

Canonical Correlation Feature Selection for Sensors With Overlapping Bands: Theory and Application

Biliana Paskaleva, Majeed M. Hayat, *Senior Member, IEEE*, Zhipeng Wang, *Student Member, IEEE*, J. Scott Tyo, *Senior Member, IEEE*, and Sanjay Krishna, *Senior Member, IEEE*

Abstract—The main focus of this paper is a rigorous development and validation of a novel canonical correlation feature-selection (CCFS) algorithm that is particularly well suited for spectral sensors with overlapping and noisy bands. The proposed approach combines a generalized canonical correlation analysis framework and a minimum mean-square-error criterion for the selection of feature subspaces. The latter induces ranking of the best linear combinations of the noisy overlapping bands and, in doing so, guarantees a minimal generalized distance between the centers of classes and their respective reconstructions in the space spanned by sensor bands. To demonstrate the efficacy and the scope of the proposed approach, two different applications are considered. The first one is separability and classification analysis of rock species using laboratory spectral data and a quantum-dot infrared photodetector (QDIP) sensor. The second application deals with supervised classification and spectral unmixing, and abundance estimation of hyperspectral imagery obtained from the Airborne Hyperspectral Imager sensor. Since QDIP bands exhibit significant spectral overlap, the first study validates the new algorithm in this important application context. The results demonstrate that proper postprocessing can facilitate the emergence of QDIP-based sensors as a promising technology for midwave- and longwave-infrared remote sensing and spectral imaging. In particular, the proposed CCFS algorithm makes it possible to exploit the unique advantage offered by QDIPs with a dot-in-a-well configuration, comprising their bias-dependent spectral response, which is attributable to the quantum Stark effect. The main objective of the second study is to assert that the scope of the new CCFS approach also extends to more traditional spectral sensors.

Index Terms—Canonical correlation (CC) analysis, classification, dot-in-a-well (DWELL), feature selection, infrared photodetectors, quantum dots, spectral imaging, spectral sensing, subspace projection.

I. INTRODUCTION

IN THE past two decades, infrared spectral imaging in the wavelength range of 4–18 μm has found many applications in night vision, battlefield imaging, missile tracking and recognition, mine detection, and remote sensing, to name a few. Examples of spectral imagers operating in the

8–12- μm atmospheric windows include the Airborne Hyperspectral Imager (AHI) and the Spatially Enhanced Broadband Array Spectrograph System, which contain, respectively, 256 and 128 narrowband channels. However, the price of offering such sophisticated spectral imaging is enormous due to the complexity of the optical systems that render the detailed spectral information. Recently, efforts have been made to develop two-color and even multicolor focal-plane arrays (FPAs) for longwave (LW) applications [1], [2]; these sensors can electronically be tuned to two or more regions of the spectrum. Clearly, such tunable sensors offer greater optical simplicity as the spectral response is controlled electronically rather than optically. However, most existing multicolor sensors are limited in that the spectral sensitivity can only be electronically switched but not continuously tuned.

More recently, a new technology has emerged for continuously tunable midwave-infrared (MWIR) and LW-infrared (LWIR) sensing that utilizes intersubband transition in nanoscale self-assembled systems; these devices are termed quantum-dot infrared photodetectors (QDIPs). QDIP-based sensors promise a less expensive alternative to the traditional hyperspectral and multispectral sensors while offering more tuning flexibility and continuity compared to multicolor sensors [2]. QDIPs are based on a mature GaAs-based processing, and they are sensitive to normally incident radiation and have lower dark currents compared to their quantum-well counterparts [3], [4]. Unfortunately, QDIPs have low quantum efficiency, and much effort is currently underway to enhance that efficiency through increasing the number of quantum dots (QD) layers as well as using new supporting structures such as photonic crystals [5], [6]. Additionally, QDIPs with a dot-in-a-well configuration exhibit a bias-dependent spectral response, which is attributable to the quantum Stark effect, whereby the detector's responsivity can be altered in shape and central wavelength by varying the applied bias. Fig. 1 shows the bias-dependant spectral responses of the QDIP device used in this paper, measured with a broadband source and a Fourier transform infrared spectrophotometer at a temperature of 30 K.¹ Bias voltages in the range of -4.2 to -1 and 1 to 2.6 V, in steps of 0.2 V, were applied to this device. As shown in Fig. 1, the central wavelength and the shape of the detector's responsivity continuously change with the applied bias voltage. Therefore, a single QDIP can be exploited as a multispectral infrared sensor;

Manuscript received April 11, 2007; revised February 26, 2008. Current version published October 1, 2008. This work was supported in part by the National Science Foundation under Award IIS-0434102 and Award ECS-401154, and in part by the National Consortium for MASINT Research through a Partnership Project led by the Los Alamos National Laboratory.

B. Paskaleva, M. M. Hayat, and S. Krishna are with the Department of Electrical and Computer Engineering, and the Center for High Technology Materials, University of New Mexico, Albuquerque, NM 87131 USA (e-mail: kotarak@unm.edu).

Z. Wang and J. S. Tyo are with the College of Optical Science, University of Arizona, Tucson, AZ 85721 USA (e-mail: tyo@optics.arizona.edu).

Digital Object Identifier 10.1109/TGRS.2008.921637

¹This QDIP was fabricated by Professor Krishna's group at the Center for High Technology Materials, University of New Mexico. Device details will be reported elsewhere.

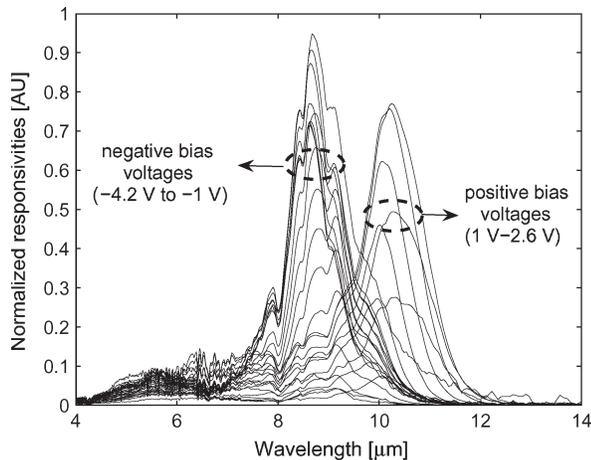


Fig. 1. Normalized spectral responses of QDIP 1780 used in this paper. The left cluster of spectral responsivities corresponds to the range of negative bias voltages between -4.2 and -1 V. The right cluster of spectral responsivities corresponds to the range of positive bias voltages between 1 and 2.6 V.

photocurrents of a single QDIP, driven by different operational biases, can be viewed as outputs of different spectrally broad and overlapping bands. While the broad spectral coverage is advantageous for broadband forward-looking infrared imaging, it is disadvantageous for applications that require narrow spectral resolution, such as chemical agent detection. Postprocessing strategies that exploit the spectral overlap in the QDIP's bands have recently been developed for continuous spectral tuning [7]–[9]. The inherent and often significant spectral overlap in the bands of a QDIP sensor produces a high level of redundancy in the output photocurrents of these bands. This redundancy, which is similar to the redundancy present in the outputs of the cones of the human eye, necessitates the development of lower-dimensional uncorrelated representations of the sensed data.

The presence of noise in the photocurrents (i.e., dark current and Johnson noise) further complicates the extraction of reliable spectral information from the highly overlapping and broad spectral bands of QDIP devices. Johnson noise results from the random motion of electrons in resistive elements and occurs regardless of any applied voltage [10]. On the other hand, current resulting from the generation and recombination process within the photoconductor will cause fluctuation in the carrier concentration and, hence, fluctuation in the conductivity of the semiconductor [10]. Generation and recombination noise, or so-called shot noise, becomes important in small bandgap semiconductors, in which the Johnson noise can also be high. Finally, at very low frequencies (e.g., less than 1 kHz), the flicker noise, also known as $1/f$ noise, also becomes an issue; it arises from surface and interface defects, and traps in the bulk of the semiconductor. However, for integration times of 1 ms or smaller, this noise is not important. Noise in QDIP detectors is dominated by the Johnson noise at temperatures less than 40 K and by the shot noise at higher temperatures (e.g., 77 K or above).

It is well known that in the presence of noise, the existing feature-reduction techniques may not always yield reliable information compression. It was shown in [11] that in the

principal component analysis (PCA) approach, the variance of the multispectral/hyperspectral data does not always reflect the actual signal-to-noise ratio (SNR) due to the unequal noise variances in different spectral bands. Therefore, it is possible that a band with a low variance may have a higher SNR than a band with a high variance. As a result, modified approaches such as the maximal noise fraction (MNF) transform were developed [11] based on maximizing the SNR; this method first whitens the noise covariance and then performs PCA. Other techniques include “higher-order methods” such as projection pursuit (PP) and independent component analysis (ICA) [12]–[14]; these methods search for “interesting” projection directions generating features that maximally deviate from “Gaussianity” or directions that maximize a certain projection index. Following the idea of the MNF transform [11], Lennon and Mercier in [15] proposed to adjust both PP and ICA to the noise in such a way that the SNRs of the noise-adjusted components are significantly increased compared to the SNRs of the components determined by the original algorithms.

In an earlier work [16], we proposed a mathematical theory for spectrally adaptive feature-selection approach for a general class of sensors with overlapping and noisy spectral bands. This theory builds upon the geometrical sensing model developed by Wang *et al.* [17], [18], in which the sensing process is viewed as a projection of the *scene space*, defined as the space of all spectral patterns of interest, onto a space spanned by the sensor bands, termed the *sensor space*. The main contributions of this paper are as follows. First, it provides a rigorous derivation of the heuristics that we reported earlier in [16], thereby providing a precise formulation of a canonical correlation feature-selection (CCFS) algorithm. The paper also provides new insights into the optimal feature-selection criterion for a class of sensors with overlapping and noisy bands. More precisely, for a specific pattern (or subspace of patterns) representing a class, a set of weights is derived that forms an optimal superposition (in the minimum mean-square-error (MMSE) sense) of the sensor bands, which we term a *superposition band*. The spectral pattern is then projected onto the direction defined by the superposition band. Thus, the superposition band can be thought of as the most informative direction for a specific pattern in the space spanned by the sensor bands in the presence of noise. Moreover, this process of selecting a superposition band is repeated in a hierarchical fashion to yield a canonical set of superposition bands that will generate, in turn, the best set of features for classes of patterns.

The rigorous validation of the proposed feature-selection algorithm in two different application contexts is another important contribution of this paper. The first application is separability and classification of rock species using laboratory spectral data and a QDIP sensor. This paper extends the preliminary results from [16] to a systematic analysis of the performance of the new CCFS algorithm for different SNR values. The results demonstrate that proper postprocessing can facilitate the emergence of QDIP-based sensors as a promising technology for MWIR and LWIR remote sensing and spectral imaging. The second, a completely new application of the CCFS algorithm, additionally validates our proposed approach in the context of spectral unmixing and abundance estimation of

hyperspectral imagery obtained from the AHI sensor. For both applications, comparison with the noise-adjusted PP shows that the CCFS can have a performance edge.

This paper is organized as follows. In Section II, we develop the theory for the proposed feature-selection technique for sensors with noisy and spectrally overlapping bands. In Section III, the theory is used to develop the CCFS algorithm for pattern classification problems. In Section IV, we study the performance of the CCFS algorithm in the two applications described above. Our conclusions are summarized in Section V.

II. MATHEMATICAL MODEL FOR SPECTRAL SENSING

A. Preliminaries

We start by reviewing germane aspects and concepts in spectral sensing drawing freely from our earlier work [16]–[18]. The spectral characteristics of bands are represented by a finite set of real-valued square-integrable spectral filters, or simply bands, $\{\hat{f}_i(\lambda)\}_{i=1}^k$, where the variable λ represents wavelength. The spectral response of the i th band is given by $\hat{f}_i(\lambda) = R_0 f_i(\lambda)$, where the unit of $\hat{f}_i(\lambda)$ is the response per watt of power incident on the detector. The scalar R_0 can be thought of as the peak responsivity and will assume the units required by $\hat{f}_i(\lambda)$, whereas the functions $\{f_i(\lambda)\}_{i=1}^k$ will be treated as dimensionless functions. Similarly, the emitted spectra of the materials of interest can be described by another set of square-integrable functions of wavelength $\{\hat{p}_i(\lambda)\}_{i=1}^m$. The emitted spectra of the i th-type material can be represented by $\hat{p}_i(\lambda) = P_0 p_i(\lambda)$, where P_0 is another constant that carries the units of the emitted radiance [W/cm²/sr/μm]. As a result, the *spectral pattern* $p_i(\lambda)$ can be assumed dimensionless. We define the universal linear space containing all the spectral patterns of interest and all spectral responses as the *spectral space* Φ . For example, Φ can be the Hilbert space $L_2([0, \infty))$ of all real-valued square-integrable functions. The subspaces spanned by the spectral bands $\{f_i(\lambda)\}_{i=1}^k$ and the spectral patterns $\{p_i(\lambda)\}_{i=1}^m$ are termed, respectively, the *sensor space* \mathcal{F} and the *pattern space* \mathcal{P} .

Ideally, the process of sensing a pattern with a spectral sensor can mathematically be represented as an inner product between the pattern and each one of the sensor bands

$$\langle p, f_i \rangle \triangleq \int_{-\infty}^{\infty} p(\lambda) f_i(\lambda) d\lambda \quad (1)$$

producing a set of photocurrents, one for each band. In actuality, however, the photocurrents are perturbed by noise, yielding the noisy photocurrent I_i for the i th band sensing the pattern p

$$I_i = \int_{\lambda_{\min}}^{\lambda_{\max}} p(\lambda) f_i(\lambda) d\lambda + N_i \quad (2)$$

where N_i represents the additive pattern-independent noise associated with the i th band, and the interval $[\lambda_{\min}, \lambda_{\max}]$ represents the common spectral support. Conceivably, different bands yield different noise levels (e.g., due to different bias

voltages in the case of a QDIP). For a given spectral pattern, the output corresponding to a single spectral band constitutes the *feature* of that pattern with respect to the band. A *spectral signature* is then defined as a k -dimensional vector in \mathbb{R}^k , whose coordinates are the measured photocurrents (features) associated with each spectral band.

B. Problem-Specific Feature Selection

We now develop the key building block for our canonical feature-selection algorithm. Specifically, we will seek to optimally replace the k -dimensional spectral signature in \mathbb{R}^k with a single spectral feature. This transformed feature \tilde{I} for the pattern p is defined as the weighted linear combination of all features, i.e., $\tilde{I} = \sum_{i=1}^k a_i I_i$, where the weights a_i are to be optimized for each pattern p . We term such a feature \tilde{I} as the *superposition current*. By using (2), the superposition current can then be expressed in the following form:

$$\tilde{I} = \sum_{i=1}^k a_i (\langle p, f_i \rangle + N_i) = \left\langle p, \sum_{i=1}^k a_i f_i \right\rangle + \sum_{i=1}^k a_i N_i. \quad (3)$$

From (3), we can deduce a useful analogy for the superposition current. Comparing this equation with (2), we see that the superposition current can be viewed as the output of an imaginary band $f = \sum_{i=1}^k a_i f_i$. We will term the band f a *superposition band* since it is a weighted superposition of the sensor's bands, and it is also associated with the superposition current. Hitherto, the problem of determining the best superposition current \tilde{I} for a given spectral pattern can be thought of as the problem of determining the optimal superposition band f in \mathcal{F} that offers the best approximation of p . Note that for a given superposition band f in \mathcal{F} , the approximation (or representation) of p rendered by this band is

$$p_f \triangleq \left(\left\langle p, \sum_{i=1}^k a_i f_i \right\rangle + \sum_{i=1}^k a_i N_i \right) f \quad (4)$$

which is a vector in \mathcal{F} that is along the direction of f but whose length is random due to noise.

Accordingly, one suitable criterion for the selection of a superposition band is to minimize the distance between the spectral pattern and its representation according to the superposition band. More precisely, we would select a set of coefficients a_1, \dots, a_k so that the L_2 norm of the error vector $\|p - p_f\|$ is minimized. Noting that $f = \sum_{i=1}^k a_i f_i$, we have

$$p_f = \sum_{i=1}^k \sum_{j=1}^k a_i a_j (\langle p, f_i \rangle + N_i) f_j.$$

Hence, for a given pattern p , we propose an optimal superposition band, represented by the vector \mathbf{a}^* , as

$$\mathbf{a}^* \triangleq \arg \min_{\mathbf{a} \in \mathbb{R}^k, \|\mathbf{a}\|=1} \mathbf{E} \left[\left\| p - \sum_{i=1}^k \sum_{j=1}^k a_i a_j (\langle p, f_i \rangle + N_i) f_j \right\|^2 \right] \quad (5)$$

where $\mathbf{a} = (a_1, \dots, a_k)^T$ is a weight vector associated with the superposition band f .

To provide a better insight into the criterion in (5) (and particularly the constraint $\|f\| = 1$), let us assume for the moment that the noise is absent. In this case, one can show that the minimization of the noiseless versions of the criterion (5) is equivalent to computing the projection $p_{\mathcal{F}}$ of p onto the subspace \mathcal{F} . By the minimum-distance property of the projection $p_{\mathcal{F}}$ (in [19, Th. 4.11]) $\inf_{g \in \mathcal{F}} \|p - g\| = \|p - p_{\mathcal{F}}\|$. The following lemma shows that $p_{\mathcal{F}}$ can be obtained (up to a sign difference) by projecting p onto unit-norm vectors in \mathcal{F} and then selecting the vector that yields the minimum error between the projection along that unit vector and p .

Lemma 1: Define $f_p \triangleq \pm(p_{\mathcal{F}}/\|p_{\mathcal{F}}\|)$. Then

$$\inf_{f \in \mathcal{F}} \|p - \langle p, f \rangle f\| = \min_{f \in \mathcal{F}, \|f\|=1} \|p - \langle p, f \rangle f\| \quad (6)$$

$$= \|p - \langle p, f_p \rangle f_p\| = \|p - p_{\mathcal{F}}\|. \quad (7)$$

The proof of this lemma is deferred to the Appendix. With this interpretation of $p_{\mathcal{F}}$ and by realizing that the inner product associated with a superposition band represented by the weight vector \mathbf{a} is corrupted by the additive noise $\sum_{i=1}^k a_i N_i$, as shown in (3), we arrive at the optimization criterion stated in (5). This justifies our selection of (5) as a criterion in the noiseless case and motivates its use as a meaningful criterion in the general case when the photocurrents are corrupted by additive noise.

The following lemma characterizes the minimization in (5).

Lemma 2: Put $f = \sum_{i=1}^k a_i f_i$, $\mathbf{a} = (a_1, \dots, a_k)^T$, and consider p_f given by (4). Without loss of generality, assume that $\|p\| = 1$, and further assume that the noise components in (4), N_1, \dots, N_k , are zero-mean and independent random variables with variances $\sigma_i^2, i = 1, \dots, k$. Then

$$\arg \min_{\mathbf{a} \in \mathbb{R}^k, \|f\|=1} \mathbf{E} [\|p_f - p\|^2] = \arg \max_{\mathbf{a} \in \mathbb{R}^k, \|f\|=1} \left\{ \langle p, f \rangle^2 - \sum_{i=1}^k a_i^2 \sigma_i^2 \right\}. \quad (8)$$

Lemma 2 provides useful information about the structure of the mean square error (MSE) in (8). The proof is deferred again to the Appendix.

If we define the SNR associated with the superposition band f represented by \mathbf{a} as

$$\text{SNR}_{\mathbf{a}} = \frac{\langle p, f \rangle^2}{\sum_{i=1}^k a_i^2 \sigma_i^2} \quad (9)$$

the criterion (8) can be written in terms of $\text{SNR}_{\mathbf{a}}$ as

$$\arg \min_{\mathbf{a} \in \mathbb{R}^k, \|f\|=1} \mathbf{E} [\|p - p_f\|^2] = \arg \max_{\mathbf{a} \in \mathbb{R}^k, \|f\|=1} \left\{ (\text{SNR}_{\mathbf{a}} - 1) \sum_{i=1}^k a_i^2 \sigma_i^2 \right\}.$$

The quantity $\langle p, f \rangle^2$ in (9) reflects how much energy from the scene is preserved during the spectral sensing process and relates this energy to the mutual position (i.e., angle) between the pattern p and any sensor band f_i that contributes to the

superposition band. More precisely, defining the interior angle θ_{p, f_i} between the spectral pattern p and any sensor band f_i as

$$\theta_{p, f_i} = \cos^{-1} \left(\frac{\langle p, f_i \rangle}{\|p\| \|f_i\|} \right)$$

if a given pattern p is ‘‘almost collinear’’ to any of the sensor bands $\{f_i\}_{i=1}^k$, then θ_{p, f_i} will nearly be zero, and the quantity $\langle p, f_i \rangle$ will attain its maximum value. In such cases, the contribution of that spectral band to the direction of the superposition band needs to be maximized to maximize the SNR for the superposition band. If $\mathcal{P} \subset \mathcal{F}$, then the angle between p and any f_i will be zero, meaning that the pattern space will completely be captured by the sensor space. On the other hand, if the angle between a given pattern $p \in \mathcal{P}$ and a spectral band $f_i \in \mathcal{F}$ is close to $\pi/2$, then this indicates the lack of correlation between the spectral pattern and the spectral band. In such a case, the pattern cannot reliably be sensed by that particular band, and the contribution of that band in the superposition band needs to be minimized.

In the presence of noise, due to the superposition process, the noise variance corresponding to the superposition band will accumulate, resulting in lower SNR and, therefore, higher approximation error. As a result, the optimal superposition band in a noisy environment may not coincide with the direction of projection of the pattern onto the sensor space, and the amount of deviation will depend upon the SNR for the individual bands.

In the next section, we use and extend the principle of optimal superposition band presented in this section to derive a canonical feature-selection algorithm. The algorithm allows us to search for a collection of weight vectors that yield the ‘‘best’’ collection of ‘‘sensing directions’’ minimizing the MSE in sensing classes of patterns.

III. CCFS

We begin by reviewing germane aspects of the canonical correlation (CC) analysis [20]–[22] of two Euclidean subspaces. In essence, based on a computed sequence of principal angles θ_k between any two finite-dimensional Euclidean spaces \mathcal{U} and \mathcal{V} , CC analysis yields the so-called CCs $\rho_k = \cos(\theta_k)$ between the two spaces. The first CC coefficient ρ_1 is computed as $\rho_1 = \max_{i,j} \mathbf{u}_i^T \mathbf{v}_j$, where the vectors \mathbf{u}_i ($i = 1, \dots, m$) and \mathbf{v}_j ($j = 1, \dots, n$) are unit length vectors that span \mathcal{U} and \mathcal{V} , respectively. The two vectors for which the maximum is attained are then removed, and ρ_2 is computed from the reduced sets of bases. This process is repeated until one of the remaining subspaces becomes null.

The CC analysis approach, however, is not applicable to cases for which the inner products between vectors are accompanied by additive noise, as in the case of the photocurrents shown in (2). In this case, a stochastic version of ‘‘principal angle’’ must be introduced and used. This new criterion was precisely introduced in Lemma 2. Thus, in our approach, we will follow the general principle of CC analysis while embracing the minimization stated in (8) as a criterion for maximal correlation.

In our formulation of the CCFS algorithm, we will restrict the attention to finite-dimensional spaces. Let us assume that

all the spectral patterns and the sensor's bands belong to an n -dimensional subspace of the Hilbert space Φ . Thus, without loss of generality, we can think of the Hilbert space Φ as \mathbb{R}^n and the functions $p \in \mathcal{P}$ and $f \in \mathcal{F}$ as Euclidean vectors \mathbf{p} and \mathbf{f} in \mathbb{R}^n , where \mathbf{p} and \mathbf{f} are the coordinate vectors of p and f , respectively. Furthermore, the inner product $\langle p, f \rangle$ can be represented by the dot product $\mathbf{p}^T \mathbf{f}$.

Further assume that \mathcal{F} is the span of k ($k \leq n$) linearly independent spectral bands represented by the columns of a matrix $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_k]$. We term \mathbf{F} as the *filter matrix*. Let \mathcal{P} denote the span of a set of m linearly independent patterns $\{\mathbf{p}_i\}_{i=1}^m$ representing the means of each one of m classes of interest. The matrix $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_m]$ is termed the *pattern matrix*. We will further assume that $m < k$.

The CCFS algorithm begins the search for the first canonical band by determining m weight vectors \mathbf{a}_i , $i = 1, \dots, m$, one for each class of interest. In particular, for the mean of the l th class, we determine a vector of weights $\mathbf{a}_l = (a_{l,1}, \dots, a_{l,k})^T$ as

$$\mathbf{a}_l = \arg \min_{\mathbf{a}_l \in \mathbb{R}^k, \|\mathbf{F}\mathbf{a}_l\|=1} \mathbf{E} \left[\|\mathbf{p}_l - (\mathbf{p}_l^T \mathbf{F}\mathbf{a}_l + \mathbf{n}^T \mathbf{a}_l) \mathbf{F}\mathbf{a}_l\|^2 \right] \quad (10)$$

where each component $a_{i,j}$ weights the corresponding sensor band \mathbf{f}_j , $j = 1, \dots, k$. Note that (10) is the equivalent matrix representation of (5), where $\mathbf{n} = (N_1, \dots, N_k)^T$ is a random vector whose components N_i are independent zero-mean random variables with variance σ_i^2 . We reiterate our earlier assertion in Section II that for each pattern \mathbf{p}_i , minimizing (10) is equivalent to selecting a direction $\sum_{j=1}^k a_{i,j} \mathbf{f}_j$ in \mathcal{F} that satisfies (8) and exhibits minimal combined noise variance and angle between the pattern and the direction.

The minimization process outlined in (10) is repeated m times as determined by the number of classes of interest, where each class is represented by its mean \mathbf{p}_i , $i = 1, \dots, m$. This process yields a set of m superposition bands, or sensing directions, $\mathbf{f}^1 = \mathbf{F}\mathbf{a}_1, \dots, \mathbf{f}^m = \mathbf{F}\mathbf{a}_m$, each one optimized with respect to the mean of each class. If the feature-selection algorithm stops here, and so the determined set of m superposition bands is used, it can be the case that these bands span a very small subspace of the sensor space since collinear patterns will determine collinear directions. The algorithm continues by selecting from this optimized set of superposition bands the one that is the most "collinear" with its corresponding mean, i.e., the superposition band that gives the minimum MSE for a particular class

$$\begin{aligned} \tilde{\mathbf{f}}^1 &= \arg \min_{\mathbf{f}^i; i=1, \dots, m} \mathbf{E} \left[\|\mathbf{p}_i - (\mathbf{p}_i^T \mathbf{f}^i + \mathbf{n}^T \mathbf{a}_i) \mathbf{f}^i\|^2 \right] \\ &= \arg \max_{\mathbf{f}^i; i=1, \dots, m} \left\{ \left(\frac{\mathbf{p}_i^T \mathbf{f}^i}{\mathbf{a}_i^T \Sigma_N \mathbf{a}_i} - 1 \right) \mathbf{a}_i^T \Sigma_N \mathbf{a}_i \right\} \end{aligned} \quad (11)$$

where the last equality follows from Lemma 2. We term the superposition band $\tilde{\mathbf{f}}^1$ as the *first canonical band*.

To ensure complete cover of the scene space within the filter space, the search for the second canonical band $\tilde{\mathbf{f}}^2$ is conducted in the orthogonal complement of $\tilde{\mathbf{f}}^1$, and it is with respect to the means of the remaining classes. More precisely, if $\tilde{\mathbf{f}}^1 = \mathbf{f}^{\ell_1}$, for

some $\ell_1 = 1, \dots, m$, then the ℓ_1 th class is excluded from the search for $\tilde{\mathbf{f}}^2$.

In general, if $\tilde{\mathbf{f}}^j$ is the j th optimal superposition band, then $\tilde{\mathbf{f}}^{j+1}$ is selected by searching in the orthogonal complement of $\tilde{\mathbf{f}}^1, \dots, \tilde{\mathbf{f}}^j$ and over all classes less the ℓ_1, \dots, ℓ_j th classes, where ℓ_i is defined through $\tilde{\mathbf{f}}^i = \mathbf{f}^{\ell_i}$. We continue in this fashion until we obtain a set of m canonical bands $\tilde{\mathbf{f}}^1, \dots, \tilde{\mathbf{f}}^m$. Note that the canonical order of the superposition bands does not depend on the presentation order of the classes of interest, since at the end of each optimization cycle, when decision is made, the algorithm always selects among all pairs (*superposition band center of a class*) the pair that yields the smallest estimation error. Each one of these canonical bands can be applied to the data to yield the so-called CC features.

The CCFS algorithm can be implemented in Matlab using the Optimization toolbox.

QR Factorization: Since the spectral bands \mathbf{f}_i , $i = 1, \dots, k$, are highly correlated, they provide a numerically ill-conditioned basis set for \mathcal{F} . Instead of directly solving (10), we may replace this problem by an equivalent one for which the minimization is carried out with respect to an orthonormal basis set for \mathcal{F} . This replacement will also speed up the numerical implementation of the optimization. More precisely, put $\mathbf{F} = \mathbf{Q}\mathbf{R}$ as the reduced QR factorization of the matrix \mathbf{F} . Then, the minimization problem

$$\arg \min_{\mathbf{a} \in \mathbb{R}^k, \|\mathbf{Q}\mathbf{R}\mathbf{a}\|=1} \mathbf{E} \left[\|\mathbf{p}_i - (\mathbf{p}_i^T \mathbf{Q}\mathbf{R}\mathbf{a} + \mathbf{n}^T \mathbf{a}) \mathbf{Q}\mathbf{R}\mathbf{a}\|^2 \right] \quad (12)$$

is equivalent to that shown in (10). Moreover, the optimization criterion in (12) can be recast in the equivalent form

$$\begin{aligned} &\arg \min_{\mathbf{b} \in \mathbb{R}^k, \|\mathbf{Q}\mathbf{b}\|=1} \mathbf{E} \left[\|\mathbf{p}_i - \mathbf{p}_i^T \mathbf{Q}\mathbf{b}\mathbf{Q}\mathbf{b} - \mathbf{n}^T \mathbf{R}^{-1} \mathbf{b}\mathbf{Q}\mathbf{b}\|^2 \right] \\ &= \arg \min_{\mathbf{b} \in \mathbb{R}^k, \|\mathbf{Q}\mathbf{b}\|=1} \left[1 - (\mathbf{p}_i^T \mathbf{Q}\mathbf{b})^2 + (\mathbf{R}^{-1} \mathbf{b})^T \Sigma_N \mathbf{R}^{-1} \mathbf{b} \right] \end{aligned}$$

whereas $\mathbf{b} = \mathbf{R}\mathbf{a}$ is the set of weights for the i th class mean derived with respect to the orthonormal basis set $\{\mathbf{q}_i\}_{i=1}^k$ for \mathcal{F} , where \mathbf{q}_i is the i th column of \mathbf{Q} .

IV. APPLICATIONS

In this section, we will describe two different applications of the CCFS algorithm. In the first application, the CCFS algorithm is applied to the spectral responses of the QDIP sensor and laboratory spectral data for the purpose of separability and classification analysis of seven classes of rocks [16], [23]. The second application is to AHI hyperspectral imagery in the context of supervised classification as well as spectral unmixing and fractional abundance estimation.

We will assume throughout this section that the noise components N_i are zero-mean normally distributed random variables. This follows from the fact that amplitude distributions for both thermal and shot noise converge to normal distributions by the central limit theorem. For the large number of electrons generating the thermal noise, the amplitude distribution of the thermal noise converges to zero-mean normal distribution. On the other hand, the actual numbers of generation-recombination

events underlying the shot noise will exhibit a Poisson distribution [10]. However, this number will become approximately normally distributed for a large average number of generation-recombination events [24]. Therefore, the amplitude distribution of the total noise will also be normal with mean equal to the mean of the shot noise and a variance equal to the sum of the variances of the two types of noise. Since the mean of the shot noise is deterministic and known (being equal to the dc value of the measured dark current), it can be subtracted from the noise without having any ramifications on the analysis or algorithm development.

A. Rock Type Classification

In the last few decades, the LWIR wavelengths have successfully been used to distinguish a number of primary silicates (feldspars, quartz, opaline silica) that are spectrally bland or have features that are nonunique at shorter wavelengths [25]. Thus, the thermal-infrared region of the spectrum is excellent for examining pure samples as well as mineralogically complex geologic materials (i.e., rocks) and is gaining popularity as a remote-sensing wavelength range for geologic applications [26], [27]. Our previous investigation of the rock type classification problem, using a Multispectral Thermal Imager (MTI) that operates in the shortwave and MWIR portions of the spectrum, has shown inadequacy of the simple minimum-distance classifier to accurately discriminate among the rock classes [23]. However, the MTI sensor in conjunction with the supervised Bayesian classifier offers much higher discrimination accuracy among the different rock types; hence, the MTI performance would serve as a good benchmark in this paper [16], [23]. (MTI was designed to be a satellite-based system for terrestrial observation with emphasis on obtaining qualitative information of the surface temperature. Currently, MTI operates with set of 15 bands, covering the broad range from 0.45 to 10.7 μm .)

1) *Definition of Training and Testing Sets:* Generally, rocks can be divided into three main geological groups: *igneous*, *metamorphic*, and *sedimentary*, which correspond to the different geological processes involved in the rock's formation. Geologists have further divided these three main rock categories into seven generic classes, which we adopted in this paper. To create the training and testing data sets, we selected a number of spectra of common rock samples in different grain sizes from the Advanced Spaceborne Thermal Emission and Reflection Radiometer hyperspectral database. Table I describes the rock classes and the endmembers included in the training set [16].

The limited number of endmembers (see Table I), however, prevented direct application of the Bayesian classifier. This fact forced us to increase the size of the training set by perturbing the endmembers in each rock class with different mixing materials. To create the perturbations, we used a simple two-component linear mixing model, where each mixture was considered as a linear combination of a representative endmember and a mixing endmember, weighted by the correspondent abundance function β . For the abundance function, we used five randomly chosen values of β between 1% and 10% for the mixing endmembers and $(100-\beta)\%$ for the representative endmembers. Using the above mixing model, we created spec-

TABLE I
ROCK TYPE GROUPS AND THEIR REPRESENTATIVE ENDMEMBERS

Group	Endmembers
Hornfelsic	hornfels (fine, coarse)
Granoblastic	pink quartzite, marble (fine, coarse) and gray quartzite (coarse)
Schistose	gray slate, chlorite schistose (fine, coarse) and chlorite
Semischistose	felstic gneiss (fine, coarse)
Igneous	andesite, basalt, diorite, gabbro, granite, rhyolite (fine, coarse), tan rhyolite and tuff (cup 8, 9)
Clastic Sedimentary	shale, siltstone, fossiliferous limestone and red sandstone (fine, coarse)
Chemical Sedimentary	limestone (fine, coarse) and dolomite

TABLE II
MIXING ENDMEMBERS USED TO CREATE RANDOM PERTURBATIONS OF THE REPRESENTATIVE ENDMEMBERS LISTED IN TABLE I

Minerals	andradite, anorthite, dolomite, quartz and topaz
Rocks	basaltic andesite, diorite gneiss, limestone and siltstone (fine and coarse)
Water, soil	distilled water, see water, dark brown loam, fine sandy loam and brown to dark brown sand
Vegetation	conifer and grass (green), spruce cellulose, citrus pectin, white peppermint, CA buckwheat, brown sycamore and brown leaf (dry)

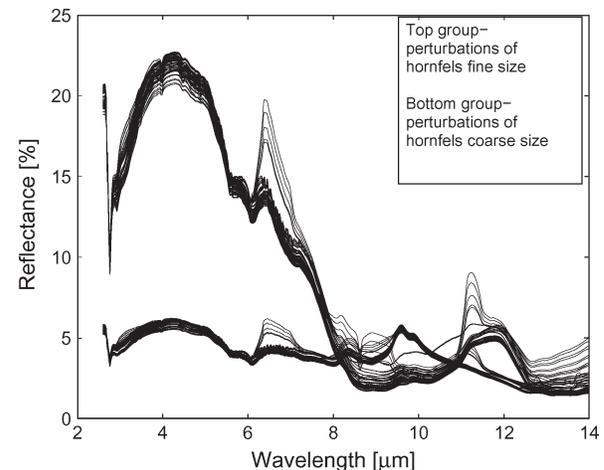


Fig. 2. Reflectivity of the hornfels showing fine (top group) and coarse size (bottom group) as well as their perturbations [16].

tral mixtures of the representative endmembers with minerals, vegetation, soil, and water [23]. We also created mixtures between fine- and coarse-size rocks, and between coarse- and fine-size rocks, according to their geological properties that make such mixtures realistic. All mixing endmembers used to enlarge the training set are presented in Table II.

Fig. 2 shows the spectral signatures of the endmembers for the class hornfelsic, fine, and coarse size, as well as their mixtures with rocks, minerals, soils, and vegetation. We created two testing sets where the mixing endmembers used to create these sets are shown in Table III. In Set-1, the representative endmembers in Table I were perturbed with the rocks listed in Table III. For the abundance function, we used five randomly chosen values within the range of 1% to 10%. Set-2 is an enlargement of Set-1 with the addition of mixtures of the representative endmembers (see Table I) with soils, minerals, and vegetation listed in Table III.

TABLE III
MIXING ENDMEMBERS USED TO CREATE RANDOM PERTURBATIONS
OF THE REPRESENTATIVE ENDMEMBERS LISTED IN TABLE I
TO CREATE TEST SET-1 AND TEST SET-2

Minerals	andradite, antigorite, erionite, fluorite, quartz and spodumene
Rocks	basalt, pink marble and black shale (fine and coarse)
Water, soil	see foam, grayish brown loam, dark grayish brown silty loam, reddish fine sandy loam and dark reddish fine sandy loam
Vegetation	deciduous (green), cotton cellulose, citrus pectin, sycamore-loer (yellow), CA brown buckwheat and sycamore (dry)

The addition of all the mixtures helped to increase the rank of covariance matrix to 13 in the case of QDIP and 11 in the case of MTI, which still failed short of full rank for 26-dimensional data in the case of QDIP and 13-dimensional data in the case of MTI. To mitigate this problem, we selected a subset of 13 arbitrary QDIP bands. The performance of this subset was averaged over different arbitrarily selected subsets of 13 bands. In the case of MTI, we were able to identify high correlation for bands C and L with their adjacent spectral bands, so they were removed without losing relevant information. A supervised Bayesian classifier was employed with the assumptions for normal class populations and equal priors [28]. The second assumption is reasonable as the training set was defined by geologists in accordance with the geological properties of rocks; thus, the number of samples in the training set for a certain group does not represent the frequency of occurrence of the rocks in nature. Instead, the number of samples per class reflects the rock diversity within a given class.

B. Separability and Classification Results

To set a benchmark for the performance of the CCFS algorithm, we begin by presenting the separability and classification results in the ideal case when noise is absent and without using the proposed CCFS algorithm [16].

We first compare separability and classification performance for QDIP and MTI sensors. Four sets of separability and classification results are summarized in Fig. 3 (left). The first set of results corresponds to using 11 out of 15 MTI bands (bands A–E, G, I, O, J, K and M) [29]. The second set corresponds to the case of 13 arbitrary bands out of the 26 QDIP bands. The third set of results is based on 7 MTI bands (bands G, I, O, J, K, M and N) selected to approximate the spectral range of the QDIP bands. The final fourth set is based on a subset of 7 arbitrary selected QDIP bands, shown in Fig. 4. The results presented in Fig. 3 (left) suggest that the MTI and QDIP bands yield comparable performance in the absence of noise [16].

1) *Effect of Noise*: In this section, we consider the presence of noise and compare the separability and classification results for the CCFS algorithm with four different cases, each using seven bands and for four different SNR values. The results are averaged over 100 independent noise realizations for each SNR value. Here, the number of selected superposition bands is determined by number of classes of interest, i.e., seven. The first case is termed *deterministic CCFS* (DCCFS), and it employs

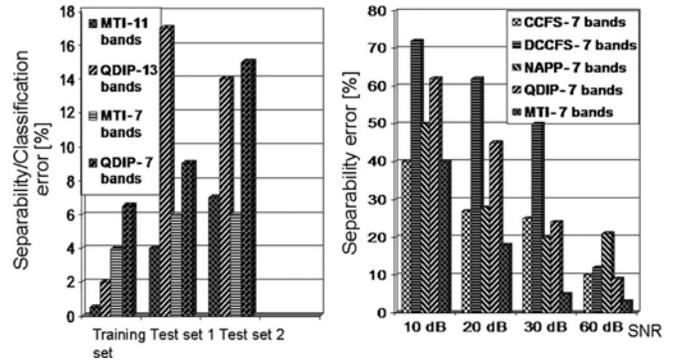


Fig. 3. (Left) Comparison in rock type separation and classification between QDIP and MTI sensors in the absence of noise [16]. (Right) Comparison in rock type separation for the training set for CCFS, DCCFS, noise-adjusted PP, seven QDIP bands, and seven MTI bands in the presence of noise with average SNR values of 10, 20, 30, and 60 dB.

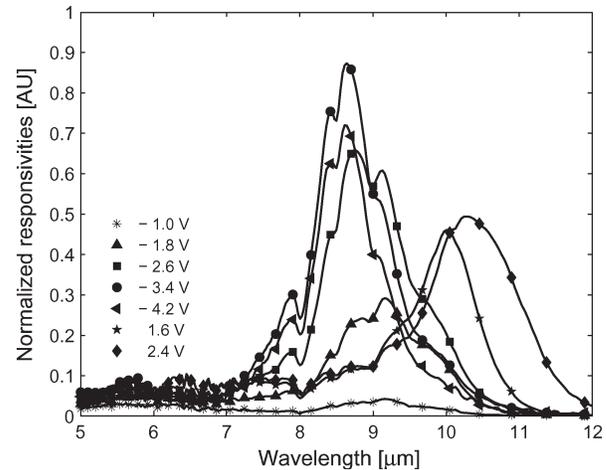


Fig. 4. Seven QDIP bands used in the rock type classification.

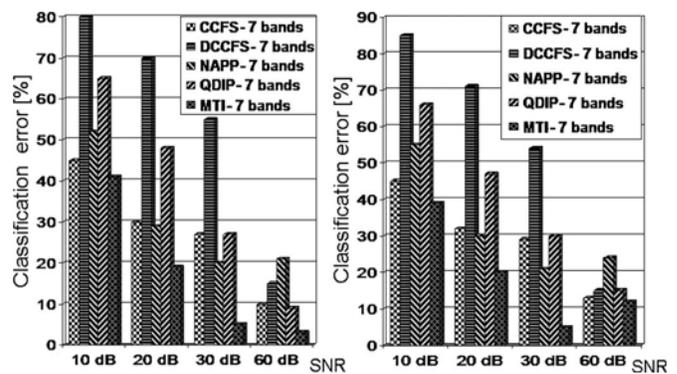


Fig. 5. Comparison in rock type classification for CCFS, DCCFS, noise-adjusted PP, QDIP bands, and MTI bands in the presence of noise with average SNR values of 10, 20, 30, and 60 dB. (Left) Test Set-1. (Right) Test Set-2.

the proposed CC feature selection but without accounting for the photocurrent noise during the selection process. In the second case, termed *noise-adjusted PP* [15], [30], we use seven features extracted using the noise-adjusted PP algorithm. Finally, the last two cases correspond to the classifiers used in Fig. 3 (left) applied to noisy data; these cases are termed *QDIP-7 bands* and *MTI-7 bands*. Figs. 3 (right) and 5 compare the

separability and classification performances, respectively, for the aforementioned five cases.

The first observation made is that embedding the noise statistics in the canonical feature selection leads to a significant improvement in the classification. As we can see from the results presented in Figs. 3 (right) and 5, for the first three SNR cases (average SNR of 10, 20, and 30 dB), the CCFS algorithm performs almost twice as good as the DCCFS algorithm. In the limiting case of a very high SNR, the performance of the CCFS and DCCFS algorithms becomes almost identical, as expected, and the classification error drops to 10%–15%.

We next compare the CCFS algorithm with the arbitrary selection of seven QDIP bands. For the average SNR of 10 dB [see Fig. 3 (right)], the separability error from the latter case is 63%, compared to 41% in the CCFS case. This result underscores the higher sensitivity of QDIP bands to significant noise levels compared to the canonical superposition bands. Notably, by using the CCFS algorithm, we were able to achieve a significant improvement in the classification performance (approximately 20%). As expected, when the average SNR increases, the performances of the two cases become comparable.

The separability and classification results also indicate that the CCFS approach offers classification capabilities comparable to those offered by the MTI bands when high levels of noise are present (10 dB). When the SNR increases to 30 dB (see Fig. 5), the classification results corresponding to the MTI bands almost reach the noiseless case classification error [see Fig. 3 (left)]; however, this trend is much slower in the case of CCFS. The results suggest that the bands designed via the CCFS approach are still more susceptible to noise compared to the MTI bands. Such a conclusion should not be surprising in view of the fact that the MTI sensor contains well-separated spectral bands with almost nonoverlapping finite supports and distinct spectral characteristics. As a result, even for high noise levels, the photocurrents obtained with MTI bands are often well separated.

2) *Comparison With the PP Approach:* We also compare the proposed CCFS algorithm with the noise-adjusted version of the PP feature-selection algorithm [12], [13], [31]. In this paper, we adopted the so-called fast ICA for the implementation of the PP algorithm and its noise-adjusted version [14], [30].

For low average SNRs of 10 dB, the separability and classification accuracy achieved with the CCFS algorithm is approximately 10% better than the one obtained with the noise-adjusted PP. As the SNR increases, the performance of the two algorithms becomes very similar, yielding almost identical separability and classification accuracy in the cases of average SNR of 20 dB (see Figs. 3 (right) and 5). However, when the SNR reaches extremely high values (see Figs. 3 (right) and 5), the CCFS algorithm once again outperforms the noise-adjusted PP approach, yielding a 10% classification error compared to the 20% error by the noise-adjusted PP for the training set and testing Set-1.

C. Application to AHI Hyperspectral Imagery

AHI is an LWIR pushbroom hyperspectral imager with a 256-by-256 element Rockwell TCM2250 HgCdTe FPA me-

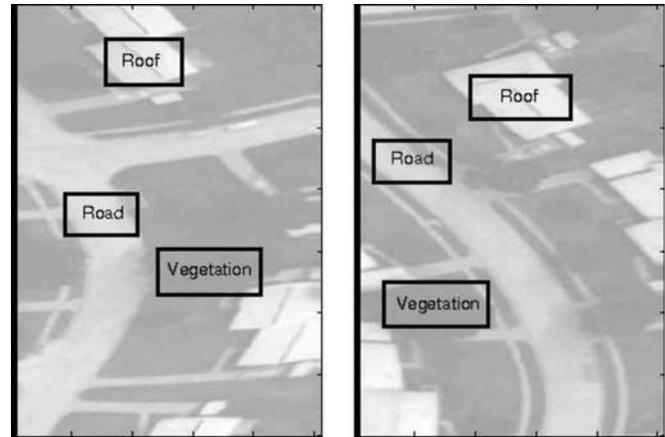


Fig. 6. (Left) Training and (right) testing areas (snapshot at $10.0967 \mu\text{m}$) selected from AHI test flight image of an urban area. The rectangular boxes indicate the approximate areas used to select the training and testing sets for the endmembers.

chanically cooled to 56 K [32]. The AHI sensor contains 256 spectral bands in the range of $7\text{--}11.5 \mu\text{m}$ with $0.1\text{-}\mu\text{m}$ spectral resolution for each spectral band. Further details on the AHI system and related data acquisition and calibration issues can be found in [32].

Here, we consider two types of problems with the CCFS used as a feature-selection algorithm: supervised Bayesian classification of three spectral classes, and spectral unmixing and abundance estimation for three endmembers. The AHI scene used in the first problem consists of roads, vegetation, and building roofs. The size of the image is 4451 by 256 pixels with 256 spectral bands. To perform supervised classification, we selected by visual examination three representative areas for each of the three classes of interest and used the spectral signatures corresponding to these areas as training sets for the classifier. We created test sets by selecting three areas that represent different spatial locations of the same image but visually correspond to the same classes. The training and testing sets contain 1250 pixels each, 450 pixels per class. The three sections of the scene, shown in Fig. 6 for $\lambda = 10.0967 \mu\text{m}$, represent the three classes of interest; these regions are used to extract the training (left) and testing (right) sets.

After the training and testing spectral sets were determined, Bayesian classification, in conjunction with CCFS, was applied to both sets, and separability and classification errors were calculated for different SNR cases. The AHI spectral bands were uniformly approximated by triangular pulses with peaks at the central frequencies and base widths of $0.1 \mu\text{m}$. As we did earlier in the rock type classification problem, four average SNR values were considered in the range of 10 to 60 dB. After the three superposition bands for each SNR case were determined, they were applied to the spectral content of each pixel in the training and testing regions shown in Fig. 6.

We also considered an application of CCFS to spectral unmixing and abundance estimation. The scene used for this application is a different AHI test-flight image, sections of which are shown in Fig. 7, which represent a snapshot of an urban area at $\lambda = 7.8267 \mu\text{m}$. The scene contains buildings, roads, vegetation, parking lots, and cars.



Fig. 7. Segments of AH1 test-flight image of an urban area at 7.8267 μm .

Spectral unmixing consists of three main stages: feature extraction, endmembers determination followed by unmixing, and fractional abundance estimation. Unmixing methods can generally be classified by the endmember determination process as automatic and interactive; the automatic methods estimate the number of endmembers, their spectral signatures, and abundance patterns using only the mixed data, the mixing model with no *a priori* information about the ground materials, and any human intervention [33]–[35]. In interactive unmixing, an analyst or expert chooses the “pure pixels” from the image or the endmember spectra from the spectral library and then estimates the fractional abundance patterns of the component materials in the image. In this paper, we used the interactive method while following the three stages described above.

First, by means of visual inspection, three main endmember categories, i.e., buildings, roads, and vegetation, were identified in the scene area part of which is captured in the image in Fig. 7. The representative spectral signatures were determined by calculating the mean of each region corresponding to the designated endmember category. Endmember determination was followed by spectral feature extraction where the CCFS was applied to determine the three most informative directions in the AH1 spectral space with respect to the three endmembers in the presence of noise. The extraction of the three superposition features, one for each endmember, follows the same approach as done in the supervised classification problem described earlier.

The last step was to estimate the abundance fraction of each endmember in every pixel from the tested area. Assuming a linear mixing model, the fractions of the endmembers can be determined by solving the problem of minimizing

$$e = \|\mathbf{x} - \mathbf{S}\mathbf{b}\|^2$$

where \mathbf{S} is the 3×3 matrix when the CCFS approach is applied to the data, whose three columns correspond to the endmembers and three rows are the superposition features, \mathbf{x} represents the mixed spectrum, and \mathbf{b} is the 3×1 fractional abundance vector. Considering the physical meaning of the mixing model, the elements of the abundance vector \mathbf{b} can be subject to two constraints: $b_i \geq 0$, $i = 1, 2, 3$, and $\sum_{i=1}^3 b_i = 1$.

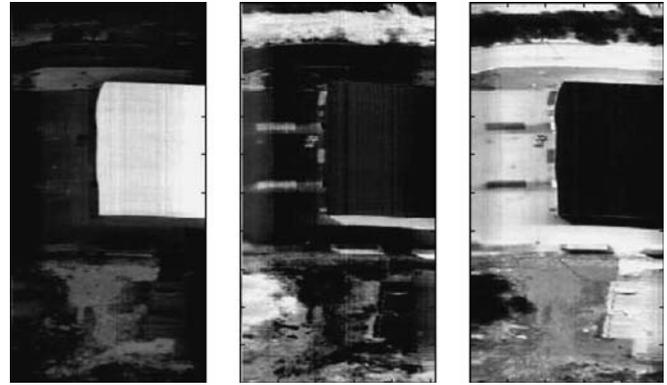


Fig. 8. (Left to right) Abundance estimation maps for endmembers building, vegetation, and road, respectively, using three uniformly spaced AH1 spectral bands in the range of 7.7 to 8.6 μm .

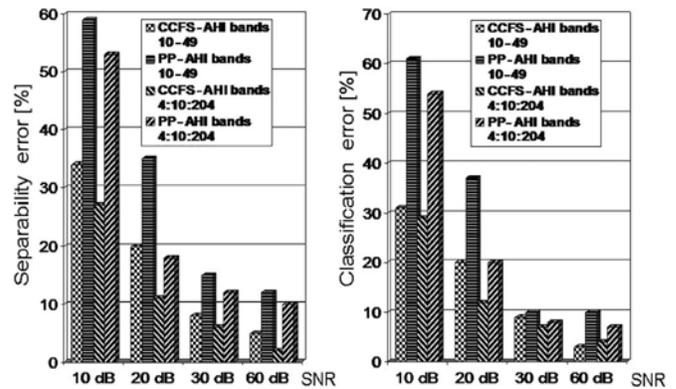


Fig. 9. (Left) Separability and (right) classification results for two subsets of AH1 bands and when CCFS and noise-adjusted PP are used.

1) *Results and Discussion:* To set a benchmark for the performance of the CCFS approach in the supervised classification and abundance estimation problems, we first discuss the results in the absence of noise. The Bayesian classification results for the three classes of interests (buildings, roads, and vegetation) for five randomly selected subsets of the AH1 spectral bands show perfect separability and classification. As for the problem of spectral unmixing and abundance estimation, Fig. 8 presents the abundance maps of the three endmembers (buildings, vegetation, and roads) when using three uniformly separated AH1 spectral bands in the range of 7.7 to 8.6 μm . The size of the tested subimage used here is 500 by 256 pixels. Fig. 8 shows that each map is able to correctly estimate the fraction of abundance of the corresponding endmember.

Next, we consider the effect of noise and compare the performance of the CCFS approach (in supervised classification and spectral unmixing) to that obtained using the noise-adjusted PP. As in the rock type classification example, four different SNR values are considered in the range of 10 to 60 dB. The search for the three optimal directions in the supervised classification problem for both CCFS and noise-adjusted PP was performed over two different subsets of the AH1 bands. The first subset consists of 40 consecutive AH1 bands in the range of 7.7 to 8.6 μm , and the second set consists of 21 uniformly spaced bands in the range of 7.7 to 11.2 μm .

The average separability and classification results for the supervised classification of road, roof, and vegetation classes,

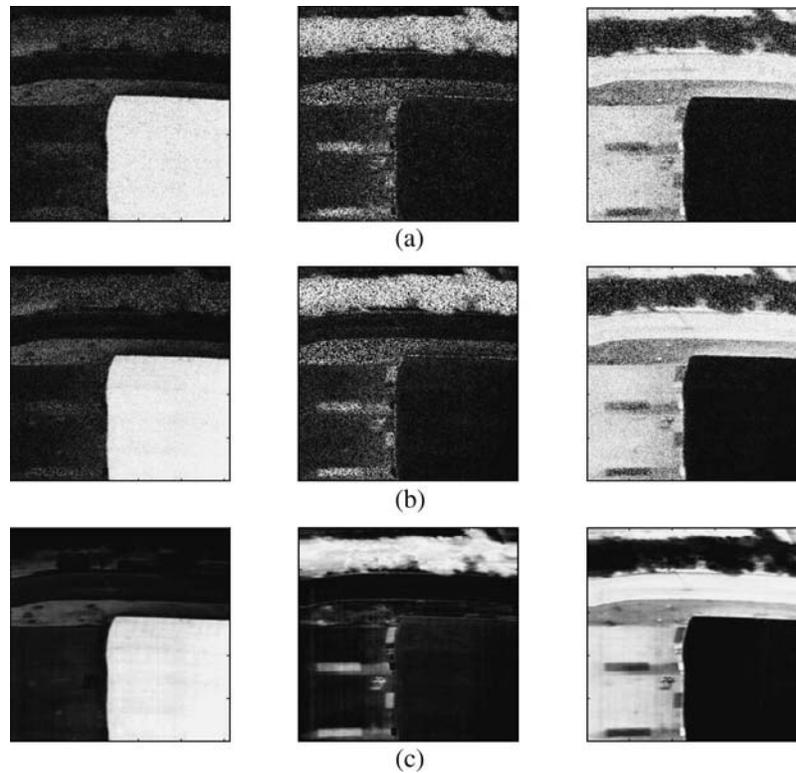


Fig. 10. (Left to right) Abundance maps for building, vegetation, and road endmembers using three superposition features selected by the CCFS algorithm from a subset of 50 bands in the range of 7.7 to 8.6 μm and for SNR levels of (a) 20 dB, (b) 30 dB, and (c) 60 dB.

averaged over 50 noise realizations, are presented in Fig. 9 for both CCFS and noise-adjusted PP approaches. The performance of CCFS in this application is consistent with that corresponding to the rock type classification problem, and it demonstrates good classification in modest SNR scenarios of 10–30 dB. Feature selection from 21 uniformly spaced AHI bands (for both CCFS and noise-adjusted PP) gives improved separability and classification than feature selection from 40 consecutive AHI bands. This result can be explained by the fact that the 40 consecutive AHI bands exhibit higher spectral correlation compared to the 21 uniformly separated bands, and thus, they are potentially more sensitive to the presence of noise. The noise-adjusted PP shows comparable performance to the CCFS algorithm; however, in this application the CCFS gives improved separability and classification compared to the noise-adjusted PP for all SNR cases. We point out that for these applications, we have observed a very high sensitivity of the performance of the fast ICA implementation of the PP to the initial guess for the projection matrix. In some cases, the classification and separability errors were low; however, in other cases, they were much higher than the averaged errors presented in the tables. One possible explanation is that the initialization of the projection matrix by random numbers may not necessarily yield a good initial guess for the hyperspectral data involved.

Fig. 10(a)–(c) shows three groups of fractional abundance maps for SNR values of 20, 30, and 60 dB, respectively, and when the CCFS is applied to 50 consecutive AHI bands in the range of 7.7 to 8.6 μm . The corresponding results for the noise-adjusted PP approach are shown in Fig. 11(a)–(b). The size of

the subimage used for this problem is 250 by 256 pixels, and it represents a subsection of the image shown in Fig. 8. It is seen that the CCFS approach once again shows good performance. The CCFS and the noise-adjusted PP similarly perform for the SNR value of 10 dB (results not shown). Figs. 10(a) and 11(a) compare the abundance maps created using the three CCFS features and three noise-adjusted PP features, respectively, for the SNR value of 20 dB. The maps show improved performance of the CCFS compared to the noise-adjusted PP, which was not able to clearly discriminate between the endmembers of vegetation and road in this SNR case. As expected, the results for both CCFS and noise-adjusted PP improve as the SNR is increased, as shown in Figs. 10(b) and 11(b). For the high SNR case of 60 dB, we compare the performance of CCFS described by the abundance maps in Fig. 10(c) to the AHI image in Fig. 7 and to the abundance maps presented in Fig. 8, representing the noiseless case when three AHI bands are used. The results show that at high SNR values, the performance of the CCFS approaches the noiseless limit.

We end this section by concluding that the examples considered suggest that the proposed CCFS method offers a noticeable improvement over the noise-adjusted PP algorithm in the cases of low and high SNR. Of course, these improvements come at a price of using numerical optimization procedures to compute the CCFS weights, which is the most expensive step in the CCFS algorithm. However, the cost of the optimization step can significantly be reduced by a judicious choice of the initial guess for the CCFS weights. Our implementation takes advantage of the fact that in the absence of noise, the optimization algorithm essentially computes the standard

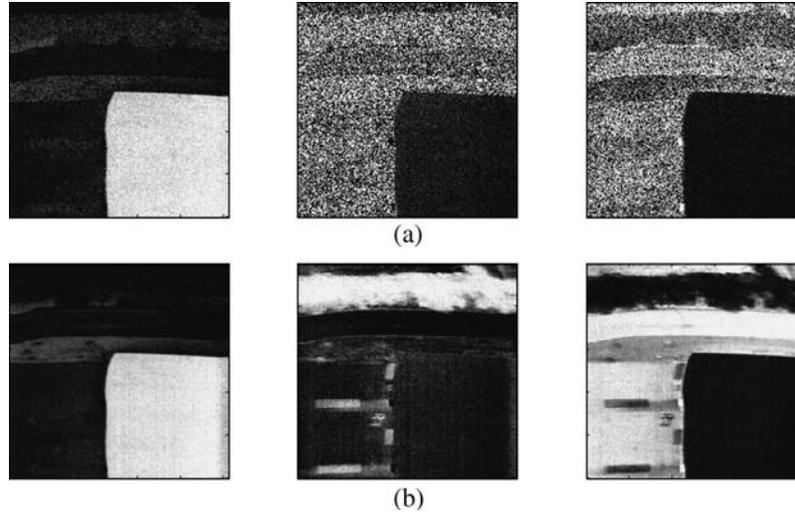


Fig. 11. (Left to right) Abundance maps for building, vegetation, and road endmembers using three spectral features selected by the noise-adjusted PP from a subset of 50 bands in the range of 7.7 to 8.6 μm and for an SNR level of (a) 20 dB and (b) 30 dB.

orthogonal projection; we, therefore, choose the coefficients of this projection as an initial guess for the optimization algorithm. In our calculations, we have observed that this choice of the initial guess results in substantial reduction in the number of optimization steps needed for convergence.

V. CONCLUSION

We have developed a problem-specific feature-selection algorithm that is appropriate for the general class of sensors whose bands are both noisy and spectrally overlapping. Our approach is based upon statistical projection-like concepts in Hilbert spaces in conjunction with CC analysis. For a given class of patterns, the algorithm seeks for a set of weights that are used to determine the optimal superposition band or sensing direction. The obtained sensing direction is optimal in the sense that it provides the best MMSE estimate of the mean of a class in the sensor space. In particular, the superposition band yields the best sensing direction, taking into account both information content and noise. The superposition-band selection procedure is sequentially repeated as many times as the number of the classes of interest, producing a canonical set of superposition bands. At each stage, the algorithm excludes from the search for the optimal direction the class that has been selected in the prior stage; moreover, every superposition band is selected from a subspace of the sensor space that is in the orthogonal complement of the previous sensing direction.

The feature-selection algorithm was applied to a QDIP LWIR sensor as a realistic representative of the class of sensors with highly overlapping and noisy spectral bands and to the AHI sensor. As demonstrated by the separability and classification results for both applications, in the presence of noise, the proposed CCFS algorithm can effectively reduce the sensor-space dimensionality while maintaining good separability and classification results. Moreover, the CCFS method provides accurate abundance fraction estimation of the endmembers in the spectral unmixing problem of the AHI hyperspectral image

data. The proposed algorithm outperforms the noise-adjusted PP technique in the cases of low and high SNR. The proposed CCFS algorithm promises robustness to the photocurrent noise by yielding sensing directions with maximal information content and minimized cumulative noise associated with each direction.

APPENDIX A PROOF OF LEMMA 1

By using the fact that $(p - p_{\mathcal{F}})$ is orthogonal to $p_{\mathcal{F}}$ (in [19, Th. 4.11]), we obtain

$$\langle p, f_p \rangle f_p = \frac{\langle (p - p_{\mathcal{F}}) + p_{\mathcal{F}}, p_{\mathcal{F}} \rangle}{\langle p_{\mathcal{F}}, p_{\mathcal{F}} \rangle} p_{\mathcal{F}} = p_{\mathcal{F}}. \quad (13)$$

Therefore

$$\|p - \langle p, f_p \rangle f_p\| = \|p - p_{\mathcal{F}}\|. \quad (14)$$

Hence, since $\inf_{g \in \mathcal{F}} \|p - g\| = \|p - p_{\mathcal{F}}\|$, (14) along with the fact that $\|f_p\| = 1$ together imply

$$\inf_{f \in \mathcal{F}, \|f\|=1} \|p - \langle p, f \rangle f\| = \|p - \langle p, f_p \rangle f_p\|. \quad (15)$$

Thus, we have proved that the infimum in (15) is achieved at $f = f_p$ or

$$\begin{aligned} \inf_{f \in \mathcal{F}, \|f\|=1} \|p - \langle p, f \rangle f\| &= \min_{f \in \mathcal{F}, \|f\|=1} \|p - \langle p, f \rangle f\| \\ &= \|p - \langle p, f_p \rangle f_p\|. \end{aligned}$$

APPENDIX B
PROOF OF LEMMA 2

Note that

$$\begin{aligned} \mathbf{E} [\|p - p_f\|^2] &= \|p\|^2 - 2 \sum_{i=1}^k \sum_{j=1}^k a_i a_j \langle p, f_i \rangle \langle p, f_j \rangle \\ &\quad + \sum_{i=1}^k \sum_{j=1}^k a_i a_j \langle p, f_i \rangle \langle p, f_j \rangle \|f\|^2 \\ &\quad + \sum_{i=1}^k \sum_{j=1}^k a_i a_j \mathbf{E}[N_i N_j] \|f\|^2 \\ &\quad - 2 \sum_{i=1}^k \sum_{j=1}^k a_i a_j \mathbf{E}[N_i] \langle p, f_j \rangle \\ &\quad + 2 \sum_{i=1}^k \sum_{j=1}^k a_i a_j \mathbf{E}[N_i] \langle p, f_j \rangle \|f\|^2. \end{aligned} \quad (16)$$

Using the stated assumptions on noise statistics and the norm of p , we obtain

$$\begin{aligned} \arg \min_{\mathbf{a} \in \mathbb{R}^k, \|f\|=1} \mathbf{E} [\|p - p_f\|^2] &= \arg \min_{\mathbf{a} \in \mathbb{R}^k, \|f\|=1} \left\{ 1 - \langle p, f \rangle^2 + \sum_{i=1}^k a_i^2 \sigma_i^2 \right\} \\ &= \arg \max_{\mathbf{a} \in \mathbb{R}^k, \|f\|=1} \left\{ \langle p, f \rangle^2 - \sum_{i=1}^k a_i^2 \sigma_i^2 \right\}. \end{aligned} \quad (17)$$

ACKNOWLEDGMENT

The authors would like to thank D. Ramirez, S. Annamalai, and Ü. Sakoglu for providing the QDIP data used in this paper and for many fruitful discussions, and T. Williams and M. Wood for providing AHI test flight hyperspectral data.

REFERENCES

- [1] J. Jiang, K. Mi, R. McClintock, M. Razeghi, G. J. Brown, and C. Jelen, "Demonstration of 256×256 focal plane array based on Al-free GaInAs-InP QWIP," *IEEE Photon. Technol. Lett.*, vol. 15, no. 9, pp. 1273–1275, Sep. 2003.
- [2] S. Krishna, S. Ragahavan, G. Winkel, A. Stinz, G. Ariawansa, S. G. Matsik, and A. Perera, "Three-color ($\lambda_{p1} \sim 3.8 \mu\text{m}$, $\lambda_{p2} \sim 8.5 \mu\text{m}$, and $\lambda_{p3} \sim 23.2 \mu\text{m}$) InAs/InGaAs quantum-dots-in-a-well detectors," *Appl. Phys. Lett.*, vol. 83, no. 14, pp. 2745–2747, Oct. 2003.
- [3] S. Krishna, "Optoelectronic properties of self-assembled InAs/InGaAs quantum dots," *III-V Semiconductor Heterostructures: Physics and Devices*, vol. 3438, pp. 234–242, 2003.
- [4] S. Krishna, "Quantum dots-in-a-well infrared photodetectors," *J. Phys. D, Appl. Phys.*, vol. 38, no. 13, pp. 2142–2150, Jul. 2005.
- [5] J. Topol'ancik, S. Pradhan, P. C. Yu, S. Chosh, and P. Bhattacharya, "Electrically injected photonic crystal edge-emitting quantum-dot light source," *IEEE Photon. Technol. Lett.*, vol. 16, no. 4, pp. 960–962, Apr. 2004.
- [6] K. T. Posani, V. Thirupati, S. Annamalai, N. Weirs-Einstein, S. Krishna, P. Perahia, O. Crisafulli, and O. J. Painter, "Nanoscale quantum dot infrared sensors with photonic crystal cavity," *Appl. Phys. Lett.*, vol. 88, no. 15, pp. 151 104-1–151 104-3, Apr. 2006.
- [7] Ü. Sakoglu, J. S. Tyo, M. M. Hayat, S. Raghavan, and S. Krishna, "Spectrally adaptive infrared photodetectors with bias-tunable quantum dots," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 21, no. 1, pp. 7–17, Jan. 2004.
- [8] Ü. Sakoglu, M. M. Hayat, J. S. Tyo, P. Dowd, S. Annamalai, K. T. Posani, and S. Krishna, "Statistical adaptive sensing by detectors with spectrally overlapping bands," *Appl. Opt.*, vol. 45, no. 28, pp. 7224–7234, Oct. 2006.
- [9] S. Krishna, M. M. Hayat, J. S. Tyo, S. Raghavan, and Ü. Sakoglu, "Detector with tunable spectral response," U.S. Patent No. 7 217 951, May 15, 2007.
- [10] P. Bhattacharya, *Semiconductor Optoelectronic Devices*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
- [11] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.
- [12] J. H. Friedman, "Exploratory projection pursuit," *J. Amer. Stat. Assoc.*, vol. 82, no. 397, pp. 249–266, Mar. 1987.
- [13] L. O. Jimenez and D. Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2653–2667, Nov. 1999.
- [14] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
- [15] M. Lennon and G. Mercier, "Noise-adjusted non orthogonal linear projections for hyperspectral data analysis," in *Proc. IGARSS*, 2003, vol. 6, pp. 3760–3762.
- [16] B. S. Paskaleva, M. M. Hayat, J. S. Tyo, Z. Wang, and M. Martinez, "Feature selection for spectral sensors with overlapping noisy spectral bands," *Proc. SPIE*, vol. 6233, pp. 623 329.1–623 329.7, 2006.
- [17] Z. Wang, B. S. Paskaleva, J. S. Tyo, and M. M. Hayat, "Canonical correlations analysis for assessing the performance of adaptive spectral imagers," *Proc. SPIE*, vol. 5806, pp. 23–34, 2005.
- [18] Z. Wang, J. S. Tyo, and M. M. Hayat, "Data interpretation for spectral sensors with correlated bands," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 24, no. 9, pp. 2864–2870, Sep. 2007.
- [19] W. Rudin, *Real and Complex Analysis*. New York: McGraw-Hill, 1986.
- [20] J. Dauxois and G. M. Nkiet, *Canonical Analysis of Two Euclidean Subspaces and its Application*, vol. 27. Amsterdam, The Netherlands: Elsevier, 1997, pp. 354–387.
- [21] A. Björck and G. H. Golub, "Numerical methods for computing angles between linear subspaces," *Math. Comput.*, vol. 27, no. 123, pp. 579–594, Jul. 1973.
- [22] A. V. Knyazev and M. E. Argentati, "Principal angles between subspaces in a A -based scalar product: Algorithms and perturbation estimates," *SIAM J. Sci. Comput.*, vol. 23, no. 6, pp. 2009–2041, 2002.
- [23] B. Paskaleva, M. M. Hayat, M. M. Moya, and R. J. Fogler, "Multispectral rock-type separation and classification," *Proc. SPIE*, vol. 5543, pp. 152–163, 2004.
- [24] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1984.
- [25] S. W. Ruff, P. R. Christensen, P. W. Barbera, and D. L. Anderson, "Quantitative thermal emission spectroscopy of minerals: A laboratory technique for measurement and calibration," *J. Geophys. Res.*, vol. 102, no. B7, pp. 14 899–14 913, 1997.
- [26] F. D. Palluconi and G. R. Meeks, *Thermal Infrared Multispectral Scanner (TIMS): An Investigator's Guide to TIMS Data*. Pasadena, CA: Jet Propuls. Lab., 1985.
- [27] K. C. Feely and P. R. Christensen, "Quantitative compositional analysis using thermal emission spectroscopy: Application to igneous and metamorphic rocks," *J. Geophys. Res.*, vol. 104, no. E10, pp. 24 195–24 210, Oct. 1999.
- [28] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Hoboken, NJ: Wiley, 2000.
- [29] W. B. Clodius, P. G. Weber, C. C. Borel, and B. W. Smith, "Multispectral thermal imaging," *Proc. SPIE*, vol. 3438, pp. 234–242, 1998.
- [30] C. B. Akgül, "Projection pursuit for optimal visualization of multivariate data," Ph.D. dissertation, Bogazici Univ., Istanbul, Turkey, 2003. [Online]. Available: <http://www.tsi.enst.fr/akgul/oldprojects/qli>
- [31] D. Landgrebe, "Information extraction principles and methods for multispectral and hyperspectral image data," *Inf. Process. Remote Sens.*, vol. 82, pp. 3–38, 1999.
- [32] P. G. Lucey, E. M. Winter, and T. J. Williams, *Two Years of Operations of the AHI: A LWIR Hyperspectral Imager*. [Online]. Available: <http://0-www.higp.hawaii.edu.pugwash.lib.warwick.ac.uk/winter/pubs/>
- [33] J. W. Boardman, "Analysis, understanding, and visualization of hyperspectral data as convex sets in n space," *Proc. SPIE*, vol. 2480, pp. 14–22.

- [34] M. Winter, "Fast autonomous spectral endmembers determination in hyperspectral data," in *Proc. 13th Int. Conf. Appl. Geologic Remote Sens.*, Vancouver, BC, Canada, 1999, vol. II.
- [35] C. Kwan, B. Ayhan, G. Chen, J. Wang, J. Baohong, and C.-I. Chang, "A novel approach for spectral unmixing, classification, and concentration estimation of chemical and biological agents," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 409–419, Feb. 2006.



Biliana Paskaleva received her B.S. in electrical engineering in 1992 from the Technical University in Varna, Bulgaria. She subsequently received her M.S. in electrical engineering in 2004 from the University of New Mexico. Biliana is currently a Ph.D. candidate in the Electrical and Computer Engineering Department and the Center for High Technology Materials at the University of New Mexico. Since 2006, Biliana has been working as an Intern at Sandia National Laboratories in Albuquerque, New Mexico. Her research interests are in the areas of remote sensing, spectro-spatial feature extraction, hyperspectral feature selection, pattern classification, image processing, detection and estimation, Bayesian statistics and neural networks.

ing, spectro-spatial feature extraction, hyperspectral feature selection, pattern classification, image processing, detection and estimation, Bayesian statistics and neural networks.



Majeed M. Hayat (S'89–M'92–SM'00) received the B.S. degree (*summa cum laude*) in electrical engineering from the University of the Pacific, Stockton, CA, in 1985 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Wisconsin, Madison, in 1988 and 1992, respectively.

From 1993 to 1996, he was with the University of Wisconsin, Madison, where he was a Research Associate and Co-Principal Investigator of a project on statistical minefield modeling and detection, which was funded by the US Office of Naval Research. In 1996, he was with the Electro-Optics Graduate Program and the Department of Electrical and Computer Engineering, University of Dayton, Dayton, OH. He is currently a Professor with the Department of Electrical and Computer Engineering and the Center for High Technology Materials, University of New Mexico, Albuquerque. His research contributions cover a broad range of topics in statistical communication theory, and signal/image processing, as well as applied probability theory and stochastic processes. Some of his research areas include queuing theory for networks, noise in avalanche photodiodes, equalization in optical receivers, spatial-noise-reduction strategies for focal-plane arrays, and spectral imaging.

Dr. Hayat was the recipient of a 1998 US National Science Foundation Early Faculty Career Award. He is a member of The International Society for Optical Engineers and Optical Society of America. He is an Associate Editor of *Optics Express* and an Associate Editor Member of the Conference Editorial Board of the IEEE Control Systems Society.



Zhipeng Wang (S'04) received the B.S. degree from Tsinghua University, China, and the M.S. degree in optical science and engineering from the University of New Mexico, Albuquerque, in 2000 and 2006, respectively. Currently, he is pursuing the Ph.D. degree in optical sciences at the University of Arizona, Tucson, AZ. Since 2003, Zhipeng has been working on spectral image processing techniques, particularly in analyzing spectral sensors with overlapping bands.



J. Scott Tyo (S'96–M'97–SM'06) received the B.S.E., M.S.E., and Ph.D. degrees from the University of Pennsylvania, Philadelphia, in 1994, 1996, and 1997, respectively, all in electrical engineering.

From 1994 to 2001, he was an officer in the U.S. Air Force, leaving service at the rank of Captain. From 1996 to 1999, he was a Research Engineer with the Directed Energy Directorate, USAF Research Laboratory, Kirtland, NM. From 1999 to 2001, he was with the Electrical and Computer Engineering (ECE) Department, U.S. Naval Postgraduate School, Monterey, CA. From 2001 to 2006, he was a faculty member with the ECE Department, University of New Mexico, Albuquerque. He is currently an Associate Professor with the College of Optical Sciences, University of Arizona, Tucson. His research interests are in the physical aspects of optical and microwave remote sensing, including ultrawideband and SAR, and polarimetric and hyperspectral imagery.

Dr. Tyo is a member of the IEEE Geoscience and Remote Sensing Society, IEEE Antennas and Propagation Society, and IEEE Laser and Electro-Optics Society, the Optical Society of America, Commissions B and E of the International Scientific Radio Union, The International Society for Optical Engineers, Tau Beta Pi, and Eta Kappa Nu.

Dr. Tyo is a member of the IEEE Geoscience and Remote Sensing Society, IEEE Antennas and Propagation Society, and IEEE Laser and Electro-Optics Society, the Optical Society of America, Commissions B and E of the International Scientific Radio Union, The International Society for Optical Engineers, Tau Beta Pi, and Eta Kappa Nu.



Sanjay Krishna (S'98–M'01–SM'08) received the M.S. degree in physics from the Indian Institute of Technology, Madras, in 1996 and the M.S. degree in electrical engineering and the Ph.D. degree in applied physics from the University of Michigan, Ann Arbor, in 1999 and 2001, respectively.

In 2001, he joined the University of New Mexico as a tenure track Faculty Member. He is currently an Associate Professor of electrical and computer engineering with the Center for High Technology Materials, University of New Mexico, Albuquerque.

He has authored/coauthored more than 50 peer-reviewed journal articles, over 50 conference proceedings, and two book chapters, and has four provisional patents. His present research interests include growth, fabrication, and characterization of self-assembled quantum dots and type-II InAs/InGaSb-based strain layer superlattices for mid-infrared lasers and detectors. In his research group, studies are also undertaken on carrier dynamics and relaxation mechanisms in quasi-zero-dimensional systems and the manipulation of these favorable relaxation times to realize high-temperature mid-infrared detectors.

Dr. Krishna received the Gold Medal from the Indian Institute of Technology in 1996 for the best academic performance in the master's program in physics. He received the Best Student Paper Award at the 16th North American Molecular Beam Epitaxy Conference, Banff, AB, Canada, in 1999, the 2002 Ralph E Powe Junior Faculty Award from Oak Ridge Associated Universities, the 2003 Outstanding Engineering Award from the IEEE Albuquerque Section, 2004 Outstanding Researcher Award from the ECE Department, the 2005 School of Engineering Junior Faculty Teaching Excellence Award, and the 2007 NCMR Chief Scientist Award for Excellence. He has also served as the Chair of the local IEEE/LEOS chapter.