Radar Target Recognition based on KLLE and a KNRD Classifier

ZHOU YUN [1], YU XUELIAN [1], LIU BENYONG [2], WANG XUEGANG [1]
[1] School of Electronic Engineering
University of Electronic Science and Technology of China
No. 4, Sec. 2, North Jianshe Road, Chengdu, 610054
CHINA
zhouyun_ee@uestc.edu.cn, iceyu23@126.com, xgwang@ee.uestc.edu.cn

[2] School of Computer Science and Technology
Guizhou University
Huaxi, Guiyang, 550025
CHINA
byliu@uestc.edu.cn

Abstract: - This paper presents a radar target recognition method using kernel locally linear embedding (KLLE) and a kernel-based nonlinear representative and discriminative (KNRD) classifier. Locally linear embedding (LLE) is one of the representative manifold learning algorithms for dimensionality reduction. In this paper, LLE is extended by using kernel technique, which gives rises to the KLLE algorithm. A KNRD classifier is a combined version of a kernel-based nonlinear representor (KNR) and a kernel-based nonlinear discriminaor (KND), two classifiers recently proposed for optimal representation and discrimination, respectively. KLLE is firstly utilized to reduce data dimension and extract features from a high resolution range profile (HRRP). Then, a KNRD classifier is employed for classification. Experimental results on measured profiles from three aircrafts indicate the relatively good recognition performance of the presented method.

Key-Words: - Radar target recognition, high resolution range profile, kernel locally linear embedding, kernel-based nonlinear representative and discriminative classifier

1 Introduction

The emergence of high resolution radar with the imaging technology enables modern radars to acquire rather detailed information about shape and structure of a target, thus providing us with a more reliable tool for target recognition. In comparison with 2-dimensional images, it is technically easy and inexpensive to obtain 1-dimensional high resolution range profiles (HRRPs) of a target, so HRRP-based target recognition has drawn much attention and been a mainstream in radar community. On the one hand, HRRPs reflect the distribution of scattering intensity and relative location of scatterers along the line of sight of radar and thus provide a lot of useful features for target recognition. On the other hand, HRRPs are sensitive to aspect variations of a target, which bring great difficulty to recognition. Therefore, one of the key problems of radar target recognition using HRRP is how to extract robust and effective features [1, 2].

The objective of feature extraction is to reduce data dimension and extract representative or discriminative features. Many classical methods have been developed and applied to radar target recognition successfully. Among the popular methods, the principal component analysis (PCA) [3] and the linear discriminant analysis (LDA) [4] are two powerful tools. But, radar target recognition is a complex nonlinear problem; sometime we cannot attain satisfactory results due to the linear nature of PCA and LDA. To overcome this weakness, corresponding nonlinear algorithms, such as kernel principal component analysis (KPCA) [5] and kernel Fisher discriminant analysis (KFDA) [6], were proposed by using kernel technique [7]. It has been demonstrated that these kernel methods are much superior to their corresponding linear counterparts in terms of recognition performance. However, they are incapable to discover the inherent geometry structure and topology relationship of data.

In the past few years, a new kind of nonlinear dimensionality reduction method, named manifold learning, has drawn much attention in the computer vision and pattern recognition community, such as Isomap [8], locally linear embedding (LLE) [9, 10], laplacian Eigenmap [11], local tangent space

alignment (LTSA) [12] and etc, which assume that data lied in a complex high-dimensional data space may reside more specifically on or nearly on a low-dimensional manifold embedded within the high-dimensional space. Thereinto, the central idea of LLE is to solve globally nonlinear problems using locally linear fitting. LLE is computationally very simple, involving closed-form linear algebraic operations and suffering no local minima problems.

In this paper, LLE is further extended by using kernel technique and thus the kernel locally linear embedding (KLLE) algorithm is formulated, which is adopted to reduce data dimension and extract features from a HRRP. Since the locality preservation is done in an implicit high dimensional feature space, in which the input data is as linearly separable as possible, the KLLE is expected to be superior to LLE in terms of recognition ability.

Classifiers also play a paramount role in radar target recognition. Generally, range profiles of different targets are overlapped in the measured space or feature space, so they are not linearly separable, and thus it is necessary to design a suitable nonlinear classifier to obtain satisfactory recognition results. Among the popular methods are the radial basis function (RBF) neural network [13] and the nonlinear support vector machine (SVM) [14]. In order to train the corresponding parameters, repeated iterative learning is needed for RBF, in which the local convergence problem is sometimes unavoidable. While SVM is based on structural risk criteria with the optimal generalization ability, in which the solution is represented as a nonlinear function in the form $f(\boldsymbol{x})=\sum_{j=1}^{M} a_j k(\boldsymbol{x},\boldsymbol{x}_j)+b$, where k is the associated kernel function, $\boldsymbol{x}_j (j=1,2,\cdots M)$ a set of training feature vectors, and b a constant being set to zero in some applications. The set of coefficients $a_j (j=1,2,\cdots M)$ is decided by the nature of the related problem. For example, when the error cost function for approximation is quadratic, it can be obtained by solving a linear system. When Vapnik's ε-insensitive cost function is adopted, it is obtained by the SVM approximation scheme, wherein a quadratic programming problem needs to be solved, so it limits the training speed [15]. In addition, the conventional SVM is formulated for two-class problems, it is extended to multiple-class problems usually by adopting one-against-one or one-against all scheme. In the former scheme, $C(C-1)/2$ SVM classifiers are designed for a C-class problem, and classification is done by voting.

Recently, two novel nonlinear classifiers named kernel-based nonlinear representor (KNR) [16] and kernel-based nonlinear discriminaor (KND) [17] were proposed for optimal representation and discrimination, respectively. Their solutions have the same form as that of SVM, wherein the constant b is set to zero for simplicity and the set of coefficients is determined by the desired outputs of the classifiers in a closed form. It is demonstrated that KNR and KND can achieve good classification ability comparative to that of SVM. Moreover, they are much less time-consuming for training than SVM since any iterative or quadratic programming procedure is avoided [18].

In this paper, a KNR and a KND are combined into a new version called a kernel-based nonlinear representative and discriminative (KNRD) classifier, which is applied to classification for radar target recognition using range profiles.

The remainder of this paper is organized as follows: in section 2, the KLLE algorithm is concretely described at first, and then the procedure of data dimension reduction and feature extraction based on the KLLE is stated. Section 3 reviews the KNR and KND at first, and then presents the criterion along with the associated derivation of a KNRD classifier. In section 4, experiments are performed on radar target recognition using HRRPs to verify the effectiveness of the presented method. Finally, conclusions are drawn in section 5.

2 KLLE

2.1 Review of LLE

LLE maps a data set $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N]$ in D-dimensional space $\boldsymbol{R}^D$ globally to a data set $\boldsymbol{Y}=[\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_N]$ in d-dimensional space $\boldsymbol{R}^d (d \ll D)$. Basically, the LLE algorithm consists of three steps:

1) Find K nearest neighbors of each data point $\boldsymbol{x}_i (i=1,2,\cdots,N)$ in $\boldsymbol{R}^D$ by using the Euclidean distance to measure similarity.

2) Compute the weights w_{ij} that best linearly reconstruct each data $\boldsymbol{x}_i$ from its K nearest neighbors $\boldsymbol{x}_j (j=1,2,\cdots,K)$ by minimizing the following cost function:

$$\varepsilon(\boldsymbol{W})=\sum_i \left\| \boldsymbol{x}_i-\sum_j w_{ij}\boldsymbol{x}_j \right\|^2 \qquad (1)$$

subject to constraints $\sum_j w_{ij} = 1$ and $w_{ij} = 0$, if $\boldsymbol{x}_j$ does not belong to the set of K nearest neighbors of $\boldsymbol{x}_i$. Then the optimal weights w_{ij} with the two constraints are found by solving least-squares problems.

3) Construct the low-dimensional embedding $\boldsymbol{Y}$ in $\boldsymbol{R}^d$, in which the local linear geometry of the high-dimensional data is preserved, via a neighborhood-preserving mapping. In particular, the same weights w_{ij} that reconstruct the data point $\boldsymbol{x}_i$ in $\boldsymbol{R}^D$ should also reconstruct its embedded coordinate $\boldsymbol{y}_i$ in $\boldsymbol{R}^d$. This is done by minimizing the following embedding cost function for the fixed weights w_{ij}:

$$\phi(\boldsymbol{Y}) = \sum_i \left\| \boldsymbol{y}_i - \sum_j w_{ij} \boldsymbol{y}_j \right\|^2 \tag{2}$$

under the constraints $\frac{1}{N}\sum_i \boldsymbol{y}_i \boldsymbol{y}_i^{\mathrm{T}} = \boldsymbol{I}$ and $\sum_i \boldsymbol{y}_i = 0$, where T denotes the complex transpose of a matrix,. The constrained minimization problem is then converted to solving the eigen-decomposition of the matrix $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{W})(\boldsymbol{I} - \boldsymbol{W})^{\mathrm{T}}$, whose eigenvectors associated with the bottom d nonzero eigenvalues form the final embedding $\boldsymbol{Y}$.

2.2 KLLE

In this section, we show how to formulate a kernel extension of LLE. To begin with, the data set $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N]$ in $\boldsymbol{R}^D$ is mapped into an implicit feature space $\boldsymbol{F}$ using a nonlinear function:

$$\Phi : \boldsymbol{x} \in \boldsymbol{R}^D \to \Phi(\boldsymbol{x}) \in \boldsymbol{F} \tag{3}$$

The objective of the KLLE is mapping the data set $\Phi(\boldsymbol{X}) = [\Phi(\boldsymbol{x}_1), \Phi(\boldsymbol{x}_2), \cdots, \Phi(\boldsymbol{x}_N)]$ in the feature space $\boldsymbol{F}$ to a new data set $\boldsymbol{Z} = [z_1, z_2, \cdots, z_N]$ in $\boldsymbol{R}^d (d \ll D)$, while the intrinsic geometry structure of the data set is preserved.

Then, in the feature space $\boldsymbol{F}$, we would like to minimize:

$$\varepsilon(\boldsymbol{W}) = \sum_i \left\| \Phi(\boldsymbol{x}_i) - \sum_j w_{ij} \Phi(\boldsymbol{x}_j) \right\|^2 \tag{4}$$

This is the same cost function as in Eq.(1), evaluated on the data set $\Phi(\boldsymbol{X})$. For each data point $\Phi(\boldsymbol{x}_i)(i = 1, 2, \cdots, N)$ in $\boldsymbol{F}$, its K nearest neighbors can be found according to the following distance measurement:

$$\begin{aligned} d_{\boldsymbol{F}} &= \left\| \Phi(\boldsymbol{x}_i) - \Phi(\boldsymbol{x}_j) \right\| \\ &= \sqrt{k(\boldsymbol{x}_i, \boldsymbol{x}_i) + k(\boldsymbol{x}_j, \boldsymbol{x}_j) - 2k(\boldsymbol{x}_i, \boldsymbol{x}_j)} \end{aligned} \tag{5}$$

where $k(\cdot,\cdot)$ is a reproducing kernel function in the Hilbert space, which satisfies:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = <\Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j)> = \Phi(\boldsymbol{x}_i)^{\mathrm{T}} \Phi(\boldsymbol{x}_j) \tag{6}$$

where $<\cdot,\cdot>$ denotes the inner product.

With the constraints mentioned in the steps of LLE, the weight matrix $\boldsymbol{W}$ can be computed in closed form. As for a particular data point $\Phi(\boldsymbol{x})$ with its K nearest neighbor points $\Phi(\boldsymbol{\eta}_j)$ and the corresponding weights w_j that sum to one, the reconstruction error can be written as:

$$\begin{aligned} \xi &= \left\| \Phi(\boldsymbol{x}) - \sum_j w_j \Phi(\boldsymbol{\eta}_j) \right\|^2 \\ &= \left\| \sum_j w_j (\Phi(\boldsymbol{x}) - \Phi(\boldsymbol{\eta}_j)) \right\|^2 \\ &= \sum_{jk} w_j C_{jk} w_k \end{aligned} \tag{7}$$

where

$$C_{jk} = (\Phi(\boldsymbol{x}) - \Phi(\boldsymbol{\eta}_j))^{\mathrm{T}} (\Phi(\boldsymbol{x}) - \Phi(\boldsymbol{\eta}_k)) \tag{8}$$

By introducing a reproducing kernel function $k(\cdot,\cdot)$, Eq.(8) can be rewritten as:

$$C_{jk} = k(\boldsymbol{x}, \boldsymbol{x}) - k(\boldsymbol{x}, \boldsymbol{\eta}_j) - k(\boldsymbol{x}, \boldsymbol{\eta}_k) + k(\boldsymbol{\eta}_j, \boldsymbol{\eta}_k) \tag{9}$$

Afterward, by solving the constrained least-squares problem, the optimal weights are given by:

$$w_j = \frac{\sum_k C_{jk}^{-1}}{\sum_{lm} C_{lm}^{-1}} \tag{10}$$

In practice, a more efficient and numerically stable way to minimize the error is simply to solve the linear equations $\sum_k C_{jk} w_k = 1$, and then to rescale the weights by:

$$w_k \leftarrow w_k \Big/ \sum_k w_k \tag{11}$$

so that they sum to one. It should be noted that sometimes the matrix $\boldsymbol{C}$ may be or near to singular, when some regularized technique is needed [10].

Now, we turn to the problem of computing a low-dimensional embedding $\boldsymbol{Z}$ which is optimally reconstructed by the weight matrix $\boldsymbol{W}$ in $\boldsymbol{R}^d$. In order to do this, we need to minimize the following cost function:

$$\phi(\boldsymbol{Z}) = \sum_i \left\| z_i - \sum_j w_{ij} z_j \right\|^2 \tag{12}$$

with

$$\frac{1}{N}\sum_i \boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}} = \boldsymbol{I} \tag{13}$$

The cost function in Eq.(12) can be simplified as:

$$\begin{aligned}\phi(\boldsymbol{Z}) &= \sum_i \left\| \boldsymbol{z}_i - \sum_j w_{ij}\boldsymbol{z}_j \right\|^2 \\ &= \left\| \boldsymbol{Z}(\boldsymbol{I}-\boldsymbol{W}) \right\|^2 \\ &= trace\left(\boldsymbol{Z}(\boldsymbol{I}-\boldsymbol{W})(\boldsymbol{I}-\boldsymbol{W})^{\mathrm{T}}\boldsymbol{Z}^{\mathrm{T}} \right) \\ &= trace\left(\boldsymbol{Z}\boldsymbol{M}\boldsymbol{Z}^{\mathrm{T}} \right)\end{aligned} \tag{14}$$

where $\boldsymbol{M} = (\boldsymbol{I}-\boldsymbol{W})(\boldsymbol{I}-\boldsymbol{W})^{\mathrm{T}}$.

Finally, similar to LLE, the constrained minimization problem above is converted to the eigenvalue problem of the matrix $\boldsymbol{M}$. That is, the embedding $\boldsymbol{Z}$ is composed of the eigenvctors corresponding to the bottom d nonzero eigenvalues of the matrix $\boldsymbol{M}$.

2.3 Feature extraction based on the KLLE

KLLE as well as LLE provides an embedding for the fixed set of training data, while there not existing an explicit mapping between the high-dimensional original (or feature) space and low-dimensional embedded space, thus the issue how to generalize the results of KLLE to the test data set in the input space remains difficult. For example, suppose computing the low dimensional embedding $\boldsymbol{z}$ for a new test vector $\boldsymbol{x}$.

For the purpose of generalization, there are two possible solutions: non-parametric model and parametric model [10]. In particular, the non-parametric method is very simple to realize, involving little computation and introducing no additional parameters. For target recognition, the low dimensional embedding of test data can be well constructed with the training data and their low dimensional embedding by using the non-parametric model, and thus the process of feature extraction of test data is completed.

Given a new test vector $\boldsymbol{x}$, we can compute its low dimensional embedding $\boldsymbol{z}$ by KLLE (with non-parametric model) according to the following steps:

1) Identify the K nearest neighbors of $\Phi(\boldsymbol{x})$ among the training data set $\Phi(\boldsymbol{X})$ in the feature space $\boldsymbol{F}$ by using the distance measurement defined in Eq.(5).

2) Compute the linear weights w_j that best reconstruct $\Phi(\boldsymbol{x})$ from its K nearest neighbors, subject to the sum-to-one constraint $\sum_j w_j = 1$, according to the method described in section 2.2.

3) Construct the low dimensional embedding $\boldsymbol{z}$ as follows:

$$\boldsymbol{z} = \sum_j w_j \boldsymbol{z}_j \tag{15}$$

where $\boldsymbol{z}_j (j = 1,2,\cdots,K)$ are the low dimensional embeddings corresponding to the K nearest neighbors of $\Phi(\boldsymbol{x})$.

3 A KNRD Classifier

We restrict our discussion to designing an appropriate nonlinear classification function, so that it has a certain desirable capability. Assume that a desirable function $f_0(\boldsymbol{x})$ is defined on a complex N-dimensional vector space $\boldsymbol{C}^N$, and that it is an element of a reproducing kernel Hilbert space (RKHS) $\boldsymbol{H}$. The reproducing kernel $k(\boldsymbol{x},\boldsymbol{x}')$ of $\boldsymbol{H}$ is a bivariate function defined on $\boldsymbol{C}^N \times \boldsymbol{C}^N$ which satisfies two conditions: for any fix $\boldsymbol{x}'$ in $\boldsymbol{C}^N$, $k(\boldsymbol{x},\boldsymbol{x}')$ is a function in $\boldsymbol{H}$; for any function $f_0(\boldsymbol{x})$ in $\boldsymbol{H}$ and any $\boldsymbol{x}'$ in $\boldsymbol{C}^N$, it holds that

$$< f_0(\boldsymbol{x}), k(\boldsymbol{x},\boldsymbol{x}') >= f_0(\boldsymbol{x}') \tag{16}$$

where $<\cdot,\cdot>$ denotes the inner product on $\boldsymbol{H}$. That is, if a function $\psi_j(\boldsymbol{x})$ is defined as:

$$\psi_j(\boldsymbol{x}) = k(\boldsymbol{x},\boldsymbol{x}_j) \tag{17}$$

then

$$f_0(\boldsymbol{x}_j) =< f_0(\boldsymbol{x}), \psi_j(\boldsymbol{x}) > \tag{18}$$

Generally, the desirable function f_0 is unknown, but its M sample values $\{y_j\}_{j=1}^M$ are known beforehand and they constitute the following training vector:

$$\begin{aligned}\boldsymbol{y} &= [y_1, y_2, \cdots, y_M]^{\mathrm{T}} \\ &= [f_0(\boldsymbol{x}_1), f_0(\boldsymbol{x}_2), \cdots, f_0(\boldsymbol{x}_M)]^{\mathrm{T}}\end{aligned} \tag{19}$$

The vector $\boldsymbol{y}$ is called the teacher signal and we assume that $\boldsymbol{y}$ is a point of an M-dimensional vector space $\boldsymbol{C}^M$. Let $\{\boldsymbol{e}_j\}_{j=1}^M$ denotes the standard basis of $\boldsymbol{C}^M$ and define a sampling operator as:

$$\boldsymbol{A} = \sum_{j=1}^{M} \boldsymbol{e}_j \otimes \overline{\psi}_j \tag{20}$$

where $\overline{\psi}$ is the complex conjugate of ψ and $(\cdot \otimes \bar{\cdot})$ is the Neuman-Schatten production defined by:

$$(\boldsymbol{e}_j \otimes \overline{\psi}_j) f_0 = < f_0, \psi_j > \boldsymbol{e}_j \tag{21}$$

Since the set of sampled values as well as the teacher signal $\boldsymbol{y}$ is uniquely determined by f_0 once the set of inputs is fixed, the following relation can be established:

$$\boldsymbol{y} = \boldsymbol{A} f_0 \tag{22}$$

Our objective is to find an approximation to the desirable function f_0 from the teacher signal $\boldsymbol{y}$.

In the viewpoint of inverse problem, the approximating can be realized by supervised learning, where a kind of inverse operator $\boldsymbol{X}$ of $\boldsymbol{A}$ is to be found under a certain criterion, so that

$$f = \boldsymbol{X}\boldsymbol{y} \tag{23}$$

becomes the optimal approximation to f_0 [19-21]. Hereinto, $\boldsymbol{X}$ is called a learning operator, and the process of obtaining f by $\boldsymbol{X}$ from $\boldsymbol{y}$ is called supervised learning.

3.1 Review of KNR and KND

3.1.1 Optimal representation and KNR

As for the optimal approximation problem discussed above, a natural criterion is to minimize the distance between f and f_0 in the metric of the space $\boldsymbol{H}$. But it is impossible to solve the problem because both f_0 in Eq.(22) and $\boldsymbol{X}$ in Eq.(23) are unknown. Ideally, for any unknown function f_0, if the estimated function f by Eq.(23) equals exactly to the original one f_0, then there is no estimation error. In this special case, Eqs.(22) and (23) yield:

$$\boldsymbol{X}\boldsymbol{A} = \boldsymbol{I} \tag{24}$$

where $\boldsymbol{I}$ is the identity operator of the Hilbert space $\boldsymbol{H}$. This result equals to:

$$\left\| \boldsymbol{I} - \boldsymbol{X}\boldsymbol{A} \right\|^2 = 0 \tag{25}$$

where $\|\cdot\|$ denotes the norm associating with the Hilbert space $\boldsymbol{H}$.

It should be pointed out that the condition led to Eq.(25) is too severe to be practically satisfied. That is, we should allow a certain deviation of $\boldsymbol{X}\boldsymbol{A}$ from $\boldsymbol{I}$. In particular, as for the problem of pattern recognition, we require that for a given class $c (c = 1, 2, \cdots, C)$, the learning operator shall satisfy:

$$\boldsymbol{X}_{\mathrm{R}}^{(c)} = \arg\min_{\boldsymbol{X}^{(c)}} \left\{ \left\| \boldsymbol{I} - \boldsymbol{X}^{(c)} \boldsymbol{A}^{(c)} \right\|^2 \right\} \tag{26}$$

so that the distance between f and f_0 is minimized, where $\boldsymbol{A}^{(c)}$ is the sampling operator corresponding to the target class c. In fact, Eq.(26) is equivalent to:

$$\boldsymbol{X}_{\mathrm{R}}^{(c)} = \arg\min_{\boldsymbol{X}^{(c)}} \left\{ tr\left[\left(\boldsymbol{I} - \boldsymbol{X}^{(c)} \boldsymbol{A}^{(c)} \right) \left(\boldsymbol{I} - \boldsymbol{X}^{(c)} \boldsymbol{A}^{(c)} \right)^* \right] \right\} \tag{27}$$

where $(\cdot)^*$ and $tr(\cdot)$ denote the adjoint operator and the trace of an operator, respectively.

It may be easily shown that a general solution to Eq.(24) is represented by:

$$\boldsymbol{X} = \boldsymbol{A}^+ + \boldsymbol{Y} - \boldsymbol{A}^+ \boldsymbol{A} \boldsymbol{Y} \boldsymbol{A} \boldsymbol{A}^+ \tag{28}$$

where $\boldsymbol{A}^+$ is the Moore-Penrose pseudoinverse of $\boldsymbol{A}$, and $\boldsymbol{Y}$ is any operator from $\boldsymbol{C}^M$ to $\boldsymbol{H}$ which satisfies:

$$(\boldsymbol{I} - \boldsymbol{A}^+ \boldsymbol{A}) \boldsymbol{Y} \boldsymbol{A} = \boldsymbol{I} - \boldsymbol{A}^+ \boldsymbol{A} \tag{29}$$

According to the projection theorem in functional analysis, if certain error is allowed to the estimated result and let:

$$\boldsymbol{X} = \boldsymbol{A}^+ \tag{30}$$

then the distance between f and f_0 is minimized. That is, the optimal solution of Eq.(27) is given by Eq.(30). Accordingly, Eqs.(22), (23) and (30) yield:

$$f = \boldsymbol{A}^+ \boldsymbol{A} f_0 = P_{\Re(\boldsymbol{A}^*)} f_0 \tag{31}$$

where $\Re(\cdot)$ denotes the range of an operator, and $P_{\Re(\boldsymbol{A}^*)}$ is the orthogonal projection operator onto the range of $\boldsymbol{A}^*$. It has been proved that $\Re(\boldsymbol{A}^*)$ is the largest subspace of $\boldsymbol{H}$, within which the best approximation to any desired function can be obtained [22].

Furthermore, Eqs.(17), (20), (21), (31) and Theorem (3.8) of Ref.[23] lead us to the following proposition [16].

Proposition 1 The kernel-based nonlinear representor (KNR) of class $c (c = 1, 2 \cdots, C)$ is represented by:

$$f_R^{(c)}(\boldsymbol{x}) = \sum_{j=1}^{M} a_{R,j}^{(c)} k(\boldsymbol{x}, \boldsymbol{x}_j^{(c)}) \tag{32}$$

with the coefficients constitute the following vector:

$$\boldsymbol{a}_R^{(c)} = [a_{R,1}^{(c)} \ a_{R,2}^{(c)} \ \cdots \ a_{R,M}^{(c)}]^{\mathrm{T}} = \boldsymbol{K}^+ \boldsymbol{y}^{(c)} \tag{33}$$

where $\boldsymbol{K}$ is the kernel matrix determined by the reproducing kernel function k of the Hilbert space $\boldsymbol{H}$ and the M training samples of class c, with

$$(\boldsymbol{K})_{ij} = k(\boldsymbol{x}_j^{(c)}, \boldsymbol{x}_i^{(c)}) \;\; i, j = 1, 2, \cdots, M \tag{34}$$

3.1.2 Optimal discrimination and KND

In the newly proposed optimal discrimination measure, the learning operator corresponding to the target pattern class is obtained by minimize the mean energy of the general outputs of all other classes. That is, for the target class $c\,(c = 1, 2\cdots, C)$ of a *C*-class problem, the optimal learning operator should satisfy:

$$\boldsymbol{X}_{\mathrm{D}}^{(c)} = \arg\min_{\boldsymbol{X}^{(c)}} \left\{ \underset{i, i\neq c}{\mathrm{mean}} \left\| \boldsymbol{X}^{(c)} \boldsymbol{y}^{(i)} \right\|^2 \right\} \tag{35}$$

where $\boldsymbol{y}^{(i)}$ is the teacher of class $i (i \neq c, i = 1, 2, \cdots, C)$. Eq.(35) establishes a discriminant criterion that for the target class *c*, its learning operator has the potentiality of suppressing the effects of all other classes by minimizing the mean energy of their outputs.

Defines an mean energy operator $\boldsymbol{Q}$ as:

$$\boldsymbol{Q} = \frac{1}{C-1} \sum_{i=1, i\neq c}^{C} \boldsymbol{Q}^{(i)} \tag{36}$$

with

$$\boldsymbol{Q}^{(i)} = \boldsymbol{y}^{(i)} \otimes \overline{\boldsymbol{y}^{(i)}} \tag{37}$$

an energy operator for the teacher signal of class $i (i \neq c, i = 1, 2, \cdots, C)$, then Eq. (35) is equivalent to

$$\boldsymbol{X}_{\mathrm{D}}^{(c)} = \arg\min_{\boldsymbol{X}^{(c)}} \left\{ tr\left(\boldsymbol{X}^{(c)} \boldsymbol{Q} \left(\boldsymbol{X}^{(c)} \right)^* \right) \right\} \tag{38}$$

According to Theorem 2.3.1 of Ref.[24], the optimal solution to Eq. (38) is given by:

$$\boldsymbol{X}_{\mathrm{D}}^{(c)} = \boldsymbol{Y} (\boldsymbol{I}_M - \boldsymbol{Q}\boldsymbol{Q}^+) \tag{39}$$

where $\boldsymbol{I}_M$ is the identity operator of space $\boldsymbol{C}^M$, $\boldsymbol{Y}$ is an arbitrary operator from $\boldsymbol{C}^M$ to $\boldsymbol{H}$, and $\boldsymbol{Q}^+$ is the Moore-Penrose pseudoinverse of $\boldsymbol{Q}$. In our study, the arbitrary operator $\boldsymbol{Y}$ in Eq.(39) is chosen as $\left(\boldsymbol{A}^{(c)} \right)^*$. In this case, the learning operator of class *c* is given by:

$$\boldsymbol{X}_{\mathrm{D}}^{(c)} = \left(\boldsymbol{A}^{(c)} \right)^* (\boldsymbol{I}_M - \boldsymbol{Q}\boldsymbol{Q}^+) \tag{40}$$

Furthermore, Eqs.(17), (20), (21), (23) and (40) result in the following proposition [17].

Proposition 2 The kernel-based nonlinear discriminator (KND) of class $c\,(c = 1, 2\cdots, C)$ is represented by:

$$f_D^{(c)}(\boldsymbol{x}) = \sum_{j=1}^{M} a_{D,j}^{(c)} k(\boldsymbol{x}, \boldsymbol{x}_j^{(c)}) \tag{41}$$

with the coefficient vector

$$\boldsymbol{a}_D^{(c)} = [a_{D,1}^{(c)} \;\; a_{D,2}^{(c)} \;\; \cdots \;\; a_{D,M}^{(c)}]^{\mathrm{T}} = (\boldsymbol{I}_M - \boldsymbol{Q}\boldsymbol{Q}^+) \boldsymbol{y}^{(c)} \tag{42}$$

Propositions 1 and 2 show that both KNR and KND have closed form solutions, and their coefficients are precisely determined by the desired outputs of the classifiers and the related operators, thus any repeated iterative learning in traditional neural networks such as RBF or quadratic programming procedure necessary for a nonlinear SVM to obtain a solution, is avoided.

3.2 Derivation of KNRD

In the discussion of section 3.1, on the one hand, a KNR is designed for optimal representation, and the criterion in Eq.(27) focuses on the outputs of the target class but ignores the effects of other classes, thus optimal representation is ensured only for the target class. On the other hand, a KND is designed for optimal discrimination, and the criterion in Eq.(38) ensures that for the target class, the effects of all other classes is optimally suppressed in the meaning of mean output energy, but no constraints is made on the outputs of the target class. Therefore, it is a natural issue to combine the above two criteria and it lead to a new criterion for both representation and discrimination of pattern features. In this section, we focus on the new criterion and the associated derivation.

Using a parameter $\rho (0 < \rho < 1)$ to control the balance between representation and discrimination, we can combine the above two criteria and obtain a new criterion, namely R&D criterion, as follows:

$$\boldsymbol{X}_{\mathrm{R\&D}}^{(c)} = \arg\min_{\boldsymbol{X}^{(c)}} \left\{ (1-\rho) \cdot tr\left[\left(\boldsymbol{I} - \boldsymbol{X}^{(c)} \boldsymbol{A}^{(c)} \right) \left(\boldsymbol{I} - \boldsymbol{X}^{(c)} \boldsymbol{A}^{(c)} \right)^* \right] + \rho \cdot tr\left[\boldsymbol{X}^{(c)} \boldsymbol{Q} \left(\boldsymbol{X}^{(c)} \right)^* \right] \right\} \tag{43}$$

For a given ρ, Eq.(43) equals to:

$$\boldsymbol{X}_{\mathrm{R\&D}}^{(c)} = \arg\min_{\boldsymbol{X}^{(c)}} \left\{ tr\left[\left(\boldsymbol{I} - \boldsymbol{X}^{(c)} \boldsymbol{A}^{(c)} \right) \left(\boldsymbol{I} - \boldsymbol{X}^{(c)} \boldsymbol{A}^{(c)} \right)^* + \lambda \boldsymbol{X}^{(c)} \boldsymbol{Q} \left(\boldsymbol{X}^{(c)} \right)^* \right] \right\} \tag{44}$$

where $\lambda = \rho/(1-\rho)$ is an equivalent control parameter. According to Ref.[25], an operator $\boldsymbol{X}_{\mathrm{R\&D}}^{(c)}$ is the optimal solution to Eq.(44) if and only if

$$\boldsymbol{X}_{\mathrm{R\&D}}^{(c)}\left[\boldsymbol{A}^{(c)}\left(\boldsymbol{A}^{(c)}\right)^{*}+\lambda\boldsymbol{Q}\right]=\left(\boldsymbol{A}^{(c)}\right)^{*} \tag{45}$$

By applying the related theorems and lemmas in Ref.[21] to this necessary and sufficient condition in Eq.(45), a general expression of the solution to the R&D criterion in Eq.(44) is given by:

$$\boldsymbol{X}_{\mathrm{R\&D}}^{(c)}=\left(\boldsymbol{A}^{(c)}\right)^{*}\left(\boldsymbol{U}^{(c)}\right)^{+}+W\left[\boldsymbol{I}_{M}-\boldsymbol{U}^{(c)}\left(\boldsymbol{U}^{(c)}\right)^{+}\right] \tag{46}$$

where $\boldsymbol{W}$ is any operator from $\boldsymbol{C}^{M}$ to $\boldsymbol{H}$, and

$$\boldsymbol{U}^{(c)}=\boldsymbol{K}+\lambda\boldsymbol{Q} \tag{47}$$

Therefore, similar to the derivation of a KND, a kernel-based nonlinear representative and discriminative (KNRD) classifier may be derived from Eqs.(17), (20), (21), (23) and (46), and it is presented in the following theorem.

Theorem 1 The kernel-based nonlinear representative and discriminative (KNRD) classifier of class $c\,(c=1,2\cdots,C)$ is represented by:

$$f_{R\&D}^{(c)}(\boldsymbol{x})=\sum_{j=1}^{M}a_{R\&D,j}^{(c)}k(\boldsymbol{x},\boldsymbol{x}_{j}^{(c)}) \tag{48}$$

with the coefficient vector

$$\begin{aligned}\boldsymbol{a}_{R\&D}^{(c)}&=[a_{R\&D,1}^{(c)}\ a_{R\&D,2}^{(c)}\ \cdots\ a_{R\&D,M}^{(c)}]^{\mathrm{T}}\\&=\left(\boldsymbol{U}^{(c)}\right)^{+}\boldsymbol{y}^{(c)}\end{aligned} \tag{49}$$

The architecture of a KNRD classifier, as well as those of a KNR and a KND, can be depicted in Fig.1 as below. Similar to KNR and KND, any repeated iterative learning or quadratic programming procedure is avoided in a KNRD classifier since the solution is in closed form.

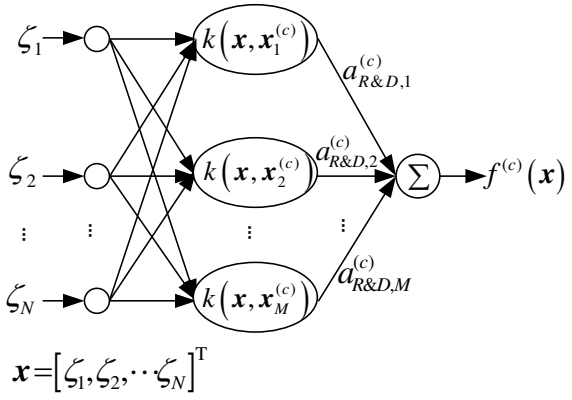
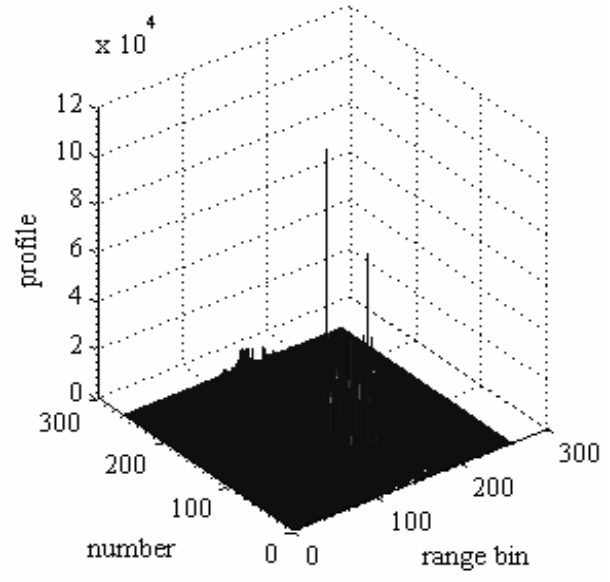


Fig.1 Architecture of a KNRD classifier of class *c*

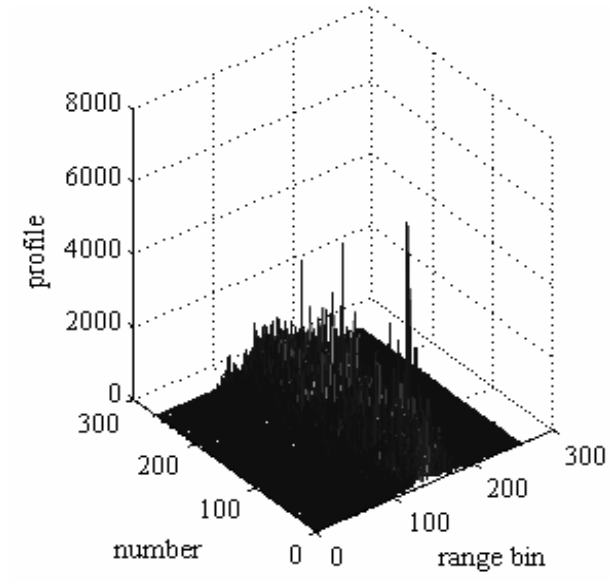
4 Application to Radar Target Recognition using HRRPs

4.1 Data description

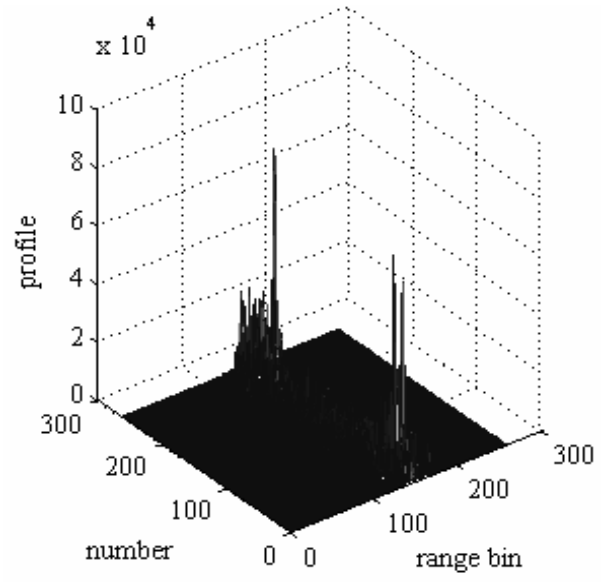
In this section, experiments are performed on radar target recognition with measured HRRPs from three flying airplanes, including An-26, Yark-42, and Cessna Citation S/II. Each profile has 256 range bins. For each target, 260 range profiles over a wide range of aspects are adopted for experiment and are shown in Fig.2. For each target, one third of all profiles are used for training and the remained ones for test.



(a) An-26



(b) Yark-42



(c) Cessna Citation S/II

Fig.2 Range profiles of three airplanes

4.2 A radar target recognition scheme

Based on the KLLE algorithm and KNRD classifiers, a radar target recognition scheme is presented and depicted in Fig.3. It includes three key modules: preprocessing, feature extraction based on the KLLE, and classification based on KNRD classifiers. In preprocessing, a HRRP is energy-normalized to reduce magnitude sensitivity and range-aligned to remove shift sensitivity. In the process of feature extraction, the KLLE is conducted on the training set to obtain the low-dimensional features of these range profiles. For a test range profile, its low-dimensional embedded feature is computed by the KLLE (with the non-parametric model) according to the steps detailed in section 2.3. In the process of classification, for each class, the training profile features are used to train a KNRD classifier, and an unknown test profile is recognized by the trained KNRD classifiers using its embedded feature.

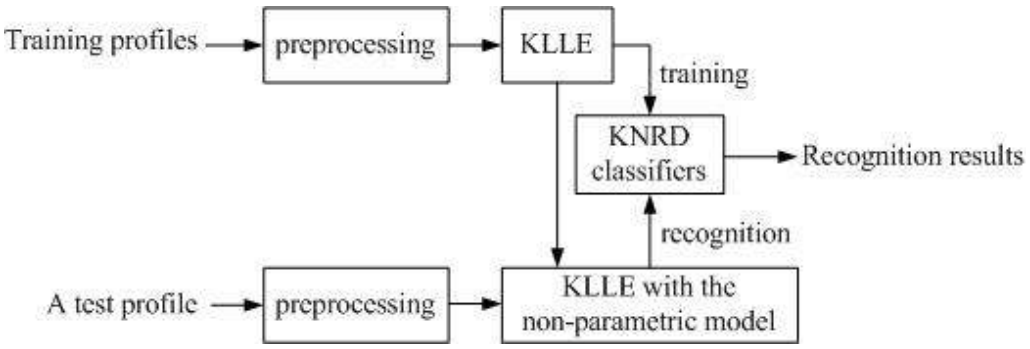


Fig.3 A radar target recognition scheme based on the KLLE and KNRD classifiers

4.3 Experimental results

Two sets of experiments are conducted. In the first one, six algorithms including PCA, KPCA, LDA, KFDA, LLE and KLLE are utilized to extract the low-dimensional feature from a HRRP for performance comparison and then KNRD classifiers are employed for classification with the control parameter λ fixed as 1. For both LLE and KLLE, the number of nearest neighbors is set as K=8. For the four kernel-based methods including KPCA, KFDA, KLLE and KNRD, the Gaussian kernel $k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp(-\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2 / \sigma)$ is adopted, and the parameter σ is empirically set as 1.

The recognition results obtained by each method at different reduced dimensionality are shown in Fig.4. Note that, the upper bound of dimensionality of LDA and KFDA is 2 for this 3-class problem.

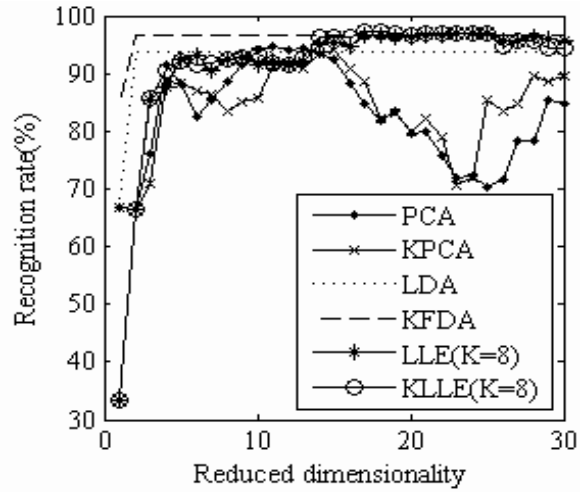


Fig.4 Recognition results of six feature extraction methods with KNRD classifiers

As can be seen from Fig.4, the recognition rates obtained by PCA and KPCA vary greatly as the increasing of reduced dimensionality. In comparison, LLE and KLLE achieve higher recognition rate with the increasing of reduced dimensionality, and get satisfactory recognition results when the dimensionality is greater than 15.

The top recognition rates achieved by each method along with the reduced dimensionality are listed in Table 1. It shows the proposed recognition scheme (that is, "KLLE+KNRD") achieves the highest recognition rate of 97.31% at the dimensionality of 18.

Table 1 Top recognition results of six methods

Method	Recognition rate (%)	Reduced dimensionality
PCA	94.63	11
LDA	93.86	2
LLE	96.93	23
KPCA	94.06	15
KFDA	96.55	2
KLLE	97.31	18

In the second experiment, only the KLLE is adopted for feature extraction with K set as 5, 8, 12 and 20, respectively, and the KNRD classifier as well as the one-against-one SVM is employed for classification. For the SVM, the Gaussian kernel is adopted and the parameter σ is also set as 1. Table 2 lists the top recognition rate achieved by the two classifiers along with the reduced dimensionality for each value of K. It can be seen that both the KNRD and SVM achieve the top recognition rate for K=8, and that the former performs better than the latter on the whole.

Table 2 Top recognition results of two classifiers

K	5	8	12	20
KNRD	95.21 (16)	97.31 (18)	96.36 (22)	95.78 (19)
SVM	95.21 (14)	95.59 (17)	95.40 (9)	93.67 (6)

Moreover, for K=8, the training time and test time (per test profile on average) taken by the SVM and the KNRD classifier to achieve the highest recognition rate, are listed in Table 3. It shows that the KNRD classifier is much less time-consuming than SVM for training. As for test, the KNRD classifier also takes less time in our experiment. Note that, the time listed in Table 3 is measured on same platforms.

Table 3 Training and test time of two classifiers

Classifier	Training time (ms)	Test time (ms)
KNRD	78	0.028
SVM	4860	0.059

Fig.5 shows the plot of recognition rates of the proposed scheme "KLLE+KNRD" versus the parameter λ, with the dimensionality fixed as 18 and the value of K fixed as 8. As can be seen, the proposed scheme performs well and stably within a wide range of λ, that is, it rather insensitive to the value of λ, which is highly desirable in practical applications since it is usually not easy to determine a very suitable parameter.

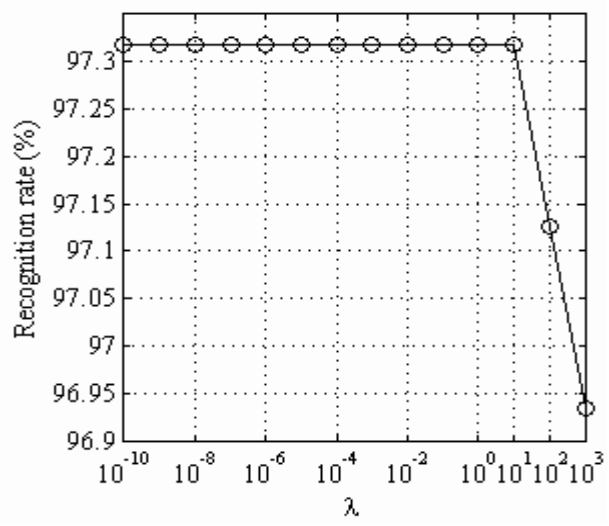


Fig.5 Recognition rates of "KLLE+KNRD" versus the parameter λ

5 Conclusions

In this paper, based on the kernel locally linear embedding (KLLE) and the kernel-based nonlinear representative and discriminative (KNRD) classifier, a radar target recognition scheme is presented. Firstly, the KLLE is utilized to extract low-

dimensional features from range profiles, which is derived by generalizing LLE into a reproducing kernel Hilbert space (RKHS) using kernel technique. Then, the KNRD classifier is employed for classification, which is obtained by combining a KNR and a KND. The closed-form solution of a KNRD classifier avoids any quadratic programming procedure and thus ensures faster training speed than SVM. The effectiveness of the presented method is demonstrated by experimental results on measured profiles from three aircrafts. In addition, our further experiments indicate that the proposed method also performs excellently in some other applications such as face recognition and handwritten number recognition.

References:

[1] Xuelian Yu, Benyong Liu, Optimal kernel discriminant analysis for radar target recognition, *Journal of University of Electronic Science and Technology of China*, Vol.37, No.6, 2008, pp.883-885.

[2] Benyong Liu, Subspace methods of radar target recognition using range profiles, *Journal of Electronics & Information Technology*, Vol.26, No.7, 2004, pp.1137-1143.

[3] P N Belhumeur, J P Hespanha, and D J Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Machine Intell*., Vol.19, No.7, 1997, pp.711-720.

[4] J W Lu, K N Platanniotis, and A N Venesanopoulos, Face recognition using LDA-based algorithm, *IEEE Trans. Neural Networks*, Vol.14, No.1, 2003, pp.195-200.

[5] B Schölkopf, A Smola, K R Muller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation*, Vol.10, No.5, 1998, pp.1299-1319.

[6] Q S Liu, R Huang, H Q Lu, and et al., Face recognition using kernel based Fisher discriminant anlyais, *Proc. Fifth IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2002.

[7] K R Muller, S Mika, G Ratsch, and et al. An introduction to kernel-based learning algorithms, *IEEE Trans. Neural Networks*, Vol.12, No.2, 2001, pp.181-201.

[8] J B Tenenbaum, V de Silva, and J C Langford, A global geometric framework for nonlinear dimensionality reduction, *Science*, Vol.290, No.5500, 2000, pp.2319-2323.

[9] T R Sam and K S Lawrence, Nonlinear dimensionality reduction by locally linear embedding, *Science*, Vol.290, No.5500, 2000, pp.2323-2326.

[10] K S Lawrence and T R Sam, Think globally, fit locally: unsupervised learning of low dimensional manifolds, *Journal of Machine Learning Research*, Vol.4, 2003, pp.119-155.

[11] M Belkin and P Niyogi, Laplacian Eigenmaps for dimensionality reduction and data representation, *Neural Computation*, Vol.5, No. 6, 2003, pp.1373-1396.

[12] Z Zhang and H Zha, Principal manifolds and nonlinear dimensionality reduction via tangent space alignment, *SIAM Journal of Scientific Computing*, Vol.26, No.1, 2004, pp.313-338.

[13] Q Zhao, Z Bao, and W Ye, Radar target recognition based on neural network, *Journal of Electronics*, Vol.13, No.1, 1996, pp.1-10.

[14] L Zhang, W D Zhou, and L C Jiao, Support vector machine for 1-d image recognition, Journal of Infrared and Millimeter Waves, Vol. 21, 2002, pp. 119-123.

[15] C Cortes and V Vapnik, Support-vector networks, *Machine Learning*, Vol. 20, No.3, 1995, pp. 273-297.

[16] J Zhang, B Y Liu and H Tan, A kernel-based nonlinear representor with application to eigenface classification, *Journal of Electronic Science and Technology of China*, Vol. 2, No.2, 2004, pp. 19-22

[17] B Y Liu, A kernel-based nonlinear discriminator with closed-form solution, *Proc. IEEE Int. Conf. on Neural Network and Signal Processing*, 2003, pp. 41-44.

[18] X L Yu, B Y Liu, and X G Wang, High resolution range profile recognition using kernel-based nonlinear classifiers, *Proc InTech'06*, pp.116-120.

[19] T Poggio and F Girosi, Netwroks for approximation and learning, *Proc. IEEE*, Vol.78, No.9, 1990, pp.1481-1479.

[20] H Ogawa, Neural network learning, generalization and over-learning, *Proc. Int. Conf. Intelligent Information Processing & System*, 1992, pp.1-6.

[21] B Y Liu, H Ogawa, An equivalent form of S-L projection learning, *Journal of Electronic Science and Technology of China*, Vol. 1, No.1, 2003, pp. 6-11

[22] S Vijayakumar, and H Ogawa, RKHS-based functional analysis for extract incremental learning, Neurocomputing, Vol.29, No.1-3, 1999, pp.85-113.

[23] A Albert, *Regression and the Moor-Penrose Pseudoinverse*, Academic Press, 1972

[24] C R Rao, and S K Mitra, *Generalized inverse of Matrices and its application*, John Wiley & Sons Press, 1971.
[25] E. Oja, and H Ogawa, Parametric projection filter for image and signal restoration, *IEEE Trans. Acoustic, Speech, and Signal Processing*, Vol.34, No.6, 1986, 1643-1653.