An oblique subspace projection approach for mixed pixel classification in hyperspectral images

Te-Ming Tu[a],*, Hsuen-Chyun Shyu[a], Ching-Hai Lee[a], Chein-I Chang[b]

[a] *Department of Electrical Engineering, Chung Cheng Institute of Technology, Tahsi, Taoyuan, Taiwan 33509, Republic of China*
[b] *Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21228-5398, USA*

Abstract

Recently oblique projection has been studied for many applications in signal processing. In this paper, the concept of oblique projection is applied to develop an algorithm for hyperspectral image classification. Compared with the orthogonal subspace projector (*OSP*), it can be found that *OSP* is a priori classifier but the oblique subspace projection classifier will be referred to a posterior. As a consequence, the oblique subspace projector (*OBP*) can be thought of as a generalized classifier including *OSP*. Furthermore, the estimation error from the OBP can be evaluated by applying the Neyman–Pearson detection theory to the corresponding receiver operating characteristic (ROC) curve so the accuracy of the classification can be calculated thereafter. Finally, some computer simulations using real airborne visible infrared image spectrometer (AVIRIS) data are accomplished to justify and compare the effectiveness of the above algorithms. © 1999 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Least-squares estimate; Orthogonal subspace projection; Oblique subspace projection; Neyman–Pearson detection; ROC curve

1. Introduction

The primary strength of multispectral/hyperspectral imagery is the utilization of multiple bands to provide information in detail on various materials contained in image pixels. In Earth remote sensing problems, the estimate of the fractions of endmembers resident in an image pixel is often important and useful, e.g., to crop production and damage assessment. Unlike being represented by light intensities for visual images, each pixel of multispectral/hyperspectral images is assumed to be a mixture of spectral reflectances resulting from different materials residing in the pixel. Detection and classification of minerals of interest from a mixed pixel are generally difficult and often present a challenging problem.

To resolve the mixed pixel classification problem, several researchers have investigated the scale and linearity of the mixture. The macrospectral mixture [1] model assumes no interaction between materials and treats a mixed pixel as a linear combination of signatures resident in the pixel with relative concentrations. In microscopic or the intimate mixture [2,3] model, the mixing is generally non-linear with a second-order effect. As

---

* Corresponding author. Tel.: 00 886 3 3803547; Fax: 00 886 3 3801407; E-mail: tutm@cco4.ccit.edu.tw

a consequence, many surface materials mix in a non-linear fashion but linear unmixing techniques, while at best an approximation, appear to work well in many circumstances [4]. In this work, the linear spectral mixture model is adopted to imply a novel linear classification technique.

Recently, an orthogonal subspace projection (*OSP*) method was proposed for hyperspectral image classification [5]. The idea is to null out or eliminate all unwanted signatures and interferences by projecting the image of interest into the orthogonal complement of the undesired signature space, then using a matched filter to extract the desired signatures from the projected image. Since there is no explicit assumption on the knowledge about the signatures [5], the *OSP* was developed as a priori linear spectral mixture model. Experimental results in Ref. [5] show the replacement of the true abundances by their estimates obtained by the *OSP* method.

The major problem for a priori model is that the signal may be corrupted by the channel or system noise. If there was no other method to deal with the noise, the estimate result could not be so correct. In order to alleviate this problem, a posterior least-squares orthogonal subspace projection approach [6] was developed where the true abundance of signatures was estimated and obtained by the criterion of the least-squares error. The least-squares estimator is an orthogonal projection matrix which maps the image of interest into both the desired and undesired signature spaces. The utilization of this matrix operator coupled with the *OSP* process can form a novel classifier to replace the original *OSP*.

In recent literatures [7,8], Behrens and Scharf showed that a least-squares orthogonal projection matrix can be further decomposed into two oblique projection matrices; one of which maps observations into the desired signature space while the other maps into the undesired signature space. However, unlike the least-squares orthogonal projection matrix [6], these two oblique projection matrices is not necessarily orthogonal.

A major difference between the *OSP* and *OBP* method is that the former utilizes complete knowledge of signature abundance whereas the latter does not. Consequently, the *OBP* needs an estimation process with errors coming from the noise. Interestingly, these two methods yield two classifiers of the same form with only a scalar difference. The quantity of this scalar will serve as an indication to the accuracy of the estimation. What is more important is that this quantity can be interpreted as the correlation between the desired signature and the undesired signatures. Restated, it tells of the similarity between the desired and undesired signatures. The higher this quantity is, the less similar the desired and undesired signatures.

To evaluate the performance and the estimation error for an *OBP*, a signal detection model is developed by making use of the Neyman–Pearson theory [9,10] and the receiver operating characteristic (ROC) curve. In this model, the false alarm probability represents a measurement of the estimation error and the detection power represents the accuracy of the estimation. Finally, some computer simulations using real airborne visible infrared image spectrometer (AVIRIS) data are accomplished to justify and compare the effectiveness of the above algorithms.

2. Problem model and formulation

In multispectral/hyperspectral image analysis, the spatial coverage of each pixel generally encompasses multiple materials. That means, the spectral signature of the pixel is an admixture of the spectral signatures of these endmembers. To study the subpixel processing techniques, the linear spectral mixture model [11] described below is generally considered.

Typographic conventions: Throughout this paper, vectors are denoted by boldface lowercase letters, matrices by boldface uppercase, and superscripts T, −1 denote transpose and matrix inverse, respectively. **I** denotes the identity matrix.

2.1. Linear mixture model for multispectral/hyperspectral images

Let $\mathbf{r}(x, y)$ be an $l \times 1$ column vector denoting the values of a pixel spatially located at the position (x, y) in a multispectral/hyperspectral scene where l is the number of bands. Assume that **M** is an $l \times p$ matrix denoted by $(\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p)$ where $\mathbf{m}_i$ is an $l \times 1$ column vector represented by the spectral signature of the ith material. Also let $a(x, y)$ be a $p \times 1$ column vector given by $(\alpha_1(x, y), \alpha_2(x, y), \ldots, \alpha_p(x, y))^\mathrm{T}$ where p is the number of materials and $\alpha_i(x, y)$ denotes the fraction of the ith signature present in $\mathbf{r}(x, y)$. A linear mixture model for the hyperspectral image pixel $\mathbf{r}(x, y)$ can be described by

$$\mathbf{r}(x, y) = \mathbf{M}\alpha(x, y) + \mathbf{n}(x, y), \tag{1}$$

where $\mathbf{n}(x, y)$ is an $l \times 1$ column vector representing additive white Gaussian noise with zero mean and variance $\sigma^2 \mathbf{I}$ and **I** is an $l \times l$ identity matrix.

2.2. A priori model versus a posterior model

The model described by Eq. (1) assumes that the spectral signatures **M** and $\boldsymbol{\alpha}$ are known as a priori. However, it is often the case where the abundance fraction of spectral signatures, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)^\mathrm{T}$, is generally not known. Therefore, in practice, $\boldsymbol{\alpha}$ must be estimated from the observation vector **r**. Here after, the spatial notation (x, y) for **r** and $\boldsymbol{\alpha}$ is suppressed and the "hat" symbol, "∧" is used to indicate an "estimate" value.

To formulate the estimator, Eq. (1) is rewritten as follows:

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n} = \mathbf{d}\,\alpha_p + \mathbf{U}\gamma + \mathbf{n} \tag{2}$$

$$= \mathbf{M}\hat{\boldsymbol{\alpha}} + \hat{\mathbf{n}} = \mathbf{d}\,\hat{\alpha}_p + \mathbf{U}\hat{\gamma} + \hat{\mathbf{n}}, \tag{3}$$

where $\mathbf{d}$ is the desired signature vector represented by the column vector $\mathbf{m}_p$, $\mathbf{U}$ is an $l \times (p-1)$ matrix given by $\mathbf{U} = (\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{p-1})$ and γ is a vector which contains the first $(p-1)$ components of $\boldsymbol{\alpha}$, $\gamma = (\alpha_1, \alpha_2, \dots, \alpha_{p-1})^{\mathrm{T}}$. The expression in Eq. (2) differs from that in Eq. (3) in the sense that the former adopts the complete knowledge of $\boldsymbol{\alpha}$. However, the latter has no prerequisite knowledge so the true $\boldsymbol{\alpha}$ is replaced by an estimate $\hat{\boldsymbol{\alpha}}$ resulting from the observation vector $\mathbf{r}$.

Using a statistical model to describe an observation is a conventional practice in the signal processing community. However, the advantage of using a priori (pr) model is significantly diminished when the number of observed samples is increased. Under this circumstance, the model described by Eq. (3), a posterior (ps) observation model, begins to show its dominance and tends to replace the pr model in Eq. (2). As the observation process proceeds with, the ps model eventually takes over the pr model. Consequently, algorithms using the ps model generally perform better than those based on a pr model. Restated, the information given by the ps model provides a better understanding than the pr model and facilitates the knowledge about the signal.

To convert the pr model to a ps model, let

$$\mathbf{P}_M = \mathbf{M}\mathbf{M}^{\#}, \tag{4}$$

$$\mathbf{P}_A = (\mathbf{I} - \mathbf{P}_M), \tag{5}$$

where $\mathbf{M}^{\#} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T$ is the pseudo-inverse of $\mathbf{M}$ and $\mathbf{I}$ is an identity matrix. Then, $\mathbf{M}\hat{\boldsymbol{\alpha}}$ and $\hat{n}$ given in Eq. (3) can be obtained as follows:

$$\mathbf{M}\hat{\boldsymbol{\alpha}} = \mathbf{P}_M \mathbf{r} \quad \text{and} \quad \hat{\mathbf{n}} = \mathbf{P}_A \mathbf{r}.$$

$\mathbf{P}_M$ and $\mathbf{P}_A$ defined by Eq. (5) are the projection operators which project the observation vector $\mathbf{r}$ into the signature and the noise space and can be referred to a ps signature and noise projector [12], respectively. The relationship between a pr model and a ps model is illustrated in Fig. 1. It is noted that these two projectors are symmetric, idempotent, and orthogonal. In other words, the signature space and the noise space projected by $\mathbf{P}_M$ and $\mathbf{P}_A$ are orthogonal complements of each other. More discussions on the characteristics of projection can be found in Ref. [12].

3. Oblique subspace projection

Recall that a ps signature projector $\mathbf{P}_M$ defined by Eq. (4) which projects an observation vector $\mathbf{r}$ into the
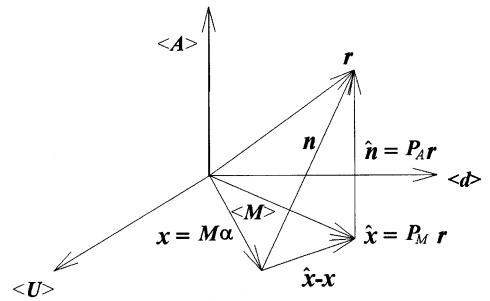

Fig. 1. Relationship between the a priori and a posterior models. $\langle d \rangle$ = desired signature space $\langle U \rangle$ = undesired signature space, $\langle M \rangle = \langle Ud \rangle$ = entire signature space, $\langle A \rangle$ = orthogonal complement space of $\langle M \rangle$.

signature space $\langle \mathbf{M} \rangle$ (i.e. $\langle \mathbf{Ud} \rangle$) is an orthogonal projector. However, a recent study [7,8] shows that the orthogonal projector $\mathbf{P}_M$ can be decomposed into a sum of two oblique projectors as follows:

$$\mathbf{P}_M = \mathbf{E}_{dU} + \mathbf{E}_{Ud}, \tag{6}$$

$$\mathbf{E}_{dU} = \mathbf{d}\,(\mathbf{d}^{\mathrm{T}} \mathbf{P}_U^{\perp} \mathbf{d})^{-1}\,\mathbf{d}^{\mathrm{T}} \mathbf{P}_U^{\perp}, \tag{7}$$

$$\mathbf{E}_{dU} = \mathbf{U}\,(\mathbf{U}^{\mathrm{T}} \mathbf{P}_d^{\perp} \mathbf{U})^{-1}\,\mathbf{U}^{\mathrm{T}} \mathbf{P}_d^{\perp} \tag{8}$$

where $\mathbf{E}$ represents oblique operations and $P_U^{\perp}$ is defined by

$$\mathbf{P}_U^{\perp} = (\mathbf{I} - \mathbf{U}\mathbf{U}^{\#}). \tag{9}$$

$P_U^{\perp}$ maps an observation vector $\mathbf{r}$ into a space which is orthogonal to that spanned by undesired signatures $\mathbf{U}$. In other words, $\mathbf{P}_U^{\perp}$ annihilates all undesired signatures.

More precisely, $\mathbf{E}_{dU}$ is a projection matrix with range space $\langle \mathbf{d} \rangle$ and null space $\langle \mathbf{U} \rangle$, but they are not necessarily orthogonal. In other words, the oblique projection operator $\mathbf{E}_{dU}$ projects the observation vector $\mathbf{r}$ into the desired signature space $\langle \mathbf{d} \rangle$ and nulls the undesired signatures space $\langle \mathbf{U} \rangle$ as depicted in Fig. 2. Thus,

$$\mathbf{E}_{dU}\mathbf{d} = \mathbf{d}; \qquad \mathbf{E}_{dU}\mathbf{U} = 0. \tag{10}$$

Similarly, $\mathbf{E}_{Ud}$ maps $\mathbf{r}$ into the undesired signature space $\langle \mathbf{U} \rangle$ and annihilates the desired signature space $\langle \mathbf{d} \rangle$. So $\mathbf{E}_{Ud}\mathbf{d} = 0$ and $\mathbf{E}_{Ud}\mathbf{U} = \mathbf{U}$. Applying $\mathbf{E}_{dU}$ to Eqs. (2) and (3) yields

$$\mathbf{E}_{dU}\mathbf{r} = \mathbf{d}\alpha_p + \mathbf{E}_{dU}\mathbf{n} \tag{11}$$

$$= \mathbf{d}\hat{\alpha}_p + \mathbf{E}_{dU}\hat{\mathbf{n}} = \mathbf{d}\hat{\alpha}_p. \tag{12}$$

Note that Eq. (12) follows from the fact that $\mathbf{E}_{dU}\hat{\mathbf{n}} = \mathbf{E}_{dU}\mathbf{P}_A\mathbf{r} = 0$.

By the detection theory [9,10], Eq. (11) represents a standard detection problem where the signal $\mathbf{d}\alpha_p$ is corrupted by the white noise $\mathbf{E}_{dU}\mathbf{n}$. The signal-to-noise
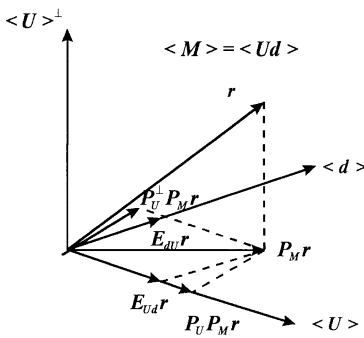
Fig. 2. Oblique subspace projection.

energy ratio (*SNR*) chosen as a criterion for classification is given by

$$SNR = \frac{E_s}{E_n},\tag{13}$$

where E_s and E_n are energies of the signal and noise, respectively.

To find a vector which maximizes the *SNR* defined by Eq. (13), a matched filter x is applied to Eq. (11) and obtains

$$\mathbf{x}^{\mathrm{T}}\mathbf{E}_{dU}\,\mathbf{r} = \mathbf{x}^{\mathrm{T}}\,\mathbf{d}\alpha_p + x^{\mathrm{T}}\mathbf{E}_{dU}\,\mathbf{n}.\tag{14}$$

Eq. (14) is indeed equivalent to maximize the following generalized eigenvalue problem:

$$\begin{aligned}\lambda_{\hat{\alpha}_p\max} &= \max_{x} SNR\\ &= \max_{x} \frac{x^{\mathrm{T}}\,\mathbf{d}\alpha_p^2\,\mathbf{d}^{\mathrm{T}}\,\mathbf{x}}{\mathbf{x}^{\mathrm{T}}\,\mathbf{E}_{dU}\,E\left[\mathbf{nn}^{\mathrm{T}}\right]\mathbf{E}_{dU}^{\mathrm{T}}\,\mathbf{x}}\\ &= \frac{\alpha_p^2}{\sigma^2}\frac{\mathbf{x}^{\mathrm{T}}\,\mathbf{dd}^{\mathrm{T}}\,\mathbf{xc}}{\mathbf{x}^{\mathrm{T}}\,\mathbf{E}_{dU}\,\mathbf{E}_{dU}^{\mathrm{T}}\,\mathbf{x}}.\end{aligned}\tag{15}$$

The solution is given by

$$\mathbf{x} = \boldsymbol{\kappa}\mathbf{d}\tag{16}$$

which yields the maximum eigenvalue

$$\lambda_{\hat{\alpha}_p,\max} = \frac{\alpha_p^2}{\sigma^2}(\mathbf{d}^{\mathrm{T}}\,\mathbf{P}_U^{\perp}\,\mathbf{d}).\tag{17}$$

Note that $\lambda_{\hat{\alpha}_p,\max}$ is determined by $\mathbf{d}^{\mathrm{T}}\mathbf{P}_U^{\perp}\mathbf{d}$ which is in turn determined by $\mathbf{P}_U^{\perp}\mathbf{d}$. However, $\mathbf{P}_U^{\perp}\mathbf{d}$ is the projection of $\mathbf{d}$ onto the space $\langle U\rangle^{\perp}$. This means that λ_{max} is simply the inner prduct of $\mathbf{d}$ and $\mathbf{P}_U^{\perp}\mathbf{d}$, $\langle\mathbf{d},\mathbf{P}_U^{\perp}\mathbf{d}\rangle$, which is the correlation of $\mathbf{d}$ and $\mathbf{P}_U^{\perp}\mathbf{d}$. Therefore, the more $\mathbf{d}$ correlates with $\mathbf{P}_U^{\perp}\mathbf{d}$, the less the correlation of $\mathbf{d}$ with $\mathbf{U}$, and the beter the discrimination of $\mathbf{d}$ from U. Later, we will shown that $\lambda_{\hat{\alpha}_p,\max}$ is also closely related to the estimation error in Eq. (24) which implies that the higher the $\lambda_{\hat{\alpha}_p,\max}$,

the smaller the estimation error. Thus, the accuracy of estimation is also determined by the magnitude of $\lambda_{\hat{\alpha}_p,\max}$.

Applying Eq. (16) with $\boldsymbol{\kappa} = 1$ to Eq. (14) results in

$$\mathbf{d}^{\mathrm{T}}\,\mathbf{E}_{dU}\,\mathbf{r} = \mathbf{d}^{\mathrm{T}}\,\mathbf{d}\alpha_p + \mathbf{d}^{\mathrm{T}}\,\mathbf{E}_{dU}\,\mathbf{n}\tag{18}$$

and

$$\mathbf{d}^{\mathrm{T}}\,\mathbf{E}_{dU}\,\mathbf{r} = \mathbf{d}^{\mathrm{T}}\,\mathbf{d}\hat{\alpha}_p.\tag{19}$$

By equating Eqs. (18) and (19), the estimated signature $\hat{\alpha}_p$ resulting from the oblique projector E_{dU} can be expressed as a sum of the true α_p and an error correction term $\mathbf{d}^{\mathrm{T}}E_{dU}\mathbf{n}/\mathbf{d}^{\mathrm{T}}\mathbf{d}$. That is,

$$\hat{\alpha}_p = \frac{\mathbf{d}^{\mathrm{T}}\,\mathbf{E}_{dU}\,\mathbf{r}}{\mathbf{d}^{\mathrm{T}}\,\mathbf{d}} = \alpha_p + \frac{\mathbf{d}^{\mathrm{T}}\,\mathbf{E}_{dU}\,\mathbf{n}}{\mathbf{d}^{\mathrm{T}}\,\mathbf{d}}.\tag{20}$$

The estimation error resulting from Eq. (20) is given by

$$\varepsilon = \hat{\alpha}_p - \alpha_p = \frac{\mathbf{d}^{\mathrm{T}}\,\mathbf{E}_{dU}\,\mathbf{n}}{\mathbf{d}^{\mathrm{T}}\,\mathbf{d}} = \frac{\mathbf{d}^{\mathrm{T}}\,P_U^{\perp}\,\mathbf{n}}{\mathbf{d}^{\mathrm{T}}\,P_U^{\perp}\,\mathbf{d}}.\tag{21}$$

Let

$$\mathbf{q}_{\mathrm{NOBP}}^{\mathrm{T}} = \frac{q_{\mathrm{OBP}}^{\mathrm{T}}}{\mathbf{d}^{\mathrm{T}}\mathbf{d}} = \frac{\mathbf{d}^{\mathrm{T}}\,E_{dU}}{\mathbf{d}^{\mathrm{T}}\,\mathbf{d}} = \frac{\mathbf{d}^{\mathrm{T}}\,\mathbf{P}_U^{\perp}}{\mathbf{d}^{\mathrm{T}}\,\mathbf{P}_U^{\perp}\,\mathbf{d}}\tag{22}$$

be the normalized oblique subspace projection (*NOBP*) which equals the oblique subspace projection (*OBP*), $\mathbf{q}_{\mathrm{OBP}}^{\mathrm{T}} = \mathbf{d}^{\mathrm{T}}\,\mathbf{E}_{dU}$, normalized by the factor $\mathbf{d}^{\mathrm{T}}\mathbf{d}$. Then Eq. (20) becomes

$$\hat{\alpha}_p = \frac{\mathbf{d}^{\mathrm{T}}\,\mathbf{P}_U^{\perp}\,\mathbf{r}}{\mathbf{d}^{\mathrm{T}}\,\mathbf{P}_U^{\perp}\,\mathbf{d}} = \mathbf{q}_{\mathrm{NOBP}}^{\mathrm{T}}\,\mathbf{r} = \boldsymbol{\alpha}_p + \mathbf{q}_{\mathrm{NOBP}}^{\mathrm{T}}\,\mathbf{n}.\tag{23}$$

The variance of ε is therefore obtained by

$$\mathrm{var}\left[\varepsilon\right] = \sigma^2\,\mathbf{q}_{\mathrm{NOBP}}^{T}\,\mathbf{q}_{\mathrm{NOBP}} = \frac{\sigma^2}{\mathbf{d}^{\mathrm{T}}\,\mathbf{P}_U^{\perp}\,\mathbf{d}}.\tag{24}$$

Eq. (24) states that the estimation error of $\hat{\alpha}_p - \alpha_p$ is inversely proportional to $\mathbf{d}^{\mathrm{T}}\mathbf{P}_U^{\perp}\mathbf{d}$. In other words, the smaller the estimation error, the higher the $\mathbf{d}^{\mathrm{T}}\mathbf{P}_U^{\perp}\mathbf{d}$ is. That also represents less correlation between the desired and the undesired signatures and hence implies a better discriminability of $\mathbf{d}$ from U. As a consequence, Eq. (24) can be used as an indication to the accuracy of the estimation of α_p. This fact will be justified by following experiments via a ROC analysis in Neyman–Pearson's detection theory.

Finally, it points out that since the *OSP* method is a priori model, there is no estimation involved in the derivation of this model. However, *OBP* is derived based on the least-squares orthogonality principle so an estimation error was introduced into Eq. (21). By comparing $\mathbf{q}_{OBP}^{\mathrm{T}}$ in Eq. (22) with $\mathbf{q}_{OSP}^{\mathrm{T}}$ in Ref. [1] and noting that $\mathbf{q}_{OSP}^{\mathrm{T}} = \mathbf{d}^{\mathrm{T}}\,\mathbf{P}_U^{\perp}$, Eq. (22) can be rewritten as

$$\mathbf{q}_{NOBP}^{\mathrm{T}} = \frac{\mathbf{q}_{OBP}^{T}}{\mathbf{d}^{\mathrm{T}}\,\mathbf{d}} = \frac{\mathbf{q}_{OSP}^{T}}{\mathbf{s}^{\mathrm{T}}\,\mathbf{P}_U^{\perp}\,\mathbf{d}}.\tag{25}$$

By Eq. (25) all three projectors $\mathbf{q}_{NOBP}^T$, $\mathbf{q}_{OBP}^T$, and $\mathbf{q}_{OSP}^T$ will have identical classification performance when $\mathbf{q}_{OBP}^T$ and $\mathbf{q}_{OSP}^T$ are normalized by a distinct normalization factor $\mathbf{d}^T \mathbf{d}$ and $\mathbf{d}^T \mathbf{P}_U^\perp \mathbf{d}$, respectively. Hence, we will discuss the performance of $\mathbf{q}_{NOBP}^T$ only in the following sections.

4. Estimation error evaluated by ROC analysis

In this section, the performance of the proposed classifier *NOBP* is evaluated by the estimation error of the abundance. In this analysis, the estimation is cast as a detection problem where the true abundance of the desired signature is viewed as a signal corrupted by a noise, i.e. the estimation error. As a result, the accuracy of estimation can be described in terms of the detection performance, which in turn can be evaluated by a receiver operating characteristics (ROC) curve in the framework of Neyman–Pearson's detection theory [9,10]. The area under ROC curve provides the measurement of detectability useful for error analysis.

Based on Eq. (23), a binary hypothesis testing problem can be formulated as follows:

$$H_0: \mathbf{z} = \mathbf{q}_{NOBP}^T \mathbf{n} \approx p_o(\mathbf{z}),$$

$$H_l: \mathbf{z} = \alpha_p + \mathbf{q}_{NOBP}^T \mathbf{n} \approx p_1(\mathbf{z}), \quad (26)$$

where the hypothesis H_0 represents the detection of the noise $\hat{\mathbf{n}} = \mathbf{q}_{NOBP}^T \mathbf{n}$ and H_l represents the detection of a constant signal specified by α_p corrupted by the noise $\hat{\mathbf{n}}$. The observable random variable is represented by z, α_p is the target to detect, and $\mathbf{n}$ is a noise.

To declare the detection of a target signal is based on an observation of z generated by one of two probability density functions, $p_o(z)$ and $p_l(z)$, depending upon which hypothesis is true. In the *NOBP* classifier, these probability density functions are given by

$$p_o(z) = N\left(0, \frac{\sigma^2}{\mathbf{d}^T \mathbf{P}_U^\perp \mathbf{d}}\right),$$

$$p_1(z) = N\left(\alpha_p, \frac{\sigma^2}{\mathbf{d}^T \mathbf{P}_U^\perp \mathbf{d}}\right), \quad (27)$$

where $N(x, y)$ represents a normal distribution of mean x and variance y. By Eq. (24) the variance of the noise is given as $\sigma^2 \mathbf{q}_{NOBP}^T \mathbf{q}_{NOBP} = \sigma^2 / \mathbf{d}^T \mathbf{P}_U^\perp \mathbf{d}$.

The Neyman–Pearson's detector for Eq. (26) is given by

$$\tilde{\delta}_{NP}(z) =$$

$$\begin{cases} 1 \ (\text{declaring } H_1) & \text{if } \mathbf{z} = \mathbf{q}_{NOBP}^T \mathbf{r} \geqslant \tau, \ \text{then } z = \mathbf{q}_{NOBP}^T \mathbf{r}, \\ 0 \ (\text{declaring } H_0) & \text{if } \mathbf{z} = \mathbf{q}_{NOBP}^T \mathbf{r} < \tau, \ \text{then } z = 0, \end{cases}$$

$$(28)$$

where

$$\tau = \left(\frac{\sigma^2}{\mathbf{d}^T \mathbf{P}_U^\perp \mathbf{d}}\right)^{1/2} \Phi^{-1}(1 - P_F) \quad (29)$$

is the threshold, Φ^{-1} is the inverse of Gaussian distribution and P_F is the probability of the false alarm.

The probability of detection in Eq. (28), denoted by P_D, can be used to represent the accuracy of the estimation in Eq. (23) in terms of estimation error given by the false alarm probability P_F.

The corresponding P_D is given by

$$P_D(\tilde{\delta}_{NP}) = 1 - \Phi\left(\frac{\tau - \alpha_p}{\sqrt{\sigma^2 \mathbf{q}^T \mathbf{P}_U^\perp \mathbf{d})^{-1}}}\right)$$

$$= 1 - \Phi\left(\Phi^{-1}(1 - P_F) - \frac{\alpha_p}{\sqrt{\sigma^2 (\mathbf{d}^T \mathbf{P}_U^\perp \mathbf{d})^{-1}}}\right).$$

$$(30)$$

Or equivalently,

$$P_D(\tilde{\delta}_{NP}) = 1 - \Phi\left(\Phi^{-1}(1 - P_F) - \sqrt{\lambda_{\hat{\alpha}_p, \max}}\right). \quad (31)$$

Eq. (31) comes from Eq. (17). By Eq. (30), the detection probability is a function of the estimation error $\sigma^2 (\mathbf{d}^T \mathbf{P}_U^\perp \mathbf{d})^{-1}$, which is also the variance of the transformed noise $\mathbf{q}_{NOBP}^T \mathbf{n}$ in Eq. (24). The detection probability can be expressed by a function of $\lambda_{\hat{\alpha}_p, \max}$ as in Eq. (31). As a result, for a fixed false alarm probability P_F, a smaller $\sigma^2 (\mathbf{d}^T \mathbf{P}_U^\perp \mathbf{d})^{-1}$ will imply higher detection probability P_D. Or, a larger $\lambda_{\hat{\alpha}_p, \max}$ can give a better detection. On the other hand, for a fixed $\sigma^2 (\mathbf{d}^T \mathbf{P}_U^\perp \mathbf{d})^{-1}$, a smaller P_F will make a lower detection probability P_D. These relationships can be well explained by the receiver operating characteristic (ROC) curve which is a plot of the detection probability P_D versus the false alarm probability P_F based on Eqs. (30) and (31).

5. Experimental results

5.1. Experiment 1 (computer simulations)

The first experiment is conducted based on the field spectrometer system (FSS) data having 60 spectral bands [13]. The major parameters of the FSS data are listed in Table 1. Since bands corresponding to the water absorption regions have no useful energy, they are removed before processing. The other 56 bands are kept in this study. In this experiment $p = 3$ is considered and $\alpha = (\alpha_1, \alpha_2, \alpha_3)^T$ represents a spectral abundance vector corresponding to a signature vector $\mathbf{M} = (\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)$. In this test, $\mathbf{m}_3$ is set to be the desired signature with its associated spectral abundance α_3. $U = (\mathbf{m}_1, \mathbf{m}_2)$ is then set to be the undesired signature vector. Three data sets are used to evaluate the performance of the estimation error

given by Eq. (24) via ROC curves based on Eqs. (30) and (31). Each data set contains three signatures from five materials as listed in Table 2. Data set 1 contains three distinct signatures as shown in Fig. 3a where spring wheat is designated as the desired signature while oats and summer fallow are undesired signatures. Data set 2 shown in Fig. 3b contains summer fallow and two other signatures, spring wheat and native grass, with similar spectral reflectances where native grass is selected as the desired signature. Data set 3 contains spring wheat, grain sorghum and native grass whose spectral reflectances are

nearly indistinguishable as shown in Fig. 3c, and native grass is chosen to be the desired signature. Each data set generates its own $\mathbf{d}^T \mathbf{P}_U^\perp \mathbf{d}$ which measures the correlation or spectral similarity between desired and undesired signatures. All these three data sets are simulated based on ground truth. One hundred pixels are simulated with each pixel containing three different signatures for various spectral reflectance abundances. In addition to these three signatures, three white Gaussian noises are also

Table 1
Parameters for the field spectrometer system

Number of bands	60
Spectral cover	0.4–2.4 μm
Altitude	60 m
IFOV (ground)	25 m

Table 2
Desired and undesired signatures for three distinct simulated data sets

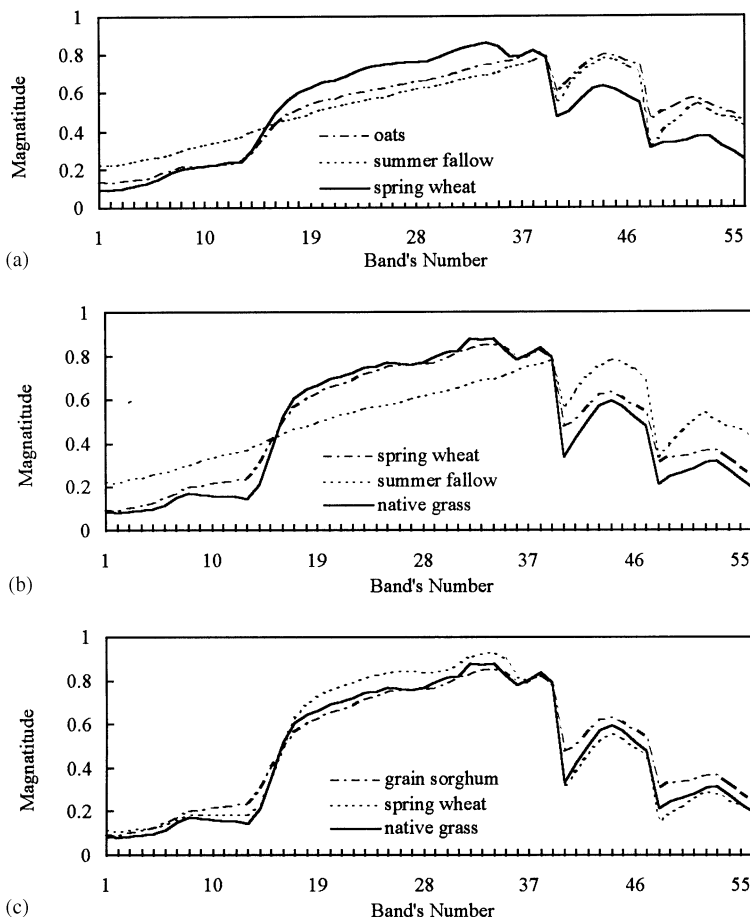| Data set	Undesired signatures	Desired signature
Data set-1	Oats and summer fallow	Spring wheat
Data set-2	Spring wheat and summer fallow	Native grass
Data set-3	Spring wheat and grain sorghum	Native grass

Fig. 3. Reflectance spectra of (a) data set 1, (b) data set 2, and (c) data set 3.

Table 3
Abundance of 100 simulated pixels for three data sets

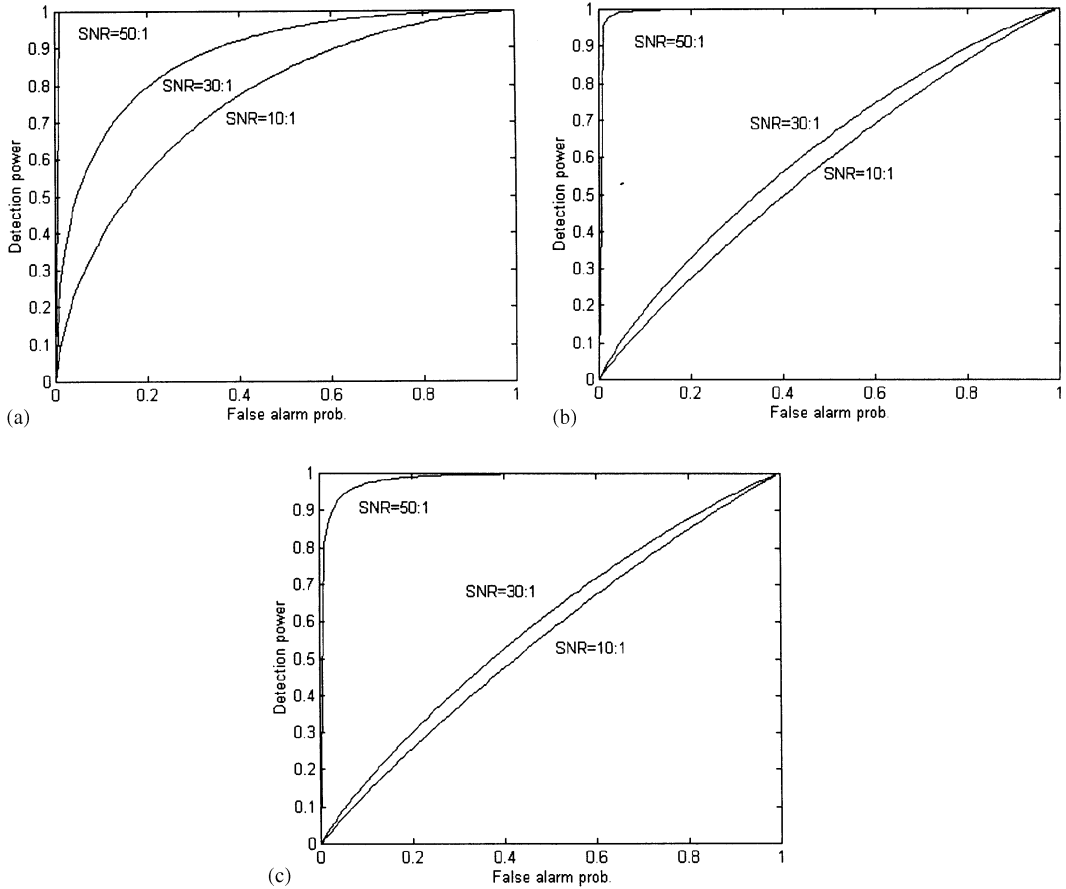|	Pixel 1–20 (%)	Pixel 21–40 (%)	Pixel 41–60 (%)	Pixel 61–80 (%)	Pixel 81–100 (%)
Desired signature	1	5	10	15	20
Undesired signature 1	49.5	47.5	45	42.5	40
Undesired signature 2	49.5	47.5	45	42.5	40

Fig. 4. ROC curves for (a) data set 1, (b) data set 2, and (c) data set 3.

simulated and added to each pixel to generate three distinct SNRs of 50:1, 30:1 and 10:1, respectively, using the SNR defined in Ref. [5]. The content of these 100 pixels for each of the three data sets is given in Table 3. In this table these 100 pixels are equally divided into five classes, each contains 20 pixels, and all pixels in each class contain the same amount of signature abundance. For example, pixels in the first class contain 1% abundance of the desired signature and 49.5% for each of the two other undesired signatures.

Experiments are conducted for each data set with distinct SNRs as stated above. The resultant ROC curves in terms of the variance of the estimation error and $\lambda_{\hat{\alpha}_p, \max}$ are shown in Fig. 4a–c, respectively. Herein, we define a measure called detection rate (DR), which calculates the area under an ROC curve as the effectiveness of the detector. Obviously, DR always lies between 0.5 and one. The worst case occurs when $DR = 0.5$, i.e. $P_D = P_F$, which implies that the detector is worthless. On the other hand, the best case occurs only when $DR = 1$, i.e. $P_D = 1$,

regardless of the false alarm probability. In these figures it can be found that the detection performance degrades as the *SNR* decreases. In addition, these figures also show that the detection capability depends upon the spectral correlation between signatures. The more correlated the spectra of signatures are, the more difficult it is to detect the signatures. As a conclusion, three factors affect the detection performance, *SNR*, false alarm probability P_F and $\mathbf{d}^T \mathbf{P}_U^{\perp} \mathbf{d}$ (or $\lambda_{\hat{\alpha}_p, \max}$) which is a measure of the spectral correlation between the desired and undesired signatures. The detail *DR*s are listed in Table 4.

5.2. Experiment 2

In this experiment, an airborne visible/infrared imaging spectrometer (AVIRIS) data set is used to compare the performance of the *NOBP* approach (given by Eq. (22)) and *NOBP* with Neyman–Pearson's detector (given by Eq. (28)). Fig. 5 shows a subsection of an AVIRIS scene of the lunar crater volcanic field (LCVF) located in Northern Nye County, NV where 158 bands were used. In order to use the Neyman–Pearson's detector to obtain the fractional images, the variance of the noise should be estimated. In this experiment, a "shift
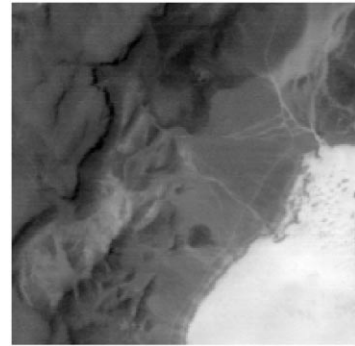
Table 4
Detection rate for three data sets with $SNR = 50$, 30 and 10

DR	$\alpha_p/\sigma = 0.5$	$\alpha_p/\sigma = 0.3$	$\alpha_p/\sigma = 0.1$
Data set-1	$DR = 1$	$DR = 1$	$DR = 1$
Data set-2	$DR = 0.94$	$DR = 0.61$	$DR = 0.58$
Data set-3	$DR = 0.81$	$DR = 0.53$	$DR = 0.52$

Fig. 5. A subsection of the upper left corner of the LCVF scene.
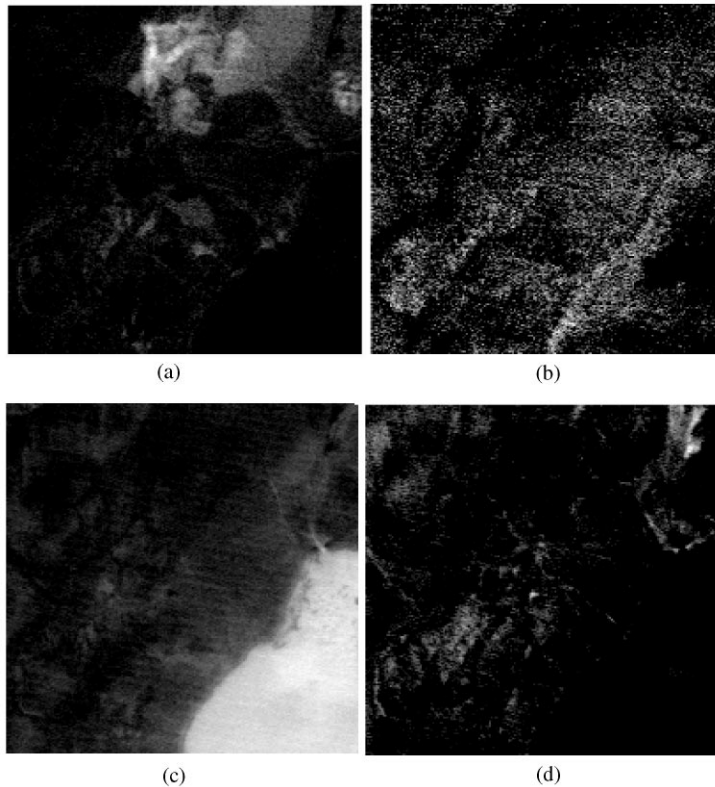

(a)   (b)

(c)   (d)

Fig. 6. *NOBP* results with Neyman–Pearson's detector and $P_F = 0.01$ for desired signature being (a) red oxidized basaltic cinders, (b) rhyolite, (c) dry playa lakebed, and (d) vegetation.
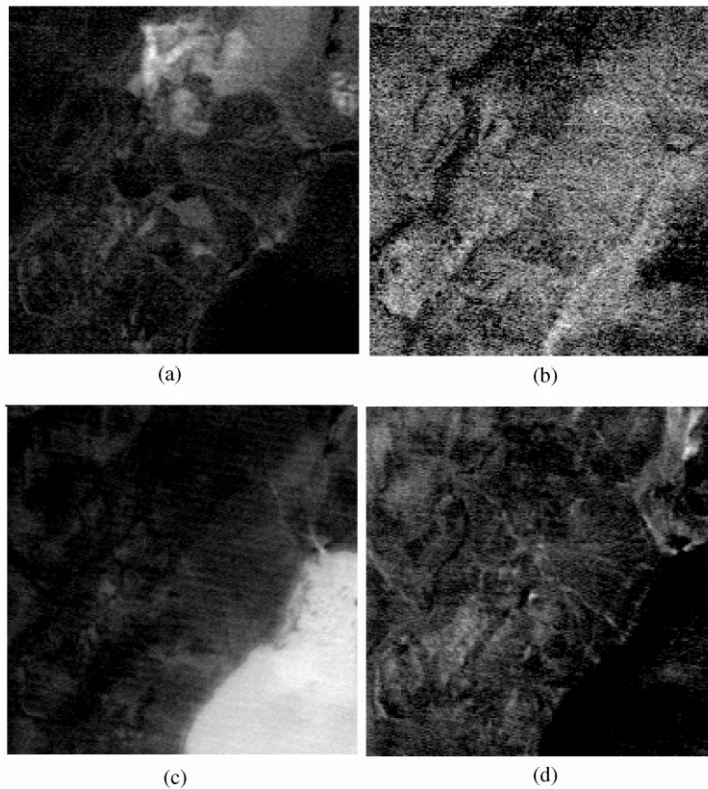
Fig. 7. *NOBP* approach result for desired signature being (a) red oxidized basaltic cinders, (b) rhyolite, (c) dry playa lakebed, and (d) vegetation.

difference" approach [14] was adopted to estimate its value and set $P_F = 0.01$. Figs. 6 and 7 summarizes the experimental results. In these two figures, the image brightness (0–255) is proportional to abundance (0–100%) of endmembers. That is, lighter pixels indicate a higher abundance of the target signatures. Comparing Fig. 6a–d to Fig. 7a–d, considerable improvements are obviously witnessed, particularly in Fig. 6a, c, and d, where the background is clearer and the estimated abundance of the desired signatures are more accurate than those in Fig. 7a, c, and d, respectively. These results are consistent with known attributes of the scene determined by field measures [15].

6. Conclusions

In this paper, the concept of oblique subspace projection (*OBP*) is applied to the classification in hyperspectral images. It improves the orthogonal subspace projection (*OSP*) method so that the abundance of signatures can be estimated by the *OBP* method more accurately. Interestingly, both *OSP* and *OBP* methods yield classifiers of the

same form with a difference of a scalar quantity determined by the estimation error only. However, this quantity is in turn mainly affected by the correlation between the desired and undesired signatures. The higher this quantity, the smaller the estimation error will be. The estimation error of the *NOBP* is further evaluated by the receiver operating characteristic (ROC) curve in the Neyman–Pearson detection theory. Experimental results show that the *NOBP* with Neyman–Pearson s detector can give accurate estimation and hence a good classifier for hyperspectral images.

References

[1] R.B. Singer, T.B. McCord, Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance, Proc. Lunar and Planetary Science Conf., 1979, pp. 1835–1848.

[2] E.B. Nash, J.E. Conel, Spectral reflectance systematics for mixtures of powdered hypersthene, labradorite, and ilmenite, J. Geophys. Res. 79 (1974) 1615–1621.

[3] R.B. Singer, Near-infrared spectral reflectance of mineral mixtures: systematic combinations of pyroxenes, olivine, and iron oxides, J. Geophys. Res. 86 (1981) 7967–7982.

[4] J.W. Boardman, F.A. Kruse, Automated spectral analysis: a geologic example using AVIRIS data, north Grapevine Mountains, Nevada, Proc. 10th Thematic Conf. on Geologic Remote Sensing, Environmental Research Institute of Michigan, Ann Arbor, MI, 1994, pp. I-407–I-418.

[5] C. Harsanyi, Chein-I Chang, Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach IEEE Trans. Geosci. Remote Sensing, 32 (4) (1994) 779–785.

[6] T.M. Tu, C.H. Chen, Chein-I Chang, A posterior least squares orthogonal subspace projection approach to desired signature extraction and detection, IEEE Trans. Geosci. Remote Sensing 35 (1) (1997) 127–139.

[7] T. Behrens, L.L. Scharf, Signal processing applications of oblique projections operators, IEEE Trans. Signal Process. 42 (1994) 1413–1423.

[8] L. Scharf, B. Friedlander, Matched subspace detectors, IEEE Trans. Signal Process. 42 (1994) 2146–2157.

[9] V. Poor, An Introduction to Signal Detection and Estimation, Springer, New York, 1988.

[10] D. Whalen, Detection of Signals in Noise, Academic Press, New York, 1971.

[11] B. Adams, M.O. Smith, Spectral mixture modeling: a new analysis of rock and soil types at the Viking Lander 1 site, J. Geophys. Res. 91 (1986) 8098–8112.

[12] L. Scharf, Statistical Signal Processing: Detection, Estimation, and Time Series Analysis, [Ch. 9] Addison Wesley, MA 1991.

[13] L. Biehl et al., A crops and soils data base for scene radiation research, Proc. Machine Processing Remotely Sensed Data Symp., 1982, pp. 169–177.

[14] A. Green, M. Berman, P. Switzer, M. Craig, A transformation for ordering multispectral data in terms of image quality with implications for noise removal, IEEE Trans. Geosci. Remote Sensing 26 (1988) 65–74.

[15] W.H. Farrand, Visible and near infrared reflectance of tuff rings and tuff cones. Ph. D. dissertation, University of Arizona, Tucson, AZ, 1991.

About the Author—TE-MING TU was born in Kaohsiung, Taiwan, in 1959. He received the B.S.E.E. degree from Chung Cheng Institute of Technology in 1986, the M.S.E.E. degree from National Sun Yat-Sen University in 1991, and the Ph.D. degree in electrical engineering from the National Cheng Kung University, in 1996. Since 1981, he has serviced in the R.O.C. Army. He was a teaching assistant from 1986–1989, an instructor from 1991 to 1993, and is currently an Associate Professor in the Department of Electrical Engineering, Chung Cheng Institute of Technology. His current research interests include multispectral/hyperspectral remote sensing, medical image, neural networks, and statistical pattern recognition.

About the Author—HSUEN-CHYUN SHYU was born in Taipei, Taiwan, R.O.C. in November 1951. He received his Bachelor, Master, and Ph.D. degrees from Chung Cheng Institute of Technology, Ta-Hsi, Tao-Yuan, National Taiwan University, Taipei, Taiwan, R.O.C., and University of Southern California, Los Angeles, California, in 1974, 1978, and 1986, respectively, all in Electrical Engineering. Currently he is an associate professor in the Department of the Electrical Engineering, Chung Cheng Institute of Technology. His research interests include digital signal processing, radar system, wavelet transforms, and neural networks.

About the Author—CHING-HAI LEE was born in Kaohsiung, Taiwan, in 1957. He obtained his B.S. and M.S. in Electrical Engineering from Chung Cheng Institute of Technology in 1979 and 1990, respectively. Since 1979, he has serviced in the R.O.C. Army. He was a supervisor in the Aeronautic Industry Development Center from 1990–1996. He is currently a Ph.D. candidate at the Chung Cheng Institute of Technology. His research interests include multispectral/hyperspectral remote sensing, parallel processing and statistical pattern recognition.

About the Author—CHEIN-I CHANG received his B.S., M.S. and M.A. degrees from Soochow University, Taipei, Taiwan, in 1973; National Tsing Hua University, Hsinchu, Taiwan, in 1975; and SUNY at Stony Brook in 1977, respectively, all in mathematics, and M.S., MSEE degree from the University of Illinois at Unbana-Champaign in 1980, 1982 and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 1987. He was a visiting assistant professor from January 1987 to August 1987, assistant professor from 1987 to 1993, and is currently an Associate Professor in the Department of Computer Science and Electrical Engineering at University of Maryland, Baltimore County. He was a visiting specialist in the Institute of Information Engineering at National Cheng Kung University, Taiwan, Republic of China from 1994–1995. His research interests include information theory and coding, signal detection and estimation, multispectral/hyperspectral remote sensing, neural networks and pattern recognition.