel detection.

REFERENCES

[1] G. Vane and A. F. H. Goetz, "Terrestrial imaging spectrocopy," *Remote Sens. Environ.*, vol. 24, pp. 1–29, Feb. 1988.
[2] R. B. Singer and T. B. McCord, "Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," in *Proc. Lunar Planet. Sci. Conf. 10th*, 1979, pp. 1835–1848.
[3] B. Hapke, "Bidirection reflectance spectrocopy. I. Theory," *J. Geophys. Res.*, vol. 86, pp. 3039–3054, Apr. 10, 1981.
[4] P. Johnson, M. Smith, S. Taylor-George, and J. Adams, "A semiempirical method for analysis of the reflectance spectra of binary mineral mixtures," *J. Geophys. Res.*, vol. 88, pp. 3557–3561, 1983.
[5] E. R. Malinowski, "Determination of the number of factors and experimental error in a data matrix," *Anal. Chem.*, vol. 49, pp. 612–617, Apr. 1977.
[6] A. R. Heute, "Separation of soil-plant spectral mixtures by factor analysis," *Remote Sens. Environ.*, vol. 19, pp. 237–251, 1986.

[7] S. K. Jenson and F. A. Waltz, "Principal components analysis and canonical analysis in remote sensing," in *Proc. Amer. Soc. Photogram. 45th Ann. Meeting*, 1979, pp. 337–348.

[8] B. H. Juang and S. Katagiri, "Discriminative learning for minimum error classification," *IEEE Trans. Signal Processing*, vol. 40, pp. 3043–3054, Dec. 1992.

[9] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 388–400, Apr. 1993.

[10] J. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 779–785, July 1994.

[11] J. W. V. Miller, J. B. Farison, and Y. Shin, "Spatially invariant image sequences," *IEEE Trans. Image Processing*, vol. 1, pp. 148–161, Apr. 1992.

[12] J. J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 1045–1046, July 1996.

[13] T. M. Tu, C.-H. Chen, and C.-I Chang, "A least squares orthogonal subspace projection approach to desired signature extraction and detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 127–139, Jan. 1997.

[14] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projections operators," *IEEE Trans. Signal Processing*, vol. 42, pp. 1413–1423, June 1994.

[15] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[16] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991, chap. 9.

[17] E. R. Malinowski, "Theory of error in factor analysis," *Anal. Chem.*, vol. 49, pp. 606–612, Apr. 1977.

[18] C.-I Chang, *Error Analysis of Least Squares Subspace Projection Approach to Linear Mixing Problems*, Remote Sensing Signal Image Processing Lab., Univ. Maryland, Baltimore County, RSSIPL-LR-96-2, 1996.

[19] X. Zhao, "Subspace projection approach to multispectral/hyperspectral image classification using linear mixture modeling," M.S. thesis, Dept. Comput. Sci. Elec. Eng., Univ. Maryland, Baltimore County, May 1996.

[20] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Berlin, Germany: Springer-Verlag, 1994.

[21] J. A. Swets and R. M. Pickett, *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. New York: Academic, 1982.

[22] C. E. Metz, "ROC methodology in radiological imaging," *Invest. Radiol.*, vol. 21, pp. 720–723, 1986.

Xiao-Li Zhao, photograph and biography not available at the time of publication.

Mark L. G. Althouse (S'90–M'92) received the B.S. degree in physics from the Pennsylvania State University, University Park, the M.S. degree in electrical engineering from Johns Hopkins University, Baltimore, MD, and the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore County (UMBC), Baltimore.

He is a part-time Faculty Member in the Department of Computer Science and Electrical Engineering, UMBC. He has been working in the area of detection and identification of chemical and biological agent clouds for the United States Army Chemical and Biological Defense Command, Aberdeen Proving Ground, Aberdeen, MD, since 1981. Lately his work has concentrated on remote infrared sensors for chemical vapor detection, both optical design and signal processing methods.

Dr. Althouse is a member of Tau Beta Pi, Sigma Xi, OSA, and SPIE.

Jeng Jong Pan received the B.S. degree in physics from National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 1975, the M.S. degree in geophysics from National Taiwan University, Taipei, in 1979, and the Ph.D. degree in geophysics from University of Connecticut, Storrs, in 1993.

He was a Geophysicist at the Prokop Exploration Inc., Texas, from 1984 to 1986. From 1987 to 1989, he worked on remote sensing at the EROS Data Center, USGS, South Dakota. From 1989 to 1992, he worked on speech recognition at Martin Marietta Laboratory, Maryland. From 1992 to 1995, he worked on the NASA's EOS project at the Research and Data Systems Corporation, Maryland. Since 1989, he has been a part time Visiting Assistant Professor in the Department of Electrical Engineering, University of Maryland, Baltimore County (UMBC). He taught digital signal, imaging, and speech processing courses at UMBC from 1989 to 1994. He joined the National Oceanic and Atmospheric Administration, National Weather Service, Office of Hydrology, in 1995, and is currently working on hydrometeorologic data analysis in the Hydrologic Research Laboratory.

Chein-I Chang (S'81–M'82–SM'92) received the B.S., M.S. and M.A. degrees from Soochow University, Taipei, Taiwan, R.O.C., in 1973, the Institute of Mathematics at National Tsing Hua University, Hsinchu, Taiwan, in 1975 and the State University of New York at Stony Brook, in 1977all in mathematics, and the M.S. and M.S.E.E. degrees from the University of Illinois at Urbana-Champaign in 1982, respectively, and the Ph.D. in electrical engineering from the University of Maryland, College Park in 1987.

He was a Visiting Assistant Professor from January 1987 to August 1987, an Assistant Professor from 1987 to 1993, and is currently an Associate Professor in the Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore. He was a Visiting Specialist in the Institute of Information Engineering, National Cheng Kung University, Tainan, Taiwan, from 1994 to 1995. His research interests include information theory and coding, signal detection and estimation, multispectral/hyperspectral image processing, neural networks, and pattern recognition.

Dr. Chang is a member of SPIE, INNS, Phi Kappa Phi, and Eta Kappa Nu.