Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 27673, 10 pages
doi:10.1155/2007/27673

Research Article

Statistical Analysis of Hyper-Spectral Data: A Non-Gaussian Approach

N. Acito, G. Corsini, and M. Diani

Dipartimento di Ingegneria dell'Informazione, Università di Pisa, Via Caruso, 14-56122 Pisa, Italy

Received 5 June 2006; Revised 9 October 2006; Accepted 24 October 2006

Recommended by Ati Baskurt

We investigate the statistical modeling of hyper-spectral data. The accurate modeling of experimental data is critical in target detection and classification applications. In fact, having a statistical model that is capable of properly describing data variability leads to the derivation of the best decision strategies together with a reliable assessment of algorithm performance. Most existing classification and target detection algorithms are based on the multivariate Gaussian model which, in many cases, deviates from the true statistical behavior of hyper-spectral data. This motivated us to investigate the capability of non-Gaussian models to represent data variability in each background class. In particular, we refer to models based on elliptically contoured (EC) distributions. We consider multivariate EC-t distribution and two distinct mixture models based on EC distributions. We describe the methodology adopted for the statistical analysis and we propose a technique to automatically estimate the unknown parameters of statistical models. Finally, we discuss the results obtained by analyzing data gathered by the multispectral infrared and visible imaging spectrometer (MIVIS) sensor.

1. INTRODUCTION

The main characteristic of hyper-spectral sensors is their ability to acquire a spectral signature of the monitored area, thus enabling a spectroscopic analysis to be carried out of large regions of terrain.

The large amount of data collected by hyper-spectral sensors can lead to an improvement in the performance of detection/classification algorithms. Within this framework, it is important to note that the spectral reflectance of the observed object is not a deterministic quantity, but is characterized by an inherent variability determined by changes in the surface of the object. In remote sensing applications, spectrum variability is emphasized by several factors, such as atmospheric conditions, sensor noise, and acquisition geometry. One possible way to properly address the spectral variability is to make use of suitable statistical models. Although the statistical approach has benefits both in classification and detection applications, in this paper, we focus on target detection problems. By using a statistical approach, the generic hyper-spectral pixel $\mathbf{x}$ is modeled as an L-dimensional random vector (where L is the number of sensor spectral channels) that is a certain multivariate probability density function (p.d.f.). Target detection reduces to a binary classification problem, where by observing $\mathbf{x}$ one must decide if it belongs to the background class (H_0 hypothesis) or to the target class (H_1 hypothesis) by using an appropriate decision rule. The availability of a multivariate model that properly accounts for the statistical behavior of hyper-spectral data leads to

(1) the derivation of the "best" decision rule,
(2) the analytical derivation of the detector's performance.

The derivation of the algorithms' performance is a critical issue in designing automatic target detection systems and is a fundamental tool for defining the criteria for a correct choice of algorithm parameters.

Most of the detection algorithms proposed in the literature (see [1, 2]) and widely used in current applications have been derived under the multivariate Gaussian assumption. The popularity of the Gaussian model is due to its mathematical tractability. In fact, it simplifies the derivation of decision rules and the evaluation of the detectors' performance. Unfortunately, the multivariate Gaussian model is not sufficiently adequate to represent the statistical behavior of each background class in real hyperspectral images. It has been proved (see [3–5]) that the Gaussian model fails in its representation of the distribution tails. In particular, current

distributions have longer tails than the Gaussian p.d.f. This is a critical issue in detection applications. In fact, the distribution tails determine the number of false alarms. Most detection applications require the algorithm test threshold to be set in order to control the probability of false alarms (P_{FA}). Generally, parameters are set on the basis of the P_{FA} predicted by the model adopted to describe the data. Since the Gaussian model underestimates the distribution tails, the parameter tuning based on such a model could be misleading in that the actual number of false alarms might exceed the desired number.

To overcome the limits of the Gaussian model in describing the statistical behavior of background classes in real hyper-spectral images, in recent years multivariate non-Gaussian models have been investigated. A very promising class of models is the family of the elliptically contoured distributions (ECD) [4, 5]. It has some statistical properties that simplify the analysis of multidimensional data and includes several distributions that have longer tails than the Gaussian one.

In this paper, we focus on three distinct probability models based on the ECD theory. ECD models were proposed in two recently published papers (see [4, 5]), where the authors applied the multivariate EC-t distribution, a particular class of ECD family, to model data gathered by the HYDICE sensor. They showed that there is a good agreement between the probability distribution estimated over HYDICE data and the theoretical one derived by assuming the EC-t model. In particular, by resorting to the properties of the EC distributions the authors compared the probability of exceedance (*PoE*) of the square of the Mahalanobis distance obtained over real data with the theoretical *PoE*. For the EC-t distribution, the *PoE* of the square of the Mahalanobis distance depends on a scalar value v. In [4, 5] the authors graphically showed that by varying v the curve corresponding to the theoretical *PoE* tends to the empirical one; they did not address the important problem of automatically estimating the value of v from the available data.

In this study, first we apply the hyper-spectral data analysis proposed in [5] and based on the EC-t distribution in order to model data collected by the MIVIS (multispectral infrared and visible imaging spectrometer) sensor. We extend the analysis procedure further by defining two different methods to estimate the parameter v. One of our proposed techniques estimates v directly from the available data. This makes the method very interesting for practical applications where the background parameters included in the algorithm decision rules must be estimated directly from the analyzed image.

Furthermore, we also analyse experimental data variability by using mixture models so as to take into account the spatial or spectral nonhomogeneity in the background classes considered. In particular, we investigate the effectiveness of mixture models whose p.d.f. is obtained as a linear combination of EC p.d.f.'s (see [6]). We consider two distinct mixture models, and we define a technique to automatically estimate their unknown parameters.

The paper is organised as follows: first, we introduce the ECD and we describe in detail the three models considered in our analysis; then, for each model we illustrate the technique used to estimate the unknown parameters. Finally, we present and discuss the results obtained by analyzing two distinct background classes in an MIVIS image.

2. NON-GAUSSIAN MODELS

2.1. Elliptically contoured distribution

The L-dimensional random vector $\mathbf{X} = [X_1, X_2, \ldots, X_L]$ is EC distributed, or equivalently it is a spherically invariant random vector (SIRV) if its p.d.f. can be expressed as

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{L/2}|\mathbf{C}|^{1/2}} h_L(d), \tag{1}$$

where we denote with d the generic realization of the random variable D corresponding to the square of the Mahalanobis distance:

$$D = (\mathbf{X} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \tag{2}$$

and $\boldsymbol{\mu}$ and $\mathbf{C}$ are the mean vector and the covariance matrix, respectively.

ECDs have some important statistical properties as follows:

(1) the isolevel curves in (1) are elliptical;
(2) each vector obtained from the element of an SIRV is also EC distributed;
(3) the p.d.f. of each set of variables $\{X_i : i \in I,\ I \in [1, \ldots, L]\}$ conditioned to $\{X_j : j \in J,\ J \cup I = [1, \ldots, L]\}$ is an EC distribution;
(4) the maximum likelihood (ML) estimates of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$ obtained from K samples $\mathbf{x}_k$ of $\mathbf{X}$ can be expressed as

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_k, \\ \hat{\mathbf{C}} &= \frac{1}{K} \sum_{k=1}^{K} (\mathbf{x}_k - \hat{\boldsymbol{\mu}}) \cdot (\mathbf{x}_k - \hat{\boldsymbol{\mu}})^T. \end{aligned} \tag{3}$$

Furthermore, on the basis of the Yao representation theorem [7], an SIRV can be expressed as

$$\mathbf{X} = A\mathbf{C}^{1/2}\mathbf{Z} + \boldsymbol{\mu}, \tag{4}$$

where $\mathbf{Z}$ is an L-dimensional Gaussian distributed random vector with zero mean and identity covariance matrix, and A is a scalar nonnegative random variable with unit squared mean value. The two variables $\mathbf{Z}$ and A are statistically independent.

According to (4), the p.d.f. of $\mathbf{X}$ is strictly related to the statistical distribution of the scalar random variable A. In particular, $\mathbf{X}$ conditioned to A has a multivariate Gaussian distribution:

$$f_{\mathbf{X}|A}(\mathbf{x} \mid \alpha) = \frac{1}{(2\pi)^{L/2}|\mathbf{C}|^{1/2}\alpha^L} \exp\left\{-\frac{d}{2\alpha^2}\right\}. \tag{5}$$

As a consequence, according to the principle of total probability, the p.d.f. of $\mathbf{X}$ can be written as

$$f_{\mathbf{x}}(\mathbf{x}) = \int_0^\infty f_{\mathbf{x}|A}(\mathbf{x} \mid \alpha) \cdot f_A(\alpha) d\alpha = \frac{1}{(2\pi)^{L/2}|\mathbf{C}|^{L/2}} \int_0^\infty \alpha^{-L} \exp\left\{-\frac{d}{2\alpha^2}\right\} f_A(\alpha) d\alpha. \quad (6)$$

The p.d.f. of A is called the SIRV *characteristic p.d.f.*

Equations (1) and (6) prove that the function $h_L(d)$ is related to the *characteristic p.d.f.* of $\mathbf{X}$ by means of the following integral equation:

$$h_L(d) = \int_0^\infty \alpha^{-L} \exp\left\{-\frac{d}{2\alpha^2}\right\} f_A(\alpha) d\alpha. \quad (7)$$

Thus, the statistical properties of $\mathbf{X}$ are uniquely determined by the mean vector $\boldsymbol{\mu}$, the covariance matrix $\boldsymbol{\Gamma}$ and the univariate p.d.f. of A.

The relationship between $h_L(d)$ and the p.d.f. $f_D(d)$ of D is (see [8, 9])

$$h_L(d) = \frac{2^{L/2} L^{L/2-1} \Gamma(L/2)}{d^{L/2-1}} f_D(d). \quad (8)$$

Equations (6) and (7) are very useful in the statistical analysis of the SIRVs. In fact, by assuming perfect knowledge of the mean and covariance matrix of $\mathbf{X}$, the analysis of the SIRV multivariate p.d.f. reduces to the study of a univariate p.d.f. In (8) the function $h_L(d)$ must be a nonnegative monotonically decreasing function (see [8]); thus, the statistical distribution of D must satisfy this constraint and cannot be chosen arbitrarily.

The class of EC distributions includes the multivariate Gaussian model. In fact, a Gaussian variable is an SIRV with

$$f_A(\alpha) = \delta(\alpha - 1), \quad h_L(d) = \exp\left\{-\frac{d}{2}\right\}. \quad (9)$$

To summarize, an EC model can be defined by specifying the multivariate p.d.f. of $\mathbf{X}$, or the p.d.f. of the scalar random variable D or by specifying the *characteristic p.d.f.* ($f_A(\alpha)$). In the latter two cases, knowledge of the mean vector and of the covariance matrix must be assumed.

2.2. Models adopted

2.2.1. Elliptically contoured t distribution model

The first model is based on multivariate EC-t distribution (see [4–6]). According to the EC-t model, the p.d.f. of $\mathbf{X}$ is expressed as

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{\Gamma[(L+\nu)/2]}{\Gamma[\nu/2](\nu\pi)^{L/2}} |\mathbf{R}|^{-1/2} \left[1 + \frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{R}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right]^{-L+\nu/2}, \quad (10)$$

where $\mathbf{R}$ is related to the covariance matrix of $\mathbf{X}$ by the following equation:

$$\mathbf{R} = \frac{\upsilon - 2}{\upsilon} \mathbf{C}. \quad (11)$$

For the EC-t distribution, the scalar variable D can be expressed as

$$D = L \frac{\upsilon - 2}{\upsilon} \Omega. \quad (12)$$

In (12) Ω denotes an F-central random variable with L and υ degrees of freedom. The parameter υ is strictly related to the shape of the distribution tails. In particular, for $\upsilon = 1$, the EC-t distribution reduces to the multivariate Cauchy distribution that has heavy tails, whereas when $\upsilon - \infty$ it tends to the multivariate Gaussian distribution characterized by lighter tails.

In [4, 5] the authors analyzed background classes including a number of pixels large enough to neglect the errors in the estimate of the mean vector and the covariance matrix. Thus, they reduced the analysis of the statistical behavior of real data to the study of the univariate distribution of D. Note that, by assuming perfect knowledge of $\boldsymbol{\mu}$ and $\mathbf{C}$, the EC-t distribution depends on the parameter υ alone. The analysis of HYDICE data was carried out in terms of a graphical comparison between the empirical *PoE* and the theoretical one. In particular, the authors showed that by varying the value of υ the theoretical *PoE* of D tends to the empirical one. They did not provide any method to automatically estimate the value of υ to obtain the best fitting.

The analysis of the statistical behavior of MIVIS data was carried out by also considering mixture models. The introduction of those models has a physical rationale in the spatial/spectral nonhomogeneity of the considered background classes. In particular, we considered models whose p.d.f.'s are expressed as a linear combination of ECD (see [6]). The models adopted are characterized by one or more parameters whose values must be set in order to obtain the best fitting between the empirical p.d.f. and the theoretical one. In mixture models, the number of parameters and the complexity of their estimation process increase with the number of component functions. One of the advantages of defining a multivariate model, that properly describes the statistical behavior of real background classes, is the ability to derive optimum detection strategies. Consequently, it is important to use models that are as simple as possible and that only have a few parameters.

For these reasons in our analysis, we considered two classes of mixture models that have few parameters and that are characterized by a high mathematical tractability. Thus, there is no physical meaning in the selected models. The models considered are denoted as Gaussian mixture model (GMM) [10] and N lognormal mixture model (N-LGM).

2.2.2. Gaussian mixture model (GMM)

The GMM exploits the fact that the distribution of hyperspectral data for a specific background class is obtained as the linear combination of a finite number N of Gaussian functions. In particular, the p.d.f. of $\mathbf{X}$ can be expressed as

$$f_{\mathrm{GMM}}(\mathbf{x}) = \sum_{i=1}^{N} \pi_i f_G(\mathbf{x}; \boldsymbol{\mu}_i, \mathbf{C}_i), \quad (13)$$

where $f_G(\mathbf{x};\boldsymbol{\mu}_i,\mathbf{C}_i)$ denotes the multivariate Gaussian p.d.f. with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\mathbf{C}_i$ and the $\pi_i \in [0,1]$ are the mixture weights subject to the sum to one constraint: $\sum_{i=1}^{N}\pi_i = 1$. Thus, the whole set of model parameters is $\Theta \equiv \{\pi_i, \boldsymbol{\mu}_i, \mathbf{C}_i,\ i = 1,\ldots,N\}$.

2.2.3. *N-lognormal mixture model (N-LGM)*

The N-LGM arises from the assumption that the p.d.f. of a background class can be expressed as the linear combination of ECD that share the same mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ and that have a lognormal *characteristic p.d.f.* The model reduces to an SIRV with mean vector $\boldsymbol{\mu}$, covariance matrix $\mathbf{C}$, and *characteristic p.d.f.* expressed as the linear combination of lognormal functions:

$$f_A(\alpha) = \sum_{i=1}^{N}\pi_i f_A^{(i)}(\alpha), \quad \pi_i \in [0,1],\ \sum_{i=1}^{N}\pi_i = 1,$$
$$f_A^{(i)}(\alpha) = \frac{1}{\sqrt{2\pi}\sigma_i\alpha}\exp\left\{-\frac{1}{2\sigma_i^2}\left[\ln\left(\frac{\alpha}{\delta_i}\right)\right]^2\right\}. \tag{14}$$

In (14) N denotes the number of mixture components and π_i the mixture coefficients. By using (8), the p.d.f. of the square of the Mahalanobis distance can be expressed as

$$f_D(d) = \frac{d^{L/2-1}}{2^{L/2}\Gamma(L/2)}\sum_{i=1}^{N}\pi_i\int_0^{\infty}\alpha^{-L}\exp\left\{-\frac{d}{2\alpha^2}\right\}f_A^{(i)}(\alpha)d\alpha. \tag{15}$$

According to the properties of the SIRV, since the variable A had a unit mean squared value, we must set the following constraints in the model (14):

$$\delta_i = -2\sigma_i^2 \quad \forall i \in [1,N]. \tag{16}$$

Thus, by assuming that $\boldsymbol{\mu}$ and $\mathbf{C}$ are known, the N-LGM is characterized by the following set of parameters:

$$\Theta \equiv \{c_1, c_2, \ldots, c_N, \pi_1, \pi_2, \ldots, \pi_{N-1}\}, \tag{17}$$

where $\pi_N = 1 - \sum_{i=1}^{N-1}\pi_i$.

3. EXPERIMENTAL DATA ANALYSIS

To analyze the statistical behavior of experimental hyperspectral data, we assume that a certain number M of pixels $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$ of a specific background class is available. Then $\mathbf{x}_i$ can be obtained by applying a classification algorithm to the image or by resorting to the ground truth if it is available. The non-Gaussian models considered in this study are characterized by one or more parameters that must be properly set in order to fit the empirical probability distribution (i.e., the distribution estimated over real data). For each of the three models, we propose a methodology to estimate the parameters from the available data.

3.1. *Elliptically contoured t distribution model: parameter estimation*

For the ECD models, we resort to (3) and (6) which represent the relationships between the multivariate p.d.f. of the data and the univariate distribution of the square of the Mahalanobis distance. The model estimates are obtained by considering the set $\{d_i : i = 1,\ldots,M; (\mathbf{x}_i - \boldsymbol{\mu})^T\mathbf{C}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\}$, where $\boldsymbol{\mu}$ and $\mathbf{C}$ are the mean vector and the covariance matrix of the background class. In practice, $\boldsymbol{\mu}$ and $\mathbf{C}$ are unknown and must be estimated from the data. In our experiments, we analyzed background classes including a large number of pixels (larger than $10L$), thus, the estimates of $\boldsymbol{\mu}$ and $\mathbf{C}$ can be reasonably considered as the exact values.

With regard to the EC-t model, the parameter v must be tuned to the empirical distribution. For this purpose, we propose two different techniques. The first one consists in setting the unknown parameter to its ML estimate from the d_is. It is obtained by looking for the value of v that maximizes the log-likelihood function defined as

$$\log\Lambda(d_1, d_2, \ldots, d_M, v) = \sum_{k=1}^{M_i}\log\left[f_D(d_k; v)\right],$$
$$f_D(d; v) = \frac{v}{v-2}\cdot\frac{1}{L}\cdot f_{\Omega}\left(\frac{d\cdot v}{v-2}\cdot\frac{1}{L}\right), \tag{18}$$

where $f_{\Omega}(\cdot)$ represents the p.d.f. of an F-central distributed random variable with L and v degrees of freedom. In evaluating the log-likelihood function, we assume the d_is are samples drawn from M random variables that are mutually independent and identically distributed. Unfortunately, the ML estimate of v cannot be obtained in closed form, so we resort to a numerical method to search for the absolute maximum of the likelihood function. For this purpose, several techniques can be adopted such as simulated annealing, stochastic sampling methods, and genetic algorithms. In this study, we adopted a genetic algorithm (GA) that uses the float representation [11]. This algorithm is efficient for numerical computations and is superior to both the binary genetic algorithm and the simulated annealing in terms of efficiency and quality of the solution (see [11]).

Note that, generally, in detection applications, in order to evaluate the test statistic in the algorithm decision rule, the background parameters must be estimated from a limited data set representing the background class where the target of interest is embedded. For this reason, the proposed estimation technique can be very useful in practical applications. In fact, it allows us to estimate the background parameter v from the samples d_is taken from the analyzed image.

In order to test the reliability of such an estimator, several computer simulations were performed. In particular, in our simulations we investigated the properties of the ML estimator for different values of the parameter v and of the number N_S of samples used to evaluate the log-likelihood function. These samples were generated according to (12), and the number of spectral bands L was set to 52 in accordance with the characteristics of the MIVIS data adopted in the experimental analysis described in Section 4. Table 1 shows the estimator mean values with respect to the number of samples and for each value of the parameter v. Whereas, Table 2 shows the estimator mean relative squared error versus the

Table 1: ML estimator: mean values obtained by simulation. Results obtained considering 10^4 realizations of the ML estimator.

	N_S			
υ	10^2	10^3	10^4	10^5
5	5,001	5	5	5
20	20,051	20,014	20	20
50	50,259	50,059	50,007	50
80	80,279	80,198	80,06	80,001

Table 2: ML estimator: mean squared error obtained by simulation. Results obtained considering 10^4 realizations of the ML estimator.

	N_S			
υ	10^2	10^3	10^4	10^5
5	10^{-5}	0	0	0
20	$3,5 \cdot 10^{-3}$	$4 \cdot 10^{-4}$	0	0
50	$6,6 \cdot 10^{-3}$	$7 \cdot 10^{-4}$	10^{-4}	$1,02 \cdot 10^{-5}$
80	$7,6 \cdot 10^{-3}$	$14 \cdot 10^{-4}$	$2 \cdot 10^{-4}$	$1,2 \cdot 10^{-5}$

number of samples. Note that for $N_S > 10^4$ the estimator mean reaches the true value of the parameter for each υ, and the estimator mean relative squared error is less than $2 \cdot 10^{-4}$. This leads us to conclude that the proposed estimator is unbiased and consistent for $N_S > 10^4$. These results are in accordance with the asymptotical properties of the ML estimators (MLE). In fact, the MLEs are asymptotically unbiased, consistent and efficient (they achieve the Cramer-Rao bound) [12].

The second technique proposed to estimate the parameter υ in the EC-t model consists in searching for the "best fitting" between the empirical and the theoretical cumulative distribution functions (c.d.f.). The goodness of fit is evaluated by a suitable cost function $J_P(\upsilon)$ calculated on P selected points (percentile) of the two c.d.f.'s and the estimate $\hat{\upsilon}$ is obtained as

$$\hat{\upsilon} = \min_{\upsilon} \{J_P(\upsilon)\},$$

$$J_P(\upsilon) = \sum_{k=1}^{P} \left\{ \frac{\log_{10}[F_{\text{emp}}(d_k)] - \log_{10}[F_{\text{th}}(d_k, \upsilon)]}{\log_{10}[F_{\text{emp}}(d_k)]} \right\}^2. \tag{19}$$

In (19) we denote with $F_{\text{emp}}(\cdot)$ the empirical c.d.f. derived from the histogram of the d_is and with $F_{\text{th}}(\cdot, \upsilon)$ the theoretical c.d.f. of the square of the Mahalanobis distance with respect to the parameter υ. The cost function evaluates the relative squared error between the logarithm of the empirical and theoretical c.d.f.'s. The logarithmic transformation is applied in order to give the same weight to the body and to the tails of the distributions. Since there is no closed form solution for the optimization problem in (19), we resort to a numerical method. In particular, we use the simplex search method described in [13]. This is a direct search method that does not use numerical or analytic gradients.

3.2. Gaussian mixture model: parameters estimation

With regard to the GMM, it is important to note that by increasing the number N of functions in the mixture, one would expect that the quality of the fitting would improve. Unfortunately, the increase in the number of mixture elements also increases the complexity of the model and limits its applicability to the analysis of the data and to the derivation of detection algorithms tuned to the statistical model. For these reasons, we considered the two distributions obtained by setting $N = 2$ (2-GMM) and $N = 3$ (3-GMM). The parameters of each multivariate Gaussian function and the mixture weights are estimated directly from $\mathbf{x}_i$ using the expectation maximization (EM) algorithm [14].

3.3. N-lognormal mixture model: parameter estimation

For the N-LGM, the parameter estimates are obtained using an approach similar to the one in (19). In this case, we search for the set of values $\hat{\Theta}$ that minimizes the cost function $J_P(\Theta)$ defined as

$$J_P(\Theta) = \sum_{k=1}^{P} \left\{ \frac{\log_{10}[f_{\text{emp}}(d_k)] - \log_{10}[f_{\text{th}}(d_k, \Theta)]}{\log_{10}[f_{\text{emp}}(d_k)]} \right\}^2, \tag{20}$$

where $f_{\text{emp}}(\cdot)$ denotes the empirical p.d.f. derived from the histogram of the d_is and $f_{\text{th}}(\cdot, \Theta)$ indicates the theoretical p.d.f. of the square of the Mahalanobis distance with respect to the parameter vector Θ:

$$f_{\text{th}}(d; \Theta) = H d^{L/2-1} \int_0^{\infty} a^{-L} \exp\left(-\frac{d}{2a^2}\right) f_A^{N-\text{LGM}}(a; \Theta) da,$$

$$H = \frac{1}{2^{L/2}\Gamma(L/2)}. \tag{21}$$

Regarding the number of elements of the mixture we can extend the remarks proposed for the GMM to the N-LGM. Thus, to limit the complexity of the model, we considered two mixture components (2-LGM).

4. EXPERIMENTAL RESULTS

The non-Gaussian models were applied to a set of real reflectance data in order to check which was the most appropriate to fit the empirical distribution. The data were collected during a measurement campaign held in Italy in 2002. The aim of the campaign was to collect data to support the development and the analysis of classification and detection algorithms. The data were gathered by the MIVIS instrument, an airborne sensor with 102 spectral channels covering the spectral region from the visible (VIS) to the thermal infrared (TIR).

In this study, we refer to a reduced data set consisting of 52 spectral channels selected by discarding the 10 TIR channels and those characterized by low signal-to-noise ratio (SNR). Furthermore, the SWIR channels were binned to enhance the SNR. The ground resolution is about 3 m.

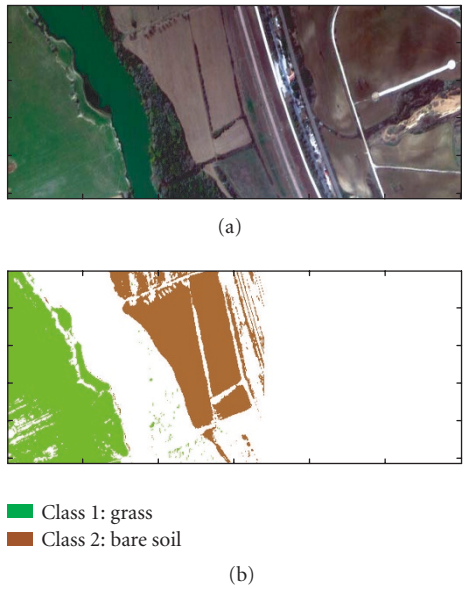

Figure 1: (a) RGB representation of the analyzed scene; (b) background classes considered.

Table 3: Number of pixels in each class.

	Class no.1	*Class no.2*
Number of pixels	369951	23482

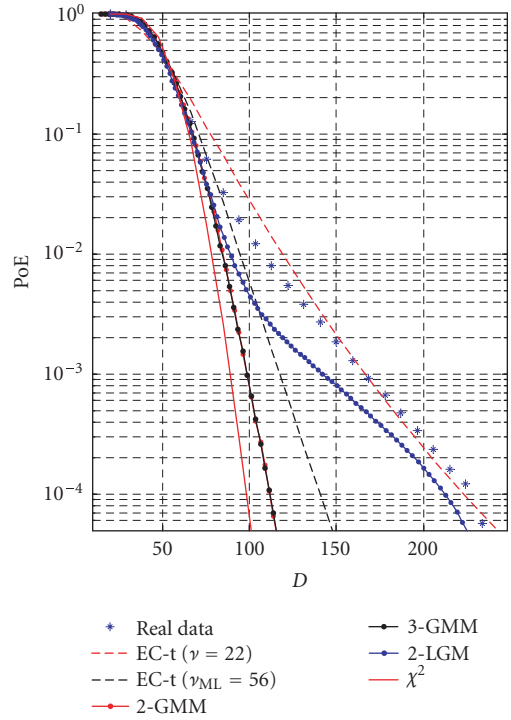

Figure 2: *Class no.1 (grass)*: *PoE* of D for the real data and for the theoretical models.

The results outlined in this paper regard two specific background classes selected from an MIVIS image using the unsupervised segmentation algorithm in [15]. The two classes are labelled as *class no.1* and *class no.2* and they correspond to two distinct regions of the scene covered by grass and bare soil, respectively. In Figure 1, we show the RGB image of the analyzed scene and we point out the two background classes considered. The number of pixels in each class is listed in Table 3. Since the number of pixels in each class is far larger than the number of sensor spectral channels, it is reasonable to assume that the errors in the mean vector and in the covariance matrix estimates from the class pixels are negligible. Thus, according to the properties of the ECDs, the analysis of the statistical behavior of real data can be reduced to the study of the distributions of the scalar variable D.

The analysis was carried out in terms of a graphical comparison between the empirical distributions and the theoretical ones. In Figures 2 and 3, the *PoE* of D estimated over real data associated with the two classes (empirical *PoE*) are compared with the *PoE* derived from each theoretical model (theoretical *PoE*). The *PoE* is defined as

$$\mathrm{PoE}(d) = 1 - \int_0^d f_D(t)\partial t, \tag{22}$$

where $f_D(\cdot)$ represents the p.d.f. of D. In plotting the *PoE*, we used the logarithmic scale in order to highlight the distribution tail.

In Figures 2 and 3, the *PoE* obtained by assuming the Gaussian model for the multivariate data has also been plotted. In this case, assuming perfect knowledge of the class mean vector and covariance matrix, the random variable D has a central χ^2 distribution with L degrees of freedom.

The results confirm that the Gaussian model does not accurately describe the statistical behavior of the data. In particular, it strongly deviates from the tails of the empirical distributions.

With regard to the EC-t model, we plotted two distributions for each class. The EC-t distributions were obtained by setting the υ parameter to the values $\hat{\upsilon}_{\mathrm{ML}}$ and $\hat{\upsilon}$ obtained by the MLE and by the procedure that minimizes the cost function in (19), respectively. In each class, the EC-t distribution derived by setting $\upsilon = \hat{\upsilon}_{\mathrm{ML}}$ does not properly account for the statistical behavior of the data. In particular, there is a good agreement between the body of the empirical distribution and the theoretical model but the distribution tail is not properly modeled. Instead, the EC-t model obtained for $\upsilon = \hat{\upsilon}$ fits the empirical distribution tail well but it is not completely appropriate for representing its body. The best performances achieved by the EC-t model with $\upsilon = \hat{\upsilon}_{\mathrm{ML}}$ in fitting the body of the empirical distributions are more evident in Figures 4 and 5. Here we plotted, for *class no.1* and *class no.2*, the empirical p.d.f. of D and the theoretical ones. In both the experiments discussed in this section the number of samples adopted to estimate the parameter υ using the MLE is larger than 10^4. Thus, according to the properties of the MLE we can state that if the pixels of each class were drawn from an EC-t distribution, $\hat{\upsilon}_{\mathrm{ML}}$ would be a reliable estimate of the model parameter. This leads us to the

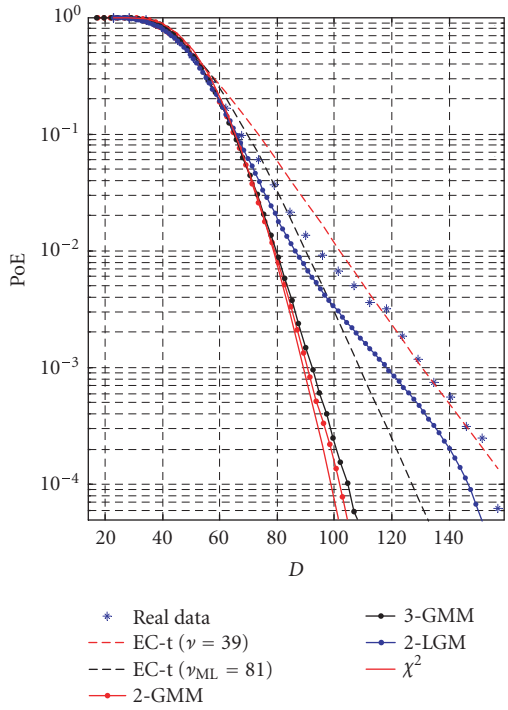

Figure 3: *Class no.2 (bare soil)*: *PoE* of D for the real data and for the theoretical models.

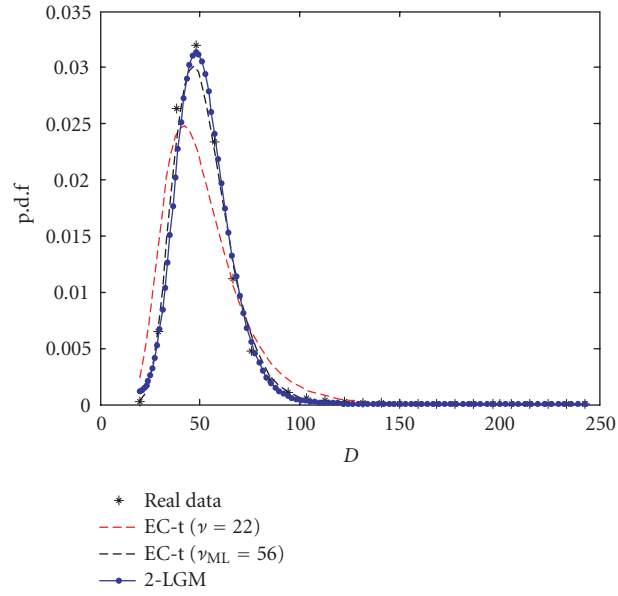

Figure 4: *Class no.1 (grass)*: p.d.f.'s for the real data and for three theoretical models.

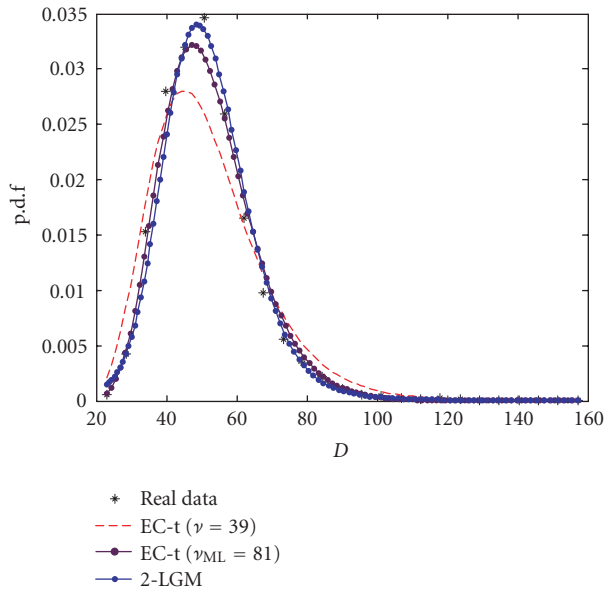

Figure 5: *Class no.2 (bare soil)*: p.d.f.'s of D for the real data and for three theoretical models.

conclusion that the statistical behavior of MIVIS data in the two considered background classes is not fully represented by means of an EC-t distribution. Furthermore, the fact that it is possible to properly describe the body and the tail of empirical distribution with two distinct EC-t models suggests that the use of mixture models is more appropriate to properly address hyper-spectral data variability. This has its physical rationale in the spectral/spatial nonhomogeneity within the observed background classes.

It is worth noting that the results suggest that the multivariate EC-t distribution cannot be adopted to derive optimum detection strategies. Nevertheless, they confirm that the tails of the empirical distribution of real hyper-spectral data can be properly represented by means of an EC-t model. The ability of EC-t models to follow the empirical distribution tails makes them very useful in assessing detection performance. In particular, since in detection applications the distribution tails are related to the number of false alarms, the EC-t models facilitate the derivation of criteria for tuning the algorithms, based on reliable predictions of the P_{FA}.

With regard to the mixture models, the 2-GMM and the 3-GMM perform better than the Gaussian model but they still do not provide a good representation of the data statistical distribution. Also note that by increasing the number of mixture elements from two to three, the results for fitting the empirical distribution do not improve significantly.

Among the statistical models considered, the 2-LGM provides the best performance in fitting the empirical distributions. In fact, it is totally suitable for representing the body of the distributions for both classes, as is proved by the results shown in Figures 4 and 5. Furthermore, Figure 3 highlights that the 2-LGM follows the behavior of the empirical distribution tail over *class no.2*. The results obtained from *class no.1* show that, except for the PoE range $[10^{-2}, 10^{-3}]$, the 2-LGM provides a good representation of the empirical distribution tail.

In order to quantify the ability of each model to address the statistical behavior of real data, we computed the fitting error index (*FEI*) defined as

$$\mathrm{FEI} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\log_{10}\left[F_{\mathrm{emp}}(d_i)\right] - \log_{10}\left[F_{\mathrm{th}}(d_i)\right]}{\log_{10}\left[F_{\mathrm{emp}}(d_i)\right]}\right\}^2. \quad (23)$$

Table 4: Fitting error index (*FEI*) values.

		EC-t ($\hat{\upsilon}$)	EC-t ($\hat{\upsilon}_{\mathrm{ML}}$)	2-LGM	2-GMM	3-GMM	χ^2
FEI	*class no.1*	0,31	0,59	0,23	0,44	0,46	0,75
	class no.2	0,25	0,37	0,19	0,47	0,50	0,60

This index is related to the relative mean squared error obtained by approximating the empirical c.d.f ($F_{\mathrm{emp}}(\cdot)$) with the theoretical one ($F_{\mathrm{th}}(\cdot)$). In computing the *FEI* we considered N different points of the two c.d.f.'s and we introduced the logarithmic transformation in order to give the same weight to the tails and to the body of the distributions. In Table 4, we report the *FEI* values for both background classes considered and for each theoretical model proposed in this manuscript.

The FEI values confirm that (1) the Gaussian model does not provide an appropriate characterization of the data variability; (2) 2-LGM has the lowest *FEI* value for both classes; (3) the EC-t model obtained with $\upsilon = \hat{\upsilon}$ gives a good representation of the empirical distribution tails, in fact it has *FEI* values close to those of the 2-LGM.

Benefits related to an accurate description of the distribution tails of real data can be obtained by predicting the detection performance of a given algorithm. In particular, improved accuracy in the estimates of the P_{FA} in real applications is expected. To give a numerical example we will now consider the well-known RX anomaly detector [16]. It is a statistical based detection algorithm and adopts as a test statistic the square of the Mahalanobis distance defined in (2). Thus, the empirical *PoE* values plotted in Figures 2 and 3 represent the P_{FA} for different values of the test threshold (λ) experienced by applying the RX detector to *class no.1* and *class no.2*, respectively. The theoretical *PoE* values in those figures are the P_{FA} predicted by applying each considered statistical model.

The availability of a model that properly accounts for the statistical behavior of each background class provides an accurate prediction of the detector P_{FA}. In Tables 5 and 6, we show the P_{FA} values, corresponding to a given test threshold, predicted by using each model presented in this study for the two classes considered. In both cases, the test threshold has been set to obtain a real P_{FA} value close to 10^{-3} (i.e., 9×10^{-4} for *class no.1* and 1.2×10^{-3} for *class no.2*). In the tables, we also show the values of the parameter η defined as

$$\eta(\overline{\lambda}) = \frac{P_{\mathrm{FA}}^{\mathrm{th}}(\overline{\lambda})}{\overline{P}_{\mathrm{FA}}^{\mathrm{emp}}} \cdot 100, \tag{24}$$

where $\overline{P}_{\mathrm{FA}}^{\mathrm{emp}}$ is the value of the false alarm probability obtained on real data, $\overline{\lambda}$ is the test threshold that allows $\overline{P}_{\mathrm{FA}}^{\mathrm{emp}}$ to be achieved, and $P_{\mathrm{FA}}^{\mathrm{th}}(\overline{\lambda})$ denotes the false alarm probability corresponding to $\overline{\lambda}$ for each considered statistical model. The values of η represent the percentage of the desired P_{FA} addressed by each theoretical model. Thus, it is a measure of the accuracy of the P_{FA} prediction task.

The results in Tables 5 and 6 show that the multivariate Gaussian model (χ^2 distribution on the test statistic) leads to serious errors in the prediction of the real P_{FA}. In fact, it only addresses the $3.38 \cdot 10^{-9}$% and the 0.0014% of $\overline{P}_{\mathrm{FA}}^{\mathrm{emp}}$ in *class no.1* and *class no.2* cases, respectively. The same conclusion can be drawn when the two multivariate Gaussian mixture models are considered. The prediction accuracy improves using the 2-LGM which allows the 48.6% and 39.6% of $\overline{P}_{\mathrm{FA}}^{\mathrm{emp}}$ to be addressed in the two cases considered. The best results were obtained by means of the EC-t model for $\upsilon = \hat{\upsilon}$ as was expected by its capacity to describe the real distribution tails. Using this model a large percentage of $\overline{P}_{\mathrm{FA}}^{\mathrm{emp}}$ is addressed both in *class no.1* and *class no.2* experiments. In fact, in the first case it is 99%, and in the second it is close to 94%.

Table 5: Second column: values of the P_{FA} predicted by using each theoretical model when the RX detector is applied to *class no.1* data and detection is accomplished with a test threshold $\overline{\lambda} = 168.61$. Third column: percentage of the P_{FA} obtained by applying the RX detector to *class no.1* data addressed by each theoretical model.

Model	$P_{\mathrm{FA}}^{(\mathrm{th})}(\overline{\lambda})$ ($\overline{\lambda} = 168.61$)	$\eta(\overline{\lambda})$ ($\overline{\lambda} = 168.61$)
χ^2	$3.09 \cdot 10^{-14}$	$3.38 \cdot 10^{-9}$
3-GMM	$4.45 \cdot 10^{-4}$	$7.96 \cdot 10^{-9}$
2-GMM	$7.30 \cdot 10^{-14}$	$8.01 \cdot 10^{-9}$
2-LGM	$7.35 \cdot 10^{-14}$	48.6
EC-t ($\hat{\upsilon}_{\mathrm{ML}}$)	$7.10 \cdot 10^{-6}$	0.77
EC-t ($\hat{\upsilon}$)	$9.12 \cdot 10^{-4}$	99.59

Table 6: Second column: values of the P_{FA} predicted by using each theoretical model when the RX detector is applied to *class no.2* data and detection is accomplished with a test threshold $\overline{\lambda} = 129.17$. Third column: percentage of the P_{FA} obtained by applying the RX detector to *class no.2* data addressed by each theoretical model.

Model	$P_{\mathrm{FA}}^{(\mathrm{th})}(\overline{\lambda})$ ($\overline{\lambda} = 129.17$)	$\eta(\overline{\lambda})$ ($\overline{\lambda} = 129.17$)
χ^2	$1.65 \cdot 10^{-8}$	0.0014
3-GMM	$4.70 \cdot 10^{-4}$	0.168
2-GMM	$2 \cdot 10^{-6}$	0.0029
2-LGM	$3.45 \cdot 10^{-8}$	39.6
EC-t ($\hat{\upsilon}_{\mathrm{ML}}$)	$7.68 \cdot 10^{-5}$	6.46
EC-t ($\hat{\upsilon}$)	$1.10 \cdot 10^{-3}$	93.98

5. CONCLUSIONS

In this paper, the ability of non-Gaussian models based on the SIRV theory to represent the statistical behavior of each background class in real hyper-spectral images has been investigated. The availability of statistical models that properly describe hyper-spectral data variability is of paramount importance in detection and classification problems. In fact, it

leads to the derivation of the best statistical decision strategies and the analytical characterization of their performance. The latter is a key element in designing automatic target detection and classification systems, in that it helps to provide criteria that can automatically set the algorithms parameters.

Three distinct non-Gaussian models have been considered: the EC-t model, the GMM, and the N-LGM both having a p.d.f. obtained as a linear combination of EC distributions. The GMM and the N-LGM were considered in order to address the multimodality of experimental data distributions due to spectral or spatial nonhomogeneity in the background classes considered. To limit the complexity of the mixture models the GMM with two (2-GMM) and three mixture components (3-GMM) and the N-LGM obtained with $N = 2$ (2-LGM) were analyzed. For each model a procedure was proposed to estimate the unknown parameters.

The analysis was performed on two distinct background classes selected on an MIVIS image. The comparison between the empirical and theoretical distributions was carried out graphically. Furthermore, for each model the *FEI* was computed to quantify the approximation errors.

The results prove that the empirical distributions cannot be represented using a unique multivariate EC-t model. In particular, they show that two distinct EC-t models must be used to properly describe the body and the tails of the empirical distributions, respectively. This leads us to conclude that mixture models must be used to properly account for MIVIS data variability. This is also confirmed by the fact that the 2-LGM, which has the lowest *FEI* values, outperforms the models considered.

It is worth noting that the low mathematical tractability of multivariate mixture models and their increasing number of parameters could complicate the derivation of decision strategies based on statistical criteria. Nevertheless, the ability to accurately describe background class variability in hyper-spectral images is crucial in characterizing the performance of the algorithms commonly used in practical applications. Within this framework, our analysis confirms that empirical distribution tails can be accurately modeled by means of an EC-t distribution. The related benefits are likely to be found in target detection applications. In particular, the ability to properly describe the distribution tails leads to accurate estimates of the P_{FA}, thus allowing the definition of criteria to automatically set the detector test threshold. In this paper, an experimental evidence of the advantages introduced by the correct modeling of real data has been provided. In particular, a case study is proposed where the accuracy of the theoretical models was quantified in terms of the P_{FA} related to the RX detector.

REFERENCES

[1] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 58–69, 2002.

[2] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29–43, 2002.

[3] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley & Sons, Hoboken, NJ, USA, 2003.

[4] D. Manolakis, D. Marden, J. Kerekes, and G. Shaw, "Statistics of hyperspectral imaging data," in *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII*, vol. 4381 of *Proceedings of SPIE*, pp. 308–316, Orlando, Fla, USA, April 2001.

[5] D. Manolakis and D. Marden, "Non Gaussian models for hyperspectral algorithm design and assessment," in *Proceedings of IEEE International Geosciences and Remote Sensing Symposium (IGARSS '02)*, vol. 3, pp. 1664–1666, Toronto, Canada, June 2002.

[6] D. Marden and D. Manolakis, "Modeling hyperspectral imaging data," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, vol. 5093 of *Proceedings of SPIE*, pp. 253–262, Orlando, Fla, USA, April 2003.

[7] K. Yao, "A representation theorem and its applications to spherically-invariant random processes," *IEEE Transactions on Information Theory*, vol. 19, no. 5, pp. 600–608, 1973.

[8] M. Rangaswamy, D. D. Weiner, and A. Ozturk, "Non-Gaussian random vector identification using spherically invariant random processes," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 1, pp. 111–124, 1993.

[9] M. Rangaswamy, D. D. Weiner, and A. Ozturk, "Computer generation of correlated non-Gaussian radar clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 1, pp. 106–116, 1995.

[10] S. G. Beaven, D. W. J. Stein, and L. E. Hoff, "Comparison of Gaussian mixture and linear mixture models for classification of hyperspectral data," in *Proceedings of IEEE International Geosciense and Remote Sensing Symposium (IGARSS '00)*, vol. 4, pp. 1597–1599, Honolulu, Hawaii, USA, July 2000.

[11] http://www.ie.ncsu.edu/mirage/GAToolBox/gaot/.

[12] S. M. Kay, *Fundamental of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 1993.

[13] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112–147, 1998.

[14] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[15] N. Acito, G. Corsini, and M. Diani, "An unsupervised algorithm for hyper-spectral image segmentation based on the Gaussian mixture model," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS '03)*, vol. 6, pp. 3745–3747, Toulouse, France, July 2003.

[16] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 38, no. 10, pp. 1760–1770, 1990.

N. Acito received the Laurea degree (cum Laude) in telecommunication engineering from University of Pisa, Pisa, Italy, in 2001, and the Ph.D. degree in methods and technologies for environmental monitoring from "Università della Basilicata," Potenza, Italy, in 2005. Since November 2004, he is a temporary Researcher with the Department of Information Engineering, University of Pisa, Italy. His research interests include signal and image processing. His current activity has been focusing on target detection and recognition in hyperspectral images.

G. Corsini received the Dr. Eng. degree in electronic engineering from the University of Pisa, Italy, in 1979. Since 1983, he has been with the Department of Information Engineering, University of Pisa, where he is currently a Full Professor of telecommunication engineering. His main research interests include multidimensional signal and image detection and processing, with emphasis on hyperspectral and multispectral data analysis of remotely sensed images. He has coauthored more than 150 technical papers published on international journals and conferences' proceedings.

M. Diani was born in Grosseto, Italy, in 1961. He received his Laurea degree (cum Laude) in electronic engineering from the University of Pisa, Italy, in 1988. He is currently an Associate Professor at the Department of Information Engineering of the University of Pisa. His main research area is in image and signal processing with application to remote sensing. His recent activity was focused in the fields of target detection and recognition in multi/hyperspectral images, and in the development of new algorithms for detection and tracking in infrared image sequences.