

# Particle Swarm Optimization-Based Band Selection for Hyperspectral Target Detection

Yan Xu, *Student Member, IEEE*, Qian Du, *Senior Member, IEEE*,  
and Nicolas H. Younan, *Senior Member, IEEE*

**Abstract**—This letter proposes particle swarm optimization (PSO)-based band selection (BS) approach for hyperspectral target detection. Due to lack of training samples in a detection problem, it is more difficult than classification-purposed BS. The objective function, called maximum-submaximum-ratio (MSR) gauging target-background separation, is proposed for target detection during PSO searching. Typical target detectors such as target-constrained interference-minimized filter and adaptive coherence estimator are studied. Experimental results demonstrate that the proposed MSR-based objective function in conjunction with PSO-based searching can select a small band set while yielding similar or even better detection performance than using all the original bands, sequential forward search-based BS, or BS relying on detection map similarity assessment.

**Index Terms**—Band selection (BS), hyperspectral imagery, particle swarm optimization (PSO), target detection.

## I. INTRODUCTION

**T**ARGET detection is one of major tasks of hyperspectral imaging. A hyperspectral image cube contains hundreds of spectral bands, leading to a high computation burden on target detection. Thus, it is necessary to do dimensional reduction on the original data for efficient analysis. Typically, there are two kinds of methods for dimensional reduction. The first one is transform-based methods (such as principal component analysis [1]), and such methods may not be preferred since they alter the physical meaning of the data. The second one is band selection (BS), which is to select a subset of bands while still generating satisfactory results using the selected bands [2]–[4].

BS algorithms have been applied to hyperspectral data analysis, and they can be implemented in either unsupervised or supervised. Unsupervised BS is to select the subset of bands without any prior information. For instance, a simple yet efficient similarity measure method was developed in [5] for BS. Supervised BS algorithms intend to select the bands producing maximum class separability with prior knowledge of training samples. Many methods (see [5], [6]) have been developed to measure the class separability for BS. However, most BS algorithms are focused on classification problems, and few have been proposed for target detection. It is more difficult to separate targets and nontargets with selected bands

due to lack of training samples for target and background modeling.

The search strategy is also an important issue. To avoid exhaustive search, which is computationally prohibitive to hyperspectral BS, sequential forward search (SFS) and sequential floating forward (SFFS) methods can be used [7]. Their basic idea is to select the best band for maximizing an objective function, then one additional band combining with the existing selected band or bands is selected to maximize the objective function. The process continues until the desired number of bands is reached. Recently, particle swarm optimization (PSO), invented by Eberhard and Kennedy [10], has been applied to hyperspectral BS with the objective of maintaining classification accuracy [8], [9]. The PSO imitates the social behavior of flocks. Each particle is a solution, and the positions of particles are randomly initialized. Then the particles fly over the problem space until certain criteria are reached. Compared with SFS and SFFS, PSO offers two major advantages: it can provide better high-dimensional solution due to global search, and it can be easily implemented in parallel. Although several other evolutionary algorithms are developed recently, such as ant colony optimization [11] and firefly algorithm [12], which have been applied to hyperspectral BS, we limit our discussion with PSO in this letter.

Here, we focus on the task of target detection where targets with known spectral signatures are to be detected from an unknown background. Targets, as small man-made objects, are often sparsely populated [13]. No training or labeled samples are available for both target and background [14]. The key of PSO-based selection is to design an effective objective function during searching. More often, labeled samples are available when evaluating an objective function. However, it is impossible for target detection. Thus, a detection-specific objective function is required. Intuitively, we can compare the detection outputs for band searching with a certain criterion, such as Euclidean distance or correlation coefficient (CC). However, a large similarity does not necessarily mean satisfactory performance, because the number of target pixels is much smaller than background pixels and similarity assessment is dominated by background. Thus, we prefer the metric that can well gauge target and background separation. Specifically, we propose the maximum-submaximum-ratio (MSR) to quantify such separation. The typical target detectors, i.e., target-constrained interference-minimized filter (TCIMF) and adaptive coherence estimator (ACE), are studied. Experimental results using the proposed MSR-based objective function in conjunction with PSO-based searching can select a small band set while yielding similar or even better detection performance than using all the original bands. The parameter setting of MSR is discussed based on the performance of detectors.

Manuscript received September 27, 2016; revised December 31, 2016; accepted January 20, 2017. Date of publication February 22, 2017; date of current version March 3, 2017.

The authors are with the Department of Electrical and Computer Engineering, Mississippi State University, MS 39762 USA (e-mail: yx131@msstate.edu; du@ece.msstate.edu; younan@ece.msstate.edu).

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

Digital Object Identifier 10.1109/LGRS.2017.2658666



Fig. 1. Cropped HyMap data.

## II. PROPOSED BAND-SELECTION METHOD

### A. Target Detectors

Assume  $N$  hyperspectral pixels  $\mathbf{X} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N]$  are obtained, where  $\mathbf{r}_i$  is an  $L \times 1$  vector. Let  $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_p]$  and  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_q]$  denote the desired and undesired target signature matrices, respectively. TCIMF [15] is designed to keep the output of the desired target at a desirable level while suppressing the energy of the output of undesired targets. It can be considered as the following linearly constrained optimization problem:

$$\min_{\mathbf{w}} \{\mathbf{w}^T \mathbf{R} \mathbf{w}\} \quad \text{s.t.} \quad [\mathbf{D}\mathbf{U}]^T \mathbf{w} = \begin{bmatrix} \mathbf{1}_{p \times 1} \\ \mathbf{0}_{q \times 1} \end{bmatrix} \quad (1)$$

with the optimal coefficient vector  $\mathbf{w}_{\text{TCIMF}}$

$$\mathbf{w}_{\text{TCIMF}} = \mathbf{R}^{-1} [\mathbf{D}\mathbf{U}] \{ [\mathbf{D}\mathbf{U}]^T \mathbf{R}^{-1} [\mathbf{D}\mathbf{U}] \}^{-1} \begin{bmatrix} \mathbf{1}_{p \times 1} \\ \mathbf{0}_{q \times 1} \end{bmatrix} \quad (2)$$

where  $\mathbf{R} = 1/N \sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^T$  is the  $L \times L$  data correlation matrix,  $\mathbf{1}_{p \times 1}$  is a  $p \times 1$  constraint column vector with all components equal to 1, and  $\mathbf{0}_{q \times 1}$  is a  $q \times 1$  constraint column vector with all components equal to 0. The output of TCIMF is presented as

$$y_{\text{TCIMF}} = \mathbf{w}_{\text{TCIMF}}^T \mathbf{r}. \quad (3)$$

For the pixel  $\mathbf{r}$ , its ACE output, derived from a binary hypothesis testing problem, can be written as [16]

$$y_{\text{ACE}} = \frac{\mathbf{r}^T \boldsymbol{\Sigma}^{-1} \mathbf{d} (\mathbf{d}^T \boldsymbol{\Sigma}^{-1} \mathbf{d})^{-1} \mathbf{d}^T \boldsymbol{\Sigma}^{-1} \mathbf{r}}{\mathbf{r}^T \boldsymbol{\Sigma}^{-1} \mathbf{r}} \quad (4)$$

where  $\boldsymbol{\Sigma}$  is the background covariance matrix. The ACE is considered as one of the most powerful detectors. Note that it cannot be simplified as a finite impulse response (FIR) filter  $\mathbf{w}$  as in TCIMF or constrained energy minimization.

### B. Criterion Functions

1) *Maximum-Submaximum-Ratio Criterion*: Let the output of the TCIMF or ACE be represented as a grayscale image. Suppose the maximum value in the output is located at  $(i, j)$ . After bands are selected, TCIMF or ACE is reconducted. Using the HyMap data in Fig. 1 to be introduced in Section III for illustration, Figs. 2 and 3 show a part of the detection outputs of TCIMF and ACE, respectively, before and after BS. Here, we generate the detection maps for F2 and the number of selected band is 30. For TCIMF, the maximum value before BS corresponds to a desired target; however, after BS, the

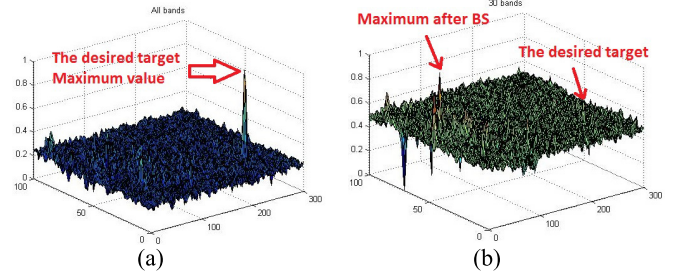


Fig. 2. TCIMF detection maps for F2. (a) All bands. (b) 30 bands.

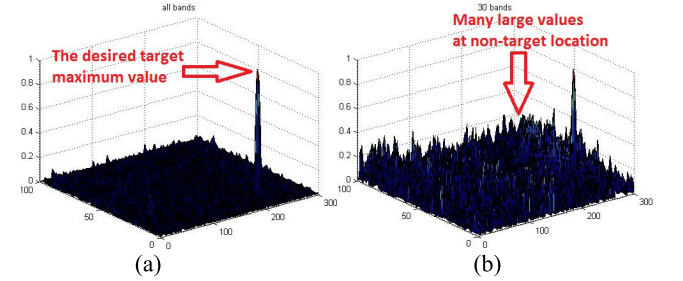


Fig. 3. ACE detection maps for F2. (a) All bands. (b) 30 bands.

location of the maximum output is shifted to another location. Moreover, before BS, the maximum value is significantly larger; after BS, there are several quite similar large values from background pixels as false alarms, which are critical to gauge the separation between target and background. For ACE, the maximum value after BS still corresponds to a desired target and is not shifted to another location, and the major change is that the overall outputs from the background becomes large.

Intuitively, desirable detection performance can be achieved when target and background are well separated. However, the majority of pixels are background and their average is always very small. Based on Figs. 2 and 3, the key to success is to obtain a large separation from those false alarm pixels with submaximum (SM) values. For TCIMF, a relatively weak detector, the average SM value is based on several background pixels yielding potential false alarms; for ACE that is relatively strong in background suppression, the average SM value may be estimated using more background pixels.

Let the average value at the maximal location  $(i, j)$  and its four nearest pixels be computed and denoted as  $M$ . Without considering these five pixels in the generated detection map after BS, the average value of the first  $K$  largest values is calculated and denoted as average SM. In the experiment, we chose  $K = 10, 20, \text{ or } 300$  (top 1% of the largest values). The objective function can be described as

$$\arg \max_{\Phi^S} \frac{M}{SM(\Phi^S)} \quad (5)$$

where  $\Phi^S$  is the selected band subset, and the SM value is its function. This criterion is called MSR. If this ratio is high, then the resulting selected bands can better separate the target and the false alarm background pixels, and most likely, maintain the detection performance.

2) *Correlation Coefficient*: Let an  $m$ -by- $n$  output image from using all the original bands be denoted as  $\mathbf{O}_1$ , and the one from the selected bands is  $\mathbf{O}_2$ . The CC of the two output

images  $\mathbf{O}_1$  and  $\mathbf{O}_2$  is calculated as

$$CC = \frac{\sum_m \sum_n (\mathbf{O}_{1mn} - \bar{\mathbf{O}}_1)(\mathbf{O}_{2mn} - \bar{\mathbf{O}}_2)}{\sqrt{(\sum_m \sum_n (\mathbf{O}_{1mn} - \bar{\mathbf{O}}_1)^2)(\sum_m \sum_n (\mathbf{O}_{2mn} - \bar{\mathbf{O}}_2)^2)}} \quad (6)$$

where  $\bar{\mathbf{O}}_1$  and  $\bar{\mathbf{O}}_2$  are the averages of the output images  $\mathbf{O}_1$  and  $\mathbf{O}_2$ , respectively. The band subset that can maximize the CC will be selected, i.e., the objective function to select the best bands is

$$\arg \max_{\Phi^S} CC(\Phi^S). \quad (7)$$

It will be shown such an image similarity measure may not work well because the majority of pixels belong to background, dominating the similarity measurement.

### C. PSO Search Strategy

Let a particle  $\mathbf{x}_{id}$  denote a possible solution. Let  $\mathbf{v}_{id}$  denote the velocity to update the current location  $\mathbf{x}_{id}$ . In each iteration, the historically local best solution is described as  $\mathbf{p}_{id}$ , and the historically global best solution among all the particles is denoted as  $\mathbf{p}_{gd}$ . The update of  $\mathbf{v}_{id}$  can be described in

$$\mathbf{v}_{id} = w \times \mathbf{v}_{id} + c_1 \times r_1 \times (\mathbf{p}_{id} - \mathbf{x}_{id}) + c_2 \times r_2 \times (\mathbf{p}_{gd} - \mathbf{x}_{id}) \quad (8)$$

which calculates the new velocity for each particle based on its previous velocity. The location update of the particles is updated as

$$\mathbf{x}_{id} = \mathbf{x}_{id} + \mathbf{v}_{id}. \quad (9)$$

In (9),  $c_1$  and  $c_2$  control the contribution of the local best and the global best solution, respectively, and  $r_1$  and  $r_2$  are two random variables within  $[0, 1]$ . The inertia weight  $w$  is applied as a scalar of the previous velocity  $\mathbf{v}_{id}$ , and it can result in better convergence in many applications [10].

In the BS problem,  $\mathbf{x}_{id}$  becomes a vector containing the selected band indices. Similarly,  $\mathbf{p}_{id}$  and  $\mathbf{p}_{gd}$  are also vectors denoting the local and global best band indices, respectively.

### D. Algorithm

The proposed BS algorithm is summarized as follows.

- 1) Randomly initialize  $P$  particles  $\mathbf{x}_{id}$ . Each particle denotes the indices of the bands to be selected.
- 2) Conduct target detection with TCIMF or ACE to generate the detection map and normalize the detection map.
- 3) Evaluate the  $P$  particles using the designed objective functions MSR or CC.
- 4) Update the velocities and the positions of the particles until a certain stopping criterion is reached.
- 5) The particle that generates the global best value contains the indices of the bands to be selected.

In the experiments, we empirically chose the number of particles  $P = 20$ , and the acceleration coefficients  $c_1 = c_2 = 2.1$ . The range of the inertia weight  $w$  was from 0.4 to 0.9. The algorithm was considered to converge if the global best  $\mathbf{p}_{gd}$  does not change after 400 iterations. The number of selected bands is changed from 30 to 50.

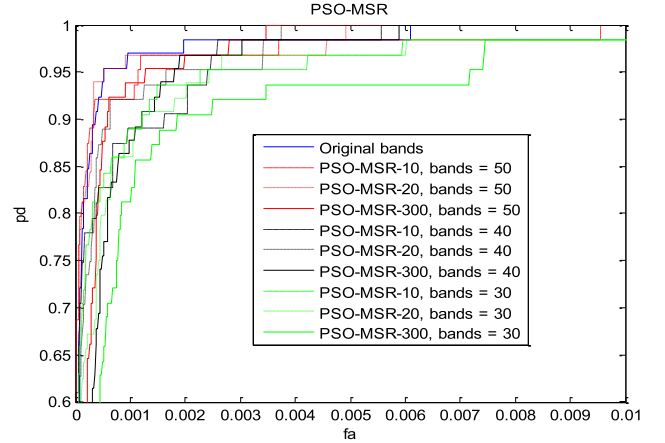


Fig. 4. ROC curve of MSR using PSO with different values of  $K$  (TCIMF).

### E. Performance Comparison

In [17], a sparseness constrained BS method is proposed for target detection. Specifically, for a detector that can be modeled as an FIR filter  $\mathbf{w}$ , a Lasso-based BS (LBS) algorithm is applicable to find a set of bands that yields the minimum Euclidean distance between the original filter output and the output with selected bands. The following objective function is to be minimized:

$$f(\hat{\mathbf{w}}) = \|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 = \|\mathbf{y} - \hat{\mathbf{w}}^T \hat{\mathbf{X}}\|_2^2 \quad \text{s.t.} \quad \|\hat{\mathbf{w}}\|_1 \leq t \quad (10)$$

where  $\hat{\mathbf{w}}$  is the detector applied to the selected bands,  $\hat{\mathbf{X}}$  denotes all the pixel vectors in selected bands only, and  $t$  is the desired sparsity. In the experiment, the LBS is compared with the MSR when the TCIMF is the detector, as the LBS is infeasible to ACE which cannot be simplified as a filter  $\mathbf{w}$ .

## III. EXPERIMENT

### A. Experimental Data

The HyMap data, obtained from Rochester Institute of Technology [18], are used in this experiment. This data have 126 spectral bands with the spatial size of  $200 \times 800$  and covers an area of Cooke City, MT, USA. The experiment is conducted based on a subimage of size  $100 \times 300$  shown in Fig. 1. The spatial resolution of the data is approximately 3 m. In the HyMap data, there are three types of vehicle (V1-V3) and four types of real fabric panels (F1-F4). There are a total of 65 target pixels. In the experiment, we use receiver operating characteristic (ROC) curve to evaluate the experiment results, which shows the tradeoff between false alarm rate (fa) and probability of detection (pd). The performance of the ROC can be measured by the area under the curve, and the larger area under a curve, the better the performance [19].

### B. Experiment on TCIMF

We first discuss the effect of the MSR parameter  $K$ . It can be concluded from Fig. 4 that averaging the first 10 submaximal values provides results comparable to averaging the first 20 submaximal values, and both cases present better results than that with  $K = 300$ . The explanation of this phenomenon may be found in Fig. 2. We can see from Fig. 2(b) that there are only a few large SM values at nontarget locations. Therefore, the objective is to select the bands suppressing these

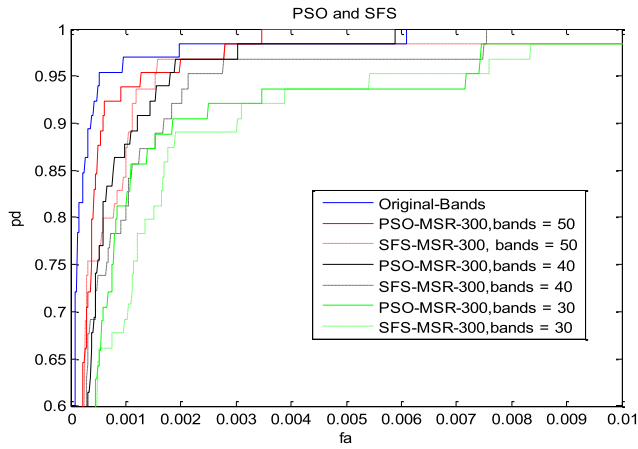
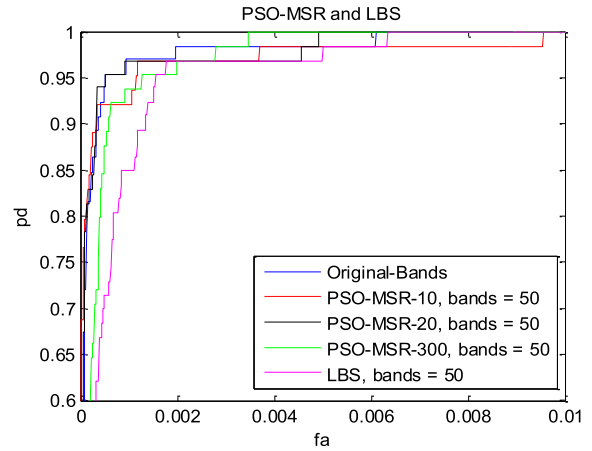
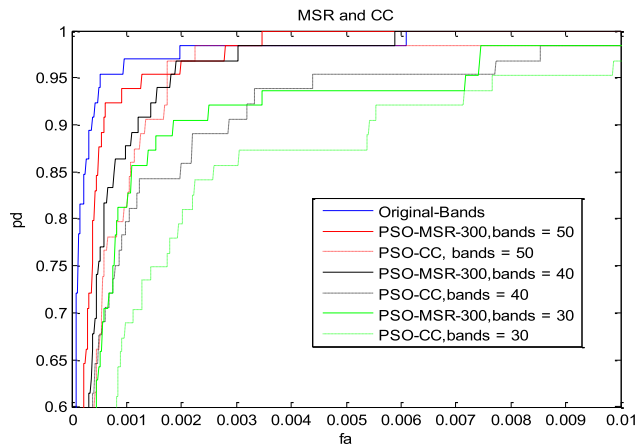
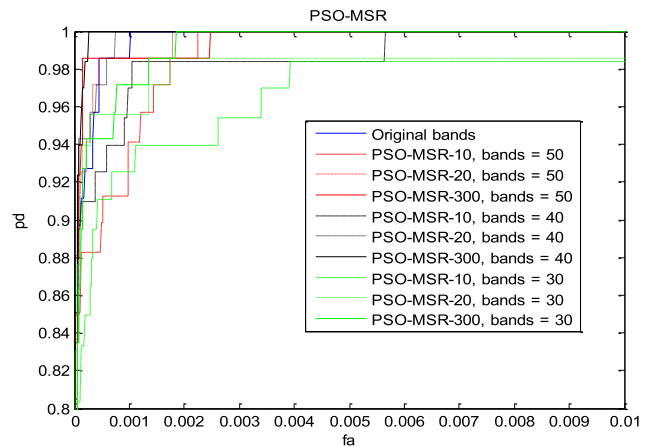

 Fig. 5. ROC curve of MSR ( $K = 300$ ) using PSO and SFS (TCIMF).


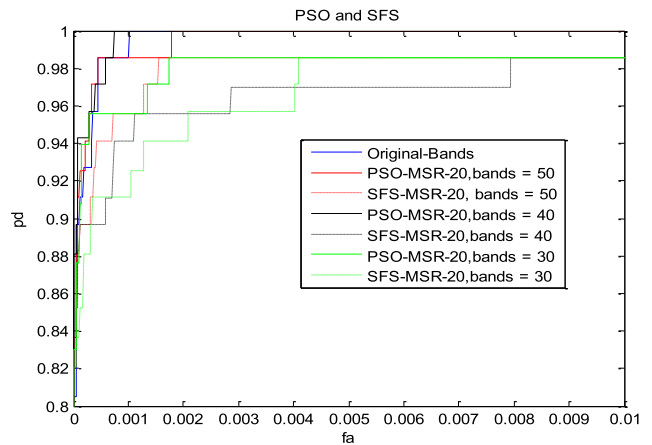
Fig. 7. Comparison with LBS when TCIMF is the detector.


 Fig. 6. ROC curve of MSR ( $K = 300$ ) and CC using PSO (TCIMF).

 Fig. 8. ROC curve of MSR using PSO with different values of  $K$  (ACE).

nontarget SM values while remaining the maximum value at the target location. When  $K$  is too large, we may also include those background pixels that can be easily suppressed. We can also see from Fig. 4 that the ROC curve is degraded when the number of selected bands decreases. This is expected since more bands provide more spectral information and thus can better separate target and background. Moreover, when the number of bands is close to 40 or 50, the detection performance is comparable to that with all the bands, since additional bands may not contain more useful information. We reach the goal to reduce the data dimensionality while keeping the detection performance.

Fig. 5 compares PSO and SFS for MSR with  $K = 300$ . Note that when  $K = 300$ , the performance of MSR is not the best. PSO can outperform SFS. Fig. 6 compares MSR ( $K = 300$ ) and CC when PSO is used for searching. MSR still provides better results than CC for different numbers of selected bands, indicating that CC may not be a good criterion. Since most of the values belong to the background, simply comparing similarity between the two image outputs cannot guarantee reliable detection performance.

Fig. 7 shows the performance comparison with the LBS when TCIMF is the detector, where the MSR can outperform even when the parameter  $K$  is not the optimum. This is mainly due to the fact that output similarity is not an appropriate metric for target detection.


 Fig. 9. ROC curve of MSR using PSO, SFS with  $K = 20$  (ACE).

### C. Experiment on ACE

The ACE performance is shown in Figs. 8–10. Several conclusions can be drawn from Fig. 8. First,  $K = 300$  and  $K = 10$  are the best and worst MSR parameters, respectively. The detection map after BS is shown in Fig. 3(b), where no significant false alarm pixels are present. If only the first ten largest values are suppressed, there are still some other similar large values at nontarget locations. Thus, it is better to suppress more SM values. Second, the ROC curve performance after

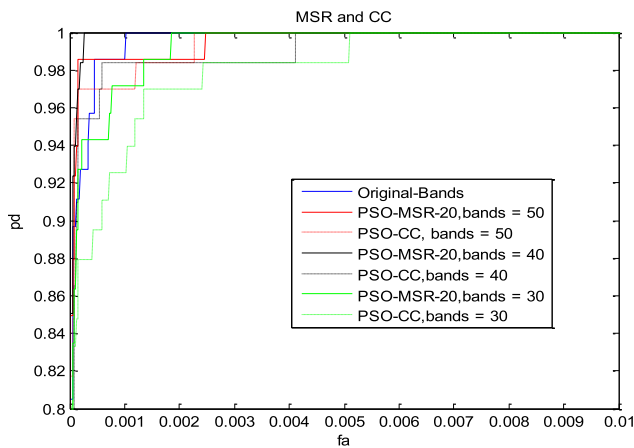


Fig. 10. ROC curve of MSR with  $K = 20$  and CC using PSO (ACE).

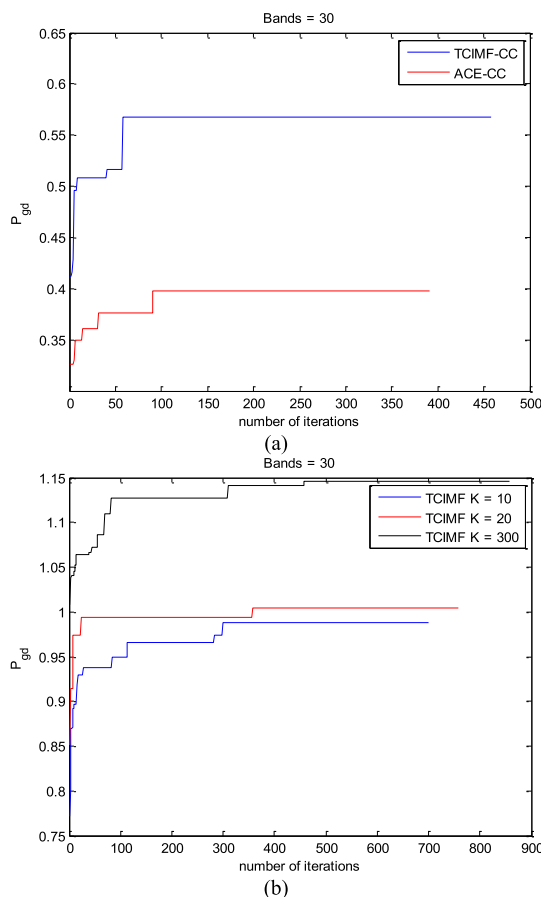


Fig. 11. Convergence curve. (a) CC. (b) TCIMF with MSR.

BS is comparable to that with all the bands. In Fig. 3(b), even after selecting 30 bands, the maximum value is still at the target location. Hence, it may not need many bands to separate target and background. Fig. 9 shows that PSO outperforms SFS. In Fig. 10, the MSR criterion is better than CC even when  $K$  is not the optimum.

#### D. Convergence

The convergence curve is shown in Fig. 11. In Fig. 11(a), when using CC as the objective function, the PSO for both TCIMF and ACE can converge after about 100 iterations.

In Fig. 11(b), the PSO with MSR searching can be terminated within 400 iterations.

#### IV. CONCLUSION

In this letter, a PSO-based BS method is proposed to select a small number of bands while maintaining the target detection performance. Typical detectors, i.e., TCIMF and ACE, are investigated. Detection output comparison with MSR or CC can serve as a simple objective function, and the MSR measuring target and false-alarm-background separation outperforms CC or other similarity-based criteria. The experiment results also demonstrated that the PSO can find better subset bands for detection compared to the traditional SFS.

#### REFERENCES

- [1] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Edu. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933.
- [2] Q. Du, "Band selection and its impact on target detection and classification in hyperspectral image analysis," in *Proc. IEEE Workshop Adv. Techn. Anal. Remotely Sensed Data*, Oct. 2003, pp. 374–377.
- [3] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138–142, Jan. 2011.
- [4] H. Su and Q. Du, "Hyperspectral band clustering and band selection for urban land cover classification," *Geocarto Int.*, vol. 27, no. 5, pp. 395–411, 2012.
- [5] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.
- [6] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.
- [7] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994.
- [8] H. Yang, Q. Du, and G. Chen, "Particle swarm optimization-based hyper-spectral dimensionality reduction for urban land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 544–554, Apr. 2012.
- [9] H. Su, Q. Du, G. Chen, and P. Du, "Optimized hyperspectral band selection using particle swarm optimization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2659–2670, Jun. 2014.
- [10] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Int. Symp. Micromach. Human Sci.*, Nagoya, Japan, 1995, pp. 39–43.
- [11] J. Gao, Q. Du, L. Gao, X. Sun, and B. Zhang, "Ant colony optimization-based supervised and unsupervised band selections for hyperspectral urban data classification," *J. Appl. Remote Sens.*, vol. 8, no. 1, p. 085094, Aug. 2014.
- [12] H. Su, B. Yong, and Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68–72, Jan. 2016.
- [13] D. Manolakis and G. S. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.
- [14] M. T. Eismann, *Hyperspectral Remote Sensing*. Washington, DC, USA: SPIE, 2012, pp. 620–624.
- [15] H. Ren and C.-I. Chang, "Target-constrained interference-minimized approach to subpixel target detection for hyperspectral images," *Opt. Eng.*, vol. 39, no. 12, pp. 3138–3145, 2000.
- [16] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2538–2541, Sep. 1999.
- [17] K. Sun, X. Geng, and L. Ji, "A new sparsity-based band selection method for target detection of hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 329–333, Feb. 2015.
- [18] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, and S. Hager, "Development of a Web-based application to evaluate target finding algorithms," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 2, Jul. 2008, pp. 915–918.
- [19] C. E. Metz, "ROC methodology in radiologic imaging," *Invest. Radiol.*, vol. 21, no. 9, pp. 720–733, 1986.