Non Zero Mean Adaptive Cosine Estimator and Application to Hyperspectral Imaging

François Vincent, Olivier Besson

Non Zero Mean Adaptive Cosine Estimator and Application to Hyperspectral Imaging

François Vincent and Olivier Besson

***Abstract*—The Adaptive Cosine Estimator (ACE) has become a popular detection scheme in many applications. Similarly to the majority of detection schemes, it assumes a zero mean noise. In some domains, such as hyperspectral imaging, this assumption no longer holds and this algorithm has to be adapted. In this paper we revisit the use of ACE in a non zero mean context. We consider the case where the data under test and the training samples differ from one scaling factor on the mean and one scaling factor on the covariance matrix. We derive two-step generalized likelihood ratio tests for both the additive model and the replacement model and show that the new detectors differ in the way the mean value is removed. A real data experiment shows that they outperform the standard version.**

***Index Terms*—Adaptive coherence estimator, detection, generalized likelihood ratio test, hyperspectral imaging.**

I. Introduction

Detecting a Signal of Interest (SoI) with a known signature in partially unknown noise is a key issue in many signal processing applications and constitutes one of the main objectives of radar systems for instance. The difficulty of this problem lies in the statistics of the noise being unknown. For instance with a Gaussian distributed noise, the mean and the covariance matrix may not be known and thus adaptive detection of the SoI requires using other samples (so called training samples) to learn the noise present in the signal under test.

Under the Gaussian assumption, many algorithms have been developed, such as Kelly's Generalized Likelihood Ratio Test (GLRT) [1], the Adaptive Matched Filter (AMF) [2] or the Adaptive Coherence/Cosine Estimator (ACE) [3], [4], to name the most popular ones. The latter algorithm, initially derived assuming a known noise covariance matrix (i.e. following a two step approach), was also shown to be the true scale invariant GLRT (i.e. following a one-step approach) [5]. It possesses very interesting properties and has become a reference in many applications, including hyperspectral imaging [6]. ACE has a striking interpretation as it simply measures the square of the cosine of the angle between the SoI and the signal under test. Thereby, ACE is robust to data with possible large amplitudes, where the level of the AMF can significantly vary while the above-mentioned angle remains the same [7].

However, ACE, just like the majority of popular detection schemes, has been derived for zero-mean signals. Unfortunately, in some applications, the data are always positive and this hypothesis is not fulfilled. This is the case in the image processing field and especially for hyperspectral applications, this domain being the main focus of this paper. In this case, ACE is simply applied after a demeaning step, i.e., after removing the mean value of the training samples.

In this paper, we revisit the use of ACE in a non-zero mean data context, with a special interest in hyperspectral image processing. More precisely, we consider two popular signal models usually used in the hyperspectral domain, namely the additive and the replacement models [8], and we derive the corresponding two-step GLRT assuming a different scaling of the mean and the covariance matrix between the data under test and the training samples, in the spirit of ACE approach. Moreover, the scaling factor affecting the mean is independent of the scaling factor affecting the covariance. We show that all solutions have similar formulations, i.e., the square of a cosine angle, but with differences in the way the mean is removed from the data. These differences induce major performance variations, as shown on a real data experiment.

The authors are with ISAE SUPAERO, Université de Toulouse, France. Email: francois.vincent@isae-supaero.fr, olivier.besson@isae-supaero.fr

II. Signal model

The problem at hand consists in deciding whether a target $\mathbf{s}$ is present in the signal under test $\mathbf{x}$ or if there is noise only $\mathbf{n}$. In order to characterize the noise, one has also access to secondary data $\mathbf{z}_k \quad k = 1, ..., K$, hopefully free of target. Unlike for the detection problem typically tackled in many signal processing applications, we consider in this paper non zero-mean data. More precisely the noise vector is assumed to be Gaussian distributed, namely $\mathbf{n} \sim \mathcal{N}(\gamma\mathbf{m}, \sigma^2\mathbf{C})$, whereas $\mathbf{z}_k \sim \mathcal{N}(\mathbf{m}, \mathbf{C})$. Doing so, we generalize the central hypothesis underlying the derivation of ACE, considering not only an unknown scaling factor σ^2 between the covariance matrices of the signal under test $\mathbf{x}$ and the secondary data $\mathbf{z}_k$, but also an unknown scaling factor γ between their means.

As for the SoI, two popular models are typically advocated in the hyperspectral domain. The more realistic one considers an opaque target that will mask a part of the background $\mathbf{n}$. This part is known as the fill factor α [9], [10], and the target will replace this fraction of the background. When $\alpha \to 0$, this so-called replacement model tends to the standard additive model, widely used in many signal processing applications. In this case, the target, if present, simply adds to the background. Therefore, the problem we tackle can be formulated as the following composite two-hypotheses test

$$\begin{cases} H_0 : & \mathbf{x} = \mathbf{n} \\ H_1 : & \mathbf{x} = \alpha\mathbf{s} + \beta\mathbf{n} \end{cases} \tag{1}$$

where β defines the model type: additive model when $\beta = 1$ or replacement model when $\beta = (1 - \alpha)$.

III. Generalization of ACE

In the sequel we derive the GLRT corresponding to the problem described in (1) assuming that $\mathbf{n} \sim \mathcal{N}(\gamma\mathbf{m}, \sigma^2\mathbf{C})$ and that $\mathbf{m}$ and $\mathbf{C}$ are known. For adaptive detection the latter will be replaced by their estimates obtained from training samples $\mathbf{z}_k$. The target signature $\mathbf{s}$ is supposed to be known and the fill factor α is deterministic and unknown. We will first derive the GLRT for the model typically assumed in the hyperspectral domain when ACE is advocated [11], namely that the mean remains the same ($\gamma = 1$). Then we will consider the more general case (γ unknown), as the rationale of choosing the same mean but a scaling factor for the covariance matrix is not obvious. Below we will say that two test statistics t_1 and t_2 are equivalent, which we denote $t_1 \equiv t_2$, if for every threshold η_1 there exists a unique threshold η_2 such that $t_1 > \eta_1 \Leftrightarrow t_2 > \eta_2$. In particular $t_1 \equiv a t_1^{2/N}$ which will enable us to discard all constant terms in the GLR and to take its $N/2$-th root.

A. Same mean ($\gamma = 1$)

First we consider that $\mathbf{n} \sim \mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$. As $\mathbf{C}$ is supposed to be known, we can whiten the data and rewrite (1) as

$$\begin{cases} H_0: & \mathbf{y} = \mathbf{b} \\ H_1: & \mathbf{y} = \alpha\mathbf{t} + \beta\mathbf{b} \end{cases} \tag{2}$$

where $\mathbf{y} = \mathbf{C}^{-\frac{1}{2}}\mathbf{x}$, $\mathbf{b} = \mathbf{C}^{-\frac{1}{2}}\mathbf{n}$, $\mathbf{t} = \mathbf{C}^{-\frac{1}{2}}\mathbf{s}$ are the whitened versions of $\mathbf{x}$, $\mathbf{n}$, $\mathbf{s}$, and $\mathbf{b} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, with $\boldsymbol{\mu} = \mathbf{C}^{-\frac{1}{2}}\mathbf{m}$.

The likelihood functions under H_0 and H_1 are

$$\begin{aligned} p_0(\mathbf{y}; \sigma^2) &\propto \sigma^{-N} \exp\left\{ -\frac{\|\mathbf{y} - \boldsymbol{\mu}\|^2}{2\sigma^2} \right\} \\ p_1(\mathbf{y}; \alpha, \beta, \sigma^2) &\propto (\beta\sigma)^{-N} \exp\left\{ -\frac{\|\mathbf{y} - \alpha\mathbf{t} - \beta\boldsymbol{\mu}\|^2}{2\beta^2\sigma^2} \right\} \end{aligned} \tag{3}$$

where $\propto$ means proportional to. The Maximum Likelihood (ML) estimates of σ^2 are known to be respectively $\hat{\sigma_0^2} = N^{-1}\|\mathbf{y} - \boldsymbol{\mu}\|^2$ and $\hat{\sigma_1^2} = (\beta^2 N)^{-1}\|\mathbf{y} - \alpha\mathbf{t} - \beta\boldsymbol{\mu}\|^2$, so that the concentrated likelihood functions are given by

$$\begin{aligned} \max_{\sigma^2} p_0(\mathbf{y}; \sigma^2) &\propto \|\mathbf{y} - \boldsymbol{\mu}\|^{-N} \\ \max_{\sigma^2} p_1(\mathbf{y}; \alpha, \beta, \sigma^2) &\propto \|\mathbf{y} - \alpha\mathbf{t} - \beta\boldsymbol{\mu}\|^{-N} \end{aligned} \tag{4}$$

In the sequel we make repeated use of the following fact:

$$\min_a \|\mathbf{z} - a\mathbf{v}\|^2 = \left\|\mathbf{P}_{\mathbf{v}}^{\perp}\mathbf{z}\right\|^2 = \left\|\mathbf{z} - \frac{\mathbf{v}^T\mathbf{z}}{\mathbf{v}^T\mathbf{v}}\mathbf{v}\right\|^2 \tag{5}$$

where $\mathbf{P}_{\mathbf{v}}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathbf{v}}$ with $\mathbf{P}_{\mathbf{v}} = (\mathbf{v}^T\mathbf{v})^{-1}\mathbf{v}\mathbf{v}^T$ the orthogonal projection on $\mathbf{v}$.

Let us consider first the case $\beta = 1$ (additive model). Then, taking the $N/2$-th root of the GLR yields

$$\begin{aligned} \mathrm{GLR} &\equiv \frac{\|\mathbf{y} - \boldsymbol{\mu}\|^2}{\left\|\mathbf{P}_{\mathbf{t}}^{\perp}(\mathbf{y} - \boldsymbol{\mu})\right\|^2} \\ &= \frac{\|\mathbf{y} - \boldsymbol{\mu}\|^2}{\|\mathbf{y} - \boldsymbol{\mu}\|^2 - \|\mathbf{P}_{\mathbf{t}}(\mathbf{y} - \boldsymbol{\mu})\|^2} \\ &\equiv \frac{\|\mathbf{P}_{\mathbf{t}}(\mathbf{y} - \boldsymbol{\mu})\|^2}{\|\mathbf{y} - \boldsymbol{\mu}\|^2} \\ &= \frac{[\mathbf{t}^T(\mathbf{y} - \boldsymbol{\mu})]^2}{\|\mathbf{y} - \boldsymbol{\mu}\|^2 \|\mathbf{t}\|^2} = \cos^2(\tilde{\mathbf{y}}, \mathbf{t}) \end{aligned} \tag{6}$$

with $\tilde{\mathbf{y}} = \mathbf{y} - \boldsymbol{\mu}$ and where $\cos(\mathbf{u}, \mathbf{v})$ represents the cosine of the angle between vectors $\mathbf{u}$ and $\mathbf{v}$.

When $\beta = 1 - \alpha$ (replacement model), we have

$$\begin{aligned} \mathrm{GLR} &\equiv \frac{\|\mathbf{y} - \boldsymbol{\mu}\|^2}{\min_\alpha \|\mathbf{y} - \boldsymbol{\mu} - \alpha(\mathbf{t} - \boldsymbol{\mu})\|^2} \\ &= \frac{\|\mathbf{y} - \boldsymbol{\mu}\|^2}{\left\|\mathbf{P}_{\mathbf{t}-\boldsymbol{\mu}}^{\perp}(\mathbf{y} - \boldsymbol{\mu})\right\|^2} \equiv \cos^2(\tilde{\mathbf{y}}, \tilde{\mathbf{t}}) \end{aligned} \tag{7}$$

where $\tilde{\mathbf{t}} = \mathbf{t} - \boldsymbol{\mu}$. Note that the mean $\boldsymbol{\mu}$ is removed both from $\mathbf{y}$ and $\mathbf{t}$, in contrast to the additive model where only $\mathbf{y}$ is centered.

B. Scaling factor on the mean (arbitrary γ)

Now we consider the more general case where the mean, in addition to the covariance matrix, is known up to a scaling factor, namely $\mathbf{n} \sim \mathcal{N}(\gamma\mathbf{m}, \sigma^2\mathbf{C})$. Again, we first whiten all vectors and the likelihood functions under H_0 and H_1 become

$$\begin{aligned} p_0(\mathbf{y}; \gamma, \sigma^2) &\propto \sigma^{-N} \exp\left\{ -\frac{\|\mathbf{y} - \gamma\boldsymbol{\mu}\|^2}{2\sigma^2} \right\} \\ p_1(\mathbf{y}; \alpha, \beta, \gamma, \sigma^2) &\propto (\beta\sigma)^{-N} \exp\left\{ -\frac{\|\mathbf{y} - \alpha\mathbf{t} - \beta\gamma\boldsymbol{\mu}\|^2}{2\beta^2\sigma^2} \right\} \end{aligned} \tag{8}$$

Substituting σ^2 by its ML estimate both under H_0 and H_1, we get

$$\begin{aligned} \max_{\sigma^2} p_0(\mathbf{y}; \gamma, \sigma^2) &\propto \|\mathbf{y} - \gamma\boldsymbol{\mu}\|^{-N} \\ \max_{\sigma^2} p_1(\mathbf{y}; \alpha, \beta, \gamma, \sigma^2) &\propto \|\mathbf{y} - \alpha\mathbf{t} - \beta\gamma\boldsymbol{\mu}\|^{-N} \end{aligned} \tag{9}$$

At this stage it is important to note that β and γ cannot be estimated individually under H_1, only the product $\beta\gamma$ can be estimated. Consequently the GLRT will be the same under the additive model and the replacement model. Pursuing, we have

$$\max_{\gamma, \sigma^2} p_0(\mathbf{y}; \gamma, \sigma^2) \propto \left\|\mathbf{y}^{\perp}\right\|^{-N} \tag{10}$$

where $\mathbf{y}^{\perp} = \mathbf{P}_{\boldsymbol{\mu}}^{\perp}\mathbf{y}$. Similarly

$$\max_{\beta, \gamma, \sigma^2} p_1(\mathbf{y}; \alpha, \beta, \gamma, \sigma^2) \propto \left\|\mathbf{y}^{\perp} - \alpha\mathbf{t}^{\perp}\right\|^{-N} \tag{11}$$

where $\mathbf{t}^{\perp} = \mathbf{P}_{\boldsymbol{\mu}}^{\perp}\mathbf{t}$. Finally, maximization with respect to α yields

$$\max_{\alpha, \beta, \gamma, \sigma^2} p_1(\mathbf{y}; \alpha, \beta, \gamma, \sigma^2) \propto \left\|\mathbf{P}_{\mathbf{t}^{\perp}}^{\perp}\mathbf{y}^{\perp}\right\|^{-N} \tag{12}$$

Therefore, the GLR for arbitrary γ is given by

$$\text{GLR} \equiv \frac{\left\|\mathbf{y}^{\perp}\right\|^2}{\left\|\mathbf{P}^{\perp}_{\mathbf{t}^{\perp}}\mathbf{y}^{\perp}\right\|^2} \equiv \cos^2(\mathbf{y}^{\perp}, \mathbf{t}^{\perp}) \tag{13}$$

It will be referred to as Mean Removal ACE (MRACE).

C. Summary

In order to get back to the original (non whitened) vectors, let us note that $\tilde{\mathbf{y}} = \mathbf{C}^{-1/2}(\mathbf{x}-\mathbf{m})$, $\mathbf{t} = \mathbf{C}^{-1/2}\mathbf{s}$, $\tilde{\mathbf{t}} = \mathbf{C}^{-1/2}(\mathbf{s}-\mathbf{m})$ and

$$\begin{aligned}\mathbf{y}^{\perp} &= \mathbf{C}^{-1/2}\left[\mathbf{x} - \frac{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{x}}{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{m}}\mathbf{m}\right] \\ \mathbf{t}^{\perp} &= \mathbf{C}^{-1/2}\left[\mathbf{s} - \frac{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{s}}{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{m}}\mathbf{m}\right]\end{aligned} \tag{14}$$

Then, substituting these expressions in (6), (7) and (13), we observe that all GLRTs share the same form, namely

$$\boxed{\cos^2(\mathbf{C}^{-1/2}(\mathbf{x}-a\mathbf{m}), \mathbf{C}^{-1/2}(\mathbf{s}-b\mathbf{m}))} \tag{15}$$

which corresponds to the square cosine between the whitened versions of the data $\mathbf{x}$ and the SoI $\mathbf{s}$, after removal of a scaled version of $\mathbf{m}$. The various detectors differ in the coefficients (a, b), i.e., in the way they remove the mean $\mathbf{m}$ from $\mathbf{x}$ and $\mathbf{s}$, see Table I for a synthesis. These seemingly minor differences will however yield significant differences in terms of performance, as will be illustrated next.

TABLE I
SUMMARY OF THE DIFFERENT DETECTORS

	Additive ($\beta = 1$)	Replacement ($\beta = 1-\alpha$)
$\gamma = 1$	$a = 1$, $b = 0$	$a = 1$, $b = 1$
arbitrary γ	$a = \frac{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{x}}{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{m}}$, $b = \frac{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{s}}{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{m}}$	

In the supplementary material provided we show that, for all three detectors above, the test statistic of (15) follows a beta distribution, which is central under H_0 and non central under H_1. Moreover, the distribution of the test statistic does not depend on γ or σ^2 under H_0.

IV. Performance evaluation

In order to assess the performance of each GLRT derived in this paper (and hence the validity of the different assumptions), we propose to test them against real hyperspectral data. More precisely, we consider the Viareggio 2013 airborne trial [12]. This benchmarking hyperspectral detection campaign took place in Viareggio (Italy), in May 2013, with an aircraft flying at 1200 meters. The open data consist of three $[450 \times 375]$ pixels maps composed of 511 samples in the Visible Near InfraRed (VINR) band $(400-1000\text{nm})$. The spatial resolution of the image is about 0.6 meters. Different kinds of vehicles as well as coloured panels served as known targets. For each of these targets, a spectral signature obtained from ground spectroradiometer measurements is available. Moreover, a black and a white cover, serving as calibration targets, were also deployed. As can be seen on Fig. 1, the scene is composed

Fig. 1. Complete RGB view of the Viareggio test scene

of parking lots, roads, buildings, sport fields and pine woods. Three images, referred here-after as im1, im2 and im3 have been acquired with different target configurations.

As for the majority of hyperspectral detection schemes, the first step of the processing aims at converting the raw measurements into a reflectance map, namely removing all atmospheric effects and non-uniform sun illumination. To this end, we use the Empirical Line Method (ELM) [13] [14], considering the black and white calibration panels. Then a spectral binning [15] is performed to reduce the vector size dimension to $N = 32$. The performance of each algorithm is assessed calculating the number of pixels having their detector's output strictly higher than the one for the target pixel. In other words, this number can be seen as a false alarm number with an optimal thresholding.

We considered three training samples configurations. A first series of experiments is conducted with a small local window of size 13×13 pixels with a 9×9 guard window, resulting in $2.75N$ training samples. Then we consider a 19×19 window to increase the number of training samples to $8.75N$. Finally, we consider the global image to estimate the background parameters. In the three corresponding tables presented hereafter, we only consider the more challenging targets to detect, as otherwise the number of false alarms is very small for any detector [16]. We compare the 3 versions of the non zero mean ACE of table I with the AMF and Kelly's GLRT. We have also added the two GLRTs derived in [17] for comparison, namely the modified FTMF and SPADE.

In light of these results, we can draw the following remarks. First of all, the popular ACE scheme simply adapted to remove the mean (additive model with $\mathbf{n} \sim \mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$) is fair only, with sometimes less good results than the basic AMF, especially when the number of training samples is small. The GLRT based on the replacement model, referred as ACE replacement (always assuming $\mathbf{n} \sim \mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$) exhibits rather better results, except for the so-called V3 target and $K = 8.75N$.

On the other hand, we observe a huge improvement when

TABLE II
FALSE ALARMS SCORE, LOCAL PROCESSING WITH $K/N = 2.75$

target, image	AMF	Kelly	Modified FTMF	SPADE	ACE additive $\mathbf{n} \sim \mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$	ACE replacement $\mathbf{n} \sim \mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$	MRACE $\mathbf{n} \sim \mathcal{N}(\gamma\mathbf{m}, \sigma^2\mathbf{C})$
V3, im1	143	158	144	150	235	255	6
V3, im2	219	303	220	304	823	169	1
P2, im1	9	11	9	17	25	1	0
P2, im2	2	2	2	7	3	0	0
V6, im3	104	97	70	35	86	38	2

TABLE III
FALSE ALARMS SCORE, LOCAL PROCESSING WITH $K/N = 8.75$

target, image	AMF	Kelly	Modified FTMF	SPADE	ACE additive $\mathbf{n} \sim \mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$	ACE replacement $\mathbf{n} \sim \mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$	MRACE $\mathbf{n} \sim \mathcal{N}(\gamma\mathbf{m}, \sigma^2\mathbf{C})$
V3, im1	68	41	69	45	38	101	13
V3, im2	39	40	44	43	58	94	0
P2, im1	7	8	7	8	15	15	0
P2, im2	6	6	6	6	17	8	0
V6, im3	142	99	167	104	17	15	0

TABLE IV
FALSE ALARMS SCORE, GLOBAL PROCESSING ($K/N = 5271$)

target, image	AMF	Kelly	Modified FTMF	SPADE	ACE additive $\mathbf{n} \sim \mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$	ACE replacement $\mathbf{n} \sim \mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$	MRACE $\mathbf{n} \sim \mathcal{N}(\gamma\mathbf{m}, \sigma^2\mathbf{C})$
V3, im1	26	26	26	26	1	1	1
V3, im2	16	16	16	16	1	1	1
P2, im1	2	2	2	2	0	0	0
P2, im2	0	0	0	0	0	0	0
V6, im3	42	42	37	37	3	3	0

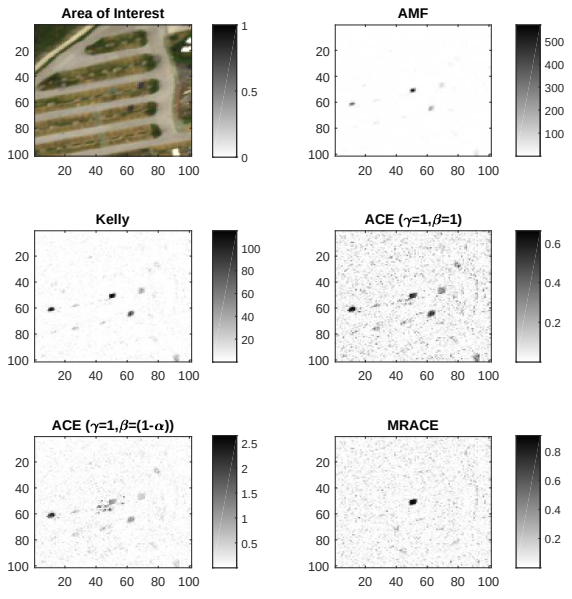

Fig. 2. Detectors output comparison

considering a scaling factor on the mean, in addition to the scaling factor on the covariance matrix ($\mathbf{n} \sim \mathcal{N}(\gamma\mathbf{m}, \sigma^2\mathbf{C})$), namely using MRACE. Indeed, in this case, the GLRT, which is the same for both the additive and the replacement models, exhibits striking performances compared to all other schemes, with a larger gap when the training samples number is small. This kind of robustness to possible bad estimation of the background parameters is certainly due to the fact that we relax the relationship between the mean of the training samples and that of the pixel under test. Finally, we also observe a false alarm reduction for target-like components when using this more general version of the non-zero mean ACE. Indeed, Figure 2 represents zooms of the detectors' output around the target, which lies at the center of the maps. Target-like parking lots splitters produce recurrent peaks in the detectors output creating possible false alarms. We clearly see that these unwanted peaks are drastically reduced when considering MRACE, increasing the so-called selectivity of the detection.

V. CONCLUSIONS

In this paper we focused on the adaptation of the popular ACE detector to the case of non zero mean data, with a special interest in hyperspectral target detection. The feature of ACE being to consider a scaling factor on the covariance matrix between the primary and the secondary data, we studied two hypothesis for the mean. The first one simply assumes that the mean remains the same, whereas the second one supposes a scaling factor on the mean, in addition to the covariance matrix. For these two cases, we derive the GLRT for both the additive and the replacement models, two models typically used in the hyperspectral context. Based on a real data experiment the new MRACE (assuming a scaling factor on the mean) seems to outperform all popular detection schemes, including the standard ACE version.

REFERENCES

[1] E. Kelly, "An adaptive detection algorithm," *IEEE Transactions Aerospace Electronic Systems*, vol. 22, no. 2, pp. 115–127, March 1986.

[2] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Transactions Aerospace Electronic Systems*, vol. 28, no. 1, pp. 208–216, January 1992.

[3] E. Conte, M. Lops, , and G. Ricci, "Asymptotically optimum radar detection in compound-Gaussian clutter," *IEEE Transactions Aerospace Electronic Systems*, vol. 31, no. 2, pp. 617–625, 1995.

[4] L. L. Scharf and T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," in *Proceedings 30th Asilomar Conference Signals Systems Computers*, Pacific Grove, CA, November 3-6 1996, pp. 1114–1117.

[5] S. Kraut, L. L. Scharf, and L. T. McWhorter, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Transactions Signal Processing*, vol. 47, no. 9, pp. 2538–2541, September 1999.

[6] D. Manolakis, M. Pieper, E. Truslow, T. Cooley, M. Brueggeman, and S. Lipson, "The remarkable success of adaptive cosine estimator in hyperspectral target detection," in *Proc. SPIE 8743, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX, 874302*, May 2013.

[7] E. Truslow, "Performance evaluation of the adaptive cosine estimator detector for hyperspectral imaging applications," Ph.D. dissertation, Northeastern University Boston, Massachusetts, 2012.

[8] D. G. Manolakis, R. B. Lockwood, and T. W. Cooley, *Hyperspectral Imaging Remote Sensing*. Cambridge University Press, 2016.

[9] A. Schaum and A. Stocker, "Spectrally-selective target detection," in *Proceedings of ISSSR*, vol. 12, April 1997, pp. 2015–2018.

[10] D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, "Is there a best hyperspectral detection algorithm?" in *Proc. of SPIE*, vol. 7334, 2009.

[11] J. Frontera-Pons, F. Pascal, and J.-P. Ovarlez, "Adaptive nonzero-mean Gaussian detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 1117–1124, February 2017.

[12] N. Acito, S. Matteoli, A. Rossi, M. Diani, and G. Corsini, "Hyperspectral airborne "Viareggio 2013 trial" data collection for detection algorithm assessment," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 6, pp. 2356–2376, June 2016.

[13] G. Ferrier, "Evaluation of apparent surface reflectance estimation methodologies," *International Journal of Remote Sensing*, vol. 16, pp. 2291–2297, 1995.

[14] G. M. Smith and E. J. Milton, "The use of the empirical line method to calibrate remotely sensed data to reflectance," *International Journal of Remote Sensing*, vol. 20, pp. 2653–2662, 1999.

[15] M. Shi and G. Healey, "Hyperspectral texture recognition using a multiscale opponent representation," *IEEE Transactions Geoscience Remote Sensing*, vol. 41, no. 5, pp. 1090–1095, May 2003.

[16] F. Vincent and O. Besson, "Generalized likelihood ratio test for subpixel target detection in hyperspectral imaging," *IEEE Transactions Geoscience Remote Sensing*, vol. 58, no. 6, pp. 4479–4489, 2020.

[17] ——, "Generalized likelihood ratio test for modified replacement model in hyperspectral imaging detection," *Signal Processing*, vol. 174, 2020.

SUPPLEMENTARY MATERIAL

The problem addressed can be formulated as the following composite two-hypotheses test

$$\begin{cases} H_0: & \mathbf{x} = \mathbf{n} \\ H_1: & \mathbf{x} = \alpha\mathbf{s} + \beta\mathbf{n} \end{cases} \tag{1}$$

where $\beta = 1$ for the additive model and $\beta = (1-\alpha)$ for the replacement model, $\mathbf{n} \sim \mathcal{N}(\gamma\mathbf{m}, \sigma^2\mathbf{C})$ where $\mathbf{m}$ and $\mathbf{C}$ are known. The test statistic for the GLRT is

$$\begin{aligned} t &= \cos^2(\mathbf{C}^{-1/2}(\mathbf{x} - a\mathbf{m}), \mathbf{C}^{-1/2}(\mathbf{s} - b\mathbf{m})) \\ &= \frac{[(\mathbf{x} - a\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{s} - b\mathbf{m})]^2}{[(\mathbf{x} - a\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x} - a\mathbf{m})]\,[(\mathbf{s} - b\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{s} - b\mathbf{m})]} \end{aligned} \tag{2}$$

where a and b are defined in Table I below.

TABLE I
SUMMARY OF THE DIFFERENT DETECTORS

	Additive ($\beta = 1$)	Replacement ($\beta = 1 - \alpha$)
$\gamma = 1$	$a = 1$, $b = 0$	$a = 1$, $b = 1$
arbitrary γ	$a = \frac{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{x}}{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{m}}$, $b = \frac{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{s}}{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{m}}$	

In this supplementary material, we provide a brief analysis of the above detectors, more precisely a stochastic representation of the test statistic in terms of well-known distributions.

Let us first consider the case $\gamma = 1$ for which one has

$$\mathbf{x} \sim \mathcal{N}(\alpha\mathbf{s} + (1 - \alpha b)\mathbf{m}, (1 - \alpha b)^2\sigma^2\mathbf{C})$$

where $b = 0$ for the additive model and $b = 1$ for the replacement model. Let us define

$$\begin{aligned} \tilde{\mathbf{x}} &= (1 - \alpha b)^{-1}\sigma^{-1}\mathbf{C}^{-1/2}(\mathbf{x} - \mathbf{m}) \\ &\sim \mathcal{N}(\alpha(1 - \alpha b)^{-1}\sigma^{-1}\mathbf{C}^{-1/2}(\mathbf{s} - b\mathbf{m}), \mathbf{I}) \end{aligned}$$

The test statistic can be written as

$$\begin{aligned} t &= \frac{[(\mathbf{x} - \mathbf{m})^T\mathbf{C}^{-1}(\mathbf{s} - b\mathbf{m})]^2}{[(\mathbf{x} - \mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x} - \mathbf{m})][(\mathbf{s} - b\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{s} - b\mathbf{m})]} \\ &= \frac{\tilde{\mathbf{x}}^T\mathbf{P}_{\mathbf{C}^{-1/2}(\mathbf{s}-b\mathbf{m})}\tilde{\mathbf{x}}}{\tilde{\mathbf{x}}^T\tilde{\mathbf{x}}} \\ &= \frac{\tilde{\mathbf{x}}^T\mathbf{P}_{\mathbf{C}^{-1/2}(\mathbf{s}-b\mathbf{m})}\tilde{\mathbf{x}}}{\tilde{\mathbf{x}}^T\mathbf{P}_{\mathbf{C}^{-1/2}(\mathbf{s}-b\mathbf{m})}\tilde{\mathbf{x}} + \tilde{\mathbf{x}}^T\mathbf{P}^{\perp}_{\mathbf{C}^{-1/2}(\mathbf{s}-b\mathbf{m})}\tilde{\mathbf{x}}} \\ &= \frac{A}{A + B} \end{aligned}$$

where

$$\begin{aligned} A &= \tilde{\mathbf{x}}^T\mathbf{P}_{\mathbf{C}^{-1/2}(\mathbf{s}-b\mathbf{m})}\tilde{\mathbf{x}} \overset{d}{=} \chi^2_1(\delta) \\ B &= \tilde{\mathbf{x}}^T\mathbf{P}^{\perp}_{\mathbf{C}^{-1/2}(\mathbf{s}-b\mathbf{m})}\tilde{\mathbf{x}} \overset{d}{=} \chi^2_{N-1}(0) \end{aligned}$$

with $\delta = \alpha^2(1-\alpha b)^{-2}\sigma^{-2}(\mathbf{s}-b\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{s}-b\mathbf{m})$. Therefore, t follows a beta distribution, which is central under H_0 and non-central under H_1. Under H_0 the distribution of t does not depend on σ^2 and the test has thus the constant false alarm rate.

Let us now consider the case $\gamma \neq 1$ for which

$$\mathbf{x} \sim \mathcal{N}(\alpha\mathbf{s} + \eta\mathbf{m}, \tau^2\mathbf{C})$$

where $\eta = \beta\gamma$ and $\tau = \beta\sigma$. For any vector $\mathbf{u}$ one can write

$$\begin{aligned} &\mathbf{C}^{-1/2}(\mathbf{u} - \frac{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{u}}{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{m}}\mathbf{m}) \\ &= \mathbf{C}^{-1/2}\mathbf{u} - \frac{\mathbf{C}^{-1/2}\mathbf{m}\mathbf{m}^T\mathbf{C}^{-1/2}}{\mathbf{m}^T\mathbf{C}^{-1}\mathbf{m}}\mathbf{C}^{-1/2}\mathbf{u} \\ &= (\mathbf{I} - \mathbf{P}_{\mathbf{C}^{-1/2}\mathbf{m}})\mathbf{C}^{-1/2}\mathbf{u} \\ &= \mathbf{P}^{\perp}_{\mathbf{C}^{-1/2}\mathbf{m}}\mathbf{C}^{-1/2}\mathbf{u} \end{aligned}$$

Therefore the test statistic is now

$$\begin{aligned} t &= \frac{[\mathbf{x}^T\mathbf{C}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{C}^{-1/2}\mathbf{m}}\mathbf{C}^{-1/2}\mathbf{s}]^2}{[\mathbf{x}^T\mathbf{C}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{C}^{-1/2}\mathbf{m}}\mathbf{C}^{-1/2}\mathbf{x}][\mathbf{s}^T\mathbf{C}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{C}^{-1/2}\mathbf{m}}\mathbf{C}^{-1/2}\mathbf{s}]} \\ &= \frac{[\mathbf{x}^T\mathbf{C}^{-1/2}\mathbf{V}\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{s}]^2}{[\mathbf{x}^T\mathbf{C}^{-1/2}\mathbf{V}\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{x}][\mathbf{s}^T\mathbf{C}^{-1/2}\mathbf{V}\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{s}]} \\ &= \frac{[\mathbf{x}^T\mathbf{C}^{-1/2}\mathbf{V}\mathbf{P}_{\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{s}}\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{x}]}{[\mathbf{x}^T\mathbf{C}^{-1/2}\mathbf{V}\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{x}]} \end{aligned}$$

where $\mathbf{V}$ is an orthogonal basis for the subspace orthogonal to $\mathbf{C}^{-1/2}\mathbf{m}$. Next note that

$$\breve{\mathbf{x}} = \tau^{-1}\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{x} \sim \mathcal{N}(\alpha\tau^{-1}\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{s}, \mathbf{I})$$

It follows that $t \overset{d}{=} C/(C + D)$ where

$$\begin{aligned} C &= \breve{\mathbf{x}}^T\mathbf{P}_{\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{s}}\breve{\mathbf{x}} \overset{d}{=} \chi^2_1(\delta) \\ D &= \breve{\mathbf{x}}^T\mathbf{P}^{\perp}_{\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{s}}\breve{\mathbf{x}} \overset{d}{=} \chi^2_{N-2}(0) \end{aligned}$$

with $\delta = \alpha^2\tau^{-2}\mathbf{s}^T\mathbf{C}^{-1/2}\mathbf{V}\mathbf{P}_{\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{s}}\mathbf{V}^T\mathbf{C}^{-1/2}\mathbf{s}$. Consequently the test statistic follows a beta distribution and under H_0 this distribution does not depend on γ (the scaling factor on the mean) and σ^2 (the scaling factor on the covariance matrix).