

Automated Target Detection and Discrimination Using Constrained Kurtosis Maximization

Qian Du, *Senior Member, IEEE*, and Ivica Kopriva, *Senior Member, IEEE*

Abstract—Exploiting hyperspectral imagery without prior information is a challenge. Under this circumstance, unsupervised target detection becomes an anomaly detection problem. We propose an effective algorithm for target detection and discrimination based on the normalized fourth central moment named kurtosis, which can measure the flatness of a distribution. Small targets in hyperspectral imagery contribute to the tail of a distribution, thus making it heavier. The Gaussian distribution is completely determined by the first two order statistics and has zero kurtosis. Consequently, kurtosis measures the deviation of a distribution from the background and is suitable for anomaly/target detection. When imposing appropriate inequality constraints on the kurtosis to be maximized, the resulting constrained kurtosis maximization (CKM) algorithm will be able to quickly detect small targets with several projections. Compared to the widely used unconstrained kurtosis maximization algorithm, i.e., fast independent component analysis, the CKM algorithm may detect small targets with fewer projections and yield a slightly higher detection rate.

Index Terms—Constrained kurtosis maximization (CKM), hyperspectral imagery, target classification, target detection.

I. INTRODUCTION

TARGET detection is one of the major tasks in hyperspectral image analysis. With very high spectral resolution, it is possible to detect targets based on the subtle spectral features. When the spatial resolution is low or the target size is relatively small compared to the spatial resolution, we must resort to spectral-analysis-based techniques. Target detection methods can be divided into two categories: supervised and unsupervised. The former relies on available target signatures, while the latter does not require prior target information.

This research focuses on unsupervised target detection, which can be achieved by finding pixels with distinct spectral features from those in their neighborhood, i.e., anomalies [1]. An anomaly usually has a small size and occupies only several pixels. In general, anomaly detection seeks to find unknown targets from an unknown background. Several approaches have been proposed for this purpose. For instance, Reed and Yu [2] developed the well-known Reed-Xiaoli (RX) algorithm to analyze an image using the second-order statistics; Ashton [3]

proposed an adaptive Bayesian classifier; the Projection Pursuit method developed by Ifarraguerri and Chang [4] employed the information divergence to find the best projector for anomaly detection; Chiang *et al.* [5] and Chang *et al.* [6] used skewness and kurtosis for target detection; Schweizer and Moura [7] presented an anomaly detection method based on Gauss–Markov random field; Robila and Varshney [8], [9] applied independent component (IC) analysis (ICA) for target detection, which is based on the minimization of mutual information. However, these techniques are relatively time-consuming and some (e.g., the RX algorithm) can only detect anomalies but cannot discriminate them from each other. The goal of this research is to develop an efficient unsupervised algorithm for hyperspectral imagery, which not only quickly detects targets but also automatically distinguishes between them.

In this research, we employ kurtosis, the most frequently used high-order moment, for unsupervised target detection. It is known that kurtosis is the normalized fourth central moment, which can measure the flatness of a probability distribution. If an image background can be modeled as a Gaussian distribution, anomalies or small man-made targets can be viewed as outliers because their sizes are relatively small and spectral features are very different compared to their surroundings. Therefore, the corresponding pixels will contribute to the tail of the distribution and make it heavier. As a result, anomalies can be detected by searching the deviation from a Gaussian distribution, which has zero kurtosis. In other words, anomaly detection can be achieved by searching the direction of non-Gaussianity. We found that kurtosis is sensitive to the outliers and works very efficiently in small target/anomaly detection. When the background cannot be modeled perfectly as a Gaussian distribution, some background classes will be detected as well. As a preprocessing step, data need to be whitened by mean removal and decorrelation to prevent the first- and second-order statistics from interfering with the following high-order statistics-based analysis [10].

A distribution having a negative kurtosis is called “sub-Gaussian,” which is flatter than a Gaussian one; a distribution with a positive kurtosis is called “super-Gaussian,” which has a sharper peak and longer tails than a Gaussian one [10]. In general, the presence of targets in a hyperspectral image makes the distribution appear to be super-Gaussian. Thus, small targets and anomalies correspond to the directions with very large positive kurtosis. Objects with large sizes are related to the kurtosis with small values, which can even be negative. To prevent classes with large kurtosis from becoming the obstacle in the target detection process, an inequality constraint on the value of kurtosis may need to be imposed. Multiple-target detection and discrimination can be achieved by using a Gram–Schmidt orthogonalization type of process.

Manuscript received February 9, 2007; revised July 9, 2007. This work was supported in part by the National Geospatial-Intelligence Agency under Grant HM15810512006 and in part by the Ministry of Science, Education and Sport, Croatia, under Grant 098-0982903-2558.

Q. Du is with the Department of Electrical and Computer Engineering and GeoResources Institute (GRI), High Performance Computing Collaboratory, Mississippi State University, Mississippi State, MS 39762 USA.

I. Kopriva is with the Division of Laser and Atomic Research and Development, Rudjer Bošković Institute, Zagreb 10002, Croatia.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LGRS.2007.907300

It is noteworthy that the developed constrained kurtosis maximization (CKM) algorithm has a close relationship with ICA, where non-Gaussianity is a metric of independence [10]. There are two types of ICA algorithms in terms of the order of ICs generated: parallel and sequential. Most ICA algorithms use the parallel mode, where all the ICs are determined simultaneously, such as the well-known joint approximate diagonalization of eigenmatrices and Bell–Sejnowski’s InfoMax algorithms [10]. For mathematical tractability, the number of generated ICs is equal to the number of bands in most cases. In general, parallel ICA algorithms have very high computational complexity. Without dimension reduction, they cannot be directly applied to hyperspectral imagery. The commonly used dimension reduction approaches are principal component analysis (PCA) [8], [9] and noise-adjusted PCA [4]. However, these processes result in lost spectral information. Another problem of the parallel mode is that when the number of ICs (which is generally unknown) is changed, all the ICs have to be regenerated. The CKM algorithm uses the sequential mode, where ICs are determined one after the other. This is computationally more efficient because: 1) dimension reduction is not a must-do step, and 2) it is less sensitive to the choice of the number of ICs since all the generated ICs remain unchanged. It is similar to Hyvärinen’s fast ICA (FastICA) algorithm [12]. The major differences are the following: 1) the FastICA does not impose any constraint, and 2) the FastICA employs the negentropy approximation (which is reduced to kurtosis when a cubic function is selected for nonlinearity measurement) as the searching criterion. Because the CKM directly uses kurtosis as the searching criterion with inequality constraints, it is more sensitive to small targets and can capture them more quickly.

Although other algorithms exist based on unconstrained kurtosis maximization (e.g., [5] and [6]), their computational efficiency cannot compete with the FastICA algorithm because they either use the parallel mode or their convergence speed is slow. Therefore, in this letter, only the FastICA algorithm is chosen as the basis of comparison for the CKM algorithm.

II. CONSTRAINED KURTOSIS MAXIMIZATION

Assume there are N original data points $\{\mathbf{u}_n\}_{n=1}^N$ with dimensionality L and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N]$ is an $L \times N$ data matrix. Let \mathbf{w} be a projector and $\mathbf{y} = \mathbf{w}^T \mathbf{U} = (y_1, y_2, \dots, y_N)$ represents the projected data. The task is to find an optimal \mathbf{w} such that $\kappa(y)$ is maximal, and then a target can be easily detected by projecting the data onto the direction of \mathbf{w} . The sample mean is prereduced and the original data \mathbf{U} are decorrelated by a data-whitening process. The resulting whitened data set is denoted as \mathbf{Z} . With a slight abuse of notation, the transformed data set is still denoted as \mathbf{y} , i.e., $\mathbf{y} = \mathbf{w}^T \mathbf{Z} = (y_1, y_2, \dots, y_N)$. If \mathbf{w} is a unity vector, the variance of y is unity as well. Then, $\kappa(y)$ becomes

$$\kappa(y) = E[(y)^4] - 3. \quad (1)$$

A. Unconstrained Kurtosis Maximization

The aforementioned optimization problem can be formulated as the following unconstrained objective function:

$$J(\mathbf{w}) = \max_{\mathbf{w}} \{|\kappa(y)|\} = \max_{\mathbf{w}} \{|\kappa(\mathbf{w}^T \mathbf{z})|\}. \quad (2)$$

Here, $|\cdot|$ represents the absolute value operation, which ensures that the absolute maximum value is searched in both super-Gaussian and sub-Gaussian cases. Taking the derivative with respect to \mathbf{w} yields

$$\Delta \mathbf{w} = \frac{\partial |\kappa|}{\partial \mathbf{w}} = \text{sign}(y) \frac{\partial E((\mathbf{w}^T \mathbf{z})^4)}{\partial \mathbf{w}} = 4 \text{sign}(y) E(y^3 \mathbf{z}) \quad (3)$$

where $\text{sign}(\cdot)$ represents the sign function.

Gradient-descent-based adaptation is a standard procedure for the adaptation of \mathbf{w} . To speed-up the convergence, the fixed-point adaptation is adopted rather than the gradient-based adaptation [11], where the new update of \mathbf{w} is the gradient value $\Delta \mathbf{w}$ followed by \mathbf{w} normalization. Therefore, the update equations for \mathbf{w} are

$$\mathbf{w} \leftarrow \Delta \mathbf{w} \quad \mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\|. \quad (4)$$

B. Constrained Kurtosis Maximization

As mentioned earlier, the kurtosis of a super-Gaussian signal can have a very large positive value (the maximum is infinity in principle), but the negative value of the kurtosis of a sub-Gaussian signal is bounded below with the minimal possible value being -2 (when variance is normalized to unity). Thus, comparing the non-Gaussianity of super-Gaussian and sub-Gaussian signals with each other using plain kurtosis is not appropriate. When searching the direction for maximum kurtosis, it is necessary to specify if it is for super-Gaussian classes ($\kappa > 0$) or for sub-Gaussian classes ($\kappa < 0$). Therefore, an inequality constraint needs to be imposed. In many cases, it is of interest to find a target/class whose kurtosis is in a certain range. For instance, anomalies ($\kappa > \kappa_{\text{ref}}$, e.g., $\kappa_{\text{ref}} = 100$) or super-Gaussian classes with large size ($0 < \kappa < \kappa_{\text{ref}}$, e.g., $\kappa_{\text{ref}} = 10$) may need to be extracted. For a general case, the constraint can be represented as $\kappa_2 \leq \kappa \leq \kappa_1$.

Now, the new objective function is to maximize the kurtosis in (2) subject to $\kappa_2 \leq \kappa \leq \kappa_1$. The constraint can be further decomposed into two inequality constraints

$$g_1(y) = \kappa(y) - \kappa_1 \leq 0 \quad g_2(y) = \kappa(y) - \kappa_2 \geq 0. \quad (5)$$

Then, the Lagrangian can be constructed as

$$L(y, \mathbf{w}) = |\kappa(y)| + \lambda_1 g_1(y) + \lambda_2 g_2(y) \quad (6)$$

where λ_1 and λ_2 are two Lagrange multipliers. The variation of \mathbf{w} is given by

$$\Delta \mathbf{w} = \frac{\partial L}{\partial \mathbf{w}} = [\text{sign}(\kappa(y)) + \lambda_1 + \lambda_2] \frac{\partial \kappa}{\partial \mathbf{w}} \quad (7)$$

where

$$\frac{\partial \kappa}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} E[(\mathbf{w}^T \mathbf{z})^4] = 4E[(\mathbf{w}^T \mathbf{z})^3 \mathbf{z}].$$

In order to derive the adaptation rules for the Lagrange multipliers λ_1 and λ_2 , the Karush–Kuhn–Tucker (KKT) conditions can be used [13], which define the necessary conditions for the optimality of general optimization problems with inequality and/or equality constraints. For a constrained optimization

problem defined on a function $f(\mathbf{x})$ with constraints $g_i(\mathbf{x})$ for $i = 1, 2, \dots, n$, the objective function is

$$\begin{aligned} & \max_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x}) \\ & \text{subject to } g_i(\mathbf{x}) \leq 0 \quad \forall i = 1, \dots, m \\ & \quad \quad \quad g_i(\mathbf{x}) \geq 0 \quad \forall i = m+1, \dots, n \end{aligned} \quad (8)$$

where \mathbf{X} is an open set and the feasible region is defined as $S = \{\bar{\mathbf{x}} \in \mathbf{X} : g_i(\bar{\mathbf{x}}) \leq 0, i = 1, \dots, m; g_i(\bar{\mathbf{x}}) \geq 0, i = m+1, \dots, n\}$. The necessary KKT conditions for a local optimum (i.e., maximum in this case) of $f(\mathbf{x})$ at the feasible point $\bar{\mathbf{x}} \in S$ state that there exists a set of Lagrange multipliers $\{\lambda_1, \dots, \lambda_n\}$ such that

$$\begin{aligned} \nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^m \lambda_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{i=m+1}^n \lambda_i \nabla g_i(\bar{\mathbf{x}}) &= 0 \\ \lambda_i \nabla g_i(\bar{\mathbf{x}}) &= 0 \quad \forall i = 1, \dots, n \\ \lambda_i &\leq 0 \quad \forall i = 1, \dots, m \\ \lambda_i &\geq 0 \quad \forall i = m+1, \dots, n \end{aligned} \quad (9)$$

where $\nabla f(\bar{\mathbf{x}})$ and $\nabla g_i(\bar{\mathbf{x}})$ represent gradients of f and g_i at $\mathbf{x} = \bar{\mathbf{x}}$, respectively. The gradients $\nabla g_i(\bar{\mathbf{x}})$ are assumed to be linearly independent. In this case, $m = 1$ and $n = 2$ and linear independence assumption of $\nabla g_i(\bar{\mathbf{x}})$ is obviously satisfied because there is only one inequality constraint of each type. The first-order KKT conditions give only necessary conditions for local maxima but not an algorithm for the adaptation of the Lagrange multipliers. Here, the same reasoning is followed since the fixed-point update is used to find \mathbf{w} . Thus, the variation of the Lagrangian with respect to multipliers λ_1 and λ_2 is given as

$$\Delta \lambda_i = \frac{\partial L}{\partial \lambda_i} = g_i(y), \quad \text{for } i = 1, 2 \quad (10)$$

and the update equations for λ_1 and λ_2 are

$$\lambda_1 \leftarrow -g_1(y) \quad \lambda_2 \leftarrow g_2(y) \quad (11)$$

where the minus sign is to penalize the Lagrangian in (6) whenever inequality constraints are not satisfied. Due to the nature of the inequality constraints, the first-order necessary KKT optimality conditions in (10) require $\lambda_1 \leq 0$ and $\lambda_2 \geq 0$ [13]. Thus, the complete learning rule for λ_1 and λ_2 becomes

$$\lambda_1 \leftarrow \min(-g_1(y), 0) \quad \lambda_2 \leftarrow \min(g_2(y), 0). \quad (12)$$

The problem defined in (5) and (6) is very general. When only the inequality constraint $g_1(y)$ is active, λ_2 in (7) is set to zero and λ_1 adaptation follows (12); when only the inequality constraint $g_2(y)$ is active, λ_1 in (7) is set to zero and λ_2 adaptation follows (12).

C. Extension to Multiple Targets

Using (3), (4) or (7), (4), and (12), only one \mathbf{w} is found, which suffices for detecting one target. To detect more targets, the whitened data \mathbf{Z} are projected onto the orthogonal subspace of \mathbf{w} using the orthogonal subspace projector (OSP) $\mathbf{P}_{\mathbf{w}}^\perp = \mathbf{I} - \mathbf{W}(\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T$, to eliminate the detected target, where \mathbf{I} is an $L \times L$ identity matrix and $\mathbf{W} = [\mathbf{w}]$. If $p-1$ targets have been detected, then $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{p-1}]$, where \mathbf{w}_i

is the projector for detecting the i th target. To detect the p th target, the OSP is used to eliminate all the $p-1$ ones. In this way, the p th target is also separated from others.

D. CKM Algorithm

- Step 1) Set ε , κ_{ref} , and p . Let $k = 1$ and $j = 0$.
- Step 2) Obtain centered and whitened data \mathbf{Z} .
- Step 3) Initialize the projector \mathbf{w}_k denoted as \mathbf{w}_k^0 .
- Step 4) Calculate $y = (\mathbf{w}_k^j)^T \mathbf{Z}$.
- Step 5) Update the Lagrange multipliers using (12).
- Step 6) Update \mathbf{w}_k^j according to (7) and (4).
- Step 7) Check if $\|\mathbf{w}_k^j - \mathbf{w}_k^{j-1}\| < \varepsilon$. If yes, go to step 8). Otherwise, $j \leftarrow j+1$, and go to step 4).
- Step 8) If $k = p$, terminate the algorithm; otherwise, $\mathbf{Z} \leftarrow$ the projection of \mathbf{Z} onto the orthogonal subspace of $\mathbf{w}_1, \dots, \mathbf{w}_k$, $k \leftarrow k+1$, and go to step 3).

E. Practical Considerations

1) *Inequality Constraint and κ_{ref} Selection:* The kurtosis value corresponding to a target is related to many factors, such as the abundance of the target component contained in a pixel, the target size relative to the image size, and the number of spectral bands. Empirically, the typical constraint for detecting small targets from hyperspectral imagery is $\kappa > \kappa_{\text{ref}} = 100$, and for objects with large size $\kappa > \kappa_{\text{ref}} = 0$. In practice, if small targets are to be detected, one can start with a small positive value as κ_{ref} ; then gradually increase this value until only small targets remain to be extracted.

2) *Initial Conditions:* λ_1 and λ_2 can be simply initialized as zero. In general, \mathbf{w} can be randomly initialized. \mathbf{w} can converge very quickly to the small targets if any. When there are only objects with large size present in an image scene, this random condition may result in a low convergence speed. One way to increase the convergence speed is to use the eigenvectors of the data covariance matrix. Since the eigenvectors represent the directions where the data energy is concentrated, using them as initial conditions can confine the search to the appropriate directions.

3) *Stopping Criterion:* The number of targets p is unknown in many cases. Even when p is known, it cannot be guaranteed that the first p extracted classes are exactly these p targets. To be conservative, one can use a reference number (e.g., the number of distinct signals in the image scene [14]) which is larger than p in the algorithm. The targets of interest can be separated in a postprocessing step with target-related prior information.

4) *Stability and Convergence Speed:* It was shown that the convergence speed of the fixed-point adaptation in the FastICA algorithm is cubic, and it converges for any non-Gaussian distribution [11]. In [15], the average convergence behavior of the FastICA was analyzed, and it was proved that the FastICA is very robust with respect to random fluctuation in initial conditions. In [16], it was shown that the FastICA is statistically efficient in achieving the Cramér–Rao lower bound for linear ICA under certain conditions. Following this analysis, it can be proved that the CKM algorithm theoretically has equally fast convergence to that of the FastICA when the constraints can be easily satisfied, wherein the Lagrange multipliers are zero in (7). Stability is preserved because the update of the Lagrange

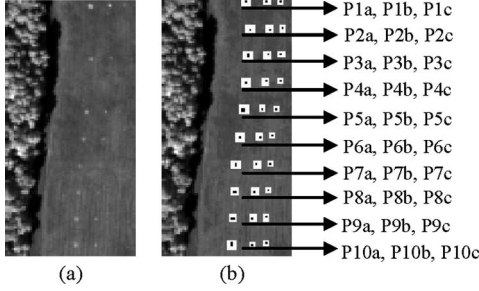


Fig. 1. (a) HYDICE scene (Band 30). (b) Spatial locations of 30 panels.

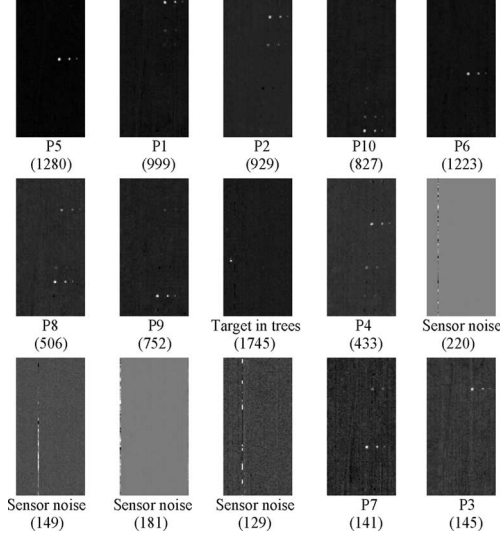


Fig. 2. Target detection and discrimination using the CKM algorithm.

multipliers using (12) satisfies the first-order KKT conditions in (9), which guarantee the convergence toward local maxima. The convergence speed of the CKM may be slow when the inequality constraints are tight.

III. EXPERIMENT

The HYDICE Forest subimage scene of size 128×64 shown in Fig. 1(a) was collected in Maryland in 1995 with about 1.5-m spatial resolution and $0.4\text{--}2.5\ \mu\text{m}$ spectral coverage. One hundred sixty-nine bands were used after bad bands were removed. This scene has 30 panels. The three panels in the same row p_{ia}, p_{ib}, p_{ic} were made from the same material of sizes $3 \times 3\text{ m}$, $2 \times 2\text{ m}$, and $1 \times 1\text{ m}$, respectively, and are considered as one class, P_i for $1 \leq i \leq 10$. The ground-truth map in Fig. 1(b) shows the precise locations of pure panel pixels (dark centers). These ten panel classes have close spectral signatures and are difficult to discriminate.

Fig. 2 shows the targets detected and discriminated using the CKM with inequality constraint $\kappa > 100$. This constraint was imposed because small man-made targets are of interest. The algorithm was randomly initialized. As shown, all the ten rows of panels were detected and separated into ten different classes (and a target in trees was also detected). Because the panel signatures are similar, sometimes another row of panels is displayed in light shade in the detection map of a specific row of panels. For instance, P3 was perceivable when detecting P8. Given the CKM's unsupervised nature, this result is satisfying because generally panels were well detected and separated.

There were four classes related to strong sensor noise, which looks like a “scratch line.” We have manually removed the bands with strong sensor noise before data analysis, but some bands may still contain sensor noise, which is why the CKM algorithm picked up this information. This demonstrates that the kurtosis-based algorithm has a higher capability of extracting outliers with low energy. After these outliers are extracted, the OSP is used to eliminate them for cleaner data.

Fig. 2 also lists the kurtosis for each target (in the parenthesis) in one realization. Overall, the kurtosis was decreasing. In other words, the targets with high kurtosis had higher chances to be extracted first. However, it is possible that the k th target may have a slightly smaller kurtosis than the $(k+1)$ th target. Targets detected in two different realizations may be in different orders. For instance, P5 was detected as the first target in the realization shown in Fig. 2 with $\kappa = 1279.6$. If in the next realization P5 was detected as the third target, then the kurtosis would be slightly decreased, for example, $\kappa = 1270$. This decrease occurred because in the second realization P5 was detected after two orthogonal subspace projections that were used to annihilate the first two targets. During these processes, some energy of P5 that is horizontal to the projections was eliminated as well, which makes the final kurtosis somewhat smaller. However, the kurtosis values of individual targets from different realizations were always in the same range.

To investigate the stability and convergence speed of the CKM algorithm, we show the convergence curves of the 11 targets (ten panel classes plus the target in trees) in Fig. 3, in comparison with Fig. 2. When the w 's were initialized using small random values as in Fig. 3(a), most of them converged after four to five iterations. The convergence speed was very fast, particularly for those distinctive targets with large kurtosis, such as the unknown target in the trees. Some targets, such as P3, P4, and P7, needed more iterations to converge. We found that these targets with relatively lower kurtosis were detected very last, as shown in Fig. 2. For a projector w to converge to a direction different from the existing directions for the detected eight targets, the algorithm has to be executed for longer adaptation. Fig. 3(b) shows the case when the eigenvectors of the data covariance matrix were used as initial conditions. Then, the speed of convergence was increased, i.e., all the targets were detected within five iterations and most within three iterations. This is because: 1) the eigenvectors represent the directions where energy is concentrated and using them as initials confines the searching along the appropriate directions; and 2) eigenvectors are mutually orthogonal, which can speed up the convergence of the late detected targets.

To compare with the FastICA algorithm without a constraint, both algorithms were run 100 times with random initials. We found that, on average, 18 projections were needed for the FastICA to detect and discriminate all the ten panel classes, while 14 projections were needed for the CKM algorithm to do the same. The detection result from the FastICA algorithm with 18 projections was compared with that from the CKM algorithm with 14 projections. The probability of detection and probability of false alarm were estimated by thresholding the normalized detection maps with the threshold being changed from 0.1 to 0.9 and comparing with the available pixel-level ground truth. The resulting receiver operating characteristic (ROC) curves are shown in Fig. 4, where the performance of the

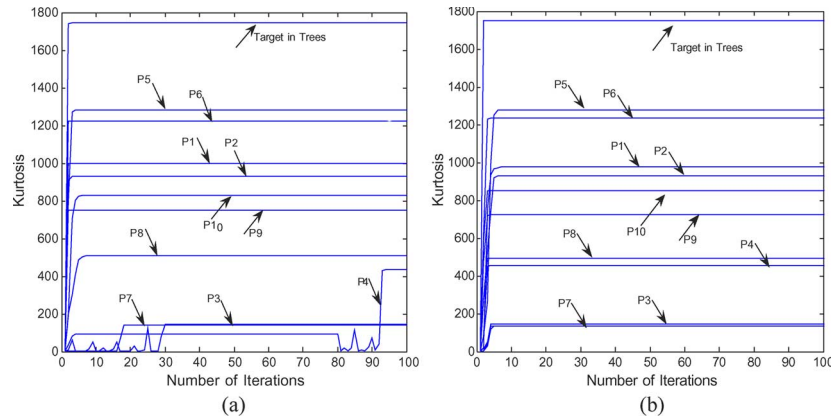


Fig. 3. Convergence curves in the HYDICE experiment. (a) Using random vectors as initials. (b) Using eigenvectors as initials.

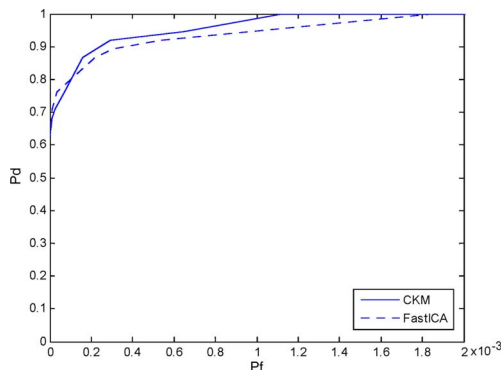


Fig. 4. ROC curves in the HYDICE experiment.

CKM algorithm is slightly better than that of the FastICA. This may be because the portion of the pixel signatures that is along the direction of the projection is sacrificed during project pursuit. This gradually reduces the energy of target pixels and may influence the final detection result. Therefore, it is desirable to extract targets with fewer projections whenever possible.

IV. CONCLUSION

The major advantages of using the CKM for automated target detection and discrimination are summarized as follows.

- 1) A very general inequality constraint is imposed when searching for a projector w that maximizes the kurtosis, which not only automatically separates the super-Gaussian and sub-Gaussian classes but also provides the flexibility for finding targets within different ranges of kurtosis.
- 2) Different from other widely used methods—such as the RX algorithm, which can only detect small targets—the CKM algorithm can separate multiple targets by incorporating the OSP concept. Compared to the unconstrained kurtosis maximization algorithms such as the FastICA, which can also detect and separate multiple targets, the CKM can complete the task with fewer projections, i.e., higher computational efficiency.
- 3) By substituting the random initial conditions with the eigenvectors of the data covariance matrix that correspond to large eigenvalues, the convergence speed of the CKM algorithm can be increased. This is particularly

important when applying the CKM algorithm to extract large objects.

ACKNOWLEDGMENT

The authors would like to thank Mr. John Brocato for his help with editing the manuscript.

REFERENCES

- [1] D. W. J. Stein *et al.*, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.
- [2] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
- [3] E. Ashton, "Detection of subpixel anomalies in multispectral infrared imagery using an adaptive Bayesian classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 2, pp. 506–517, Mar. 1998.
- [4] A. Ifarraguerri and C.-I. Chang, "Unsupervised hyperspectral image analysis with projection pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 6, pp. 2529–2538, Jun. 2000.
- [5] S.-S. Chiang, C.-I. Chang, and I. W. Ginsberg, "Unsupervised target detection in hyperspectral images using projection pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1380–1391, Jul. 2001.
- [6] C.-I. Chang, S.-S. Chiang, J. A. Smith, and I. W. Ginsberg, "Linear spectral random mixture analysis for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 2, pp. 375–392, Feb. 2002.
- [7] S. M. Schweizer and J. M. F. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 584–597, Apr. 2001.
- [8] S. A. Robila and P. K. Varshney, "Target detection in hyperspectral images based on independent component analysis," *Proc. SPIE*, vol. 4726, pp. 173–182, 2002.
- [9] S. A. Robila and P. K. Varshney, "Further results in the use of independent components analysis for target detection in hyperspectral images," *Proc. SPIE*, vol. 5094, pp. 186–195, 2003.
- [10] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley-Interscience, 2001.
- [11] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1483–1492, Oct. 1997.
- [12] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
- [13] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999.
- [14] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
- [15] S. C. Douglas, Z. Yuan, and E. Oja, "Average convergence behavior of the FastICA algorithm for blind source separation," in *Proc. 6th Int. Conf. Ind. Compon. Anal. Blind Signal Separation*, Mar. 2006, pp. 790–798.
- [16] P. Tichavský, Z. Koldovský, and E. Oja, "Performance analysis of the FastICA algorithm and Cramér–Rao bounds for linear independent component analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1189–1203, Apr. 2006.