

# The basic equation for target detection in remote sensing

Xiurui Geng, Luyan Ji and Yongchao Zhao

**Abstract**—Our research has revealed a hidden relationship among several basic components, which leads to the best target detection result. Further, we have proved that the matched filter (MF) is always superior to the constrained energy minimization (CEM) operator, both of which were originally of parallel importance in the field of target detection for remotely sensed image.

**Index Terms**—Target detection, basic equation, constrained energy minimization, matched filter, remote sensing.

## I. INTRODUCTION

TARGET detection is an important research area in the applications of remote sensing, which aims to find an effective detector that can enhance a signal output while suppressing the background. The design of the target detection algorithm must be based on a sufficient understanding of the data. Different criteria will lead to different types of target detection algorithm. Energy and maximum likelihood criteria are the most commonly used methods for target detection, and they are represented by constrained energy minimization (CEM) [1], [2], [3], [4], [5], [6] and match filter (MF) [7], [8], [9]. CEM was originally derived from a linearly constrained minimum variance adaptive beam-forming in the field of signal processing. It keeps the output of the CEM detector for the target as a constant while suppressing output energy in the background to a minimum level. CEM is widely used for target detection[2], [10], [11], [12] and has been embedded in the most frequently used remote sensing software, (the Environment for Visualizing Images (ENVI)) since Version 4.6. On the other hand, the MF detector, (the minimum variance distortionless response (MVDR) [13], [14], [15]) in array processing, is the most commonly utilized technique for various kinds of applications[16], [17], [18], [19], and has been included in the EVNI software in a very early version. MF is at optimal performance in the Neyman-Pearson sense only when the target and background classes follow multivariate normal distributions with the same covariance matrix, which is an unlikely situation for real-world images.

CEM and MF detectors are designed based on entirely different theoretical foundations. It is interesting to find that

X. Geng and Y. Zhao are with the Key Laboratory of Technology in Geo-Spatial information Processing and Application System, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China; School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, 100049, China  
E-mail: gengxr@sina.com

L. Ji is with the Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University.

Manuscript received ??, ??; revised ??, ??.

their mathematical forms are very similar, except that the MF detector needs the data to be centralized first. The result of CEM can be considered as the inner product of a detector vector and all pixel vectors. Since the inner product is sensitive to the origin position, the performance of CEM will vary with different origins. Yet, as is known, the data distribution is independent of the selection of origin. From this angle, the CEM is deficient. MF considers the mean vector as origin, which “accidentally” avoids the influence of the origin position on the target detection performance. Yet, is the mean vector the optimal data origin? Do better origins exist other than zero and mean vector? Clever eye (CE) [20] provides a strategy to find the best origins, which minimizes the output energy of the background while maintaining a constant response to the signature of interest. Since this is also based on the energy criterion, it can be considered as an optimized version of CEM.

In this paper, we prove that in order to achieve the optimal target detection performance, the optimal origins must satisfy a basic equation. This basic equation contains four components in target detection: target vector, mean vector, covariance matrix and origin. Based on this equation, we further prove that MF is always superior to CEM. That is to say, of the two benchmark target detection methods, CEM can now be considered obsolete.

## II. BACKGROUND

In this section, we briefly introduce the formation of the CEM, MF and CE detectors, and then present the main problem which exists in CE.

### A. CEM

CEM was proposed by Harsanyi in 1993, which exists originally derived from the linearly constrained minimized variance adaptive beam-forming in the field of digital signal processing. It uses a finite impulse response (FIR) filter to constrain the desired signature by a specific gain while minimizing the filter output energy [1].

Assume that the observed data matrix is given by  $\mathbf{X} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N]$ , where  $\mathbf{r}_i = [r_{i1}, r_{i2}, \dots, r_{iL}]^T$  for  $1 \leq i \leq N$  is a sample pixel vector,  $N$  is the total number of pixels, and  $L$  is the number of bands. Suppose that the desired signature  $\mathbf{d}$  is also known. The objective of CEM is to design an FIR linear filter  $\mathbf{w} = [w_1, w_2, \dots, w_L]^T$  to minimize the filter output energy subject to the constraint,  $\mathbf{d}^T \mathbf{w} = \sum_{l=1}^L d_l w_l = 1$ . Then the problem yields

$$\begin{cases} \min_{\mathbf{w}} \frac{1}{N} \left( \sum_{i=1}^N y_i^2 \right) = \min_{\mathbf{w}} \mathbf{w}^T \mathbf{R} \mathbf{w}, \\ \mathbf{d}^T \mathbf{w} = 1 \end{cases}, \quad (1)$$

where  $y_i = \mathbf{w}^T \mathbf{r}_i$  is the output of CEM for the pixel  $\mathbf{r}_i$ , and  $\mathbf{R} = (\sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^T)/N$ , which is firstly referred to as sample correlation matrix by Harsanyi [1], [2], and later renamed autocorrelation matrix by some other researchers [5], [21], [4], [3]. In this paper, we will adopt Harsanyi's nomination. The solution to this constrained minimization problem (1) is the CEM operator,  $\mathbf{w}_{CEM}$ , which is given by [1]

$$\mathbf{w}_{CEM} = \frac{\mathbf{R}^{-1} \mathbf{d}}{\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}}. \quad (2)$$

### B. MF

Though MF stems from the binary hypothesis test, it can also be explained from the perspective of the filter output energy. That is, the MF detector can be considered as the optimal solution of target detection when the data origin is positioned at the mean vector. The normalized expression of an MF detector can be written as [7], [22]

$$\mathbf{w}_{MF} = c_{MF} \mathbf{K}^{-1} (\mathbf{d} - \mathbf{m}) = \frac{\mathbf{K}^{-1} (\mathbf{d} - \mathbf{m})}{(\mathbf{d} - \mathbf{m})^T \mathbf{K}^{-1} (\mathbf{d} - \mathbf{m})} \quad (3)$$

where  $\mathbf{m} = (\sum_{i=1}^N \mathbf{r}_i)/N$  is the mean vector,  $\mathbf{K} = [\sum_{i=1}^N (\mathbf{r}_i - \mathbf{m})(\mathbf{r}_i - \mathbf{m})^T]/N$  is the covariance matrix,  $c_{MF} = 1/[(\mathbf{d} - \mathbf{m})^T \mathbf{K}^{-1} (\mathbf{d} - \mathbf{m})]$  is a scalar.

### C. CE

CEM and MF have the same expression except that MF first requires data centralization. However, this difference leads to a difference in the target detection results. That is to say, the selection of the data origin could directly affect the performance of the detector. CE introduces the data origin as a new variable in the objective function, and the corresponding optimization model becomes

$$\begin{cases} \min_{\mathbf{w}, \boldsymbol{\mu}} \mathbf{w}^T \mathbf{R}_{\boldsymbol{\mu}} \mathbf{w} \\ (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{w} = 1 \end{cases}, \quad (4)$$

where

$$\mathbf{R}_{\boldsymbol{\mu}} = \frac{1}{N} \left[ \sum_{i=1}^N (\mathbf{r}_i - \boldsymbol{\mu})(\mathbf{r}_i - \boldsymbol{\mu})^T \right] = \mathbf{K} + (\mathbf{m} - \boldsymbol{\mu})(\mathbf{m} - \boldsymbol{\mu})^T. \quad (5)$$

Apparently, when  $\mathbf{m} = \boldsymbol{\mu}$ ,  $\mathbf{R}_{\boldsymbol{\mu}}$  is equal to the covariance matrix  $\mathbf{K}$ . And when  $\mathbf{m} = \mathbf{0}$ ,  $\mathbf{R}_{\boldsymbol{\mu}}$  is equal to the sample correlation matrix  $\mathbf{R}$ .

Then, similar to CEM, the corresponding detector can be calculated by firstly fixing  $\boldsymbol{\mu}$  as follows:

$$\mathbf{w}_{\boldsymbol{\mu}} = \frac{\mathbf{R}_{\boldsymbol{\mu}}^{-1} (\mathbf{d} - \boldsymbol{\mu})}{(\mathbf{d} - \boldsymbol{\mu})^T \mathbf{R}_{\boldsymbol{\mu}}^{-1} (\mathbf{d} - \boldsymbol{\mu})}. \quad (6)$$

From (6) we can see that a different  $\boldsymbol{\mu}$  can lead to a different detector. It is interesting to find that when  $\boldsymbol{\mu} = \mathbf{m}$ ,  $\mathbf{w}_{\boldsymbol{\mu}} = \mathbf{w}_{MF}$ ,

and when  $\boldsymbol{\mu} = \mathbf{0}$ ,  $\mathbf{w}_{\boldsymbol{\mu}} = \mathbf{w}_{CEM}$ . However, both of these may not be the optimal solution for (4).

Substitute (6) into the cost function of (4), and the average filter output energy of CE becomes

$$f(\boldsymbol{\mu}) = \frac{1}{(\mathbf{d} - \boldsymbol{\mu})^T \mathbf{R}_{\boldsymbol{\mu}}^{-1} (\mathbf{d} - \boldsymbol{\mu})}. \quad (7)$$

Note that the numerator of (7) is a constant, so minimizing  $f(\boldsymbol{\mu})$  is equivalent to maximizing the denominator. We denote the denominator as  $g(\boldsymbol{\mu})$ , and then the original optimization problem (4) can be transformed into [20]

$$\max_{\boldsymbol{\mu}} g(\boldsymbol{\mu}) = (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{R}_{\boldsymbol{\mu}}^{-1} (\mathbf{d} - \boldsymbol{\mu}). \quad (8)$$

It should be noted here that  $g(\boldsymbol{\mu})$  corresponds to the reciprocal of the average filter output energy. Since the filter output of the target is fixed to 1, the minimization on the average output energy is equivalent to the minimization on the background output energy. As a result, the larger the  $g(\boldsymbol{\mu})$ , the smaller the background output energy, and thus the better the detector performance [20]. Table I tabulates the expressions of the detectors, the filter output of the target, and the corresponding output energy for the three target detectors. It is easy to find their similarity in the mathematical expression. And the only difference is the selection of the data origin. It should be noted here that, when we describe a detector as 'best/optimal' or 'superior' in this study, it means that the filter output energy of the detector is minimum or smaller.

According to (8),  $g(\boldsymbol{\mu})$  varies with the origin position  $\boldsymbol{\mu}$ , which indicates the target detection result is influenced by the selection of origin. In Ref [20], the optimal origin  $\boldsymbol{\mu}^*$  is solved by applying the gradient ascent method to (8). Then the CE detector,  $\mathbf{w}_{\boldsymbol{\mu}^*}$  is calculated accordingly. The derivative of  $g(\boldsymbol{\mu})$  with respect to  $\boldsymbol{\mu}$  is expressed as

$$\begin{aligned} g'(\boldsymbol{\mu}) &= -2\mathbf{K}^{-1} (\mathbf{d} - \boldsymbol{\mu}) - \\ &\quad 2(\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{m} - \boldsymbol{\mu}) \left( 1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{m} - \boldsymbol{\mu}) \right) \mathbf{K}^{-1} (-\mathbf{m} + 2\boldsymbol{\mu} - \mathbf{d}) \\ &\quad \left( 1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{m} - \boldsymbol{\mu}) \right)^2 \\ &\quad + \frac{2 \left( (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K} (\mathbf{m} - \boldsymbol{\mu}) \right)^2 \mathbf{K}^{-1} (\mathbf{m} - \boldsymbol{\mu})}{\left( 1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{m} - \boldsymbol{\mu}) \right)^2}. \end{aligned} \quad (9)$$

However,  $g(\boldsymbol{\mu})$  is not a convex function, so the local extreme problem seems inevitable for CE. One possible way to determine the global maximum of  $g(\boldsymbol{\mu})$  is by finding out all local maxima, which is apparently an impossible task in practice.

## III. BASIC EQUATION FOR TARGET DETECTION

### A. Analytical solution of CE

The CEM and MF detectors have a similar expression in mathematics, but their theoretical foundations are completely different. As a result, it is hard to theoretically compare their advantages and disadvantages. The CE algorithm can be regarded as the bridge between CEM and MF.

TABLE I

THE DETECTOR, OUTPUT OF TARGET AND THE AVERAGE OUTPUT ENERGY OF CEM, MF AND CE

Method	Detector	Output of target	Average output energy
MF	$\frac{\mathbf{K}^{-1}(\mathbf{d}-\mathbf{m})}{(\mathbf{d}-\mathbf{m})^T \mathbf{K}^{-1}(\mathbf{d}-\mathbf{m})}$	1	$\frac{1}{(\mathbf{d}-\mathbf{m})^T \mathbf{K}^{-1}(\mathbf{d}-\mathbf{m})}$
CEM	$\frac{\mathbf{R}^{-1}\mathbf{d}}{\mathbf{d}^T \mathbf{R}^{-1}\mathbf{d}}$	1	$\frac{1}{\mathbf{d}^T \mathbf{R}^{-1}\mathbf{d}}$
CE	$\frac{\mathbf{R}_{\mu}^{-1}(\mathbf{d}-\mu)}{(\mathbf{d}^T-\mu)\mathbf{R}_{\mu}^{-1}(\mathbf{d}-\mu)}$	1	$\frac{1}{(\mathbf{d}^T-\mu)\mathbf{R}_{\mu}^{-1}(\mathbf{d}-\mu)}$

Fortunately, we find that all local maximal points of  $g(\mu)$  are mathematically equivalent to the solution of a system of linear equations, which can be stated as the following theorem.

**Theorem 1.** Assume that the target vector  $\mathbf{d}$  is known. To achieve the optimal target detector from the perspective of output energy of the detector, the target vector,  $\mathbf{d}$ , data mean vector,  $\mathbf{m}$ , the data covariance matrix,  $\mathbf{K}$  and origin,  $\mu$  must satisfy

$$(\mathbf{d} - \mathbf{m})^T \mathbf{K}^{-1} (\mathbf{m} - \mu) = 1. \quad (10)$$

See Appendix A for details about the proof. (10) is an interesting equation. The general solution set of  $\mu$  in (10) corresponds to all the local maxima of (8). Therefore, the optimal origin,  $\mu^*$  can be directly determined by the solution of (10), instead of applying the gradient ascent method as in Ref [20]. (10) has an infinite number of solutions, with the dimensionality of a solution set equal to the number of bands minus 1. By inspection, it can be found that (10) includes the basic elements in target detection: target, mean vector, covariance matrix and origin. As a result, (10) is named the basic equation of target detection, and also the clever eye (CE) equation.

### B. Equivalence between CE and MF

Theorem 1 indicates that the optimal origins can be acquired by solving (10). Based on this fact, we further find that though (10) has an infinite number of solutions, they all correspond to the same target detector, which is equivalent to the MF detector. We are now about to establish the equivalence between CE and MF detector.

**Theorem 2.** The CE detector is equivalent to the MF detector. That is, for any solution of (10),  $\mu^*$ , there exists a constant  $c$  such that

$$\mathbf{R}_{\mu^*}^{-1} (\mathbf{d} - \mu^*) = c \mathbf{K}^{-1} (\mathbf{d} - \mathbf{m}) \quad (11)$$

The proof of Theorem 2 is demonstrated in Appendix B. This theorem indicates that all local extrema of (8) lead to the same detector. Therefore, the local extremum problem of (8) does not exist! It is a very interesting conclusion. Since CE is an optimized version of CEM from the perspective of energy, this equivalence between CE and MF detectors indirectly shows that MF is always superior to CEM! That is to say, as one of the two classical algorithms in the field of remote sensing and signal processing, CEM can now be considered redundant.

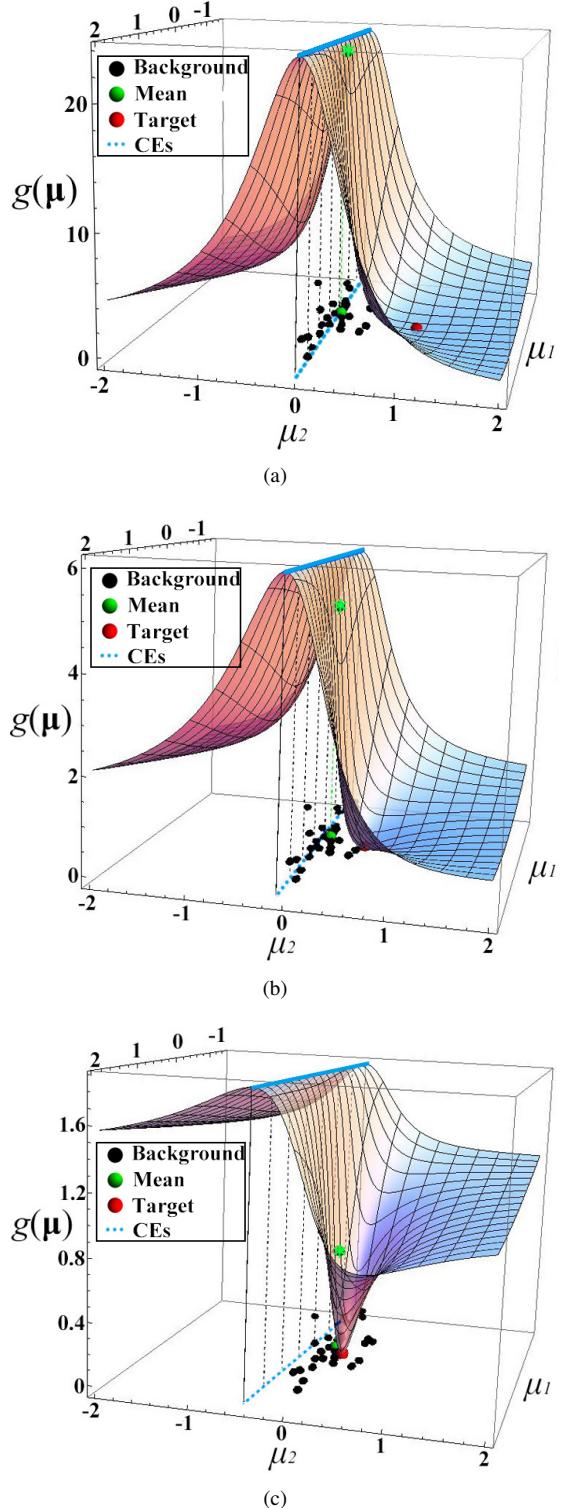


Fig. 1. Influence of data origin on the performance of target detection. 30 background points (marked in black), one target point (marked in red) and the mean vector (marked in green) are distributed on the bottom plane, and the CE line (marked in blue and dotted) are also plotted on the bottom plane. The three sub-figures represent the cases where the target point is far away from (a), close to (b) and within (c) the background points. When  $\mathbf{d}$  is far from the background points,  $g(\mu)$  has a larger value with a maximum larger than 20, indicating that the separability between target and background is the best. When  $\mathbf{d}$  is close to background points, the value of  $g(\mu)$  declines compared with that in (a) (with the maximum value less than 6). Accordingly, the separability between target and background is reduced. The last case is that the target point lies within the background, which makes it difficult to differentiate the target from the background points. In this case, the value of  $g(\mu)$  is the lowest with the maximum value less than 2. That is to say, the separability between target and background is the worst. In all, wherever  $\mathbf{d}$  is,  $g(\mu)$  can always reach the maxima at the solution of the CE equation.

There is an elegant geometry behind (8) and (10). Thus, in the following, we will further illustrate the relationship between the objective function in (8) and the CE equation in (10) from a geometrical view. First of all, 30 2-dimensional random points, which are scattered in the bottom of Fig. 1 and marked in black, are generated as background by the "randn" function in Matlab. Then, an extra point is manually set as the target and marked in red. In addition, the surface of  $g(\mu)$  as the function of  $\mu$  is depicted in Fig. 1. Seemingly, the data origin can be placed in any position on the plane, where the target and background points are located. Yet, different data origins indicate different filter output energies, and thereby different target detection performances. From (10), only those data origins, which satisfy the CE equation, can make the filter output energy reach the minimum (or make  $g(\mu)$  maximum), and they are located in the blue dotted straight line in the bottom. On the other hand,  $g(\mu)$  reaches the minimum point at  $\mu = d$  with  $g(d) = 0$ .

From Fig. 1 one can notice that the shape of surface  $g(\mu)$  varies with the target  $d$ . However, the function always reaches the maxima at the solution of the CE equation. Another fact should be pointed out here that the mean vector (green point) is not corresponding to the solution of the CE equation. Yet amazingly, MF has the same target detection performance as CE.

#### IV. EXPERIMENTS

For ease of understanding, we compared the performance of the three algorithms (CEM, MF, CE) using synthetic and real data. Two metrics, the output energy and correlation, together with the receiver operating characteristic (ROC) curve are adopted in the following experiments.

##### A. Experiments with synthetic data

A synthetic image is designed as follows: 1) generate a  $50 \times 50 \times 3$  background image following a 3-dimension normal distribution; 2) add a normally distributed  $5 \times 5 \times 3$  target image in the middle of the background image. The distribution of the simulated data set in the three dimensional feature space is shown in Fig. 2(a). The data cloud of the background is relatively flat with the target points positioned above.

The true image is shown in Fig. 2(b) and the target detection outputs of CEM, MF and CE are given in Figs. 2(c) to 2(e). Visually, the performance of CEM is worst, since the contrast between the target and background of CEM is lowest. Also the output energy of CEM is the largest, as listed in Table II. The reason that CE outperforms CEM can be attributed to the fact that CE can search the best data origin for target detection. As revealed by the basic equation in (10), all the CE points are located on a plane, as shown in Fig. 2(a), and they always lead to the same detector,  $w_{CE}$ , which has the same direction as the MF detector,  $w_{MF}$ . However, CEM has a quite different detection direction. Clearly, a better separability between target and background can be achieved along the direction derived from CE and MF.

Yet, it is much harder to compare the results of MF and CE visually, although CE can achieve a lower output energy.

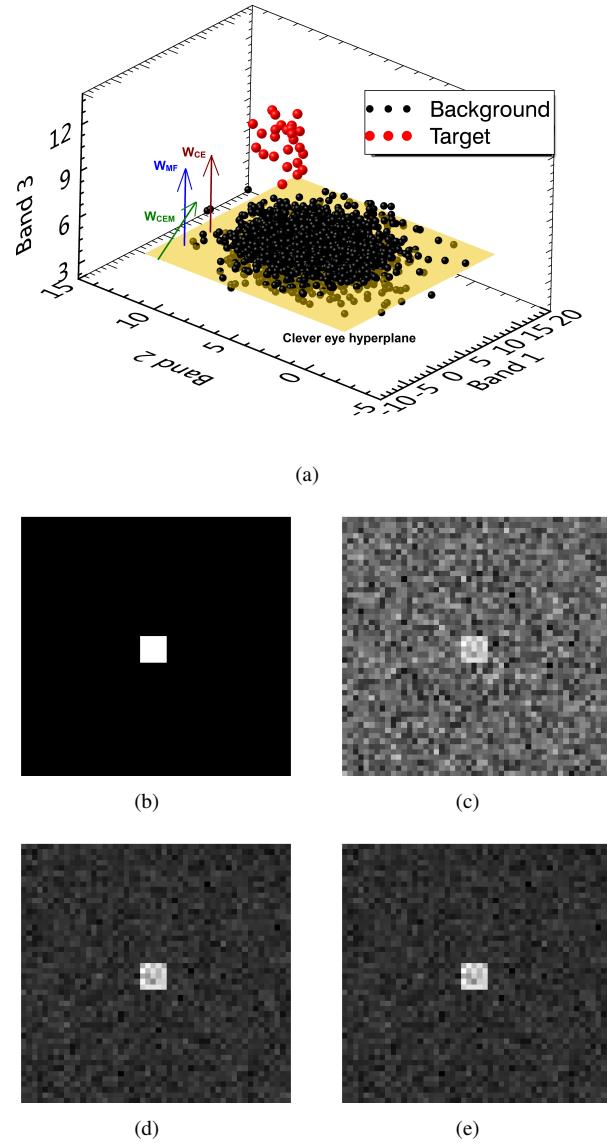


Fig. 2. Distribution of the synthetic image in 2-dimensional feature space (a), true image (b) and the target detection results of CEM (c), MF (d) and CE (e).

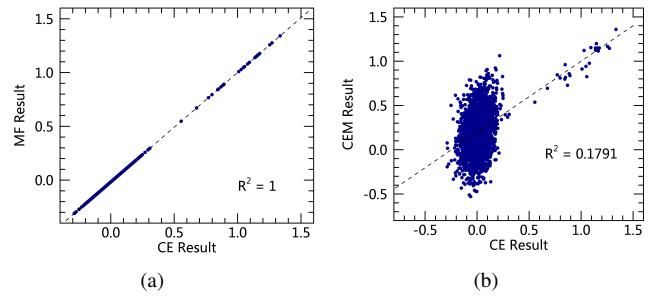


Fig. 3. Comparison of the target detection results of MF and CE (a), and CEM and ranCE(b).

Fig. 3(a) plots the comparison of the output results of MF and CE. It can be seen that their outputs are completely linearly dependent ( $R^2 = 1$ )! This result directly indicates

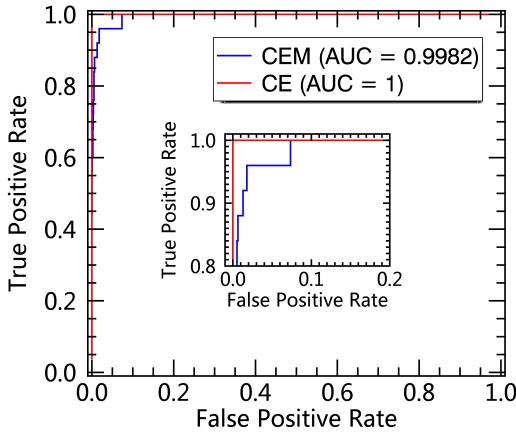


Fig. 4. Comparison of the ROC curves for CEM and CE detectors using synthetic data.

the equivalence between CE and MF, which has been proved in Theorem 2. On the other hand, the correlation coefficient between the CE and CEM output is very low ( $R^2 = 0.1791$ ) as shown in Fig. 3(b), which infers that changing the data origin can greatly alter the CEM output. The ROC curves of CEM and CE plotted in Fig. 4 show an advantage of CE over CEM in the target detection of this data set.

TABLE II  
OUTPUT ENERGY OF CEM, MF AND CE FOR SYNTHETIC AND REAL DATA.

Method	CEM	MF	CE
Synthetic data	0.1048	0.0173	0.01701
Real data	9.1895e-04	9.1303e-04	9.1220e-04

### B. Experiments with real data

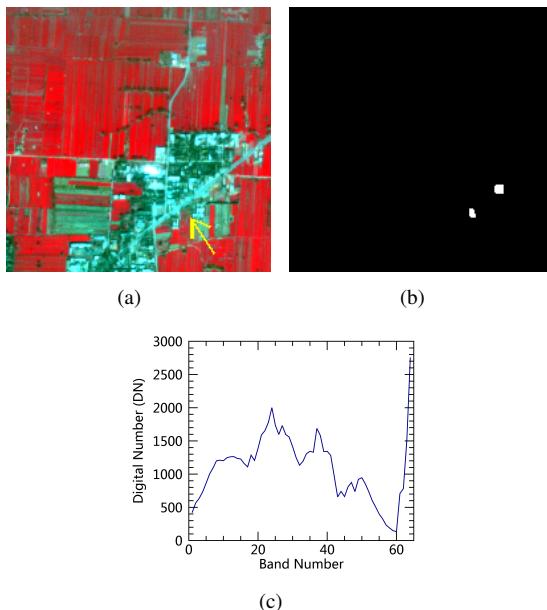


Fig. 5. False color image (R:857.9nm, G:669.6nm, B:559.6nm) of the real data (a), the ground truth image (b) and spectral signature of the target (c).

In this section, we will demonstrate the equivalence between CE and MF by using a real hyperspectral image of  $200 \times 200$  pixels, as shown in Fig. 5(a). It was collected by the Operational Modular Imaging Spectrometer-II, which is a hyperspectral imaging system developed by the Shanghai Institute of Technical Physics, Chinese Academy of Sciences. The data were acquired in the city of Xian, China in 2003, and are composed of 64 bands from 460nm to 10250 nm at a spatial resolution of 3.6m. A small man-made target, marked by an arrow in the scene (Fig. 5(a)) is selected as the target of interest, and the spectral signature is shown in Fig. 5(c). The manually determined ground truth map based on field survey is shown in Fig. 5(b).

The target detection results of the three methods are quite similar and no significant differences can be observed visually (Fig. 6). We present the correlation curves between the outputs of the three detectors in Figs. 7(a) and 7(b), and it can be seen that the MF output is completely linearly correlated to the CE output ( $R^2 = 1$ ). Again, this result reveals the equivalence between CE and MF.

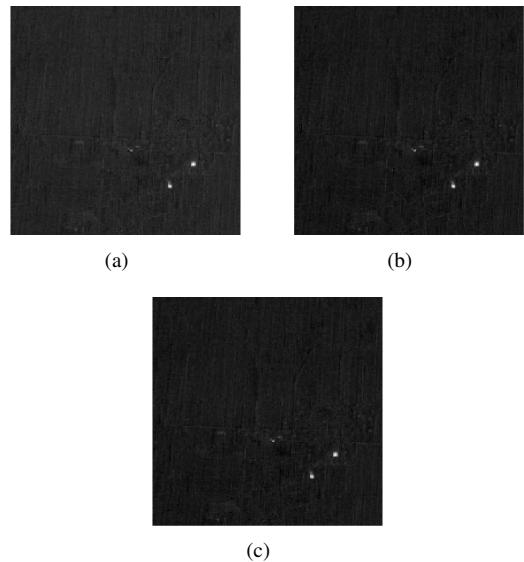


Fig. 6. Target detection results of CEM (a), MF (b) and CE (c).

The target results by CEM and CE are also very similar with a correlation as high as  $R^2 = 0.9926$ . The reason for this can be explained as follows. According to Ref [20], the advantage of MF over CEM decreases with an increase in the number of the bands involved. Since the number of bands for hyperspectral data is usually very high, CEM can derive very similar result as MF and ACEM. That is why CEM has been widely applied for hyperspectral target detection and rarely questioned. However, the output energy of CEM is still higher than that of both MF and CE as shown in Table II. Moreover, according to the ROC curves in Fig. 8, CE still surpasses CEM, though they both have very good detection performance.

### V. CONCLUSION

MF is the best target detector from the perspective of maximum likelihood, while CEM is the representative one

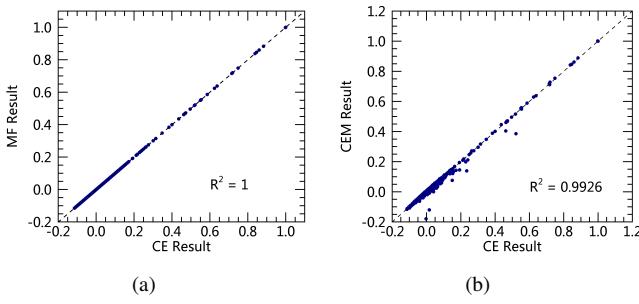


Fig. 7. Comparison of the target detection results of MF and CE (a), and comparison result between CEM and CE (b) using real data.

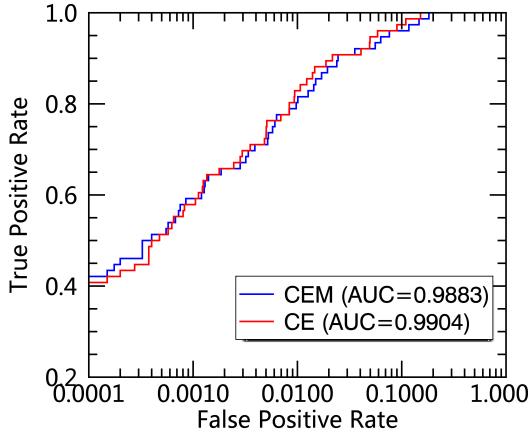


Fig. 8. Comparison of the ROC curves for CEM and CE detectors using real data.

from the perspective of energy. Usually, it is difficult to compare algorithms developed from different criteria, so MF and CEM are considered as two benchmark methods in target detection and both are embedded in the ENVI software, which is one of the most frequently used software packages in the remote sensing community. When the origin is introduced as a variable, it is found that there exists an inherent relationship between the two criteria [20]. Specifically, we have proved that the basic elements in target detection (including target vector, mean vector, covariance matrix and origin) should follow a fundamental equation to acquire the best detector from the angle of energy. In addition, we prove the equivalence between CE and MF, which indirectly demonstrates that MF is always superior to CEM. Thus, we suggest that the classical target detection CEM should be considered redundant.

On the other hand, the MF detector is not optimal from the perspective of output energy. Yet it is equivalent to the CE detector, which indicates that the energy criterion is not a perfect one. Therefore, we will focus on searching for a more reasonable criterion for target detection in future studies.

#### APPENDIX A PROOF OF THEOREM 1

**Theorem 1.** Assume that the target vector  $\mathbf{d}$  is known. To achieve the optimal target detector from the perspective of

output energy, the target vector,  $\mathbf{d}$ , data mean vector,  $\mathbf{m}$ , the data covariance matrix,  $\mathbf{K}$  and origin,  $\boldsymbol{\mu}$  must satisfy

$$(\mathbf{d} - \mathbf{m})^T \mathbf{K}^{-1} (\mathbf{m} - \boldsymbol{\mu}) = 1 \quad (\text{A.1})$$

*Proof.* To prove the theorem we first take the derivative of  $g(\boldsymbol{\mu})$  with respect to  $\boldsymbol{\mu}$ ,

$$\begin{aligned} g'(\boldsymbol{\mu}) &= -2\mathbf{K}^{-1}(\mathbf{d} - \boldsymbol{\mu}) - \\ &\frac{2(\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu})(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))\mathbf{K}^{-1}(-\mathbf{m} + 2\boldsymbol{\mu} - \mathbf{d})}{(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))^2} \\ &+ \frac{2((\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}(\mathbf{m} - \boldsymbol{\mu}))^2 \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu})}{(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))^2}. \end{aligned} \quad (\text{A.2})$$

After a little algebra, we can have

$$\begin{aligned} g'(\boldsymbol{\mu}) &= \frac{2(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}) - (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))}{(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))^2} \\ &\cdot \left[ -(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))\mathbf{K}^{-1}\mathbf{b} \right. \\ &+ (1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}) - (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))\mathbf{K}^{-1}\boldsymbol{\mu} \\ &\left. + (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu})\mathbf{K}^{-1}\mathbf{m} \right]. \end{aligned} \quad (\text{A.3})$$

$g(\boldsymbol{\mu})$  reaches its extremum when  $g(\boldsymbol{\mu})' = \mathbf{0}$ . To see this, let

$$1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}) = (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}), \quad (\text{A.4})$$

or

$$\begin{aligned} &(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}) - (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))\mathbf{K}^{-1}\boldsymbol{\mu} \\ &= (1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))\mathbf{K}^{-1}\mathbf{b} - (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu})\mathbf{K}^{-1}\mathbf{m}. \end{aligned} \quad (\text{A.5})$$

If (A.5) holds, then

$$\begin{aligned} \boldsymbol{\mu} &= \frac{(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))}{(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}) - (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))}\mathbf{d} \\ &- \frac{(\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu})}{(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}) - (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))}\mathbf{m} \end{aligned} \quad (\text{A.6})$$

It is easy to observe that the sum of the two coefficients of  $\mathbf{d}$  and  $\mathbf{m}$  is equal to 1. For simplification, set

$$\alpha = \frac{(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))}{(1 + (\mathbf{m} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}) - (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{m} - \boldsymbol{\mu}))}.$$

Then, (A.6) can be rewritten as

$$\boldsymbol{\mu} = \alpha\mathbf{d} + (1 - \alpha)\mathbf{m}. \quad (\text{A.7})$$

It follows that

$$(\mathbf{m} - \boldsymbol{\mu}) = \alpha(\mathbf{m} - \mathbf{d}), \quad (\text{A.8})$$

$$(\mathbf{d} - \boldsymbol{\mu}) = (1 - \alpha)(\mathbf{d} - \mathbf{m}). \quad (\text{A.9})$$

Substitute (A.7), (A.8) and (A.9) into (A.6), then we can derive  $\mu = d$ . In this case, the objective function  $g(\mu) = 0$  and reaches the minimum. Yet, in this study, we only focus on the maximum point, so  $\mu$  must satisfy (A.4). By simplification, (A.4) becomes

$$(d - m)^T K^{-1} (m - \mu) = 1 \quad (\text{A.10})$$

which is the clever eye equation and therefore the theorem holds.  $\square$

## APPENDIX B PROOF OF THEOREM 2

**Theorem 2.** *The CE detector is equivalent to the MF detector. That is, for any solution of (10),  $\mu^*$ , there exists a constant c such that*

$$R_{\mu^*}^{-1} (d - \mu^*) = cK^{-1} (d - m) \quad (\text{B.1})$$

*Proof.* Using Sherman–Morrison formula [23], the inverse matrix of  $R_\mu$  can be calculated by

$$R_\mu^{-1} = (K + (m - \mu)(m - \mu)^T)^{-1} = K^{-1} - \frac{K^{-1}(m - \mu)(m - \mu)^T K^{-1}}{1 + (m - \mu)^T K^{-1} (m - \mu)}. \quad (\text{B.2})$$

Substitute (B.2) into (B.1), then we have

$$\left( K^{-1} - \frac{K^{-1}(m - \mu)(m - \mu)^T K^{-1}}{1 + (m - \mu)^T K^{-1} (m - \mu)} \right) (d - \mu) = cK^{-1} (d - m), \quad (\text{B.3})$$

and so

$$K^{-1} (d - \mu) = \frac{K^{-1}(m - \mu)(m - \mu)^T K^{-1}}{1 + (m - \mu)^T K^{-1} (m - \mu)} K^{-1} (d - \mu) + cK^{-1} (d - m) \quad (\text{B.4})$$

Apparently, to prove the theorem is equivalent to prove (B.4). If  $\mu$  satisfies  $g'(\mu) = 0$ , then according to (A.3) we have

$$\begin{aligned} & K^{-1} (d - \mu) = \\ & \frac{(d - \mu)^T K^{-1} (m - \mu) (1 + (m - \mu)^T K^{-1} (m - \mu)) K^{-1} (-m + 2\mu - d)}{\left(1 + (m - \mu)^T K^{-1} (m - \mu)\right)^2} \\ & + \frac{\left((d - \mu)^T K^{-1} (m - \mu)\right)^2 K^{-1} (m - \mu)}{\left(1 + (m - \mu)^T K^{-1} (m - \mu)\right)^2}, \end{aligned} \quad (\text{B.5})$$

which can be further turned into

$$\begin{aligned} & K^{-1} (d - \mu) = \frac{K^{-1}(m - \mu)(m - \mu)^T K^{-1}}{1 + (m - \mu)^T K^{-1} (m - \mu)} K^{-1} (d - \mu) \\ & + \frac{(d - \mu)^T K^{-1} (m - \mu) K^{-1} (d - \mu)}{1 + (m - \mu)^T K^{-1} (m - \mu)} \\ & - \frac{\left((d - \mu)^T K^{-1} (m - \mu)\right)^2 K^{-1} (m - \mu)}{\left(1 + (m - \mu)^T K^{-1} (m - \mu)\right)^2}, \end{aligned} \quad (\text{B.6})$$

Considering (B.4) and (B.7), to prove the theorem we need to verify the equivalence of the right sides of both equations as follows,

$$\begin{aligned} & \frac{(d - \mu)^T K^{-1} (m - \mu) K^{-1} (d - \mu)}{1 + (m - \mu)^T K^{-1} (m - \mu)} \\ & - \frac{\left((d - \mu)^T K^{-1} (m - \mu)\right)^2 K^{-1} (m - \mu)}{\left(1 + (m - \mu)^T K^{-1} (m - \mu)\right)^2} \\ & = cK^{-1} (d - m) \end{aligned} \quad (\text{B.7})$$

By theorem 1, (B.7) can be simplified as

$$K^{-1} (d - m) = cK^{-1} (d - m) \quad (\text{B.8})$$

(B.8) holds when  $c = 1$ . Thus CE is equivalent to MF.  $\square$

## REFERENCES

- [1] J. C. Harsanyi, *Detection and classification of subpixel spectral signatures in hyperspectral image sequences*. PhD thesis, University of Maryland, 1993.
- [2] W. H. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailings in the coeur d'alene river valley, idaho, through the use of a constrained energy minimization technique," *Remote Sensing of Environment*, vol. 59, no. 1, pp. 64 – 76, 1997.
- [3] C.-I. Chang and D. C. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, pp. 1144–1159, May 2000.
- [4] H. Ren, Q. Du, C.-I. Chang, and J. O. Jensen, "Comparison between constrained energy minimization based approaches for hyperspectral imagery," in *Advances in Techniques for Analysis of Remotely Sensed Data, 2003 IEEE Workshop on*, pp. 244–248, Oct 2003.
- [5] J. M. Liu, C. M. Wang, B. C. Chieu, C. Chang, H. Ren, and C. W. Yang, "Generalized constrained energy minimization approach to subpixel detection for multispectral imagery," in *Remote Sensing*, pp. 125–135, 1999.
- [6] G.-C. Lin, C.-M. Wang, W.-J. Wang, and S.-Y. Sun, "Automated classification of multispectral {MR} images using unsupervised constrained energy minimization based on fuzzy logic," *Magnetic Resonance Imaging*, vol. 28, no. 5, pp. 721 – 738, 2010.
- [7] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 79–116, 2003.
- [8] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, pp. 29–43, Jan 2002.
- [9] D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, "Robust matched filters for target detection in hyperspectral imaging data," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 1, pp. I-529–I-532, April 2007.
- [10] M. S. Alam, "Mine detection in multispectral imagery data using constrained energy minimization," *Proceedings of SPIE - The International Society for Optical Engineering*, 2008.
- [11] M. Monaco, R. Camilli, F. D'Ambrosio, G. M. Del, and A. Pantosti, "High spectral and spatial resolution hyperspectral imagery for quantifying russian wheat aphid infestation in wheat using the constrained energy minimization classifier," *Journal of Applied Remote Sensing*, vol. 8, no. 1, pp. 271–279, 2014.
- [12] G. C. Lin, W. J. Wang, and C. M. Wang, "Feature selection algorithm for classification of multispectral mr images using constrained energy minimization," in *Hybrid Intelligent Systems (HIS), 2010 10th International Conference on*, pp. 43–46, Aug 2010.
- [13] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, pp. 1408–1418, Aug 1969.
- [14] C. D. Seligson, "Comments on "high-resolution frequency-wavenumber spectrum analysis"," *Proceedings of the IEEE*, vol. 58, pp. 947–949, June 1970.
- [15] B. D. Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, pp. 4–24, April 1988.
- [16] A. P. Williams and E. H. Jr., "Estimation of leafy spurge cover from hyperspectral imagery using mixture tuned matched filtering," *Remote Sensing of Environment*, vol. 82, no. 23, pp. 446 – 456, 2002.

- [17] M. Wolfel, "Warped-twice minimum variance distortionless response spectral estimation," in *Signal Processing Conference, 2006 14th European*, pp. 1–4, Sept 2006.
- [18] M. Wolfel and J. McDonough, "Minimum variance distortionless response spectral estimation," *IEEE Signal Processing Magazine*, vol. 22, pp. 117–126, Sept 2005.
- [19] A. W. Yarbrough, M. J. Mendenhall, R. K. Martin, and S. T. Fiorino, "Hyperspectral-based adaptive matched filter detector error as a function of atmospheric water vapor estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 2029–2039, April 2014.
- [20] X. Geng, L. Ji, and K. Sun, "Clever eye algorithm for target detection of remote sensing imagery," *{ISPRS} Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 32 – 39, 2016.
- [21] C. Chang and H. Ren, "Generalized constrained energy minimization approach to subpixel detection for multispectral imagery," *Optical Engineering*, vol. 39, no. 5, pp. 1275–1281, 2000.
- [22] D. Manolakis, "Detection algorithms for hyperspectral imaging applications: a signal processing perspective," in *Advances in Techniques for Analysis of Remotely Sensed Data, 2003 IEEE Workshop on*, pp. 378–384, Oct 2003.
- [23] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to changes in the elements of a given column or given row of the original matrix," *Annals of Mathematical Statistics*, vol. 20, 1949.