RESEARCH **Open Access**

CrossMark

Orthogonal maximum margin projection subspace for radar target HRRP recognition

Daiying Zhou

Abstract

In this paper, a novel target recognition method, namely orthogonal maximum margin projection subspace (OMMPS), is proposed for radar target recognition using high-resolution range profile (HRRP). The core of OMMPS is to maximize the between-class margin by increasing the between-class scatter distance and reducing the within-class scatter distance simultaneously. By introducing the nonlinear mapping function, we also derive the kernel version of OMMPS, namely orthogonal kernel maximum margin projection subspace (OKMMPS). Compared with maximum margin criterion (MMC) method, OMMPS are optimal in meaning of maximum margin due that the coordinate axes of OMMPS are obtained sequentially by solving the constrained optimization problem, thus improves the recognition performance. In addition, the number of efficient features for OMMPS is not limited by the number of pattern classes, and the appropriate features can still be obtained for separating the classes, even in high-dimensional space with only a few classes. Moreover, the coordinate axes of OMMPS are mutually orthogonal, and the features extracted by OMMPS reduce the redundancy. The extensive experimental results show that the proposed method has better recognition performance than the other methods such as MMC and LDA.

Keywords: Radar target recognition, HRRP, Maximum margin criterion, Orthogonal maximum margin projection subspace

1 Introduction

We are able to obtain the high-resolution range profile (HRRP) by the wideband radar. The HRRP is the amplitude of radar-returned echoes from target as a function of range cell, which represents the distribution of projection of target radar scattering centers along the radar line of sight. It can provide geometric structure information such as target size and information of scattering centers, which is very useful in target classification. Therefore, the radar target recognition using HRRP has intensively been focused by radar target recognition community [1–7]. K. T. Kim et al. propose some invariant features for HRRP [8–9]. Y. Shi et al. [10] use a novel neural network classifier for HRRP recognition. S. K. Wong [11] presents a feature selection method in frequency domain. D. E. Nelson et al. [12] study a new iterated wavelet feature for HRRP classification. R. A. Mitchell et al. [13] extract some robust statistical

features from HRRP for radar target recognition. X. J. Liao et al. [14] use sequential HRRPs to identify the ground targets. C. Y. Wang et al. [15] model the radar echoes for radar HRRP recognition by T-mixture model. M. Li et al. [16] propose a sparse representation-based denoizing method for improving recognition performance using HRRP. L. Du et al. [17] apply the statistical model for recognizing the radar HRRP. L. Shi et al. [18] use the local factor analysis to model the non-Gaussianity of the radar HRRP data. J. S. Fu et al. [19] extract the between-class discriminant information and among-class discriminant information for improving the classification performance. However, HRRP is sensitive to target aspect, time-shift, and amplitude-scale. These factors increase the between-class ambiguities which must be resolved and degrade the classification accuracy. Moreover, HRRP is typically high-dimensional, non-Gaussian, and interdimension dependently distributed and increases the difficulties in statistical modeling of pattern objects. Thus, the radar target recognition using HRRP is still a challenging task.

Correspondence: dyzhou2011@163.com
School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China

Springer

Many previous works have shown that subspace methods are very effective in pattern recognition task. For example, the principal component analysis (PCA) can preserve the large variance directions [20]. The linear discriminant analysis (LDA) is able to maximize the between-class distance and minimize the within-class distance simultaneously [21]. PCA and LDA are widely applied for feature extraction and dimension reduction. In order to process the nonlinear problem, KPCA [22] and KFDA [23] are proposed based on kernel trick. However, the performance of these methods cannot be improved further when the objects such as HRRPs are usually high-dimensional vectors and do not satisfy the assumption of the Gaussian distribution.

Because they only capture the global geometric structure of dataset and do not consider the local geometric structure information that is very important for target recognition.

To capture the local structure information, several manifold learning methods have been proposed. X. F. He et al. [24] present the locality preserving projections (LPP) by means of a weight matrix (called heat kernel). H. T. Chen et al. [25] propose the local discriminant embedding (LDE) using the neighbors and class relations of data. D. Cai et al. [26] study the orthogonal Laplacian-faces (OLPP) by computing a set of orthogonal basis functions. L. Zhu et al. [27] propose the orthogonal discriminant locality preserving projections (ODLPP) by orthogonalizing the basis vectors. S. J. Wang et al. [28] present an exponential locality preserving projections (ELPP) via introducing the matrix exponential function. The above methods can obtain impressive results. However, they only emphasize the compactness between the neighbors or same-class data points and do not consider the optimal separation between different-class data points. Therefore, the discriminative power may be improved by combining the manifold learning and discriminant analysis.

Motivated by the idea, S. Yan et al. [29] present margin fisher analysis (MFA) method. MFA uses the intrinsic graph and the penalty graph to characterize the local structure in discriminant analysis and thus increases the intraclass compactness and interclass separability. M. Sugiyama [30] proposes the local fisher discriminant analysis (LFD) approach by taking local structure of the data into account, and the multimodal data can be embedded appropriately. D. Cai at al. [31] study the locality sensitive discriminant analysis (LSDA) method, which utilizes local geometry structure of the data manifold and discriminant information at the same time. T. Zhang et al. [32] present a discriminative locality alignment (DLA) algorithm by imposing discriminative information in the part optimization stage. DLA can attack the distribution nonlinearity of measurements and preserve the discriminative ability while avoiding the

small sample size problem. B. Li et al. [33] propose the locally linear discriminant embedding (LLDE) method. LLDE apply the constrained weights to strengthen the classification ability. Y. Chen et al. [34] present a local coordinate factorization (NLCF) method by adding a local coordinate constraint into the standard NMF objective function. Q. Gao et al. [35] propose the stable orthogonal local discriminate embedding algorithm by introducing the orthogonal constraint on the basis vectors. C. Hou et al. [36] propose a unified framework that explicitly unfolds the manifold and reformulate local approaches as the semi-definite programs and thus improves the performance of some algorithms such as locally linear embedding (LLE), laplacian eigenmaps (LE), and local tangent space alignment (LTSA). Although the above methods are successful in many applications, their recognition performance may decrease when objects as HRRPs suffer from the large within-class variation due to the fact that these methods are often in lack of robustness and generalization.

Inspired by the maximum margin of SVM, A. Kocsor et al. [37] propose the margin maximizing discriminant analysis (MMDA) approach. The core of MMDA is to maximize the between-class margin on the decision boundary by applying the normals of a set of pairwise orthogonal margin maximizing hyperplanes to construct a projection subspace. But MMDA is only fit for binary classification problem and cannot be applied for multi-class classification problem directly. Based on the similar idea, H. F. Li et al. [38] present the maximum margin criterion (MMC) method. The aim of MMC is to maximize the trace of the difference of the between-class scatter matrix and within-class scatter matrix. It can be applied for multiclass classification directly and avoid the small sample size (SSS) problem. However, the coordinate axes of MMC subspace are not optimal in meaning of maximum margin due to the fact that they are solved by (SVD) on the difference between the between-class scatter matrix and within-class scatter matrix without exerting the constraints. Thus, its performance can be improved further.

In this paper, a novel target recognition method, namely orthogonal maximum margin projection subspace (OMMPS), is proposed for radar target HRRP recognition. The aim of OMMPS is to maximize the between-class margin by increasing the between-class scatter distance and reducing the within-class scatter distance simultaneously. By exerting the orthogonality constraint on the objective function, we can solve the OMMPS. The OMMPS has three advantages: First, the number of features does not depend on the number of classes. As a result, the appropriate features can still be obtained for separating the classes, even in high-dimensional space with only a few classes. Second, the coordinate axes

of OMMPS are optimal in meaning of maximum margin because the coordinate axes are solved sequentially by exerting the orthogonality constraint on the objective function. Third, the coordinate axes of OMMPS are mutually orthogonal and the features extracted by OMMPS reduce the redundancy, thus improves the recognition performance.

2 OMMPS

Let $\mathbf{X} = [\mathbf{x}_{11} \cdots \mathbf{x}_{1N_1} \cdots \mathbf{x}_{C1} \cdots \mathbf{x}_{CN_C}]$ denotes a training sample set, $\mathbf{x}_{ij}$ is the jth n-dimensional HRRP vector of ith class. Each class contains N_i training samples, and the number of total training sample for C classes is $N(N = N_1 + N_2 + \cdots + N_C)$. Let $\mathbf{A}$ represents a $n \times m$-dimensional matrix $(m < n)$. Projecting $\mathbf{x}_{ij}$ into m-dimensional feature subspace below

$$\mathbf{y}_{ij} = \mathbf{A}^T \mathbf{x}_{ij} \tag{1}$$

where T denotes transposition, and $\mathbf{y}_{ij}$ is a m-dimensional vector, namely subprofile of $\mathbf{x}_{ij}$. Firstly, computing the between-class scatter distance d_B in subprofile space

$$
\begin{aligned}
d_B &= \frac{1}{2} \sum_{i=1}^{C} \sum_{k=1}^{C} \frac{N_i N_k}{N^2} (\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_k)^T (\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_k) \\
&= \sum_{i=1}^{C} \frac{N_i}{N} (\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^T (\bar{\mathbf{y}}_i - \bar{\mathbf{y}}) \\
&= Tr \left(\sum_{i=1}^{C} \frac{N_i}{N} (\bar{\mathbf{y}}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^T \right)
\end{aligned}
\tag{2}
$$

where $Tr(\cdot)$ is the trace of matrix. $\bar{\mathbf{y}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{y}_{ij}$, $\bar{\mathbf{y}}_k = \frac{1}{N_k} \sum_{k=1}^{N_k} \mathbf{y}_{kj}$, and $\bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^{C} \sum_{j=1}^{N_i} \mathbf{y}_{ij}$ are the mean vectors of ith class' training subprofiles, kth class' training subprofiles, and total training subprofiles, respectively. Substituting Eq. (1) into Eq. (2), it follows that

$$
\begin{aligned}
d_B &= Tr \left(\mathbf{A}^T \left(\sum_{i=1}^{C} \frac{N_i}{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \right) \mathbf{A} \right) \\
&= Tr(\mathbf{A}^T \mathbf{S}_B \mathbf{A})
\end{aligned}
\tag{3}
$$

where $\bar{\mathbf{x}}_i$ is the mean vector of ith class' training samples, and $\bar{\mathbf{x}}$ is the mean vector of total training samples. $\mathbf{S}_B$ is the between-class scatter matrix in original sample space

$$\mathbf{S}_B = \sum_{i=1}^{C} \frac{N_i}{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \tag{4}$$

Secondly, computing the within-class scatter distance d_W in subprofile space

$$
\begin{aligned}
d_W &= \sum_{i=1}^{C} \left(\frac{N_i}{N} \sum_{j=1}^{N_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)^T (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i) \right) \\
&= Tr \left(\sum_{i=1}^{C} \left(\frac{N_i}{N} \sum_{j=1}^{N_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)^T \right) \right)
\end{aligned}
\tag{5}
$$

Substituting Eq. (1) into Eq. (5), we can get

$$
\begin{aligned}
d_W &= Tr \left(\mathbf{A}^T \sum_{i=1}^{C} \left(\frac{N_i}{N} \sum_{j=1}^{N_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T \right) \mathbf{A} \right) \\
&= Tr(\mathbf{A}^T \mathbf{S}_W \mathbf{A})
\end{aligned}
\tag{6}
$$

where $\mathbf{S}_W$ is the within-class scatter matrix in original sample space

$$\mathbf{S}_W = \sum_{i=1}^{C} \left(\frac{N_i}{N} \sum_{j=1}^{N_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T \right) \tag{7}$$

According to the geometric structure in subprofile space, we define the between-class margin in subprofile space below

$$d_M = d_B - d_W = Tr(\mathbf{A}^T (\mathbf{S}_B - \mathbf{S}_W) \mathbf{A}) \tag{8}$$

where d_M is the between-class margin.

The aim of OMMPS is to seek an orthogonal projection subspace by maximizing the between-class margin based on the orthogonality constraint, i.e., solving the following maximization problem

$$
\begin{aligned}
\mathbf{a}_r &= \arg \max_{\{\mathbf{a}_r\}} d_M \\
&= \arg \max_{\{\mathbf{a}_r\}} \{ Tr(\mathbf{a}_r^T (\mathbf{S}_B - \mathbf{S}_W) \mathbf{a}_r) \} \\
&= \arg \max_{\{\mathbf{a}_r\}} \{ \mathbf{a}_r^T (\mathbf{S}_B - \mathbf{S}_W) \mathbf{a}_r \} \\
& \quad r = 1, 2, \cdots, m
\end{aligned}
\tag{9}
$$

and

$$\mathbf{a}_r^T \mathbf{a}_r = 1 \tag{10}$$

$$\mathbf{a}_l^T \mathbf{a}_r = 0 \quad l = 1, 2, \cdots, r-1 \tag{11}$$

where $\mathbf{a}_r$ is the column vector of matrix $\mathbf{A}$, i.e., $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2 \dots \mathbf{a}_m]$, namely orthogonal maximum margin projection subspace (OMMPS). Although the objective function in Eq. (9) is similar to that of MMC [38], the objective function of MMC does not include the orthogonality constrains. Besides, MMC obtains the projection subspace using the eigenvectors corresponding to the first largest eigenvalues of matrix $(\mathbf{S}_B - \mathbf{S}_w)$, and thus, the projection vectors of MMC are not optimal in meaning of maximum margin. We solve the above optimization problem by following steps.

To solve $\mathbf{a}_1$, we construct a Lagrangian function using Eqs. (9) and (10)

$$J(\mathbf{a}_1, \lambda_1) = \mathbf{a}_1^T (\mathbf{S}_B - \mathbf{S}_W) \mathbf{a}_1 - \lambda_1 (\mathbf{a}_1^T \mathbf{a}_1 - 1) \qquad (12)$$

where λ_1 is a Lagrangian multiplier. Taking the vector derivative of $J(\mathbf{a}_1, \lambda_1)$ with respect to $\mathbf{a}_1$ and set the resultant equation to zero, we can get the generalized eigenvector equation

$$(\mathbf{S}_B - \mathbf{S}_W) \mathbf{a}_1 = \lambda_1 \mathbf{a}_1 \qquad (13)$$

Let $\lambda_1^{\max}$ is the largest eigenvalue of matrix $(\mathbf{S}_B - \mathbf{S}_W)$ and $\boldsymbol{\mu}_1^{\max}$ is the corresponding eigenvector; then, we may set

$$\mathbf{a}_1 = \boldsymbol{\mu}_1^{\max} \qquad (14)$$

After obtaining the $\mathbf{a}_1$, combining Eqs. (9)–(11), we can form the Lagrangian function

$$J(\mathbf{a}_r, \lambda_1, \lambda_2 \cdots \lambda_{r-1}, \lambda_r) = \mathbf{a}_1^T (\mathbf{S}_B - \mathbf{S}_W) \mathbf{a}_1 - \lambda_1 \mathbf{a}_1^T \mathbf{a}_r - \lambda_2 \mathbf{a}_2^T \mathbf{a}_r \cdots$$
$$- \lambda_{r-1} \mathbf{a}_{r-1}^T \mathbf{a}_r - \lambda_r (\mathbf{a}_r^T \mathbf{a}_r - 1)$$
$$(15)$$

where $\lambda_1, \lambda_2, \lambda_{r-1}$, and λ_r are Lagrangian multipliers. In a similar way, taking the derivative of $J(\mathbf{a}_r, \lambda_1, \lambda_2 \cdots \lambda_{r-1}, \lambda_r)$ in Eq. (15) with respect to $\mathbf{a}_r$ and $\lambda_l (l = 1, 2, \cdots, r)$, solving the resultant equation leads to

$$(\mathbf{I} - (\mathbf{a}_1 \mathbf{a}_1^T + \mathbf{a}_2 \mathbf{a}_2^T + \cdots + \mathbf{a}_{r-1} \mathbf{a}_{r-1}^T))(\mathbf{S}_B - \mathbf{S}_W) \mathbf{a}_r = \lambda_r \mathbf{a}_r \quad (16)$$

Let $\lambda_r^{\max}$ is the largest eigenvalue of Eq. (16) and $\boldsymbol{\mu}_r^{\max}$ is the corresponding eigenvector; then, we can set

$$\mathbf{a}_r = \boldsymbol{\mu}_r^{\max} \qquad (17)$$

According to the above discussion, it is obvious that the basis vectors of OMMPS are solved sequentially by exerting the orthogonality constraint on the objective function. As a result, they are mutually orthogonal and optimal in meaning of maximum margin. Therefore, OMMPS has better discriminative power than MMC. The steps of feature extraction based on OMMPS are shown in Algorithm 1.

Algorithm 1. The feature extraction based on OMMPS

Task: Solve the linear subprofile features using the training data set

$$\mathbf{X} = [\mathbf{x}_{11} \cdots \mathbf{x}_{1N_1} \cdots \mathbf{x}_{C1} \cdots \mathbf{x}_{CN_C}]$$

Step 1) Determine the subprofile's dimensionality m
Step 2) Compute the matrix $\mathbf{S}_B$ and $\mathbf{S}_W$ by equation (4) and (7)
Step 3) SVD to matrix $(\mathbf{S}_B - \mathbf{S}_W)$, and obtain the $\mathbf{a}_1$ by equation (14)
Step 4) SVD to matrix $(\mathbf{I} - (\mathbf{a}_1 \mathbf{a}_1^T + \cdots + \mathbf{a}_{r-1} \mathbf{a}_{r-1}^T))$ $(\mathbf{S}_B - \mathbf{S}_W)$ for $r = 2$, and obtain the $\mathbf{a}_2$ by equation (17)

Step 5) Set $r = r + 1$ repeat Step 4 until $\mathbf{a}_m$ is obtained. Then $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 \cdots \mathbf{a}_m]$
Step 6) Obtain the linear subprofile of HRRP vector $\mathbf{x}$ using equation (1)

3 Orthogonal kernel maximum margin projection subspace (QKMMPS)

When nonlinear variations in HRRPs are very serious, the HRRPs of different classes may not be separable linearly. We introduce the nonlinear mapping to solve this problem. A nonlinear function ϕ is used to map $\mathbf{x}_{ij}$ into a high-dimensional feature space F below

$$R^n : \mathbf{x}_{ij} \rightarrow F : \phi(\mathbf{x}_{ij}) \qquad (18)$$

where the dimensionality of feature space F is n'; here, n' may be any value or infinite. Let $\mathbf{A}_\phi$ denotes a $n' \times m^\phi$-dimensional transformation matrix, namely orthogonal kernel maximum margin projection subspace; then, $\phi(\mathbf{x}_{ij})$ is projected into m^ϕ dimensional space as follow

$$\mathbf{y}_{ij}^\phi = \mathbf{A}_\phi^T \phi(\mathbf{x}_{ij}) \qquad (19)$$

where $\mathbf{y}_{ij}^\phi$ is m^ϕ-dimensional column vector, namely non-linear subprofile of HRRP vector $\mathbf{x}_{ij}$ in low-dimensional feature space. In a similar way, we can compute the between-class margin d_M^ϕ in nonlinear subprofile space

$$d_M^\phi = Tr\left(\mathbf{A}_\phi^T (\mathbf{S}_B^\phi - \mathbf{S}_W^\phi) \mathbf{A}_\phi \right) \qquad (20)$$

where $\mathbf{S}_B^\phi$ and $\mathbf{S}_W^\phi$ are the between-class scatter matrix and within-class scatter matrix in high-dimensional feature space F, respectively.

$$\mathbf{S}_B^\phi = \sum_{i=1}^{C} \frac{N_i}{N} \left(\bar{\mathbf{x}}_i^\phi - \bar{\mathbf{x}}^\phi \right) \left(\bar{\mathbf{x}}_i^\phi - \bar{\mathbf{x}}^\phi \right)^T \qquad (21)$$

$$\mathbf{S}_W^\phi = \sum_{i=1}^{C} \left(\frac{N_i}{N} \sum_{j=1}^{N_i} \left(\mathbf{x}_{ij}^\phi - \bar{\mathbf{x}}_i^\phi \right) \left(\mathbf{x}_{ij}^\phi - \bar{\mathbf{x}}_i^\phi \right)^T \right) \qquad (22)$$

where $\mathbf{x}_{ij}^\phi = \phi(\mathbf{x}_{ij})$, $\bar{\mathbf{x}}_i^\phi = (1/N_i) \sum_{j=1}^{N_i} \phi(\mathbf{x}_{ij})$, and $\bar{\mathbf{x}}^\phi = (1/N) \sum_{i=1}^{C} \sum_{j=1}^{N_i} \phi(\mathbf{x}_{ij})$. Based on the aim of orthogonal kernel maximum margin projection subspace (OKMMPS), we may get OKMMPS by solving the following constrained maximization problem

$$\begin{aligned}
\mathbf{a}_r^\phi &= \arg\max_{\{\mathbf{a}_r^\phi\}} d_M^\phi \\
&= \arg\max_{\{\mathbf{a}_r^\phi\}} \{ Tr((\mathbf{a}_r^\phi)^T (\mathbf{S}_B^\phi - \mathbf{S}_W^\phi)^W \mathbf{a}_r^\phi) \} \\
&= \arg\max_{\{\mathbf{a}_r^\phi\}} \{ (\mathbf{a}_r^\phi)^T (\mathbf{S}_B^\phi - \mathbf{S}_W^\phi)^W \mathbf{a}_r^\phi \} \\
& \quad r = 1, 2, \cdots, m
\end{aligned}$$

(23)

and

$$(\mathbf{a}_r^\phi)^T \mathbf{a}_r^\phi = 1 \tag{24}$$

$$(\mathbf{a}_l^\phi)^T \mathbf{a}_r^\phi = 0 \quad l = 1, 2, \cdots, r-1 \tag{25}$$

where $\mathbf{a}_r^\phi$ is the column vector of matrix $\mathbf{A}_\phi$, i.e., $\mathbf{A}_\phi = [\mathbf{a}_1^\phi \mathbf{a}_2^\phi \cdots \mathbf{a}_m^\phi]$, namely OKMMPS. Because the expression of nonlinear mapping $\phi(\cdot)$ is not defined explicitly, it is impossible to solve the Eq. (23) for obtaining OKMMPS directly. We use kernel trick to solve this problem.

Let

$$\mathbf{a}_r^\phi = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \alpha_{rij} \phi(\mathbf{x}_{ij}) \tag{26}$$

and

$$k(\mathbf{x}_{ij}, \mathbf{x}_{lk}) = \phi^T(\mathbf{x}_{ij}) \phi(\mathbf{x}_{lk}) \tag{27}$$

where α_{rij} is a coefficient, $\mathbf{x}_{ij}$ and $\mathbf{x}_{lk}$ are n-dimensional column vectors, and $k(\mathbf{x}_{ij}, \mathbf{x}_{lk})$ is a kernel function. Substituting Eqs. (26) and (27) into Eqs. (23)–(25), it follows that

$$\boldsymbol{\alpha}_r = \max_{\{\alpha_r\}} \left\{ (\boldsymbol{\alpha}_r)^T (\mathbf{S}_B^\alpha - \mathbf{S}_W^\alpha)^W \boldsymbol{\alpha}_r \right\} \qquad r = 1, 2, \cdots, m \tag{28}$$

and

$$\boldsymbol{\alpha}_r^T \mathbf{K} \boldsymbol{\alpha}_r = 1 \tag{29}$$

$$\boldsymbol{\alpha}_l^T \mathbf{K} \boldsymbol{\alpha}_r = 0 \quad l = 1, 2, \cdots, r-1 \tag{30}$$

where

$$\boldsymbol{\alpha}_r = [\alpha_{r11} \ \alpha_{r12} \ \cdots \ \alpha_{rCN_C}]^T \tag{31}$$

$$\mathbf{S}_B^\alpha = \sum_{i=1}^{C} \frac{N_i}{N} (\mathbf{P}_i - \mathbf{P})(\mathbf{P}_i - \mathbf{P})^T \tag{32}$$

$$\mathbf{S}_W^\alpha = \sum_{i=1}^{C} \frac{N_i}{N} \sum_{j=1}^{N_i} \left((\mathbf{K})_{ij} - \mathbf{P}_i \right) \left((\mathbf{K})_{ij} - \mathbf{P}_i \right)^T \tag{33}$$

where

$$(\mathbf{P}_i)_{lj} = \frac{1}{N_i} \sum_{k=1}^{N_i} k(\mathbf{x}_{lj}, \mathbf{x}_{ik}) \quad i, l = 1, 2, \cdots, C \quad j = 1, 2, \cdots, N_l \tag{34}$$

$$\begin{aligned}
\left((\mathbf{K})_{ij} \right)_{lk} &= k(\mathbf{x}_{lk}, \mathbf{x}_{ij}) \\
i, l &= 1, 2, \cdots, C \\
j &= 1, 2, \cdots, N_i \quad k = 1, 2, \cdots, N_l
\end{aligned} \tag{35}$$

Combining Eqs. (28) and (29), we can construct the following function for getting $\boldsymbol{\alpha}_1$

$$J(\boldsymbol{\alpha}_1, \gamma_1) = \boldsymbol{\alpha}_1^T (\mathbf{S}_B^\phi - \mathbf{S}_W^\iota) \boldsymbol{\alpha}_1 - \lambda_1 (\boldsymbol{\alpha}_1^T \mathbf{K} \boldsymbol{\alpha}_1 - 1) \tag{36}$$

where γ_1 is a Lagrangian multiplier. Taking the vector derivative of $J(\boldsymbol{\alpha}_1, \gamma_1)$ with respect to $\boldsymbol{\alpha}_1$ and set the resultant equation to zero, we can get the generalized eigenvector equation

$$\mathbf{K}^{-1} (\mathbf{S}_B^\phi - \mathbf{S}_W^\phi) \boldsymbol{\alpha}_1 = \gamma_1 \boldsymbol{\alpha}_1 \tag{37}$$

Similar to the observation in Section 2, we set

$$\boldsymbol{\alpha}_1 = \boldsymbol{\mu}_1^{\alpha, \max} \tag{38}$$

where $\boldsymbol{\mu}_1^{\alpha, \max}$ is the eigenvector corresponding to the largest eigenvalue $\gamma_1^{\max}$ of matrix $\mathbf{K}^{-1}(\mathbf{S}_B^\phi - \mathbf{S}_W^\phi)$.

Combining Eqs. (28), (29), and (30), the function is constructed using Lagrangian multipliers to solve $\boldsymbol{\alpha}_r (2 \leq r \leq m)$

$$\begin{aligned}
J(\boldsymbol{\alpha}_r, \gamma_1, \gamma_2 \cdots \gamma_{r-1}, \gamma_r) &= \boldsymbol{\alpha}_1^T (\mathbf{S}_B^\phi - \mathbf{S}_W^\phi) \boldsymbol{\alpha}_1 - \gamma_1 \boldsymbol{\alpha}_1^T \mathbf{K} \boldsymbol{\alpha}_r - \gamma_2 \boldsymbol{\alpha}_2^T \mathbf{K} \boldsymbol{\alpha}_r \cdots \\
& \quad - \gamma_{r-1} \boldsymbol{\alpha}_{r-1}^T \mathbf{K} \boldsymbol{\alpha}_{r-1} - \gamma_r (\boldsymbol{\alpha}_r^T \mathbf{K} \boldsymbol{\alpha}_r - 1)
\end{aligned} \tag{39}$$

where γ_1, γ_2, γ_{r-1}, and γ_r are Lagrangian multipliers. In a similar way, we can get the following eigenvector equation

$$\mathbf{K}^{-1} (\mathbf{I} - (\mathbf{K} \boldsymbol{\alpha}_1 \boldsymbol{\alpha}_1^T + \mathbf{K} \boldsymbol{\alpha}_2 \boldsymbol{\alpha}_2^T + \cdots + \mathbf{K} \boldsymbol{\alpha}_{r-1} \boldsymbol{\alpha}_{r-1}^T)) (\mathbf{S}_B^\alpha - \mathbf{S}_W^\alpha) \boldsymbol{\alpha}_1 = \gamma_r \boldsymbol{\alpha}_r \tag{40}$$
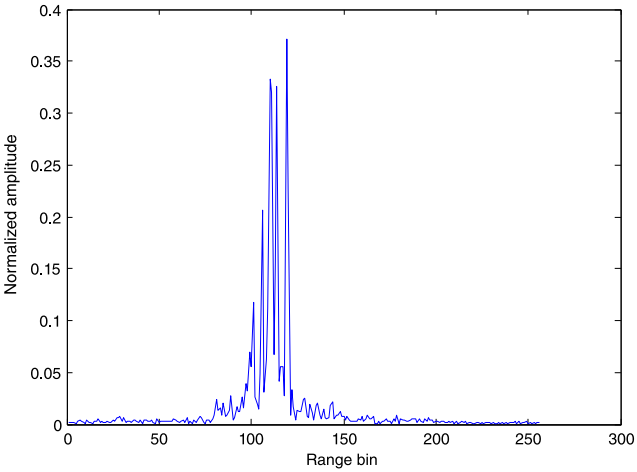
Let $\gamma_r^{\max}$ is the largest eigenvalue of Eq. (40) and $\boldsymbol{\mu}_r^{\alpha, \max}$ is the corresponding eigenvector; then, we set

$$\boldsymbol{\alpha}_r = \boldsymbol{\mu}_r^{\alpha, \max} \tag{41}$$

After obtaining $\mathbf{S}_B^\alpha \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \cdots, \boldsymbol{\alpha}_m$, $\phi(\mathbf{x})$ is projected into the nonlinear subprofile space according to Eq. (19); it follows that

$$\mathbf{y}^\phi = [\boldsymbol{\alpha}_1 \quad \boldsymbol{\alpha}_2 \cdots \boldsymbol{\alpha}_m]^T \begin{bmatrix} k(\mathbf{x}_{11}, \mathbf{x}) \\ k(\mathbf{x}_{12}, \mathbf{x}) \\ \vdots \\ k(\mathbf{x}_{CN_C}, \mathbf{x}) \end{bmatrix} \tag{42}$$
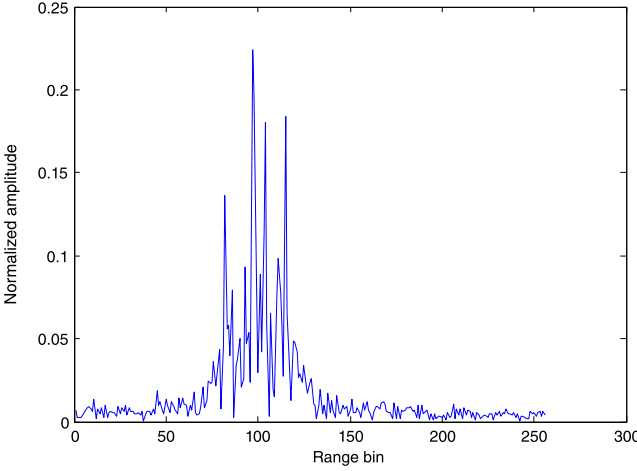
where $\mathbf{y}^\phi$ is the corresponding nonlinear subprofile of $\mathbf{x}$. The steps of feature extraction based on OKMMPS are shown in Algorithm 2.
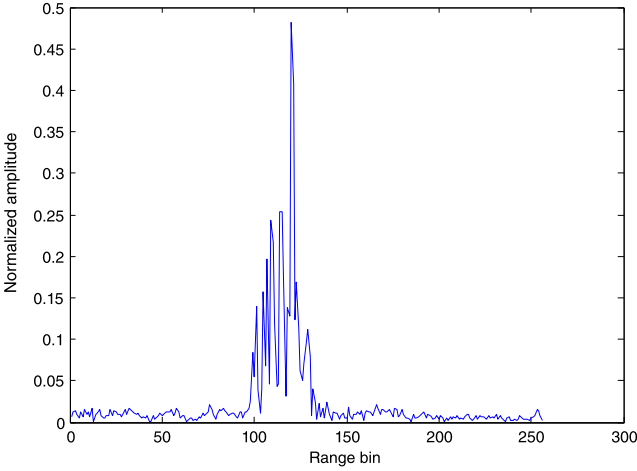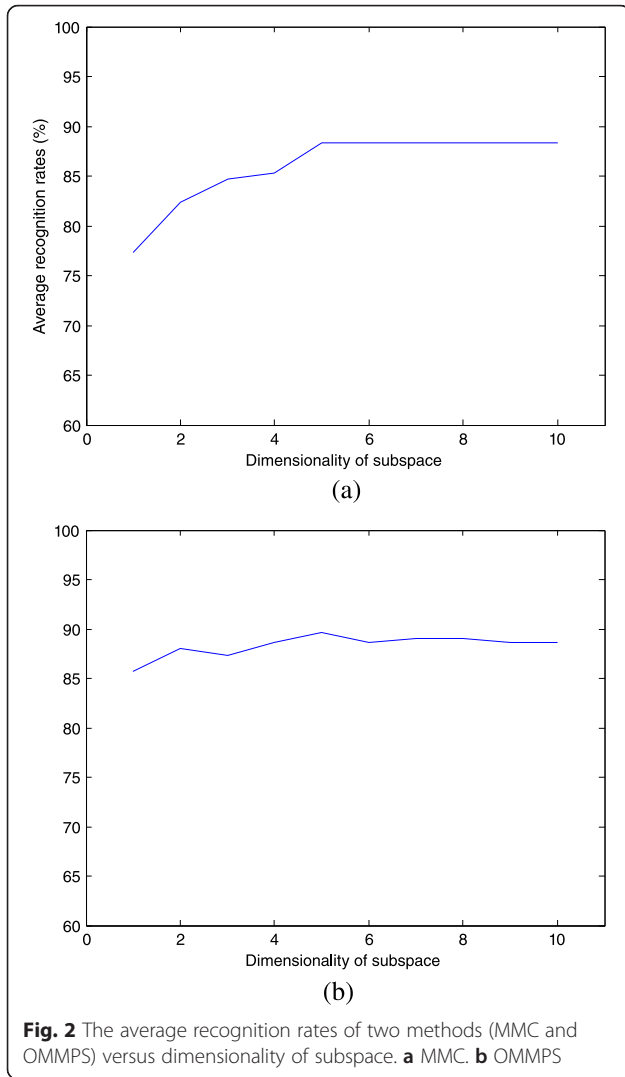
Fig. 1 The HRRPs of the three airplanes. **a** An-26. **b** Jiang. **c** Yark-42

Fig. 2 The average recognition rates of two methods (MMC and OMMPS) versus dimensionality of subspace. **a** MMC. **b** OMMPS

Algorithm 2. The nonlinear feature extraction based on OKMMPS

Task: Solve the nonlinear subprofile features using the training data set

$$\mathbf{X} = [\mathbf{x}_{11} \cdots \mathbf{x}_{1N_1} \cdots \mathbf{x}_{C1} \cdots \mathbf{x}_{CN_C}]$$

Step 1) Determine the subprofile's dimensionality m
Step 2) Select the kernel function
Step 3) Compute the matrix $\mathbf{K}$, and $\mathbf{S}_W^\alpha$ by equation (27), equation (32) and (33)

Step 4) SVD to matrix $\mathbf{K}^{-1}(\mathbf{S}_B^\alpha - \mathbf{S}_W^\alpha)$, and obtain the $\boldsymbol{\alpha}_1$ by equation (38)
Step 5) SVD to matrix $\mathbf{K}^{-1}(\mathbf{I} - (\mathbf{K}\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^T + \cdots + \mathbf{K}\boldsymbol{\alpha}_{r-1}\boldsymbol{\alpha}_{r-1}^T))(\mathbf{S}_B^\alpha - \mathbf{S}_W^\alpha)$ for $r = 2$ and obtain the $\boldsymbol{\alpha}_2$ by equation (41)
Step 6) Set $r = r + 1$ repeat Step 5 until $\boldsymbol{\alpha}_m$ is obtained
Step 7) Obtain the nonlinear subprofile of HRRP vector $\mathbf{x}$ using equation (42)

4 Experimental results

To show the effectiveness of the proposed method, we perform the extensive experiments on the measured data of three kinds of airplanes.

4.1 Data description

The data used in experiments are HRRPs measured from three airplanes, including An-26, Jiang, and Yark-42. For each airplane, 240 HRRPs over a wide range of aspects are adopted. For each airplane, one quarter of all HRRPs are used for training and the rest are used for testing. Before running experiments, each HRRP is preprocessed by energy normalization. The HRRPs of three airplanes are illustrated in Fig. 1.

4.2 The dimensionality of subspace

In this experiment, we consider the effect of subspace's dimensionality on recognition performance. The training data and testing data are as described above. The subspace's dimensionality is set from 1 to 10. The nearest-neighbor classifier is applied for classification. Two kernels are used, i.e., radial basis function kernel (RBFK)

$$k(\mathbf{x}, \mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\sigma^2}} \tag{43}$$

and polynomial function kernel (PFK)

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^d \tag{44}$$

where the kernel parameters σ and d are set by the cross-validation method.

Figure 2 shows the average recognition rates of two methods (MMC [38] and OMMPS) versus the subspace' dimensionality. From Fig. 2a, it can be seen that there is a big rise in the average recognition rate when the subspace' dimensionality is increased from 1 to 5, and the average recognition rates keep same approximate when the subspace' dimensionality is above 5. Thus, the proper

Table 1 The average recognition rates along with the dimensionalities using radial basis function kernel (%)

Method	$\sigma = 5$	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$	$\sigma = 40$	$\sigma = 50$
KPCA	69.33 (5)	82.00 (15)	83.00 (7)	83.33 (10)	*84.00 (10)*	84.00 (10)
KFDA	76.33 (2)	*85.42 (2)*	75.67 (2)	78.33 (2)	82.00 (2)	81. 00 (2)
OKMMPS	73.00 (10)	75.67 (2)	*90.33 (50)*	88.67 (60)	88.00 (75)	87.33 (80)

The italicized numbers represent the highest recognition rates of one method for different parameter value

Table 2 The average recognition rates using polynomial function kernel (%)

Method	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$
KPCA	82.33 (8)	83.00 (7)	83.00 (10)	81.67 (10)	80.33 (10)	80.00 (5)
KFDA	79.67 (2)	77.00 (2)	75.33 (2)	78.00 (2)	84.33 (2)	83.00 (2)
OKMMPS	88.67 (67)	87..67 (74)	87.67 (80)	87.67 (80)	86.67 (80)	86.33 (80)

The italicized numbers represent the highest recognition rates of one method for different parameter value

dimensionality of MMC is set as 5 in the following experiments. From Fig. 2b, the appropriate dimensionality of OMMPS can also be set as 5. In a similar way, the proper dimensionality of OKMMPS with RBFK and PFK are set as 50 and 67, respectively.

4.3 Kernel parameters
In this experiment, we set the appropriate parameters for kernel methods such as OKMMPS, KPCA [22], and KFDA [23] by the cross-validation method. For radial basis function kernel, the parameter σ is set as 5, 10, 20, 30, 40, and 50. For polynomial function kernel, the parameter d varies from 1 to 10. The training data and test data are the same as previous experiments. The nearest-neighbor classifier is applied for classification. The experiment is run for each parameter. Tables 1 and 2 illustrate the average recognition rates along with the dimensionalities of three kernel methods for varying value of parameters. As can be seen in Tables 1 and 2, OKMMPS achieves the best recognition results when the radial basis function kernel with $\sigma = 20$ and the polynomial function kernel with $d = 1$ are selected for OKMMPS. We choose the parameters for other kernel methods in similar way. The best kernel parameters chosen for the methods mentioned above are shown in Table 3. In addition, it can also be observed that the methods with the radial basis function kernel have higher recognition rates than those with polynomial function kernel. It shows that the radial basis function kernel can well represent the nonlinearity appearing in HRRP samples for these data.

4.4 The variation of target aspect
The HRRPs change largely when the target aspect varies with a few degrees, which increase the difficulty in classifying the targets. In this experiment, we consider the robustness of MMC, OMMPS, and OKMMPS to the variation of target aspect. The training data is the same as the previous experiments. Three subsets of testing data are selected, including 300, 420, and 540 HRRPs,

Table 3 The best kernel parameters for kernel-based methods

Method	Radial basis function kernel	Polynomial function kernel
	σ	d
KPCA	40	2
KFDA	10	5
OKMMPS	20	1

respectively. For each class, 100, 140, and 180 HRRPs are chosen for three subsets of testing data, respectively. It is obvious that the variation of target aspect becomes large when the number of HRRPs increases. The radial basis function kernel is used. The parameters of these methods are set according to the above experiments. The dimensionality of MMC, OMMPS, and OKMMPS is 5, 5, and 50, respectively. The parameter of radial basis function kernel for OKMMPS is set as 20. The nearest-neighbor classifier is applied for classification. The recognition results of three methods for three subsets of testing data are illustrated in Fig. 3. From Fig. 3, it is shown that the average recognition rates decrease when the number of testing samples is increased from 100 to 180, i.e., the variation of target aspect becomes large. However, the recognition rates of OMMPS and OKMMPS are still better than those of MMC for three subsets. This means that OMMPS and OKMMPS are more robust to variation of aspect than MMC. The reason is that the basis vectors of OMMPS and OKMMPS are obtained by solving the optimization problem sequentially, and they are optimal in meaning of maximum margin. Thus, the high classification accuracy can be obtained when the within-class scatter is large due to big change of HRRPs.

4.5 Performance comparison
To show the effectiveness of the proposed method further, we evaluate the performance of OMMPS and OKMMPS compared with MMC [38], PCA [20], LDA [21], KPCA [22], and KFDA [23] under different SNR. The SNR is set as 5, 10, 15, 20, 25, and 30 dB. For each SNR, the recognition results are averaged for 50 run. The dimensionality of subspace for MMC, OMMPS, OKMMPS, PCA, LDA, KPCA, and KFDA is 5, 5, 50, 26, 2, 10, and 2, respectively. The radial basis function kernel is used. According to the experimental results of subsection 4.3, the kernel parameter for OKMMPS, KPCA, and KFDA is set as 20, 40, and 10, respectively. The nearest-neighbor classifier is applied for classification. Figure 4 shows the average rates of seven methods versus SNR. Some interesting observations can be seen from Fig. 4.

(1) When SNR is above 15 dB, the kernel methods such as OKMMPS, KFDA, and KPCA outperform the corresponding linear methods such as OMMPS, LDA, and PCA. At SNR = 15 dB, the average recognition rates of OKMMPS, KFDA, KPCA, OMMPS, LDA, and PCA are
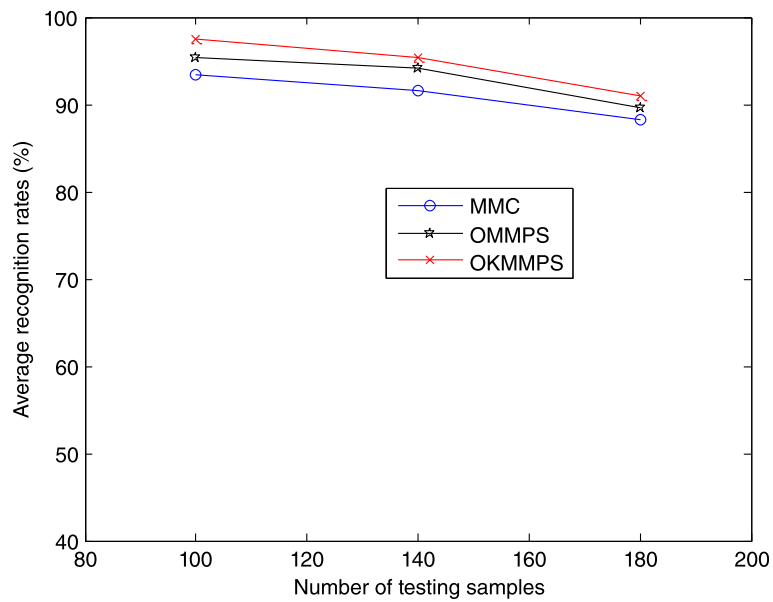
Fig. 3 The average recognition rates of three methods versus the number of testing samples

86.52, 81, 79.67, 85.33, 80, and 79.33 %. It shows that the kernel methods are more robust to noise than the linear methods. This is because the nonlinearity in HRRPs is very obvious due to the effect of noise, and the kernel methods can well represent the nonlinearity variation appearing in HRRP samples by nonlinear mapping. Thus, the separability between the different classes can be improved.

(2) MMC has better recognition performance than LDA for all SNR level when the number of training data

is much less than the dimensionality of HRRP. At SNR = 15 dB, the average recognition rates of MMC and LDA are 83.33 and 80 %, respectively. This demonstrates that MMC has better discriminative power than LDA for small size of training data. The reason is that LDA suffers from small sample size (SSS) problem in the case of small size of training data. However, MMC does not need the inversion of the within-class scatter matrix and may avoid the SSS problem. As a result, the features extracted by MMC are more robust.
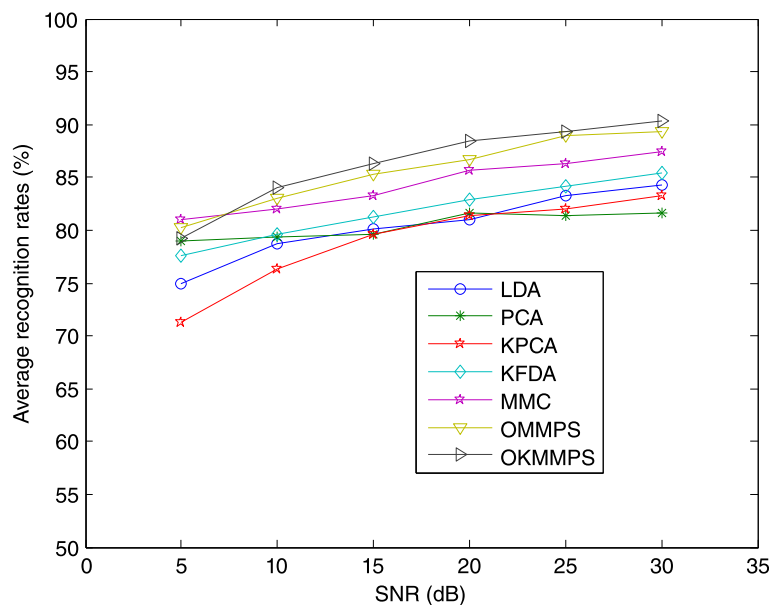

Fig. 4 The average recognition rates of seven methods versus SNR

(3) The discriminative ability of OMMPS and OKMMPS is superior to that of MMC when the SNR is from 5 to 30 dB. At SNR = 15 dB, the average recognition rates of OMMPS, OKMMPS, and MMC are 85.23, 86.42, and 83.33 %, respectively. The reason is that the basis vectors of OMMPS and OKMMPS are obtained by solving the optimization problem sequentially and they are optimal in meaning of maximum margin. Especially, the basis vectors of OKMMPS are still orthogonal in high-dimensional feature space. It means that the features extracted by OMMPS and OKMMPS are more discriminative than those extracted by MMC.

5 Conclusions

In this paper, we propose a novel radar target recognition method using HRRP, namely orthogonal maximum margin projection subspace (OMMPS). The kernel version, called as orthogonal kernel maximum margin projection subspace (OKMMPS), is also derived. The proposed method is able to maximize the between-class margin by increasing the between-class scatter distance and reducing the within-class scatter distance simultaneously. The experimental results on the measured data of three kinds of planes show that

(1) OMMPS and OKMMPS can still obtain the appropriate dimensionality of subspace for high-dimensional HRRP vector with three classes.
(2) The radial basis function kernel can better represent the nonlinearity appearing in HRRP samples than the polynomial function kernel.
(3) OMMPS and OKMMPS are more robust to the variation of target aspect than MMC method.
(4) OMMPS and OKMMPS have higher recognition performance than the other methods.

Abbreviations
HRRP: high-resolution range profile; MMC: maximum margin criterion; OKMMPS: orthogonal kernel maximum margin projection subspace; OMMPS: orthogonal maximum margin projection subspace.

Competing interests
The authors declare that they have no competing interests.

References
1. HJ Li, SH Yang, Using range profiles as feature vectors to identify aerospace objects. IEEE Trans. Antennas. Propag. **41**(March), 261–268 (1993)
2. KB Eom, R Chellappa, Noncooperative target classification using hierarchical modeling of high-range resolution radar signatures. IEEE Trans. Signal Process. **45**(September), 2318–2326 (1997)
3. SP Jacobs, JA Sullivan, Automatic target recognition using sequences of high range resolution radar range profiles. IEEE Trans. Aerosp. Electron. Syst **36**, 364–381 (2000)
4. A Zyweck, RE Bogner, Radar target classification of commercial aircraft. IEEE Trans. Aerosp. Electron. Syst. **32**(February), 598–606 (1996)
5. AK Shaw, R Vasgist, R Williams, HRR-ATR using eigen-templates with observation in unknown target scenario. Proc. SPIE **4053**, 467–478 (2000)
6. BM Huther, SC Gustafson, RP Broussad, Wavelet preprocessing for high range resolution radar classification. IEEE. Trans. Aerosp. Electron. Syst **37**, 1321–1331 (2001)
7. R Wu, Q Gao, J Liu, H Gu, ATR scheme based on 1-D HRR profiles. Electronics Letters. **38**(December), 1586–1587 (2002)
8. KT Kim, DK Seo, HT Kim, Efficient radar target recognition using the MUSIC algorithm and invariant feature. IEEE Trans. Antennas. Propag. **50**(March), 325–337 (2002)
9. J Zwart, R Heiden, S Gelsema, F Groen, Fast translation invariant classification of HRR range profiles in a zero phase representation. IEE Proc. Radar Sonar Navig. **150**(June), 411–418 (2003)
10. Y Shi, XD Zhang, A Gabor atom network for signal classification with application in radar target recognition. IEEE Trans. Signal. Process. **49**(December), 2994–3004 (2001)
11. SK Wong, Non-cooperative target recognition in the frequency domain. IEE Proc. Radar Sonar Navig. **151**(February), 77–84 (2004)
12. DE Nelson, JA Starzyk, DD Ensley, Iterated wavelet transformation and discrimination for HRR radar target recognition. IEEE Trans. System Man and Cybernetics-part: system and humans **33**(January), 52–57 (2003)
13. RA Mitchell, JJ Westerkamp, Robust statistical feature based aircraft identification. IEEE Trans. Aerosp. Electron. Syst. **35**(March), 1077–1093 (1999)
14. XJ Liao, P Runkle, L Carin, Identification of ground targets from sequential high-range-resolution radar signatures. IEEE Trans. Aerosp. Electron. Syst. **38**(April), 1230–1242 (2002)
15. CY Wang, JL Xie, The T-mixture model approach for radar HRRP target recognition. Int. J. Comput. Electr. Eng. **5**(5), 500–503 (2013)
16. M Li, GJ Zhou, B Zhao, TF Quan, Sparse representation denoising for radar high resolution range profiling. Int. J. Antennas Propag. **2014**(3), 1–8 (2014)
17. L Du, HW Liu and Z Bao, Radar HRRP statistical recognition: parametric model and model selection, IEEE Transactions on Signal Processing, 56 (5), 1931–1944 (2008).
18. L Shi, PH Wang, HW Liu, L Xu, Z Bao, Radar HRRP statistical recognition with local factor analysis by automatic Bayesian Ying-Yang harmony learning. IEEE Trans. Signal Processing **59**(2), 610–617 (2011)
19. JS Fu, XH Deng, WL Yang, Radar HRRP recognition based on discriminant information analysis. WSEAS Trans. Inf. Sci. Appl. **8**(4), 185–201 (2011)
20. LM Novak and GJ Owirka, Radar target recognition using an eigen-image approach. IEEE Int. Radar Conf., 129-131 (1994).
21. BY Liu, WL Yang, Radar target recognition using canonical transformation to extract features. Proc. SPIE **3545**, 368–371 (1998)
22. B Chen, HW Liu and Z Bao, PCA and kernel PCA for radar high range resolution profiles recognition. 2005 IEEE International Radar conference, Virginia, USA, 2005, pp. 528–533.
23. S Mika, G Ratsch, J Weston, B Scholkopf and KR Muler, Fisher discriminant analysis with kernels. IEEE International Workshop on Neural networks for signal processing, Wisconsin, USA, 1999, pp. 41–48.
24. XF He, P Niyogi, *Locality preserving projections* (Proc. Conf. Advances in Neural Information Processing System 16, Vancouver, Canada, 2003)
25. HT Chen, HW Chang, TL Liu, Local discriminant embedding and its variants. IEEE Computer Society Conference on Computer Vision & Pattern Recognition, San Diego, California, USA, 2005, 2(2):846–853.
26. D Cai, XF He, JW Han, HJ Zhang, Orthogonal Laplacian faces for face recognition. IEEE Transactions on Image Processing, 2006, 15(11):3608–3614.
27. L Zhu, SN Zhu, Face recognition based on orthogonal discriminant locality preserving projections. Neurocomputing **70**, 1543–1546 (2007)
28. SJ Wang, HL Chen, XJ Peng, CG Zhou, Exponential locality preserving projections for small sample size problem. Neurocomputing **74**, 3654–3662 (2011)
29. S Yan, D Xu, B Zhang, H Zhang, Q Yang, S Lin, Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 40–51 (2007)
30. M Sugiyama, Local fisher discriminant analysis for supervised dimensionality reduction, Proceedings of the International Conference on Machine Learning (ICML), Las Vegas, Nevada, USA, 2006, pp. 905–912.
31. D Cai, X He, K Zhou, J Han, H Bao, Locality sensitive discriminant analysis, Proceedings of the 20th International Joint Conference Artificial Intelligence (IJCAI), Hyderabad, India, 2007, pp. 708–713.

32. T Zhang, D Tao, X Li, J Yang, Patch alignment for dimensionality reduction. IEEE Trans. Knowl. Data Eng. **21**(9), 1299–1313 (2009)
33. B Li, C Zheng, DS Huang, Locally linear discriminant embedding: an efficient method for face recognition. Pattern Recog. **41**(12), 3813–3821 (2008)
34. Y Chen, J Zhang, D Cai, W Liu, X He, Nonnegative local coordinate factorization for image representation. IEEE Trans. Image Process. **22**(3), 969–979 (2013)
35. Q Gao, J Ma, H Zhang, X Gao, Y Liu, Stable orthogonal local discriminant embedding for linear dimensionality reduction. IEEE Trans. Image Process. **22**(7), 2521–2530 (2013)
36. C Hou, C Zhang, Y Wu, Y Jiao, Stable local dimensionality reduction approaches. Pattern Recog. **42**(9), 2054–2066 (2009)
37. A Kocsor, K Kovacs, C Szepesvari, Margin maximizing discriminant analysis. Proc. 15th Eur. Conf. Mach. Learn. **32**(1), 227–238 (2004)
38. H Li, T Jiang, K Zhang, Efficient and robust feature extraction by maximum margin criterion. IEEE Trans. Neural Netw. **17**, 157–165 (2006)