Does Independent Component Analysis Play a Role in Unmixing Hyperspectral Data?

José M. P. Nascimento, *Student Member, IEEE,* and José M. Bioucas Dias, *Member, IEEE*

***Abstract*—Independent component analysis (ICA) has recently been proposed as a tool to unmix hyperspectral data. ICA is founded on two assumptions: 1) the observed spectrum vector is a linear mixture of the constituent spectra (endmember spectra) weighted by the correspondent abundance fractions (sources); 2) sources are statistically independent. Independent factor analysis (IFA) extends ICA to linear mixtures of independent sources immersed in noise. Concerning hyperspectral data, the first assumption is valid whenever the multiple scattering among the distinct constituent substances (endmembers) is negligible, and the surface is partitioned according to the fractional abundances. The second assumption, however, is violated, since the sum of abundance fractions associated to each pixel is constant due to physical constraints in the data acquisition process. Thus, sources cannot be statistically independent, this compromising the performance of ICA/IFA algorithms in hyperspectral unmixing. This paper studies the impact of hyperspectral source statistical dependence on ICA and IFA performances. We conclude that the accuracy of these methods tends to improve with the increase of the signature variability, of the number of endmembers, and of the signal-to-noise ratio. In any case, there are always endmembers incorrectly unmixed. We arrive to this conclusion by minimizing the mutual information of simulated and real hyperspectral mixtures. The computation of mutual information is based on fitting mixtures of Gaussians to the observed data. A method to sort ICA and IFA estimates in terms of the likelihood of being correctly unmixed is proposed.**

***Index Terms*—Independent component analysis (ICA), independent factor analysis (IFA), mixture of Gaussians, unmixing hyperspectral data.**

I. Introduction

THE DEVELOPMENT of high spatial resolution airborne and spaceborne sensors has improved the capability of ground-based data collection in the fields of agriculture, geography, geology, mineral identification, detection, and classification of targets activities [1]–[7].

Hyperspectral sensors use many contiguous bands of high spectral resolution covering the visible, near-infrared, and short-wave infrared spectral bands (0.3–2.5μm) [8], [9]. Letting alone the effects of the atmosphere, the signal read by a hyperspectral sensor at a given band and from a given pixel is a mixture of the energies scattered by the constituent substances located in the respective pixel spatial coverage [10].

Manuscript received July 22, 2003; revised October 3, 2004. This work was supported by the Fundação para a Ciência e Tecnologia under Projects POSI/34071/CPS/2000 and PDCTE/CPS/49967/2003.

J. M. P. Nascimento is with the Instituto de Telecomunicações and Instituto Superior de Engenharia de Lisboa, 1949-001 Lisbon, Portugal (e-mail: zen@isel.pt).

J. M. B. Dias is with the Instituto de Telecomunicações and Instituto Superiror Técnico, 1949–001 Lisbon, Portugal (e-mail: bioucas@lx.it.pt).

Digital Object Identifier 10.1109/TGRS.2004.839806

Hyperspectral unmixing is the decomposition of the pixel spectra into a collection of constituent spectra, or spectral signatures, and their corresponding fractional abundances that indicates the proportion of each endmember present in the pixel. Depending on the mixing scales at each pixel, the observed mixture is either linear or nonlinear [4], [11]. The linear mixing model holds when the mixing scale is macroscopic [12]. The nonlinear model holds when the mixing scale is microscopic (or intimate mixtures) [13], [14]. The linear model assumes negligible interaction among distinct endmembers [15], [16]. The nonlinear model assumes that incident solar radiation is scattered by the scene through multiple bounces involving several endmember [17].

Under the linear mixing model, and assuming that the number of endmembers and their spectral signatures are known, hyperspectral unmixing is a linear problem, which can be addressed, for example, under the maximum-likelihood setup [18], the constrained least squares approach [19], the spectral signature matching [20], the spectral angle mapper [21], and the subspace projection methods [19], [22], [23]. Orthogonal subspace projection was introduced in [22]. This technique reduces the data dimensionality, suppresses undesired spectral signatures, and detects the presence of a spectral signature of interest. The basic concept is to project each pixel onto a subspace that is orthogonal to the undesired signatures. As shown in [18], the orthogonal subspace projection technique is equivalent to the maximum-likelihood estimator. This projection technique was extended by three unconstrained least squares approaches [23] (signature space orthogonal projection, oblique subspace projection, target signature space orthogonal projection). Other works using maximum *a posteriori* probability framework [24] and projection pursuit [25], [26] have also been applied to hyperspectral data.

In most cases, the number of endmembers and their signatures are not known. Independent component analysis (ICA) is an unsupervised source separation process [27] that has shown success in blind source separation, feature extraction, and unsupervised recognition [28]. ICA consists in finding a linear decomposition of observed data into statistically independent components.

Given that hyperspectral data are, in given circumstances, linear mixtures, ICA comes to mind as a possible tool to unmix this class of data. In fact, the application of ICA to hyperspectral data has been proposed in [29], where endmember signatures are treated as sources and the mixing matrix is composed by the abundance fractions, and in [24], [30]–[36], where sources are the abundance fractions of each endmember. However, ICA is based on the assumption of mutually independent sources, which is not the case of hyperspectral data, since the sum of the abundance fractions is constant, implying dependence among

abundances. This dependence compromises ICA applicability to hyperspectral images. In addition, hyperspectral data is immersed in noise, which degrades the ICA performance.

Independent factor analysis (IFA) [37] was introduced as a method for recovering independent hidden sources from their observed noisy mixtures. IFA is divided in two steps. First, source densities and noise covariance are estimated from the observed data by maximum likelihood. Second, sources are reconstructed by an optimal nonlinear estimator. Although IFA is a well suited technique to unmix independent sources under noisy observations, the dependence among abundance fractions in hyperspectral imagery compromises, as in the ICA case, the IFA performance.

This paper addresses hyperspectral data source dependence and its impact on ICA and IFA performances. The study considers simulated and real data and is based on mutual information minimization. Hyperspectral observations are described by a generative model. This model takes into account the degradation mechanisms normally found in hyperspectral applications, namely signature variability [38]–[40], abundance constraints, topography modulation, and system noise. The computation of mutual information is based on fitting mixtures of Gaussians (MOG) to data. The MOG parameters (number of components, means, covariances, and weights) are inferred using the minimum-description-length (MDL)-based algorithm [41]. We study the behavior of the mutual information as function of the unmixing matrix. The conclusion is that the unmixing matrix minimizing the mutual information might be very far from the true one. Nevertheless, some abundance fractions might be well separated, mainly in the presence of strong signature variability, large number of endmembers, and a high signal-to-noise ratio (SNR).

This paper is organized as follows. Section II presents a spectral radiance model and formulates the spectral unmixing as a linear problem accounting for abundance constraints, signature variability, topography modulation, and system noise. Section III presents a brief resume of ICA and IFA algorithms. Section IV illustrates the performance of IFA and of some well-known ICA algorithms with experimental data. Section V studies the impact of hyperspectral source dependence on the mutual information. Section VI studies the performance of ICA and IFA on hyperspectral data, based on simulated data. Section VII presents results of ICA based on real data. Section VIII concludes with some remarks.

II. Spectral Radiance Model

Fig. 1 schematizes a typical passive remote sensing scenario. The sun illuminates a random media formed by the earth surface and by the atmosphere; a sensor (airborne or spaceborne) reads, within its *instantaneous field of view* (IFOV), the scattered radiance in the solar-reflectance region extending from 0.3–2.5 μm, encompassing the visible, near-infrared, and shortwave infrared bands. Angles θ and ϕ, with respect to the normal $\mathbf{n}$ on the ground, are the colatitude and the longitude, respectively. The solar and sensor directions are (θ_0, ϕ_0) and (θ_s, ϕ_s), respectively.

The total radiance at the surface level is the sum of three components, as schematized in Fig. 1: the sunlight (ray 1), the skylight (ray 2), and the light due to the *adjacency effect* (ray 3),

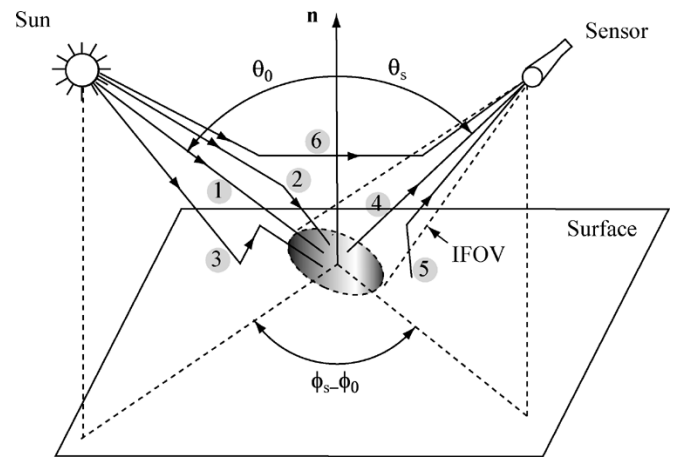


Fig. 1. Schematic diagram of the main contributions to the radiance read by the sensor in the solar spectrum.

i.e., due to the successive reflections and scattering between the surface and the atmosphere. Following [42] and [43], the spectral radiances of these components, are, at a given wavelength λ, respectively, given by the following.

1) $L_1 = \mu_0 E_0 T_\downarrow$, where E_0 is the solar flux at the top of the atmosphere, $\mu_0 = \cos(\theta_0), T_\downarrow = T_\downarrow(\theta_0)$ is the downward transmittance.
2) $L_2 = \mu_0 E_0 t_\downarrow$, where $t_\downarrow = t_\downarrow(\theta_0)$ is the downward diffuse transmittance factor.
3) $L_3 = \mu_0 E_0 T'_\downarrow([\rho_t S + (\rho_t S)^2 + (\rho_t S)^3 \ldots]$, where $T'_\downarrow = [T_\downarrow + t_\downarrow]$, ρ_t is the mean reflectance of the surroundings with respect to the atmospheric point spread function, and S is the spherical albedo of the atmosphere.

The total radiance incident upon the sensor location is the sum of three components: the light scattered by the surface (ray 4), the light scattered by the surface and by the atmosphere (ray 5), and light scattered by the atmosphere (ray 6), the so-called *path radiance*. Assuming a Lambertian surface, and again following [42] and [43], these radiances at the top of the atmosphere are, at wavelength λ, respectively, given by the following.

1) $L_4 = (\mu_0 E_0/\pi)(T'_\downarrow T_\uparrow/(1-\rho_t S))\rho$, where ρ is the surface reflectance and $T_\uparrow = T_\uparrow(\theta_0)$ is the upward transmittance.
2) $L_5 = (\mu_0 E_0/\pi)(T'_\downarrow t_\uparrow/(1-\rho_t S))\rho_t$, where $t_\uparrow = t_\uparrow(\theta_s)$ is the upward diffuse transmittance factor.
3) $L_6 = (\mu_0 E_0/\pi)\rho_a$, where $\rho_a(\theta_0, \theta_s, \phi_s - \phi_0)$ is the atmosphere reflectance.

The total radiance L incident upon the sensor location is thus

$$L = a\rho + b$$

where

$$a = \frac{\mu_0 E_0}{\pi} \frac{T'_\downarrow T_\uparrow}{1-\rho_{tS}} \tag{1}$$

$$b = \frac{\mu_0 E_0}{\pi} \left(\frac{T'_\downarrow t_\uparrow}{1-\rho_{tS}} \rho_t + \rho_a \right). \tag{2}$$

Let us assume that the sensor has B channels (wavebands). Assuming linear receivers and narrow wavebands, the signal at

the output the ith channel (waveband centered at wavenumber λ_i) is given by

$$r_i = c_i\rho + d_i + n_i$$

where c_i and d_i are proportional to $a(\lambda_i)$ and $b(\lambda_i)$, respectively, and n_i denotes the receiver electronic noise at channel i plus the Poisson (photonic) signal noise (e.g., see [44]).

Terms a and b in (1) and (2) depend in a complex way of the sun and sensor directions, of the atmosphere composition, of the topography, and of the scene materials and configurations [42], [43], [45]. The compensation for this terms, the so-called atmospheric correction, is a necessary step in many quantitative algorithms aimed at extracting information from multispectral or hyperspectral imagery [43], [46], [47].

This paper addresses linear unmixing of fractional abundances at the pixel level. The term linear means that the observed entities are linear combinations of the endmembers spectral signatures weighted by the correspondent fractional abundances. Therefore, we assume that atmospheric correction has been applied to a degree assuring a linear relation between the radiance L and the reflectance ρ, i.e., for each channel, the relation between the radiance and the reflectivity is linear with coefficients not depending on the pixel.

The details of the atmospheric correction necessary to achieve such a linear relation are beyond the scope of the paper. Notice, however, that it may happen that no correction is necessary. That is the case when the scene is a surface of approximately constant altitude, the atmosphere is horizontally homogeneous, and ρ_t, the mean reflectance of the surroundings, exhibits negligible variation.

A. Linear Spectral Mixture Model

In spectral mixture modeling, the basic assumption is that the surface is made of a few number of endmembers of relatively constant spectral signature or, at least, constant spectral shape. If the multiple scattering among distinct endmembers is negligible and the surface is partitioned according to the fractional abundances, then the spectral radiance upon the sensor location is well approximated by a linear mixture of endmember radiances weighted by the correspondent fractional abundances [4], [10], [11], [33], [48].

Under the linear mixing model and assuming that the sensor radiation pattern is ideal, i.e., constant in the IFOV and zero outside, the output of channel i from a given pixel is

$$r_i = c_i \sum_{j=1}^{p} \rho_{ij}\alpha_j + d_i + n_i \tag{3}$$

where ρ_{ij} denotes the reflectance of endmember j at wavenumber λ_i, α_j denotes the fractional abundance of endmember j at the considered pixel, and p is the number of endmembers.

Fractional abundances are subject to

$$\sum_{j=1}^{p} \alpha_j = 1 \quad \alpha_j \geq 0, \qquad j = 1, \ldots, p. \tag{4}$$

For a real sensor, the output of channel i is still formally given by (3), but α_j depends on sensor point spread function (PSF) $h_{x,y}(u,v)$ according to

$$\alpha_j = \frac{\int_{A_i} h_{x,y}(-u,-v)dudv}{\int_{\Re^2} h_{x,y}(-u,-v)dudv}$$

where A_i denotes the set of points on the surface belonging to the ith endmember. The PSF $h_{x,y}(u,v)$ may be spatially variant (i.e., it depends on surface coordinates x and y), and it includes the effect of finite aperture and receiver impulse response. Notice that the meaning of α_j depends not only on the true fractional abundance, but also on the endmember distribution inside the IFOV and on the surface point (x,y). Wu and Schowengerdt [49] propose an image-restoration-based approach to mitigate the errors introduced by the nonideal nature of the PSF.

Herein, and for simulation purposes, we assume that the signal at the output of the channel i is given by (3). Endmember radiances were extracted from a hyperspectral subimage of Indian Pine Test Site in Northwestern Indiana acquired by an Airborne Visible Infrared Imaging Spectrometer (AVIRIS) instrument in June 1992. Noisy channels and water absorption channels were removed (channels 1–4, 107–113, 150–166, and 221–224). Since only high SNR channels are considered, Poisson noise is neglected. Concerning atmospheric correction, this image has been processed to remove path radiance (ray 5 in Fig. 1) and the light scattered by interaction between surface and the atmosphere (ray 6 in Fig. 1). Notice, however, that the corrected image is still in radiance units.

Let $\mathbf{r}$ be an $B \times 1$ vector, where B is the total number of bands, and $\mathbf{m}_j^o \equiv [c_1\rho_{1j}, c_2\rho_{2j}, \ldots, c_B\rho_{Bj}]^T$ is the so-called signature of the jth endmember. Equation (3) can be written as

$$\mathbf{r} = \mathbf{M}\alpha + \mathbf{w}' \tag{5}$$

where $\mathbf{M} \equiv [\mathbf{m}_1^o, \mathbf{m}_2^o, \ldots, \mathbf{m}_p^o]$ is a matrix with the signatures of the endmembers present in the covered area, $\boldsymbol{\alpha} \equiv [\alpha_1, \alpha_2, \ldots, \alpha_p]^T$, and $\mathbf{w}'$ models additive receiver electronic noise. The notation $(\cdot)^T$ indicates vector transposed.

Model (5) is an oversimplification of reality, as it does not take into account signature variability (from pixel to pixel) due to changes in the configuration and in the composition of substances, surface contaminants, variation in the substances such as age-induced color fading due to oxidation or bleaching, uncompensated atmospheric and environmental effects, and uncompensated errors in the sensor. Signature variability has been studied and accounted for in a few unmixing algorithms (e.g., see [50]–[52]).

Signature variability is primarily characterized by spectral shape invariance [38], i.e., while the spectral shapes of the endmembers are fairly consistent, their amplitude varies considerably over the scene. Based on this rationale, we model spectral variability of the ith endmember at a given pixel as

$$\mathbf{m}_i = \psi_i \mathbf{m}_i^o + \mathbf{w}_i \tag{6}$$

where $\psi_i \geq 0$ is a scale factor, and $\mathbf{w}_i$ is a zero-mean random vector. Noise $\mathbf{w}_i$ accounts for signature variability not modeled by ψ_i. Introducing (6) into (5), we obtain, for a given pixel

$$\mathbf{r} = \mathbf{M}\boldsymbol{\psi}\boldsymbol{\alpha} + \sum_{i=1}^{p} \alpha_i \mathbf{w}_i + \mathbf{w}' \tag{7}$$

where $\boldsymbol{\psi} \equiv \mathrm{diag}(\psi_1, \ldots, \psi_p)$ is a $p \times p$ diagonal matrix.

Having in mind that illumination variability due to surface topography affects equally all bands and that we are assuming additive system noise, then we can write

$$\mathbf{r} = \mathbf{M}\underbrace{\gamma\boldsymbol{\psi}\boldsymbol{\alpha}}_{\mathbf{s}} + \underbrace{\gamma\sum_{i=1}^{p}\alpha_i\mathbf{w}_i + \mathbf{w}'}_{\mathbf{w}} = \mathbf{M}\mathbf{s} + \mathbf{w} \tag{8}$$

where γ models surface topography. Model (8) is still linear. If $\mathbf{M}$ is known, estimating $\mathbf{s}$ is an ordinary linear problem. If $\mathbf{M}$ is unknown, the problem is more difficult. Considering independent sources $\mathbf{s}$ and no noise, unmixing (8) can be solved under the ICA framework. This has been addressed in [29]–[33]. Linear unmixing of independent sources in the presence of noise has been addressed in [37], where the IFA algorithm was introduced. Herein, and in the presence of noise we adopt this algorithm.

Signature variability introduces randomness among the sources and thereby attenuates their statistical dependencies. This can be understood by computing the correlation factor between sources $s_i = \psi_i\alpha_i$ and $s_j = \psi_j\alpha_j$. Assuming that ψ_i is independent of ψ_j and that $\boldsymbol{\psi}$ is independent of $\boldsymbol{\alpha}$, we have (9)–(11), shown at the bottom of page, where $E[\psi_k^2] \geq E^2[\psi_k]$ was invoked to obtain the right hand side of (10) and (11). We conclude then that signature variability does not increase source correlation. Of course, decorrelation does not imply independency. It is, however, plausible that increasing decorrelation means increasing independency. In fact, in the next sections we give experimental evidence that increasing signature variability improves ICA and IFA results.

III. ICA AND IFA

ICA [28], [53], [54] is an unsupervised source separation process, which has been applied to linear blind separation problems [55]–[57]. The goal of ICA is to recover independent sources, given only sensor observations that are unknown linear mixtures of the unobserved independent sources.

Let $\mathbf{r}$ be an $B \times 1$ observation column vector, such as

$$\mathbf{r} = \mathbf{M}\mathbf{s} \tag{12}$$

where $\mathbf{M}$ is an unknown $B \times p$ $(B \geq p)$ mixing matrix and $\mathbf{s} \equiv [s_1\ s_2\ \ldots\ s_p]^T$ is an unknown random data vector of mutually independent sources having unknown distributions, although, at most, one might be Gaussian distributed. ICA finds a $p \times B$ separating matrix $\mathbf{W}$, such that

$$\mathbf{y} = \mathbf{W}\mathbf{r} = \mathbf{P}\mathbf{C}\mathbf{s} \tag{13}$$

where $\mathbf{y}$ is a vector of independent components, and $\mathbf{P}$ and $\mathbf{C}$ are permutation and scale matrices, respectively.

ICA looks for a linear representation that maximizes a nongaussianity measure [58]. A commonly objective function used in ICA algorithms is the mutual information [28] of vector $\mathbf{y} \equiv [y_1\ y_2\ \ldots\ y_p]^T$ given by

$$\begin{aligned} I(y_1, y_2, \ldots, y_p, \mathbf{W}) &\equiv \sum_i H(y_i) - H(\mathbf{y}) \\ &= \sum_i H(y_i) - H(\mathbf{r}) - \log|\mathbf{W}| \end{aligned} \tag{14}$$

where $H(y_i)$, $H(\mathbf{y})$, and $H(\mathbf{r})$ are the entropy of random variable y_i, of random vector $\mathbf{y}$, and of random vector $\mathbf{r}$, respectively (e.g., see [56] and [57]). The mutual information is a measure of dependence between random variables. It is nonnegative and equals to zero if and only if variables are statistically independent.

Most ICA algorithms find the separating matrix by minimizing (14), or an equivalent objective function, with respect to $\mathbf{W}$. The *Negentropy* (e.g., see [55] and [58]), an entity closely related with the mutual information, has also been used as an objective function to obtain $\mathbf{W}$. It is defined as

$$J(\mathbf{y}) \equiv H(\mathbf{y}_{\mathrm{gauss}}) - H(\mathbf{y}) \tag{15}$$

where $\mathbf{y}_{\mathrm{gauss}}$ is a Gaussian random vector with the same mean and covariance as $\mathbf{y}$ [59]. *Negentropy* is nonnegative and is equal to zero if and only if $\mathbf{y}$ has Gaussian distribution. Assuming that components y_i, $i = 1, \ldots, p$, are uncorrelated, it follows that

$$I(y_1, y_2, \ldots, y_p) = J(\mathbf{y}) - \sum_i J(y_i) \tag{16}$$

which means that finding maximum of *Negentropy* directions, i.e., maximizing $\sum_i J(y_i)$ with respect to $\mathbf{W}$, is equivalent to minimize the mutual information.

Well-known ICA methods are *fastica* [55], *jade* [56], and the Bell and Sejnowski algorithm [57]. *Fastica* is based on a fixed-point procedure and uses the absolute value of kurtosis as a measure of nongaussianity. *Jade* uses the fourth-order cross cumulants of the data to separate sources. The Bell and Sejnowski algorithm use the stochastic gradient ascent learning rule to minimize the mutual information.

IFA [37] was proposed as a method for recovering independent hidden sources from their observed mixtures immersed in additive noise. IFA is divided in two steps. First, source densities and noise covariance are estimated from the observed data by maximum likelihood. Second, sources are reconstructed by an optimal nonlinear estimator. IFA assumes the observation model

$$\mathbf{r} = \mathbf{M}\mathbf{s} + \mathbf{w} \tag{17}$$

$$\left|\varrho_{s_i s_j}\right| = \frac{|E[\psi_i]E[\psi_j]|\left|C_{\alpha_i\alpha_j}\right|}{\sqrt{E\left[\psi_i^2\right]E\left[\alpha_i^2\right] - E^2[\psi_i]E^2[\alpha_i]}\sqrt{E\left[\psi_j^2\right]E\left[\alpha_j^2\right] - E^2[\psi_j]E^2[\alpha_j]}} \tag{9}$$

$$\leq \frac{|E[\psi_i]E[\psi_j]|}{\sqrt{E\left[\psi_i^2\right]E\left[\psi_j^2\right]}}\frac{\left|C_{\alpha_i\alpha_j}\right|}{\sigma_{\alpha_i}\sigma_{\alpha_j}} \tag{10}$$

$$\leq \left|\varrho_{\alpha_i\alpha_j}\right| \tag{11}$$

where sources are independent with unknown distributions (at most one is Gaussian), and $\mathbf{w}$ is Gaussian noise with covariance C_w, a matrix not necessarily diagonal. To make the model analytically tractable, each source density is modeled by a MOG. An expectation–maximization (EM) algorithm [60], [61] is applied to compute the maximum-likelihood estimate of the noise covariance and the Gaussian mixture parameters.

The classical principal component analysis (PCA) [62] seeks for a linear decomposition that best represents data in a least squares sense. PCA finds a linear transformation $\mathbf{y} = \mathbf{W}\mathbf{r}$ where each row vector of $\mathbf{W}$ corresponds to the normalized orthogonal eigenvector of the data covariance matrix. While PCA uses only second-order statistics, ICA looks for components which are statistically independent rather than uncorrelated; thus, it requires statistics of orders higher than the second [58].

In the next section, we give evidence that the most well-known ICA and IFA algorithms do not correctly unmix hyperspectral data. This study is based on several experiments, where the degradation mechanisms normally found in hyperspectral applications, namely signature variability, abundance constraints, topography modulation, and system noise, are simulated in our model.

IV. ICA and IFA Evaluation With Simulated Data

In this section, we apply ICA (*fastica* [55], *jade* [56], and Bell and Sejnowski [57]) and the IFA [37] algorithms to simulated data. Four experiments are conducted: in the first, modeling a canonical ICA scenario, the abundance fractions are independent; in the second, modeling an ideal hyperspectral scenario, we enforce only constraint (4), meaning that abundance fractions are dependent; in the third, modeling a real hyperspectral scenario, we generate abundance fractions according to (8), thus modeling abundance fraction dependence, signature variability, topography modulation, and system noise; in the fourth, we generate sources according to (8) and evaluate the mean magnitude of the cross-correlation factor between sources and their estimates by *fastica* algorithm, as function of the SNR, of signature variability, and of the number of sources. We have adopted the cross-correlation factor as a performance measure, because ICA and IFA unmixes abundance fractions up to a constant factor.

In all experiments, the scene dimension is of 30×30 pixels and endmember signatures were extracted from a hyperspectral subimage of Indian Pine Test Site in northwestern Indiana acquired by an AVIRIS instrument in June 1992.[1] Noisy channels and water absorption channels were removed (channels 1–4, 107–113, 150–166, and 221–224). Concerning atmospheric correction, this image has been processed to remove path radiance and the light scattered by the interaction between surface and the atmosphere. Table I presents the name of the substances extracted. Fig. 2 shows the angle between pairs of extracted signatures. The lowest and the highest angles are, approximately, 4° (between second and seventh signatures) and 48° (between fourth and ninth signatures), respectively. Endmember 9 has the highest angle with respect to the closest endmember. In the first three experiments, three endmembers were selected [see Fig. 2(a)]; in the fourth experiment, the number of endmembers vary from three to ten.

[1]Available at http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/.

TABLE I
Substances Extracted From the Dataset

endmember	substance
1	Corn-notill
2	Grass/Trees
3	Oats
4	Grass/Pasture
5	Hay-windrowed
6	Wheat
7	Bldg-Grass-Tree-drives
8	Stone-steel towers
9	House roof
10	Steel

In the first experiment, the abundance fractions are mutually independent following a Dirichlet distribution given by $p(\alpha_i) = \alpha_i^{\mu_i - 1}/\Gamma(\mu_i)$, where $\mu_i = E[\alpha_i]$ is the expected value of the ith endmember fraction, $E[\cdot]$ denotes the expectation operator, and $\Gamma(\cdot)$ denotes the Gamma function. Note that $\mu_i > 0$ and $\sum_i \mu_i = 1$. Mean values are set to $\mu_1 = 0.4$, $\mu_2 = 0.3$, and $\mu_3 = 0.3$. In the remaining experiments, abundance fractions follow a joint Dirichlet distribution given by

$$p(\alpha_1, \alpha_2, \ldots, \alpha_p) = \frac{\Gamma(\mu_1 + \mu_2 + \ldots + \mu_p)}{\Gamma(\mu_1)\Gamma(\mu_2)\ldots\Gamma(\mu_p)} \times \alpha_1^{\mu_1 - 1}\alpha_2^{\mu_2 - 1}\ldots\alpha_p^{\mu_p - 1}. \quad (18)$$

Table II presents the sample mean of the cross-correlation coefficients between the abundance fractions and their estimates and the sample cross-correlation coefficients between the endmember signatures and their estimates. These coefficients are based on 256 Monte Carlo runs. It shows that under this condition (independent abundance fractions) the first three algorithms work very well; they unmix the abundance fractions and extract the signatures of each endmember. The Bell and Sejnowski algorithm did not correctly unmix the endmembers.

In the second experiment, abundance fractions are dependent, following a Dirichlet distribution with parameters $\mu_1 = 0.4$, $\mu_2 = 0.3$, and $\mu_3 = 0.3$. Such distribution constrains abundance fractions to $0 \leq \alpha_i \leq 1$ and $\sum_i \alpha_i = 1$. It is clear that none of the algorithms correctly unmix the original dependent data. IFA and *fastica* algorithms only estimate two abundance fractions because they implement a preprocessing step to whiten the observed data and to reduce the dimension, where only two endmember were correctly found (note that the constraint $\sum_i \alpha_i = 1$ decreases by one the dimension of the observed data).

In the third experiment, abundance fractions are dependent following a Dirichlet distribution with parameters $\mu_1 = 0.4$, $\mu_2 = 0.3$, and $\mu_3 = 0.3$ for each endmember. Scale ψ_i, controlling signature variability, is uniformly distributed with in the interval [0.9, 1.1]; parameter γ is Beta distributed[2] with parameters $\beta_1 = 1, \beta_2 = 0.8$. Noise in (8) is zero-mean white Gaussian

[2]The Beta density is $p(\gamma) = (\Gamma(\beta_1 + \beta_2)/\Gamma(\beta_1)\Gamma(\beta_2))\gamma^{\beta_1 - 1}(\gamma - 1)^{\beta_2 - 1}$ for $\gamma \geq 0$.

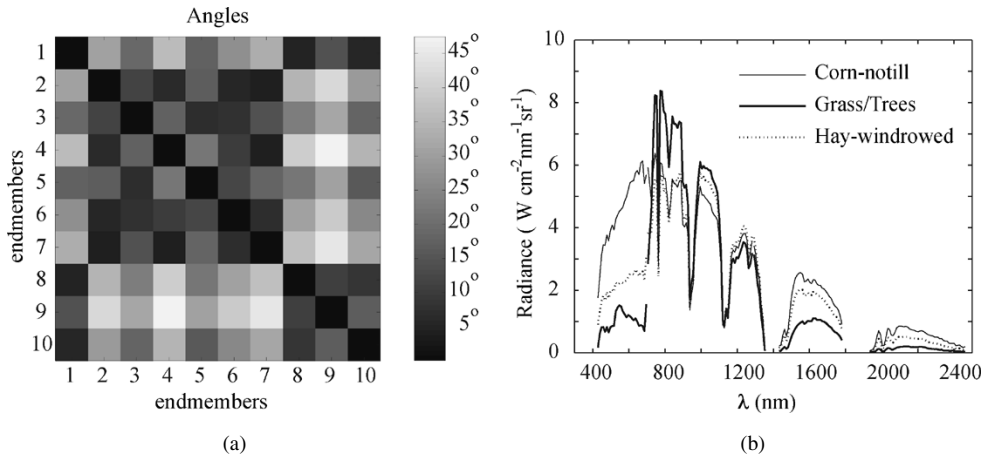


Fig. 2. (a) Angle between pairs of radiances spectra used to generate simulated scenes. (b) Radiance spectra of the first second and fifth substances used in experiments I, II, and III.

TABLE II
SAMPLE CROSS CORRELATION BETWEEN ABUNDANCE FRACTIONS AND THE CORRESPONDENT ESTIMATES ($\overline{\varrho_\alpha}$) AND SAMPLE CROSS CORRELATION BETWEEN ENDMEMBER SIGNATURES AND THE CORRESPONDENT ESTIMATES ($\overline{\varrho_m}$). RESULTS BASED ON 256 MONTE CARLO RUNS

		IFA		*fastica*		*jade*		Bell and Sejnowski	
	source	$\overline{\varrho_\alpha}$	$\overline{\varrho_m}$	$\overline{\varrho_\alpha}$	$\overline{\varrho_m}$	$\overline{\varrho_\alpha}$	$\overline{\varrho_m}$	$\overline{\varrho_\alpha}$	$\overline{\varrho_m}$
Experiment I	1	0.9767	0.9725	0.9978	0.9971	0.9994	0.9997	0.7817	0.7119
	2	0.9926	0.9777	0.9970	0.9921	0.9985	0.9952	0.8553	0.7071
	3	0.9797	0.9765	0.9895	0.9921	0.9911	0.9954	0.5738	0.5803
Experiment II	1	-	-	-	-	0.1978	0.1623	0.3332	0.8836
	2	0.8310	0.9259	0.7902	0.7123	0.4778	0.5794	0.8352	0.9812
	3	0.4913	0.9144	0.3281	0.7366	0.2482	0.5529	0.4490	0.9815
Experiment III	1	0.9755	0.4534	0.8780	0.6432	0.9304	0.6965	0.9727	0.7843
	2	0.9670	0.8619	0.9641	0.7631	0.9752	0.7668	0.9692	0.7830
	3	0.9065	0.8618	0.8454	0.7729	0.8981	0.7645	0.8778	0.6238

with variance σ^2 in each band such that the SNR defined as the ratio between the signature power and the noise power, i.e.,

$$\mathrm{SNR} = 10\log_{10}\frac{E\left[(\mathbf{Ms})^T(\mathbf{Ms})\right]}{E\left[\mathbf{w}^T\mathbf{w}\right]} \tag{19}$$

is set to 30 dB. As in the previous experiments, ICA and IFA algorithms do not correctly unmix the three abundance fractions. IFA, however, yields the best results being able to approximately unmix two abundance fractions.

In the fourth experiment, we compute the cross-correlation factor between abundance fractions and their estimates as function of the SNR, number of endmembers, and signature variability. Two abundance fraction distributions are considered: 1) symmetric Dirichlet distributions ($\mu_k = 1/p$, $k = 1, \ldots, p$); 2) asymmetric Dirichlet distributions [$\mu_k = 5/(p+8)$, $k = 1$, 2; $\mu_k = 1/(p+8)$, $k = 3, \ldots, p$]. Signature variability is controlled by the distribution of the scale random vector ψ_i. In this experiment, we assume that ψ_i is uniformly distributed within the interval $[\eta, 1]$, where $0 \leq \eta \leq 1$. Thus $\eta = 1$ means absence of variability, whereas $\eta = 0$ means maximum variability.

Although IFA was conceived to recover independent sources from linear mixtures immersed in noise, this algorithm was not considered in this experiment, because IFA computational complexity increases exponentially with the number of emdmembers. In this experiment, only the *fastica* algorithm was applied.

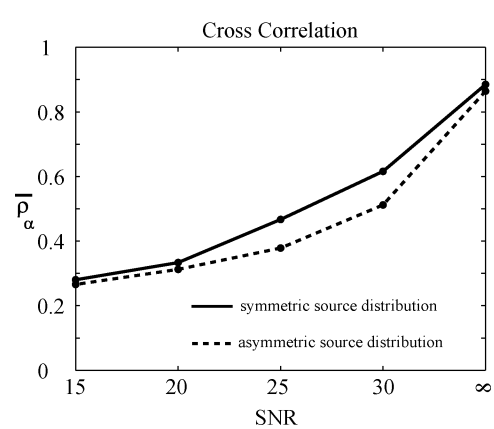


Fig. 3. Mean magnitude of the cross-correlation factors between abundance fractions and their estimates $\overline{\varrho_\alpha}$ as function of the SNR, for $p = 10$ and $\eta = 0.8$. Results are based on 256 Monte Carlo runs.

Figs. 3–5 present the sample mean cross-correlation factors, $\overline{\varrho_\alpha}$, between each abundance fraction and the correspondent estimate based on 256 Monte Carlo runs.

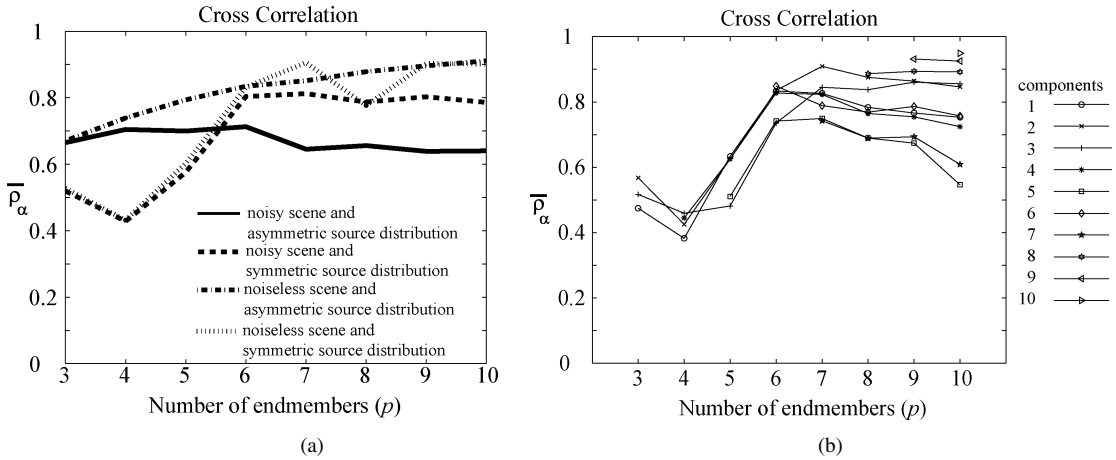


Fig. 4. Mean magnitude of the cross-correlation factors between abundance fractions and their estimates. Results are based on 256 Monte Carlo runs. (a) $\overline{\varrho_\alpha}$ as function of the number of endmembers p in the scene for $\eta = 0.6$. (b) $\overline{\varrho_\alpha}$ as function of p, for each individual source ($\eta = 0.6$, symmetric sources, SNR $= 30$ dB).

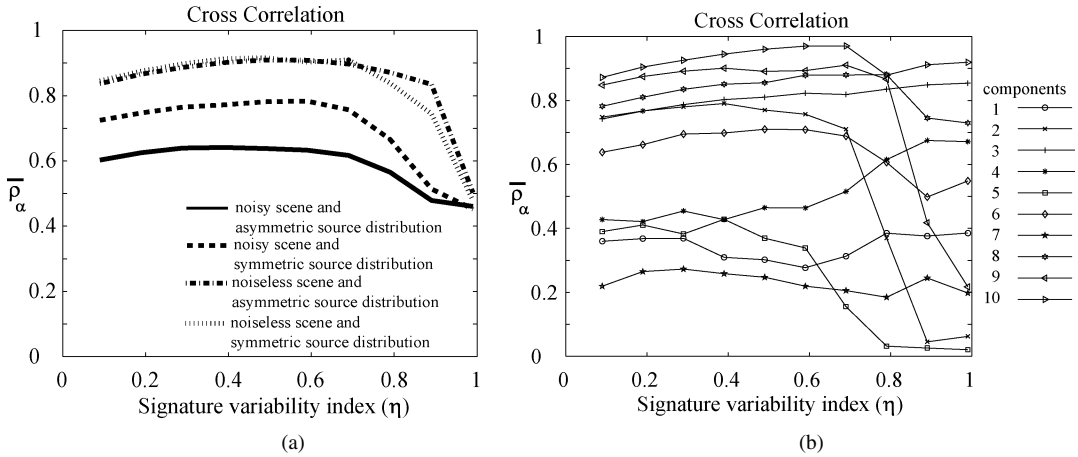


Fig. 5. Mean magnitude of the cross-correlation factors between abundance fractions and their estimates. Results are based on 256 Monte Carlo runs. (a) $\overline{\varrho_\alpha}$ as function of parameter η, for $p = 10$. (b) $\overline{\varrho_\alpha}$ as function of parameter η, for all sources ($p = 10$, asymmetric sources, SNR $= 30$ dB).

Fig. 3 shows an increasing $\overline{\varrho_\alpha}$ as function of the SNR (15, 20, 25, 30, ∞ dB), for $p = 10$ and $\eta = 0.8$. Asymmetry of abundance fraction distributions affects little the unmixing results. Fig. 4(a) shows $\overline{\varrho_\alpha}$ as function of the number of endmembers, for $p = 3, \ldots, 10$, SNR $= (30, \infty)$ dB, and $\eta = 0.6$. As the number of endmembers increases, the statistical dependence among sources decreases and we expect a better performance of ICA algorithms. This trend can be observed in Fig. 4(a), at least for high SNR. In Fig. 4(b), the sample mean of the magnitude of the cross-correlation factor is shown for each source separately, with SNR $= 30$ dB, $\eta = 0.6$, and symmetric source distribution. In this figure, we can see that there is always endmembers correctly unmixed and others incorrectly unmixed, regardless the number of endmembers. Fig. 5(a) shows $\overline{\varrho_\alpha}$ as a function of signature variability, for SNR $= (30, \infty)$ dB, and $p = 10$. Unmixing performance is quasi constant for $\eta \in [0.1, 0.6]$ and takes higher values in noiseless scenes. As η approaches to 1, meaning smaller signature variability and higher statistical dependence among sources, ICA performance decays as expected. In Fig. 5(b), we can see $\overline{\varrho_\alpha}$ for each sources separately, with SNR $= 30$ dB, $p = 10$, and asymmetric source distribution. Second, fifth, and seventh endmembers are clearly incorrectly unmixed. Note that with respect to Fig. 2(a), endmember 7, which jointly with endmember 2 form the closest pair, shows the worst unmixing result, whereas endmember 9, which has the highest angle with respect to the closest endmember, displays a good unmixing result.

The pattern of behavior exhibited in experiments I–IV was systematically replicated regardless the source statistics (constraint (4) is understood). Basically, we conclude that the accuracy of ICA applied to hyperspectral data tends to increase with the signature variability, the number of sources, and the SNR. There are, however, always endmembers incorrectly unmixed, regardless the unmixing scenario.

In the next sections, the estimation of the unmixing matrix is considered. This study is based on the minimization of the mutual information, which give some evidence on the reasoning underlying ICA and IFA limitations in unmixing hyperspectral data.

V. Minimization of Mutual Information

This section addresses the behavior of the mutual information in the neighborhood of the true unmixing matrix. The aim is

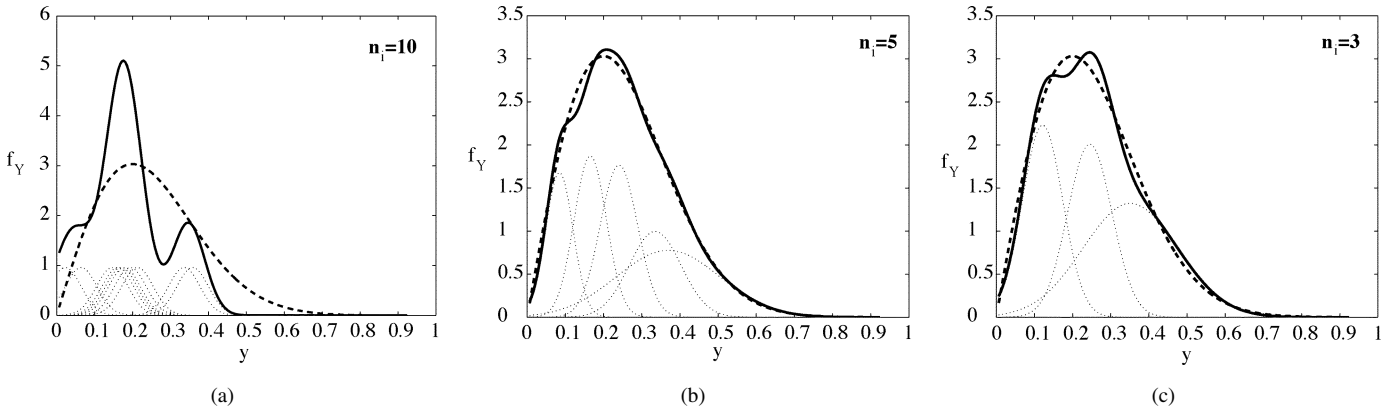


Fig. 6. Rayleigh probability density function (dashed line) fitted with a Gaussian mixture (solid line) using the algorithm [41]. (a) Initial parameters (ten Gaussian modes). (b) Solution with five Gaussian modes. (c) Solution constrained to three Gaussian modes.

to show that when sources are dependent, minimizing mutual information does not yield the true unmixing matrix.

A brief introduction to ICA and IFA was presented in Section III. Given an unknown linear mixture of the unobserved independent sources $\mathbf{r} = \mathbf{Ms}$, ICA and IFA look for a matrix $\mathbf{W}$ that maximizes a nongaussianity measure of the vector $\mathbf{y} = \mathbf{Wr}$. An objective function commonly used in ICA algorithms is the mutual information [28].

Assuming $|\mathbf{W}|$ constant, the minimization of the mutual information reduces to finding [see (14)]

$$\widehat{\mathbf{W}} = \arg\min_{\mathbf{W}} I(y_1, y_2, \ldots, y_p, \mathbf{W}) = \arg\min_{\mathbf{W}} \sum_i H(y_i) \tag{20}$$

where

$$H(y_i) \equiv -\int_{-\infty}^{+\infty} f_{Y_i}(u) \log f_{Y_i}(u) du \tag{21}$$

with f_{Y_i} being the probability density function of y_i. To compute (21), we need to estimate f_{Y_i}, for $i = 1, \ldots, p$. This estimate is based on fitting a mixture of Gaussians [63]

$$f_{Y_i}(y_i) = \sum_{q_i=1}^{n_i} \sigma_{q_i} \left(2\pi \left|\Sigma_{q_i}\right|\right)^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}\left(y_i - \mu_{q_i}\right)^T \Sigma_{q_i}^{-1} \left(y_i - \mu_{q_i}\right)\right] \tag{22}$$

where n_i is the number of Gaussians modes to fit the ith source and σ_{q_i}, μ_{q_i}, and Σ_{q_i} are the weight, the mean, and the covariance of the q_ith Gaussian mode, respectively. The number of Gaussians modes, and respective parameters (means, covariances and weights) are obtained via the MDL-based expectation-maximization algorithm (MDL-EM) [41]. The entropy (21) is computed via numerical integration.

Fig. 6 shows an example of a Rayleigh probability density function fitted with a Gaussian mixture using algorithm [41]. Fig. 6(a)–(c) presents the probability density functions obtained with, respectively, the initial parameters, the solution, and the solution constrained to three Gaussian modes.

Various authors [37], [55] have referred to the fact that the maxima of $I(y_1, y_2, \ldots, y_p, \mathbf{W})$ with respect to $\mathbf{W}$ are not very sensitive to the shape of f_{Y_i}. For example, Attias [37] uses only three Gaussian modes to fit whatever density shape. Herein, however, we use all modes given by the MDL-EM algorithm [41], as we are interested, not only in the unmixing matrix $\mathbf{W}$, but also in computing the mutual information $I(y_1, y_2, \ldots, y_p, \mathbf{W})$ as function of $\mathbf{W}$.

VI. Experimental Results

In the next five experiments, we study the behavior of the mutual information $I(y_1, y_2, \ldots, y_p, \mathbf{W})$, for $\mathbf{W}$ in the neighborhood of the true unmixing matrix $\mathbf{W} = \mathbf{M}^{-1}$. In all experiments, we assume constraint $|\mathbf{W}| = 1$. This setting does not constraint the unmixed results, as they are defined up to a constant.

Experiment I: The first experiment considers independent abundance fractions with uniform distribution to test our setup under canonical mixing conditions. This experiment assumes $p = 3$ (number of endmembers), $B = 3$ (number of bands), $\mathbf{M} = \mathbf{I}$, $\boldsymbol{\psi} = \mathbf{I}$, $\gamma = 1$, and $\mathbf{w} = \mathbf{0}$.

Fig. 7(a) shows the mutual information as function of ϕ_1 and ϕ_2 in a grayscale (ϕ_1 and ϕ_2 define a rotation in $\Re^3$. We term this angles azimuth and elevation, respectively). The minimum is global and occurs for $\phi_1 = 0$ and $\phi_2 = 0$, i.e., $\mathbf{W} = \mathbf{I}$.

As mentioned above, abundance fractions in hyperspectral data are not independent. In order to test ICA with this constraint, we generate abundance fractions according to the Dirichlet distribution [see (18)] parameterized with $\mu_1 = 1/3$, $\mu_2 = 1/3$, and $\mu_3 = 1/3$ (recall that μ_i is the expected value of the ith abundance fraction).

Experiment II: In this experiment, we set $p = 3$, $B = 3$, $\mathbf{M} = \mathbf{I}$, $\boldsymbol{\psi} = \mathbf{I}$, $\gamma = 1$, $\mathbf{w} = \mathbf{0}$.

In Fig. 7(b), we present the mutual information as function of angles ϕ_1 and ϕ_2. No ICA algorithm could ever correctly unmix the original dependent data, since ($\hat{\phi}_1 \simeq -\pi/4$, $\hat{\phi}_2 \simeq -\pi/5$), far from the true unmixing matrix, i.e., ($\phi_1 = 0$, $\phi_2 = 0$).

Experiment III: In this experiment, the abundance fractions are dependent and Dirichlet distributed ($\mu_1 = 1/3$, $\mu_2 = 1/3$,

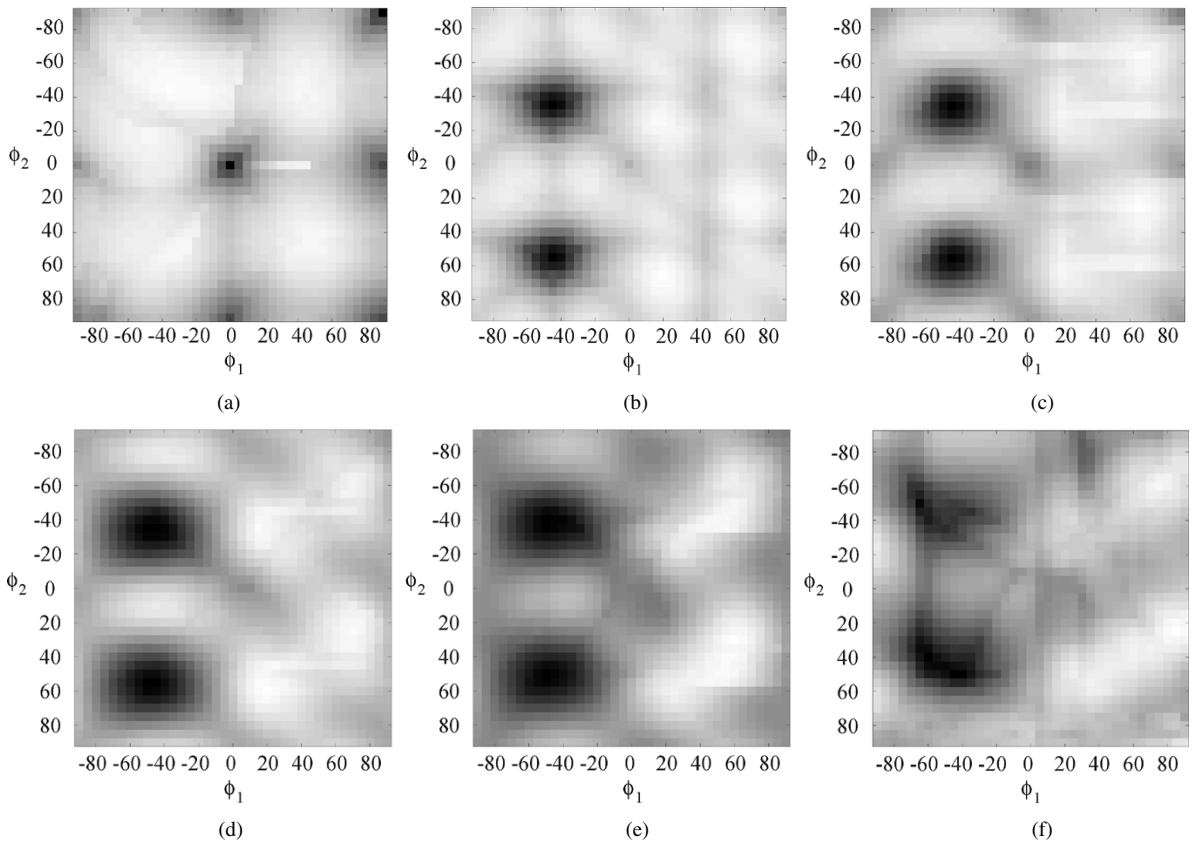


Fig. 7. Mutual information as function of parameters ϕ_1 (azimuth angle) and ϕ_2 (elevation angle). (a) Three independent sources. (b) Dependent sources. (c) Dependent sources with illumination perturbations and signature variability. (d) Noisy scene (SNR $=$ 25 dB). (e). Noisy scene (SNR $=$ 20 dB). (f) Noisy scene (SNR $=$ 15 dB).

$\mu_3 = 1/3$). The remaining parameters are $p = 3$, $B = 3$, $\mathbf{M} = \mathbf{I}$, $\mathbf{w} = \mathbf{0}$, ψ_i uniformly distributed in the interval [0.9, 1.1] and γ Beta distributed with parameters $\beta_1 = 1$ and $\beta_2 = 0.8$.

Fig. 7(c) presents the mutual information as function of angles ϕ_1 and ϕ_2. Although there is a local minimum at $\phi_1 = 0$ and $\phi_2 = 0$, the absolute minimum occurs at $\hat{\phi}_1 \simeq -\pi/4$ and $\hat{\phi}_2 \simeq -\pi/5$.

Experiment IV: This experiment is similar to experiment III, but now we add zero-mean white Gaussian noise. Herein we considered three scenarios: SNR $= 15, 20, 25$ dB.

Fig. 7(d)–(f) shows the obtained mutual information for SNR $= 15, 20, 25$ dB, respectively. The surface exhibits a random pattern as the noise level increases. The global minimum occurs at $\hat{\phi}_1 \simeq -\pi/5$, $\hat{\phi}_2 \simeq -\pi/4$ for SNR $= 25$ dB [see Fig. 7(d)], $\hat{\phi}_1 \simeq 10\pi/36$, $\hat{\phi}_2 \simeq -10\pi/36$ for SNR $= 20$ dB [see Fig. 7(e)], and $\hat{\phi}_1 \simeq 10\pi/36$, $\hat{\phi}_2 \simeq -2\pi/9$ for SNR $= 15$ dB [see Fig. 7(f)]. All these global minima are far from ($\phi_1 = 0$, $\phi_2 = 0$). Notice, however, the presence of a local minimum at $\hat{\phi}_2 \simeq 0$ and $\hat{\phi}_2 \simeq 0$.

Experiment V: The last simulation considers ten endmembers with asymmetric Dirichlet distributions ($\mu_1 = 5/18$, $\mu_2 = 5/18$, $\mu_k = 1/18$; $k = 3, \ldots, 10$). The remaining parameters are $\mathbf{M} = \mathbf{I}$, $\mathbf{w} = \mathbf{0}$, ψ_i uniformly distributed in the interval [0.9, 1.1], and γ Beta distributed with parameters $\beta_1 = 1$ and $\beta_2 = 0.8$. The mixture is immersed in zero-mean white Gaussian noise, corresponding to SNR $= 20$ dB.

Fig. 8(a) shows the mutual information (up to a constant) as function of the rotation matrix

$$\mathbf{W} = \begin{bmatrix} \cos\phi_{12} & \sin\phi_{12} & & & \mathbf{0} \\ -\sin\phi_{12} & \cos\phi_{12} & & & \\ & & 1 & & \\ & & & \ddots & \\ \mathbf{0} & & & & 1 \end{bmatrix}.$$

Notice that ϕ_{12} has the meaning of an angle between the first and second components. Note that the minimum occurs at $\hat{\phi}_{12} \simeq 0$, which means that first and second component could be correctly unmixed; on the right, the mutual information is presented as function of ϕ_{24}. We can observe a local minima at the origin but the global minimum is at $\hat{\phi}_{24} \simeq -\pi/4$. We conclude that, under these conditions, although might exist local minima corresponding to the true unmixing matrix, the global minimum of the mutual information might be very far from true one.

The pattern of behavior described in experiments II, III, IV, and V was systematically observed in a series of experiments with different abundance fraction distributions. Basically, we concluded that in linear hyperspectral data unmixing, the unmixing matrix $\mathbf{W}$ minimizing the mutual information might be very far from the true one, at least for a few number of endmembers. This is basically in agreement with conclusions drawn in Section IV.

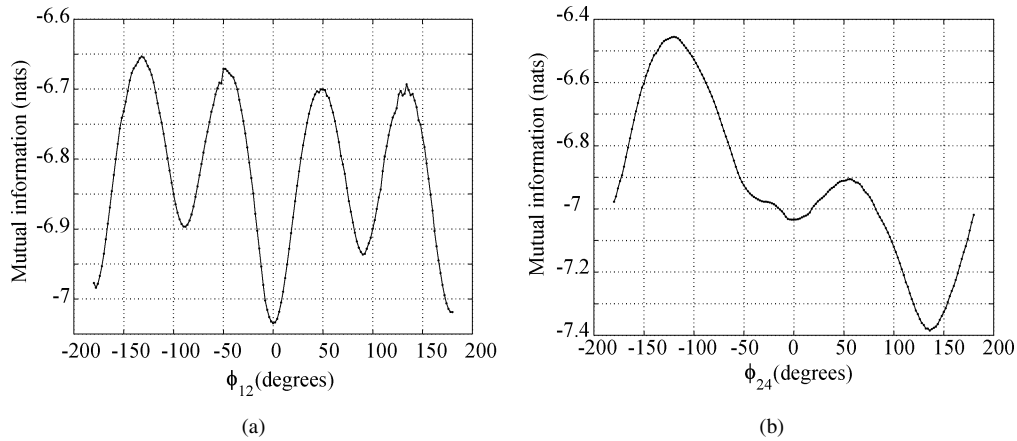


Fig. 8. Ten dependent components are mixed with noise added. (a) Mutual information (up to a constant) as function of rotation angle ϕ_{12} angle between first and second components. (b) Mutual information (up to a constant) as function of rotation angle ϕ_{24} angle between second and fourth components.

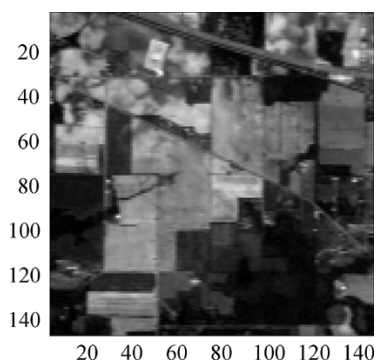


Fig. 9. Indiana Pines hyperspectral dataset, band 29 ($\lambda = 667.3$ nm).

VII. Fastica Algorithm Applied to Real Data

In this section, we consider a subimage of the hyperspectral dataset from Indian Pine Test Site in Northwestern Indiana acquired by an AVIRIS in June 1992.[3] The dataset is composed of 220 spectral channels with 10 nm bandwidth acquired in the 0.4–2.5-μm region. It contains 145 $\times$ 145 pixels (21 025 pixels) with a ground pixel resolution of 17 m [64]. This region contains a mixture of agriculture and forestry. However, due to the early season date of data collection, the cultivated land appears to have very little canopy cover yet. There is a major dual lane highway (U.S. 52 and U.S. 231), a rail line crossing near the top, a major secondary road (Jackson Highway) near the middle, several other county roads, and houses (Fig. 9 shows band 29 of the dataset). The ground truth of the region can be found in [65]. It classifies the ground covered area to 16 classes and ignores many small variations within fields that can be seen in the image data (see Fig. 9). With respect to atmospheric correction, this dataset has been processed to remove path radiance and the light scattered by interaction between surface and the atmosphere. Notice, however, that the corrected image is still in radiance units.

A PCA preprocessing step was implemented to whiten the observed data and to reduce dimension: every pixel vector $\mathbf{r}$ is transformed into a vector of dimension 16 with zero-mean and identity covariance matrix: $\mathbf{x} = \mathbf{Q}(\mathbf{r} - E[\mathbf{r}])$, such that $E[\mathbf{x}] = 0$ and $E[\mathbf{x}\mathbf{x}^T] = \mathbf{I}$, where $\mathbf{I}$ the identity matrix.

The *Fastica* algorithm was applied to the processed dataset. Fig. 10 presents the first 16 components. First, second, fourth, tenth, and eleventh components correspond to houses or man made materials that exist in a few locations in the scene [see Fig. 10(b), (d), (j), and (k)]. We can see that maximum contrast of sources occurs in house locations, stone-steel towers, Jackson highway, and rail line. Third and fifth components are a mixture of grass with pasture and trees, respectively [see Fig. 10(c) and (e)]. The sixth component [see Fig. 10(f)] represents a mixture of grass, soybeans, and corn. Apparently the seventh component is hay-windrowed [see Fig. 10(g)]. Ninth component represents vegetation mowed [see Fig. 10(i)]. Components presented in Fig. 10(a), (h), (l), and (m) do not represent any class of the available ground truth [65]. Fig. 10(n)–(p) are mainly noise. We conclude that six sources are unmixed and ten are a mixture of several materials present in the scene or are manly noise. This is basically in accordance with our findings based on simulated data.

A pertinent question is what sources are correctly unmixed. To address this question we computed the entropy of each component, normalized to unit variance, following the procedure described in Section V. The underlying idea is that a mixed source tends to exhibit higher entropy, the maximum value being $H_{\max} = 1.419$ achieved by a Gaussian source. Table III shows the entropy for each component shown in Fig. 10(a)–(p). By inspection of Table III, we can identify two subsets: the first, with smaller values of entropy, corresponds to components correctly unmixed [see Fig. 10(b)–(g)]; the second, with larger values of entropy, corresponds to components with mixed sources [see Fig. 10(a) and (h)–(p)]. The larger values, near $H_{\max}$, correspond to the last three component which are mostly noise. Observing Table III, we conclude that it might be possible to design a threshold to define which components are correctly unmixed and which ones are incorrectly unmixed. This question is to be addressed in future work.

[3]Available at http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/.

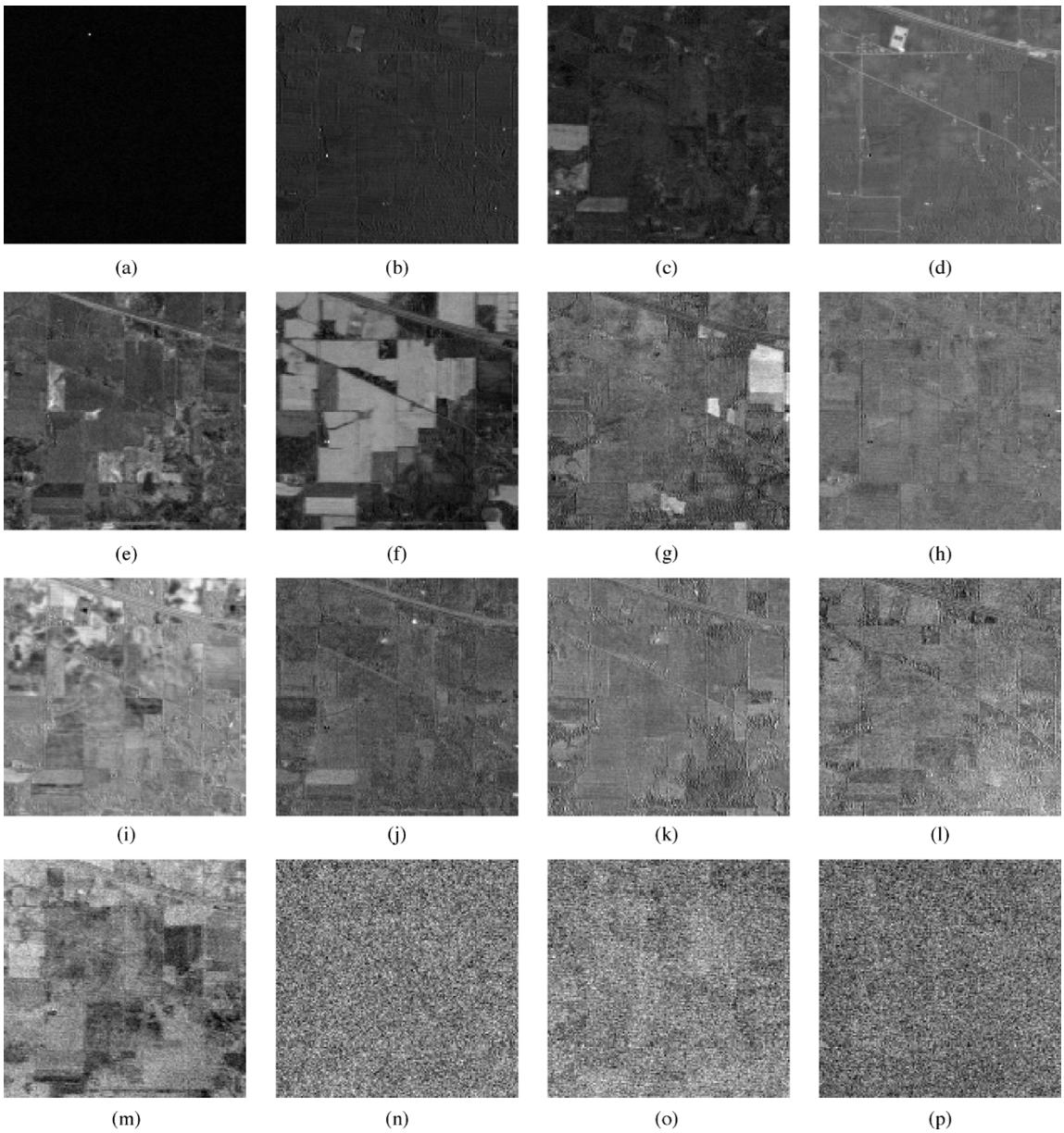


Fig. 10. Independent components extracted from a subimage of the Indiana Pine test site, with *fastica* algorithm. (a) Not classified. (b) Houses. (c) Grass/pasture. (d) Stone steel towers, rail line, and highways. (e) Grass/trees. (f) Grass, soybeans, and corn. (g) Hay-windrowed. (h) Not classified. (i) Vegetation mowed. (j) Houses. (k) Houses. (l) Not classified. (m) Not classified. (n) Noise. (o) Noise. (p) Noise.

VIII. Concluding Remarks

Blind hyperspectral linear unmixing aims at estimating the number of reference substances, also called endmembers, their spectral signatures, and their fractions at each pixel, called abundance fractions, using only the observed data (mixed pixels). Geometric approaches have been used whenever pure pixels are present in data [66]–[70]. In most cases, however, pure pixels can not be found in data. In such cases, unmixing procedures become a difficult task.

In the recent past, ICA has been proposed as a tool to unmix hyperspectral data [24], [29]–[36]. ICA consists in finding a linear decomposition of data into statistically independent components. IFA extends ICA concepts when noise is present. Crucial assumptions of ICA and IFA are that each pixel is a linear mixture of endmember signatures weighted by the correspondent abundance fractions and these abundances are independent. Concerning hyperspectral data, the first assumption is valid whenever the multiple scattering among the distinct endmembers is negligible and the surface is partitioned according to the fractional abundances. The second assumption, however, is not valid due to physical constraints on the acquisition process.

This paper addresses the impact of the abundance fraction (sources) dependence on unmixing hyperspectral data with ICA/IFA. The study considers simulated and real hyperspectral data. Hyperspectral observations are described by a generative model which includes degradation mechanism such as signature variability, abundance constraints, topography modulation, and system noise.

TABLE III
ENTROPY OF EACH COMPONENT

Component	Entropy	Unmixed
a)	1.4020	
b)	1.3290	⋆
c)	1.1834	⋆
d)	1.2002	⋆
e)	1.3399	⋆
f)	1.1300	⋆
g)	1.2811	⋆
h)	1.4031	
i)	1.3781	
j)	1.3945	
k)	1.3842	
l)	1.4069	
m)	1.4172	
n)	1.4188	
o)	1.4187	
p)	1.4189	

IFA and three well-known ICA algorithms were tested on simulated data. Our main findings were the following.

1) ICA/IFA performance increases with the SNR.
2) ICA/IFA performance tends to increase with the signature variability and/or with the number of endmembers. The underlying reason is that by increasing the signature variability and/or the number of endmembers the statistical dependence among endmembers is attenuated.
3) There are always endmembers incorrectly unmixed, regardless the unmixing scenario.

In order to assess the impact of hyperspectral abundance fraction dependence on the ICA/IFA algorithms, we studied the behavior of the mutual information of the unmixed sources in the neighborhood of the true unmixed data. We conclude that in hyperspectral linear unmixing, the unmixing matrix minimizing the mutual information might be vary far from the true one, at least for a few number of endmembers.

Finally, ICA and IFA algorithms were tested in a subimage of the hyperspectral dataset from Indian Pine test site in Northwestern Indiana acquired by an AVIRIS in June 1992. According to the available ground truth of the region, we conclude that six sources are correctly unmixed and ten are uncorrectly unmixed. This is in line with the conclusion drawn from simulated data. A method based on the source entropy to sort the output of ICA or IFA algorithms according to the likelihood of being correctly separated was proposed.

As future work, we intend to extend ICA/IFA concepts under a Bayesian framework [24] to linear mixtures of dependent sources, which is the case of hyperspectral data.

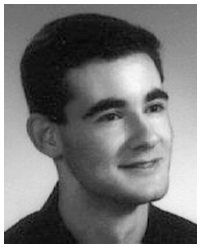
ACKNOWLEDGMENT

The authors acknowledge M. A. T. Figueiredo for providing the MDL-EM code [41]. The authors would like to thank to the reviewers for the many valuable comments and suggestions.

REFERENCES

[1] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002.
[2] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
[3] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging aplications," *IEEE Signal Process.Mag.*, vol. 19, no. 1, pp. 29–43, 2002.
[4] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, 2002.
[5] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, and A. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, 2002.
[6] M. O. Smith, P. E. Johnson, and J. B. Adams, "Quantitative determination of mineral types and abundances from reflectance spectra using principal component analysis," in *Proc. 15th Lunar Planetary Sci. Conf., Part 2, Geophysical Research*, vol. 90, 1985, pp. C797–C804.
[7] A. R. Gillespie, M. O. Smith, J. B. Adams, S. C. Willis, A. F. Fisher, and D. E. Sabol, "Interpretation of residual images: Spectral mixture analysis of AVIRIS images, Owens Valley, California," in *Proc. 2nd AVIRIS Workshop*, vol. 90-54, R. O. Green, Ed., 1990, pp. 243–270.
[8] T. Lillesand and R. Kiefer, *Remote Sensing and Image Interpretation*, 3rd ed. New York: Wiley, 1994.
[9] G. Vane, R. Green, T. Chrien, H. Enmark, E. Hansen, and W. Porter, "The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 44, pp. 127–143, 1993.
[10] J. B. Adams and M. O. Smith, "A new analysis of rock and soil types at the viking lander 1 site," *J. Geophys. Res.*, vol. 91, no. B8, pp. 8098–8112, 1986.
[11] S. Liangrocapart and M. Petrou, "Mixed pixels classification," *Proc. SPIE*, vol. 3500, pp. 72–83, 1998.
[12] R. B. Singer and T. B. McCord, "Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," in *Proc. 10th Lunar Planetary Sci. Conf.*, 1979, pp. 1835–1848.
[13] R. Singer, "Near-infrared spectral reflectance of mineral mixtures: Systematic combinations of pyroxenes, olivine, and iron oxides," *J. Geophys. Res.*, vol. 86, pp. 7967–7982, 1981.
[14] B. Nash and J. Conel, "Spectral reflectance systematics for mixtures of powdered hypersthene, labradoride, and ilmenite," *J. Geophys. Res.*, vol. 79, pp. 1615–1621, 1974.
[15] B. Hapke, "Bidirection reflectance spectroscopy. I. theory," *J. Geophys. Res.*, vol. 86, pp. 3039–3054, 1981.
[16] R. N. Clark and T. L. Roush, "Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications," *J. Geophys. Res.*, vol. 89, no. B7, pp. 6329–6340, 1984.
[17] C. C. Borel and S. A. Gerstl, "Nonlinear spectral mixing models for vegetative and soils surface," *Remote Sens. Environ.*, vol. 47, no. 2, pp. 403–416, 1994.
[18] J. J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1045–1046, Jul. 1996.
[19] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York: Kluwer, 2003.
[20] A. S. Mazer and M. Martin *et al.*, "Image processing software for imaging spectrometry data analysis," *Remote Sens. Environ.*, vol. 24, no. 1, pp. 201–210, 1988.
[21] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembres using the spectral angle mapper (SAM) algorithm," in *Summaries of the 3rd Annu. JPL Airborne Geosci. Workshop*, vol. 1, R. O. Green, Ed., 1992, Pub. 92–14, pp. 147–149.
[22] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.
[23] C. Chang, X. Zhao, M. L. G. Althouse, and J. J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 898–912, May 1998.
[24] L. Parra, K.-R. Mueller, C. Spence, A. Ziehe, and P. Sajda, "Unmixing hyperspectral data," *Adv. Neural Inf. Process. Systems*, vol. 12, pp. 942–948, 2000.
[25] A. Ifarraguerri and C. Chang, "Unsupervised hyperspectral image analysis with projection pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 6, pp. 2529–2538, Nov. 2000.

[26] L. O. Jimenez and D. A. Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2653–2664, Nov. 1999.
[27] P. Common, C. Jutten, and J. Herault, "Blind separation of sources, part II: Problem statement," *Signal Process.*, vol. 24, pp. 11–20, 1991.
[28] P. Common, "Independent component analysis: A new concept," *Signal Process.*, vol. 36, pp. 287–314, 1994.
[29] J. Bayliss, J. A. Gualtieri, and R. Cromp, "Analysing hyperspectral data with independent component analysis," *Proc. SPIE*, vol. 3240, pp. 133–143, 1997.
[30] C. Chen and X. Zhang, "Independent component analysis for remote sensing study," *Proc. SPIE*, vol. 3871, pp. 150–158, 1999.
[31] T. M. Tu, "Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent component analysis approach," *Opt. Eng.*, vol. 39, no. 4, pp. 897–906, 2000.
[32] S.-S. Chiang, C.-I Chang, and I. W. Ginsberg, "Unsupervised hyperspectral image analysis using independent component analysis," in *Proc. IGARSS*, vol. 7, 2000, pp. 3136–3138.
[33] N. Keshava, J. Kerekes, D. Manolakis, and G. Shaw, "An algorithm taxonomy for hyperspectral unmixing," *Proc. SPIE*, vol. 4049, pp. 42–63, 2000.
[34] M. Lennon, M. Mouchot, G. Mercier, and L. Hubert-Moy, "Spectral unmixing of hyperspectral images with the independent component analysis and wavelet packets," in *Proc. IGARSS*, vol. 6, 2001, pp. 2896–2898.
[35] N. Kosaka and Y. Kosugi, "ICA aided linear spectral mixture analysis of agricultural remote sensing images," *Proc. 4th Int. Symp. Independent Component Analysis and Blind Signal Separation*, pp. 221–226, 2003.
[36] V. Botchko, E. Berina, Z. Korotkaya, J. Parkkinen, and T. Jaaskelainen, "Independent component analysis in spectral images," in *Proc. 4th Int. Symp. Independent Component Analysis and Blind Signal Separation*, 2003, pp. 203–207.
[37] H. Attias, "Independent factor analysis," *Neural Comput.*, vol. 11, no. 4, pp. 803–851, 1999.
[38] G. Shaw and H. Burke, "Spectral imaging for remote sensing," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 3–28, 2003.
[39] J. S. Tyo, J. Robertson, J. Wollenbecker, and R. C. Olsen, "Statistics of target spectra in HSI scenes," in *Proc. SPIE Conf. Imaging Spectrometry VI*, 2000.
[40] G. Healey and D. Slater, "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2706–2717, Nov. 1999.
[41] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 44, no. 3, pp. 381–396, Mar. 2002.
[42] D. Tanre, M. Herman, P. Deschamps, and A. de Leffe, "Atmospheric modeling for space measurements of ground reflectances, including bidirectional properties," *Appl. Opt.*, vol. 18, pp. 3587–3594, 1979.
[43] E. Vermote, D. Tanr, J. Deuz, M. Herman, and J. Morcette, "Second simulation of the satellite signal in the solar spectrum 6S: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 675–686, May 1997.
[44] A. K. Jain, *Fundamentals of Digital Image Processing*, E. Cliffs, Ed. Upper Saddle River, NJ: Prentice-Hall, 1989.
[45] K. Liou, *An Introduction to Atmospheric Radiation*, 2nd ed. Orlando, FL: Academic, 2002.
[46] D. Roberts, Y. Yamaguchi, and R. Lyon, "Calibration of airborne imaging spectrometer data to percent reflectance using field spectral measurements," in *Proc. 19th Int. Symp. Remote Sensing of Environment 2*, Ann Arbor, MI, 1985, pp. 679–688.
[47] A. Berk, L. Bernstein, G. Anderson, P. Acharya, D. Robertson, J. Chetwynd, and S. Adler-Golden, "MODTRAN cloud and multiple scattering upgrades with application to AVIRIS," *Remote Sens. Environ.*, vol. 65, pp. 367–375, 1998.
[48] B. Hapke, *Theory of Reflectance and Emmittance Spectroscopy*. Cambridge, U.K.: Cambridge Univ. Press, 1993.
[49] H.-H. Wu and R. A. Schowengerdt, "Improved estimation of fraction images using partial image restoration," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, no. 4, pp. 771–778, Jul. 1993.
[50] C. Bateson, G. Asner, and C. Wessman, "Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1083–1094, Mar. 2000.
[51] F. Kruse, "Spectral identification of image endmembers determined from AVIRIS data," in *Summaries of the VII JPL Airborne Earth Science Workshop*, 1998.
[52] J. Boardman and F. Kruse, "Automated spectral analysis: A geological example using AVIRIS data, northern grapevine mountains, Nevada," in *Proc. 10th Thematic Conf., Geologic Remote Sensing*, 1994.
[53] A. Hyvärinen, "Survey on independent component analysis," *Neural Comput. Surv.*, vol. 2, pp. 94–128, 1999.
[54] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
[55] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, no. 4–5, pp. 411–430, 2000.
[56] J. Cardoso, "Infomax and maximum likelihood of source separation," *IEEE Signal Process. Lett.*, vol. 4, no. 4, pp. 112–114, Apr. 1997.
[57] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 10, pp. 215–234, 1995.
[58] T.-W. Lee, M. Girolami, A. J. Bell, and T. J. Sejnowski, "An unifying information-theoretic framework for independent component analysis," *Comput. Math. Appl.*, vol. 31, no. 11, pp. 1–21, Mar. 2000.
[59] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[60] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood estimation from incomplete data via EM algorithm," *J. R. Stat. Soc.*, vol. 39, no. B, pp. 1–38, 1977.
[61] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.
[62] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.
[63] G. McLachlan and D. Peel, *Finite Mixture Models*. New York: Wiley, 2000.
[64] P.-F. Hsieh and D. Landgrebe, "Classification of high dimensional data," School of Electr. Comput. Eng., Purdue Univ., West Lafayette, IN, Ph.D. thesis and Tech. Rep. TR-ECE 98-4, 1998.
[65] D. Landgrebe, "Multispectral data analysis: A signal theory perspective," School of Electr. Comput. Eng., Purdue Univ., West Lafayette, IN, 1998.
[66] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 1, pp. 99–109, Jan. 1994.
[67] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Summaries 4th Annu. JPL Airborne Geoscience Workshop, AVIRIS Workshop*, vol. 1, 1993, JPL Pub. 93-26, pp. 11–14.
[68] J. Theiler, D. Lavenier, N. Harvey, S. Perkins, and J. Szymanski, "Using blocks of skewers for faster computation of pixel purity index," in *Proc. SPIE Int. Conf. Optical Science and Technology*, 2000.
[69] M. E. Winter, "N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. SPIE Conf. Imaging Spectrometry V*, 1999, pp. 266–275.
[70] J. M. P. Nascimento and J. M. B. Dias, "Fast unsupervised extraction of endmembers spectra from hyperspectral data," in *Proc. 10th Int. Symp. Remote Sensing. Remote Sensing for Environmental Monitoring, GIS Applications, and Geology III*, vol. 5239, 2003, pp. 314–321.



José M. P. Nascimento (S'03) received the B. S. and E.E. degree from Instituto Superior de Engenharia de Lisboa, Politechnic Institute of Lisbon, Lisbon, Portugal, and the M.Sc. degree in electrical and computer engineer from Instituto Superior Técnico (IST), Technical University of Lisbon, in 1993, 1995, and 2000, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at Instituto Superior Técnico, Technical University of Lisbon.

He is currently a Professor in the Department of Electronics, Telecommunications and Computer's Engineering, Instituto Superior de Engenharia de Lisboa. He is also Researcher with the Institute of Telecommunications. His research interests include remote sensing, signal and image processing, pattern recognition, and communications.

José M. Bioucas Dias (S'87–M'95) received the E.E., M.Sc., and Ph.D. degrees in electrical and computer engineering from the the Instituto Superior Técnico (IST), Technical University of Lisbon, Lisbon, Portugal, in 1985, 1991, and 1995, respectively.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, IST. He is also a Researcher with the Communication Theory and Pattern Recognition Group, IST. His research interests include remote sensing, signal and image processing, pattern recognition, and communications.