Radar HRRP Recognition based on Supervised Exponential Sparsity Preserving Projection with Small Training Data Size

Xueling YANG [1,2,3], Gong ZHANG [1,2], Hu SONG [3]

[1] Key Lab of Radar Imaging and Microwave Photonics, Ministry of Education, Nanjing, China
[2] College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[3] Nanjing Marine Radar Institute, Nanjing, China

yangxueling@nuaa.edu.cn, gzhang@nuaa.edu.cn, andysonghu@126.com

Submitted October 10, 2024 / Accepted February 2, 2025 / Online first March 21, 2025

Abstract. *The echo signals from ships and sea clutter are coherently accumulated. Therefore, it is difficult to capture and distinguish the features within the signals. In addition, due to poor measurement conditions, the radar system can only collect data from a limited number of non-cooperative ships. In this article, a method termed supervised exponential sparsity preserving projection (E-MMC-SPP) is proposed for recognizing ship classes based on high-resolution range profile (HRRP). The method consists of three parts: First, to extract richer features from sea clutter, a maximum margin criterion sparse reconstructive relationship is constructed, which maximally preserves the sparse reconstruction of data and enhances class separability. Second, matrix exponential is utilized to ensure the positive definiteness of the coefficient matrices, thereby addressing the small-sample-size (SSS) problem. Finally, an efficient numerical method is presented for solving the corresponding large-scale matrix exponential eigenvalue problem. Experimental results on measured radar data demonstrate that the proposed method effectively reduces feature dimensionality and enhances target recognition performance with limited training data.*

Keywords

Supervised exponential sparsity preserving projection, high-resolution range profile, ship recognition, small-sample-size (SSS) problem

1. Introduction

High-resolution range profile (HRRP) is a real vector, which is the amplitude of coherent summations of the complex time return from target scatterers in each range resolution cell. HRRP can reveal the detailed physical structure characteristics of the target, such as target size, scatterer distribution and other abundant information. During the past couple of years, HRRP has been extensively studied in the area of radar automatic target recognition (RATR) [1–3]. Feature extraction, a fundamental problem in HRRP recognition, is a key aspect of pattern recognition. Usually, feature extraction is achieved via raw data dimensionality reduction (DR), which is transformation of high-dimensional data into low-dimensional data. Currently, the most widely used DR methods are subspace transformation and deep learning (DL). Subspace transformation methods, such as principal component analysis (PCA) [4], linear discriminant analysis (LDA) [5], Laplacian eigenmaps (LE) [6], are especially useful for high-dimensional data modeling. They can be formulated as eigenproblems by offering great potential for efficient learning of nonlinear and linear models without local minima. DL methods, including Convolutional Neural Networks (CNN) [7], Long Short-Term Memory Networks (LSTM) [8], and their derived architectures [9–11], help reduce reliance on manually designed feature extraction rules. Instead, they automatically extract deep descriptive features of the target.

At present, the recognition technology based on HRRP is mainly used in air targets. Due to most of the research results obtained in the environment of Gaussian white noise, the HRRPs of air targets are mainly affected by noise. Different from air targets, the echo signals of ships and sea clutter are coherently accumulated. The sea clutter is related to many factors, usually showing the obvious non-stationary and non-Gaussian characteristics. In addition, due to the poor measurement conditions, the radar system cannot guarantee the detection and tracking of non-cooperative ship targets for a long time, which severely limits the collection of HRRP data. In summary, HRRP of ship has the following three characteristics: i) It contains sea clutter, ii) It has a higher data dimension, iii) The training data is small. However, DR methods usually require substantial data for estimating parameters, and their recognition performance generally degrades severely with the decrease of the training samples, especially for the complicated statistical model with numerous unknown parameters. Therefore, one of the most challenging

DOI: 10.13164/re.2025.0143

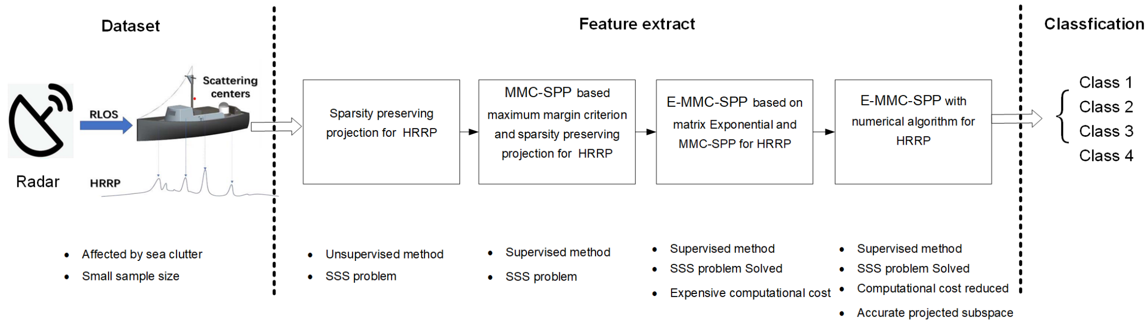


Fig. 1. The idea architecture of the overall ship recognition under small training data size.

tasks in ship recognition based on HRRP is the small-sample-size (SSS) problem.

In the last decades, many approaches have been proposed to deal with the SSS problem. Even there are some researchers who discuss theoretical aspects of the small training data size problem [12]. The first type of these approaches is the preprocessed method which employs PCA as a preprocessor before executing the DR method. This type has been widely used in the primitive articles such as neighborhood preserving embedding (NPE) [13] and sparsity preserving projection (SPP) [14]. However, the PCA stage may discard some valuable information contained in the joint null-space of the coefficient matrices, and be computationally expensive due to the use of the singular value decomposition (SVD). By using the regularized method, the second type includes regularized LDA [15] and regularized locality preserving projection [16] transforms the original generalized eigenvalue problem (GEP) to a solvable one by using regularization techniques. Yet, the performance of this type is strongly dependent on the regularization parameters, and there is no direct way of evaluating the parameters. The third is the kernel trick such as multiple kernel projection subspace fusion (MKPSF) [17], kernel joint discriminant analysis (KJDA) [18] and multiple kernel learning (MKL) [19]. They map a low-dimensional space to a linearly divisible higher-dimensional space through a kernel transformation However, choosing a proper kernel function for a specific real-world problem is challenging. The fourth type focuses directly on the DL methods such as sparse auto encoder (SAE) [20], discriminant deep autoencoders (DDAEs) [21], generative adversarial network (GAN) [22] and lightweight transformer network (LTN) [23]. Yet, the balance between network depth and model generalization performance, as well as the problem of network overfitting should be considered. Another type of these approaches is the exponential method such as exponential LDA (EDA) [24], exponential discriminant locality preserving projection [25], exponential NPE (ENPE) [26] and exponential SPP (ESPP) [27]. while, this type involves huge computational cost due to computing matrix exponentials and solving the matrix exponential eigenvalue problem.

In recent years, some linear and nonlinear supervised DR techniques based on the maximum margin criterion (MMC) [28] have been proposed, such as neighborhood-preserving discriminant projections (NPDP) [29]. In [30], we propose a method termed as multi-scale fusion kernel sparse preserving projection based on Kernel SPP to extract the richer feature information of ship from sea clutter. In this paper, a supervised method of MMC-SPP is introduced based on SPP and MMC. However, MMC-SPP suffers from the SSS problem. When the sample dimension is larger than the training sample size, the projection coefficient matrix of MMC-SPP is non-positive, which will seriously affect the recognition accuracy. To overcome the SSS problem, motivated by the work in Wu et al. [31], Wang et al. [32], a supervised exponential sparsity preserving projection (E-MMC-SPP) technique is proposed for ship recognition based on HRRP. Further, to overcome the huge computational and solving the corresponding large-scale matrix exponential eigenvalue problem, an efficiently numerical method is presented. Here, a two-step strategy is taken. First of all, the maximum margin criterion sparse reconstructive relationship is established. And then, an optimal projection direction is automatically learned based on the E-MMC-SPP. The idea architecture of ship recognition under small training data size is depicted in Fig. 1. Experiments on measured data show that the proposed method can effectively reduce the feature dimensionality and improve the recognition performance under the condition of small samples.

The major contributions can be summarized as follows.

i) A maximum margin criterion sparse reconstructive relationship is constructed, which can make full use of data information and simultaneously utilize the label information. The proposed method can maximally preserve the sparse reconstruct of data and maximize the class by separability resorting to the thought of MMC and SPP.

ii) A novel framework, E-MMC-SPP, is mainly constructed for HRRP recognition with small training data. Specially, the proposed method uses matrix exponential to ensure the positive definiteness of the coefficient matrices, which can overcome the SSS problem.

iii) In the proposed method, we develop a numerical algorithm to solve the large-scale matrix exponential eigenvalue problem which involves huge computational cost and storage requirement since the exponential of a matrix is often dense even if the matrix is sparse.

iv) The proposed method in this paper is evaluated based on the measured HRRP data. Extensive studies demonstrate that the proposed method exhibits much stable recognition performance when the training data size is small.

The rest of this paper is organized as follows. In Sec. 2, we briefly analyze the mechanism of DR with the SPP method. In Sec. 3, we propose the MMC-SPP technique, and discuss the HRRP RATR procedures via E-MMC-SPP in detail, and then develop a numerical algorithm to solve the matrix exponential eigenvalue problem. Experiments on measured data are conducted in Sec. 4 and conclusions are given finally in Sec. 5.

2. DR of HRRP Based on SPP

In this section, we briefly discuss the mechanism of DR with the SPP method [14, 33, 34]. In most existing DR methods, we know that constructing an affinity weight matrix plays a key role. The sparsity of the weight matrix is an important means to encode the structures of the data, which is helpful to improve the performance of the DR methods. Based on sparse representation, SPP aims to preserve the sparse reconstruction relationship of the data set.

Assuming that C is the number of HRRP classes in the training set, and $N = \sum_{k=1}^{C} N_k$ represents the total number of training HRRP samples, where N_k $(k = 1, 2, \ldots, C)$ is the number of training HRRP samples for each class. $\mathbf{X}_k = \left[\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \ldots, \mathbf{x}_{k,l_k}, \ldots, \mathbf{x}_{k,N_k}\right]$ is the training HRRP samples matrix of the kth class, where, $\mathbf{x}_{k,l_k} \in \mathrm{R}^n$ is a column vector of the l_kth HRRP sample in the kth class and n is the dimension of HRRP. $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_C] \in \mathrm{R}^{n\times N}$ is all training HRRP samples matrix. Each $\mathbf{x}_{k,l_k}$ can be sparsely represented by the rest of HRRP by solving the following optimization problem [14]:

$$\begin{aligned} \mathbf{s}_{k,l_k} &= \arg\min_{h_{k,l_k}} \left\|\mathbf{s}_{k,l_k}\right\|_1, \\ \text{s.t. } \mathbf{x}_{k,l_k} &= \mathbf{X}\mathbf{s}_{k,l_k}, \\ \mathbf{e}^{\mathrm{T}}\mathbf{s}_{k,l_k} &= 1 \end{aligned} \tag{1}$$

where, $\mathbf{s}_{k,l_k} = [s_{k,l_k 1,1}, \ldots, 0, s_{k,l_k k,l_{k+1}}, \ldots, s_{k,l_k C,N_C}]^{\mathrm{T}}$ is the sparse representation coefficient vector, the kl_kth element 0 represents that the sparse representation problem has nothing to do with $\mathbf{x}_{k,l_k}$ itself, and $\mathbf{e}$ represents the column vector where all elements are 1.

In the case of noise, Equation (1) can be represented as [14]:

$$\begin{aligned} \mathbf{s}_{k,l_k} &= \arg\min_{h_i} \left\|\mathbf{s}_{k,l_k}\right\|_1, \\ \text{s.t. } &\left\|\mathbf{X}\mathbf{s}_{k,l_k} - \mathbf{x}_{k,l_k}\right\| \le \delta, \\ \mathbf{e}^{\mathrm{T}}\mathbf{s}_{k,l_k} &= 1 \end{aligned} \tag{2}$$

where δ is the noise tolerance.

Then SPP seeks a transform matrix $\mathbf{w} \in \mathrm{R}^{n\times d}$ to project HRRP from a high-dimensional space into a d-dimensional space, where $n > d$ [14]. SPP aims to preserve the sparse reconstruction relationship and minimize the following objective function [33]:

$$\begin{aligned} &\min \sum_{k=1}^{C}\sum_{l_k=1}^{N_k} \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_{k,l_k} - \mathbf{w}^{\mathrm{T}}\mathbf{X}\mathbf{s}_{k,l_k}\right)^2 \\ &\text{s.t. } \mathbf{w}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{w} = 1. \end{aligned} \tag{3}$$

The sparse reconstructive weight matrix $\mathbf{S}$ can be given as $\mathbf{S} = [\mathbf{s}_{1,1}, \mathbf{s}_{1,2}, \ldots, \mathbf{s}_{k,N_k}, \ldots, \mathbf{s}_{C,N_C}]$, then, the optimization criteria of SPP method is obtained as [34]:

$$\min_{V} \frac{\mathbf{w}^{\mathrm{T}}\mathbf{X}(\mathbf{I} - \mathbf{S} - \mathbf{S}^{\mathrm{T}} + \mathbf{S}\mathbf{S}^{\mathrm{T}})\mathbf{X}^{\mathrm{T}}\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{w}}. \tag{4}$$

Let d denote the dimension of the embedding subspace, then the projection matrix $\mathbf{w}'s$ are the eigenvectors corresponding to smallest d positive eigenvalues of the following GEM [34]:

$$\mathbf{X}(\mathbf{I} - \mathbf{S} - \mathbf{S}^{\mathrm{T}} + \mathbf{S}\mathbf{S}^{\mathrm{T}})\mathbf{X}^{\mathrm{T}}\mathbf{w} = \lambda\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{w}. \tag{5}$$

As suggested by Qiao et al. [14], in some cases the maximum formulation can get a more numerically stable solution, so we will focus on solving the GEM (5). If $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_d$ are the smallest d eigenvalues of the GEM (5), $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d$ are the corresponding orthonormal eigenvectors, then $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d$ form a basis of the projected subspace, and the optimal projection matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d]$ is the solution of the following minimization problem [34]:

$$\begin{cases} \min \ \mathrm{tr}[\mathbf{W}^{\mathrm{T}}\mathbf{X}\mathbf{L}\mathbf{X}^{\mathrm{T}}\mathbf{W}] \\ \text{s.t. } \mathbf{W}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{W} = \mathbf{I}. \end{cases} \tag{6}$$

where $\mathbf{L} = \mathbf{I} - \mathbf{S} - \mathbf{S}^{\mathrm{T}} + \mathbf{S}\mathbf{S}^{\mathrm{T}}$, and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix.

Figure 2 shows the sparse representation coefficients of all training HRRP samples. It can be seen that the decomposition coefficients are sparse indeed and the primary sparse coefficients assemble within the range of corresponding object class.

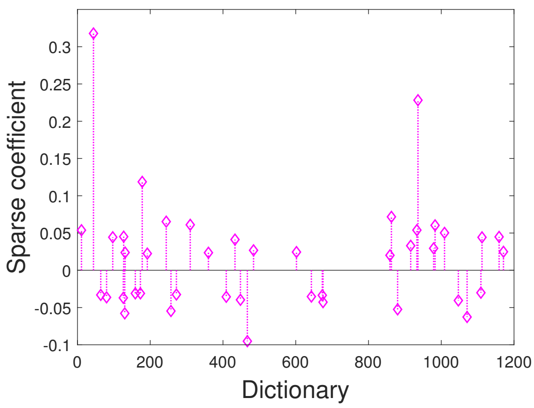


Fig. 2. Sparse coefficients of HRRP.

3. E-MMC-SPP and Its Solver

3.1 Principles of MMC-SPP

To maximize the preservation of data structure in the dimension-reduced space while improving classification performance, the strengths of SPP and MMC are combined. In this section, we learn an optimal projection direction automatically based on the MMC-SPP. The optimal projection direction makes full use of data information and labels information.

Similar to SPP, MMC aims to find the optimal projection direction W to maximize the margin between interclass samples. The objective function can be expressed as follows:

$$\begin{cases} \min\ \mathrm{tr}[\mathbf{W}^{\mathrm{T}}\mathbf{XLX}^{\mathrm{T}}\mathbf{W}] \\ \max\ \mathrm{tr}[\mathbf{W}^{\mathrm{T}}(\mathbf{S}_{\mathrm{b}} - \eta\mathbf{S}_{\mathrm{w}})\mathbf{W}] \\ s.t.\ \mathbf{W}^{\mathrm{T}}\mathbf{XX}^{\mathrm{T}}\mathbf{W} = \mathbf{I}, \end{cases} \tag{7}$$

and

$$\mathbf{S}_{\mathrm{b}} = \frac{1}{N}\sum_{i=1}^{C} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^{\mathrm{T}},$$

$$\mathbf{S}_{\mathrm{w}} = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{n_i}(\mathbf{x}_{i,j} - \mathbf{m}_i)(\mathbf{x}_{i,j} - \mathbf{m}_i)^{\mathrm{T}}$$

where $\mathbf{S}_{\mathrm{b}}$ and $\mathbf{S}_{\mathrm{w}}$ are the between-class scatter matrix and within-class scatter matrix, respectively, η is positive constant, $\mathbf{m}_i$ and $\mathbf{m}$ are the mean vectors of the class and training HRRP samples, respectively.

The solution to the multi-objective constrained optimization problem in (7) is to find a subspace which maximizes the margin between different classes simultaneously and preserves the sparsity property. So, we can change (7) into the following constrained problem, like:

$$\begin{array}{l} \min\ \mathrm{tr}[\mathbf{W}^{\mathrm{T}}[\mathbf{XLX}^{\mathrm{T}} - (\mathbf{S}_{\mathrm{b}} - \eta\mathbf{S}_{\mathrm{w}})]\mathbf{W}] \\ \mathrm{s.t.}\ \mathbf{W}^{\mathrm{T}}\mathbf{XX}^{\mathrm{T}}\mathbf{W} = \mathbf{I}. \end{array} \tag{8}$$

Here, we set $\mathbf{e}_{i,j} = [0, \ldots, 0, 1, 0, \ldots, 0]^{\mathrm{T}} \in \mathrm{R}^N$ which is the canonical basis vector of dimensions, then we can obtain $\mathbf{x}_{i,j} = \mathbf{X}\mathbf{e}_{i,j}$.

Thus, we can obtain:

$$\mathbf{m}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{x}_{i,j} = \mathbf{X}\frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{e}_{i,j} = \frac{1}{n_i}\mathbf{XB},$$

and

$$\mathbf{m} = \mathbf{X}\frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{n_i}\mathbf{e}_{i,j} = \frac{1}{N}\mathbf{XD}0$$

where $\mathbf{B} = [0, 0, \ldots, 1, 1, 1, \ldots, 0, 0]^{\mathrm{T}} \in \mathrm{R}^N$ and $\mathbf{D}$ is the N dimensional column vector of $1s$.

$\mathbf{S}_{\mathrm{b}}$ and $\mathbf{S}_{\mathrm{w}}$ can be expressed as:

$$\begin{aligned} \mathbf{S}_{\mathrm{b}} &= \tfrac{1}{N}\sum_{i=1}^{C} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^{\mathrm{T}} \\ &= \mathbf{X}[\tfrac{1}{N}\sum_{i=1}^{C}\tfrac{1}{n_i}\mathbf{BB}^{\mathrm{T}} - \tfrac{2}{N^2}\sum_{i=1}^{C}\mathbf{BD}^{\mathrm{T}} + \tfrac{1}{N^2}\mathbf{DD}^{\mathrm{T}}]\mathbf{X}^{\mathrm{T}}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{S}_{\mathrm{w}} &= \tfrac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{n_i}(\mathbf{x}_{i,j} - \mathbf{m}_i)(\mathbf{x}_{i,j} - \mathbf{m}_i)^{\mathrm{T}} \\ &= \mathbf{X}[\tfrac{1}{N}I - \tfrac{2}{N}\sum_{i=1}^{C}\sum_{j=1}^{n_i}\tfrac{1}{n_i}\mathbf{e}_{i,j}\mathbf{B}^{\mathrm{T}} \\ &\quad + \tfrac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{n_i}(\tfrac{1}{n_i{}^2}\mathbf{BB}^{\mathrm{T}})]\mathbf{X}^{\mathrm{T}} \end{aligned}$$

where $\mathbf{I}$ is the identity matrix.

Then

$$\begin{aligned} &\mathbf{S}_{\mathrm{b}} - \eta\mathbf{S}_{\mathrm{w}} \\ &= \mathbf{X}[\tfrac{1}{N}\sum_{i=1}^{C}\tfrac{1}{n_i}\mathbf{BB}^{\mathrm{T}} - \tfrac{2}{N^2}\sum_{i=1}^{C}\mathbf{BD}^{\mathrm{T}} + \tfrac{1}{N^2}\mathbf{DD}^{\mathrm{T}} - \eta\tfrac{1}{N}\mathbf{I} \\ &\quad + \eta\tfrac{2}{N}\sum_{i=1}^{C}\sum_{j=1}^{n_i}\tfrac{1}{n_i}\mathbf{e}_{i,j}\mathbf{B}^{\mathrm{T}} - \eta\tfrac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{n_i}(\tfrac{1}{n_i{}^2}\mathbf{BB}^{\mathrm{T}})]\mathbf{X}^{\mathrm{T}} \\ &= \mathbf{XKX}^{\mathrm{T}} \end{aligned}$$

where $\mathbf{K}$ is the symmetric matrices.

So, we can change (8) into the following constrained problem:

$$\begin{array}{l} \min\ \mathrm{tr}[\mathbf{W}^{\mathrm{T}}[\mathbf{XLX}^{\mathrm{T}} - (\mathbf{S}_{\mathrm{b}} - \eta\mathbf{S}_{\mathrm{w}})]\mathbf{W}] \\ = \min\ \mathrm{tr}[\mathbf{W}^{\mathrm{T}}[\mathbf{X}(\mathbf{L} - \mathbf{K})\mathbf{X}^{\mathrm{T}}]\mathbf{W}] \\ = \min\ \mathrm{tr}[\mathbf{W}^{\mathrm{T}}(\mathbf{XMX}^{\mathrm{T}})\mathbf{W}] \\ \mathrm{s.t.}\ \mathbf{W}^{\mathrm{T}}\mathbf{XX}^{\mathrm{T}}\mathbf{W} = \mathbf{I} \end{array} \tag{9}$$

where $\mathbf{M} = \mathbf{L} - \mathbf{K}$ is the symmetric matrices.

Equation (9) can be solved by Lagrange multiplier method. The minimization criterion (9) can be transformed into solving the following GEM:

$$\mathbf{XMX}^{\mathrm{T}}\mathbf{w} = \lambda\mathbf{XX}^{\mathrm{T}}\mathbf{w}. \tag{10}$$

3.2 Ship Target Recognition Based on E-MMC-SPP

To overcome the SSS problem arising in MMC-SPP, we consider the matrix exponential approach, and propose a supervised exponential sparsity preserving projection (E-MMC-SPP) technique for the ship recognition. And then a method for solving the exponential eigenvalue problem is presented.

i) Supervised exponential sparsity preserving projection

Given an $n \times n$ matrix $\mathbf{A}$, its exponential is defined as follow [35]:

$$\exp(\mathbf{A}) = \mathbf{I}_n + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \ldots + \frac{\mathbf{A}^k}{k!} + \ldots$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix.

The following results about the matrix exponential are essential and useful for analyzing E-MMC-SPP.

i) $\exp(\mathbf{A})$ is a finite matrix.

ii) If square matrix $\mathbf{A}$ commutes with $\mathbf{B}$, i.e., $\mathbf{AB} = \mathbf{BA}$, then

$$\exp(\mathbf{A} + \mathbf{B}) = \exp(\mathbf{A}) \exp(\mathbf{B}).$$

iii) For an arbitrary square matrix $\mathbf{A}$, there exists the inverse of $\exp(\mathbf{A})$. This is given by

$$\exp(\mathbf{A})^{-1} = \exp(-\mathbf{A}).$$

iv) If $\mathbf{T}$ is a nonsingular matrix, then

$$\exp(\mathbf{T}^{-1}\mathbf{AT}) = \mathbf{T}^{-1} \exp(-\mathbf{A})\mathbf{T}.$$

v) If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are eigenvectors of $\exp(\mathbf{A})$ that are related to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are also eigenvectors of $\exp(\mathbf{A})$ that are related to the eigenvalues $e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_n}$ of $\exp(\mathbf{A})$.

Similar to the minimization problem for MMC-SPP, the criterion of E-MMC-SPP can be transformed to:

$$\begin{array}{l} \min \ \mathrm{tr}[\mathbf{W}^{\mathrm{T}} \exp(\mathbf{XMX}^{\mathrm{T}})\mathbf{W}] \\ \text{s.t. } \mathbf{W}^{\mathrm{T}} \exp(\mathbf{XX}^{\mathrm{T}})\mathbf{W} = \mathbf{I}. \end{array} \tag{11}$$

Then the optimal $\mathbf{w}'s$ are the eigenvectors corresponding to the smallest eigenvalues of the following GEP:

$$\exp(\mathbf{XMX}^{\mathrm{T}}) = \mu \exp(\mathbf{XX}^{\mathrm{T}})\mathbf{w}. \tag{12}$$

We can follow from the property of matrix exponential that $\exp(\mathbf{XMX}^{\mathrm{T}})$ and $\exp(\mathbf{XX}^{\mathrm{T}})$ are both positive definite. Then E-MMC-SPP does not suffer from the SSS problem and the GEP is equivalent to the following standard eigenvalue problem:

$$\begin{array}{l} \exp(\mathbf{XX}^{\mathrm{T}})^{-1} \exp(\mathbf{XMX}^{\mathrm{T}}) \\ = \exp(-\mathbf{XX}^{\mathrm{T}}) \exp(\mathbf{XMX}^{\mathrm{T}}) = \mu\mathbf{w}. \end{array} \tag{13}$$

Algorithm 1. Computing E-MMC-SPP (base version).

Input: Training data matrix $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_C]$ and the corresponding labels C; The dimension d of the projected subspace; Testing data $\mathbf{X}'$.

Output: Transformation matrix $\mathbf{W}$ and representation matrix $\mathbf{Y}$.

1: Construct the weight matrix $\mathbf{S}$ by solving the SPP problem, and form the matrix $\mathbf{L}$.

2: Construct the between-class scatter matrix $\mathbf{S}_{\mathrm{b}}$ and within-class scatter matrix $\mathbf{S}_{\mathrm{w}}$, and form the matrix $\mathbf{K}$ and the matrix $\mathbf{M}$.

3: Compute matrices $\mathbf{XX}^{\mathrm{T}}$, $\mathbf{XMX}^{\mathrm{T}}$, $\exp(-\mathbf{XX}^{\mathrm{T}})$ and $\exp(\mathbf{XMX}^{\mathrm{T}})$.

4: Solve the SEP for eigenvectors $\{\mathbf{w}_i\}_{i=1}^{d}$ corresponding to the smallest d eigenvalues.

5: Set $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d]$ which spans the projected subspace and compute $\mathbf{Y} = \mathbf{W}^{\mathrm{T}}\mathbf{X}$.

Extra computations of $\exp(-\mathbf{XX}^{\mathrm{T}})$ and $\exp(\mathbf{XMX}^{\mathrm{T}})$ are needed in Algorithm 1. When n is large, computing the matrix exponential is very time-consuming. Therefore, we should design an effective way to evaluate the matrix exponential.

ii) Solution to E-MMC-SPP

As discussed above, Algorithm 1 can not only obtain more valuable information than SPP, but also keep active even if the SSS problem occurs. However, it still suffers from heavy computational cost and storage requirement since one has to compute $\exp(-\mathbf{XX}^{\mathrm{T}})$ and $\exp(\mathbf{XMX}^{\mathrm{T}})$, and then solve the large-scale matrix exponential eigenvalue problem (13).

For large-scale eigenvalue problem, the Krylov subspace methods [36], [37] are the power tool, which is used to evaluate the matrix exponential-vector products. As we known, in a certain Krylov subspace method, the main cost is the computation of matrix vector products. Therefore, the key to deal with the eigenvalue problem of large-scale matrix exponential is the computation of matrix exponential-vector products, meaning that there is no need to compute the matrix exponential $\exp(-\mathbf{XX}^{\mathrm{T}})$ and $\exp(\mathbf{XMX}^{\mathrm{T}})$. In the rest of this subsection, we will show a numerical algorithm to evaluate the matrix exponential-vector products based on the structures of the matrix $\mathbf{XX}^{\mathrm{T}}$ and $\mathbf{XMX}^{\mathrm{T}}$.

For simplicity, we denote $\mathbf{A} = \mathbf{XX}^{\mathrm{T}}$, $\mathbf{B} = \mathbf{XMX}^{\mathrm{T}}$, then

$$\begin{array}{l} \exp(\mathbf{XX}^{\mathrm{T}})^{-1} \exp(\mathbf{XMX}^{\mathrm{T}}) \\ = \exp(-\mathbf{A}) \exp(\mathbf{B}) = \mu\mathbf{w}. \end{array} \tag{14}$$

By the matrix exponential property, if the matrix equation $\mathbf{AB} = \mathbf{BA}$ holds, we have $\exp(\mathbf{A} + \mathbf{B}) = \exp(\mathbf{A}) \exp(\mathbf{B})$. But it is easily to prove that there is $\mathbf{AB} = (\mathbf{BA})^{\mathrm{T}}$, not $\mathbf{AB} = \mathbf{BA}$, so the above equation does not hold. However, in fact, the difference of $\exp(-\mathbf{A}) \exp(\mathbf{B})$ and $\exp(\mathbf{B} - \mathbf{A})$ is little. Let

$$\mathbf{AB} = \mathbf{BA} + \Delta\mathbf{T}$$

and

$$\exp(\mathbf{B} - \mathbf{A}) = \exp(-\mathbf{A}) \exp(\mathbf{B}) + \Delta\mathbf{U}.$$

According to theorem presented in [26], we have

$$\|\Delta\mathbf{U}\| \le \frac{1}{2} \exp(\|\mathbf{B}\| + \|\mathbf{A}\|) \, \|\Delta\mathbf{T}\| .$$

Contrasted to the matrix $\exp(-\mathbf{A}) \exp(\mathbf{B})$ and $\exp(\mathbf{B} - \mathbf{A})$, $\Delta\mathbf{U}$ occupies a very small proportion. And then the GEP may be replaced with:

$$\begin{array}{l} \exp(\mathbf{XMX}^{\mathrm{T}} - \mathbf{XX}^{\mathrm{T}}) \\ = \exp[\mathbf{X}(\mathbf{M} - \mathbf{F})\mathbf{X}^{\mathrm{T}}] = \mu\mathbf{w} \end{array} \tag{15}$$

where $\mathbf{F}$ represents the square matrix and all elements are 1.

Next, we will evaluate the matrix-vector product. Denote the QR decomposition of $\mathbf{X}$ by

$$\mathbf{X} = \mathbf{Q} \begin{pmatrix} \mathbf{R} \\ 0 \end{pmatrix}$$

where $\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2] \in \mathrm{R}^{m \times m}$ is an orthogonal matrix with $\mathbf{Q}_1 \in \mathrm{R}^{m \times n}$ and $\mathbf{R} \in \mathrm{R}^{n \times n}$ is an upper triangular matrix.

Since $\mathbf{Q}_2$ is not needed for actual computation, the modified Gram-Schmidt orthogonalization can be used to perform QR decomposition. Then we have

$$\mathbf{X}(\mathbf{M} - \mathbf{F})\mathbf{X}^{\mathrm{T}} = \mathbf{Q}_1\mathbf{R}(\mathbf{M} - \mathbf{F})\mathbf{R}^{\mathrm{T}}\mathbf{Q}_1{}^{\mathrm{T}}.$$

Using the properties of matrix exponential, we get

$$\begin{aligned}&\exp[\mathbf{X}(\mathbf{M}-\mathbf{F})\mathbf{X}^{\mathrm{T}}]\\ &= \mathbf{Q}_1 \exp[\mathbf{R}(\mathbf{M}-\mathbf{F})\mathbf{R}^{\mathrm{T}}]{\mathbf{Q}_1}^{\mathrm{T}} + \mathbf{Q}_2{\mathbf{Q}_2}^{\mathrm{T}}.\end{aligned}$$

Here, $\mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}_m$. Then we get

$$\mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}_1{\mathbf{Q}_1}^{\mathrm{T}} + \mathbf{Q}_2{\mathbf{Q}_2}^{\mathrm{T}} = \mathbf{I}_m.$$

Further

$$\begin{aligned}&\exp[\mathbf{X}(\mathbf{M}-\mathbf{F})\mathbf{X}^{\mathrm{T}}]\\ &= \mathbf{Q}_1 \exp[\mathbf{R}(\mathbf{M}-\mathbf{F})\mathbf{R}^{\mathrm{T}}]{\mathbf{Q}_1}^{\mathrm{T}} + \mathbf{I}_m - \mathbf{Q}_1{\mathbf{Q}_1}^{\mathrm{T}}.\end{aligned}$$

Denote the spectral decomposition of $\mathbf{R}(\mathbf{M}-\mathbf{F})\mathbf{R}^{\mathrm{T}}$ by

$$\mathbf{R}(\mathbf{M}-\mathbf{F})\mathbf{R}^{\mathrm{T}} = \mathbf{V}\mathbf{D}V^{\mathrm{T}}$$

where $\mathbf{V} \in \mathrm{R}^{n\times n}$ is orthogonal, $\mathbf{D} \in \mathrm{R}^{r\times r}$ is diagonal.

According to the properties of matrix exponential, we have:

$$\begin{aligned}&\exp[\mathbf{X}(\mathbf{M}-\mathbf{F})\mathbf{X}^{\mathrm{T}}]\\ &= \mathbf{Q}_1\mathbf{V}\exp(\mathbf{D})\mathbf{V}^{\mathrm{T}}{\mathbf{Q}_1}^{\mathrm{T}} + \mathbf{I}_m - \mathbf{Q}_1{\mathbf{Q}_1}^{\mathrm{T}}.\end{aligned} \tag{16}$$

Matrix exponential of $\mathbf{D}$ can be easily calculated since it is diagonal. To avoid overflow of the matrix exponential calculation when the biggest value in $\mathbf{D}$ is too large, we can choose a proper scaling parameter $\tau > 0$ replace $\mathbf{R}(\mathbf{M}-\mathbf{F})\mathbf{R}^{\mathrm{T}}$ by $\frac{1}{\tau}\mathbf{R}(\mathbf{M}-\mathbf{F})\mathbf{R}^{\mathrm{T}}$.

The selection of $\tau > 0$ is critical, in this paper we choose it as $\tau = \lambda_D$, where λ_D is the biggest value in $\mathbf{D}$.

The whole procedure of performing classification by E-MMC-SPP can be formally summarized as follows.

Algorithm 2. Computing E-MMC-SPP (optimized version).

Input: Training data matrix $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_C]$ and the corresponding labels C; The dimension d of the projected subspace; Testing data $\mathbf{X}'$.
Output: Transformation matrix $\mathbf{W}$ and representation matrix $\mathbf{Y}$.
1: Construct the weight matrix $\mathbf{S}$ by solving the SPP problem, and form the matrix $\mathbf{L}$.
2: Construct the between-class scatter matrix $\mathbf{S}_{\mathrm{b}}$ and within-class scatter matrix $\mathbf{S}_{\mathrm{w}}$, and form the matrix $\mathbf{K}$ and the matrix $\mathbf{M}$.
3: Compute the economic QR decomposition of $\mathbf{X}$ for $\mathbf{Q}_1$ and $\mathbf{R}$.
4: Compute the spectral decomposition of $\mathbf{R}(\mathbf{M}-\mathbf{F})\mathbf{R}^{\mathrm{T}}$ for $\mathbf{V}$ and $\mathbf{D}$.
5: Choose a proper scaling parameter τ.
6: Solve the SEP for eigenvectors $\{\mathbf{w}_i\}_{i=1}^{d}$ corresponding to the smallest d eigenvalues by a certain Krylov subspace method using (16) to generate the matrix-vector products.
7: Set $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d]$ which spans the projected subspace and compute $\mathbf{Y} = \mathbf{W}^{\mathrm{T}}\mathbf{X}$.

After the mapping matrix is obtained, both the training samples and the testing samples can be mapped into a low-dimensional feature space

$$\mathbf{Y} = \mathbf{W}^{\mathrm{T}}\mathbf{X}$$

and

$$\mathbf{Y}' = \mathbf{W}^{\mathrm{T}}\mathbf{X}'.$$

The proposed method belongs to subspace transformation, which is one of the pattern recognition. In this paper, the SVM classifier is adopted in the testing stage.

4. Experimental and Results

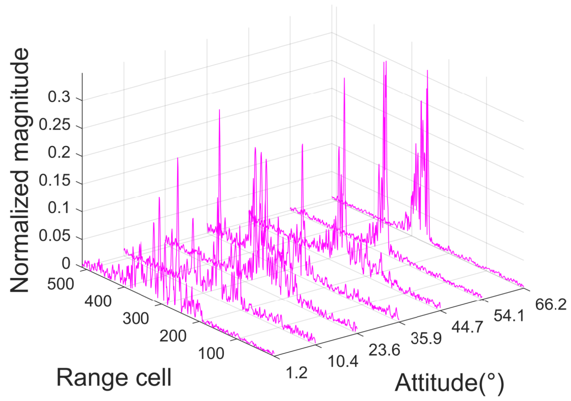
In this section, several experimental results are presented to demonstrate the effectiveness of the proposed method for classification tasks on the measured datasets. The real data experiment includes two sub-experiments: one is the data dimensionality reduction ability and algorithm optimization performance verification experiment, another is the recognition accuracy comparison experiment with small training data.

4.1 HRRP Data Set of Ship

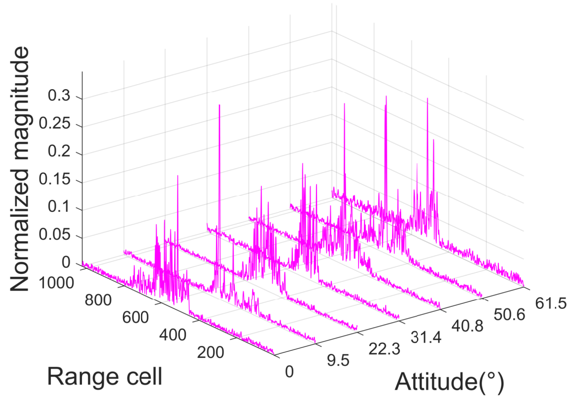
To verify the robustness and effectiveness of the proposed method, several key experiments were conducted using measured datasets. The datasets with three common radar bands are balanced and composed of four classes of ships respectively. The datasets are measured by numerous homogeneous coastal surveillance radars: each band belongs to L, X, or S. The bandwidth of L band is 50 MHz, there are 225 targets of each class, 900 in total, and the size of each raw HRRP is 512. The bandwidth of X band is 200 MHz, there are 360 targets of each class, 1400 in total, and the size of each raw HRRP is 1024. The bandwidth of S band is 300 MHz, there are 500 targets of each class, 2000 in total, and the size of each raw HRRP is 2048. The labels of the data samples are derived from intelligence support and the labels are confirmed by the operator.

Generally, the antenna erection height of coastal surveillance radar ranges from tens of meters to more than one thousand meters above sea level, and the maximum detection distance is tens of kilometers. Ignoring the inclination of the ship caused by the waves, it can be calculated that the range of the radar grazing angle is approximately 0.5° to 5°. Relative to the azimuth angle that can be changed in the range of 0° to 360°, the influence of the change of the grazing angle is negligible. Therefore, the azimuth sensitivity is mainly considered in this paper. Here we define attitude angle, which is the included angle between radar line of sight and target heading, and its value ranges from 0° to 180°. When the attitude is greater than 70° or less than 110°, the sea surface ship target is in tangential or approximate a tangential motion state. At this time, HRRPs cannot completely represent the inherent structural characteristics of the target.

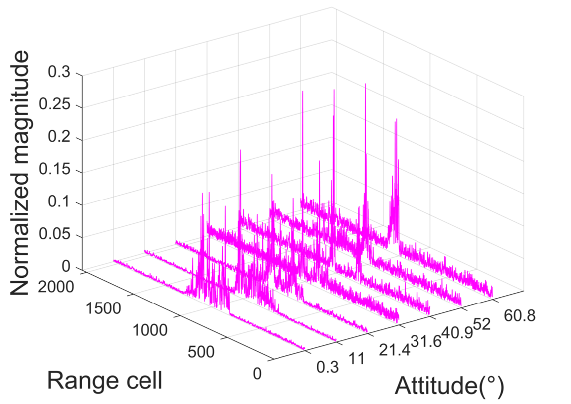
Hence, HRRPs with the attitude range of 0° to 70° are selected for experimental analysis. The datasets are balanced and consist of four classes of range profiles. A characteristic example HRRP of each class with L-band, X-band and S-band radar can be seen in Fig. 3. For the sake of ensuring the sample balance of the training samples and the testing samples, we divide the samples into 7 small parts according to 10° posture interval, and randomly extracts data from the 7 parts according to the proportion to form the training samples and the testing samples. The number of HRRPs for training and testing are tabulated in Tab. 1. So as to reduce the impact of amplitude sensitivity on the recognition results, the HRRPs are normalized.



(a)



(b)



(c)

Fig. 3. (a) Characteristic example HRRPs of ship target with L-band radar; (b) Characteristic example HRRPs of ship target with X-band radar; (c) Characteristic example HRRPs of ship target with S-band radar.

4.2 Recognition Accuracy Comparison with Different Feature Dimensions

In the experiments, the dimensionality reduction capability and optimization performance of the proposed algorithm is compared and analyzed with the classical PCA [38] and SPP [33]. The experiments are made from two aspects: one is to show the CPU time of implementing each algorithm, another is to evaluate the ship recognition performance of the proposed E-MMC-SPP method. All experiments are performed in MATLAB R2017b on an Intel Core 4 GHz PC with 16GB memory under Windows 10 system.

To compare the CPU time of implementing each algorithm, we conduct the experiment with different cropped sizes, that is, $n = 512$, $n = 1024$ and $n = 2048$. As comparison, the performance of SPP by keeping 98% energy in the PCA preprocessing step, abbreviated as PCA (0.98) + SPP, is also given. We run PCA, PCA (0.98) + SPP, Algorithm 1 and Algorithm 2. Table 2 shows the CPU times in seconds, barring the procedure of constructing weight matrix $\mathbf{S}$ since the procedure is run once for PCA (0.98) + SPP, Algorithm 1 and Algorithm 2.

From Tab. 2, we observe that the CPU times spent by all four algorithms increase as the cropped size increases. Compared with PCA and PCA (0.98) + SPP, Algorithm 1 and Algorithm 2 have higher CPU times, and they increase much faster with the data dimension increasing. This is mainly because the calculation of matrix exponents is time consuming. Besides, Algorithm 2 takes slightly less CPU times than Algorithm 1 with an order of magnitude difference. In fact, Algorithm 1 needs to compute the matrix exponential $\exp(-\mathbf{X}\mathbf{X}^{\mathrm{T}})$ and $\exp(\mathbf{X}\mathbf{M}\mathbf{X}^{\mathrm{T}})$. Algorithm 2 only needs to compute the matrix exponential $\exp(\mathbf{D})$. Then, it can be seen that the proposed numerical algorithm (Algorithm 2) can solve the large-scale matrix exponential eigenvalue problem.

Target	Training samples	Testing samples	Band
Class 1	125	100	L
Class 2	125	100	L
Class 3	125	100	L
Class 4	125	100	L
Class 1	200	150	X
Class 2	200	150	X
Class 3	200	150	X
Class 4	200	150	X
Class 1	300	200	S
Class 2	300	200	S
Class 3	300	200	S
Class 4	300	200	S

Tab. 1. The number of HRRP for training and testing for the four classes targets.

n	512	1024	2048
PCA	0.55	0.84	1.22
PCA(0.98) + SPP	0.91	1.53	5.14
Algorithm 1	45.37	659.21	4302.58
Algorithm 2	8.65	49.87	355.96

Tab. 2. CPU times of four algorithms on the measured HRRP.

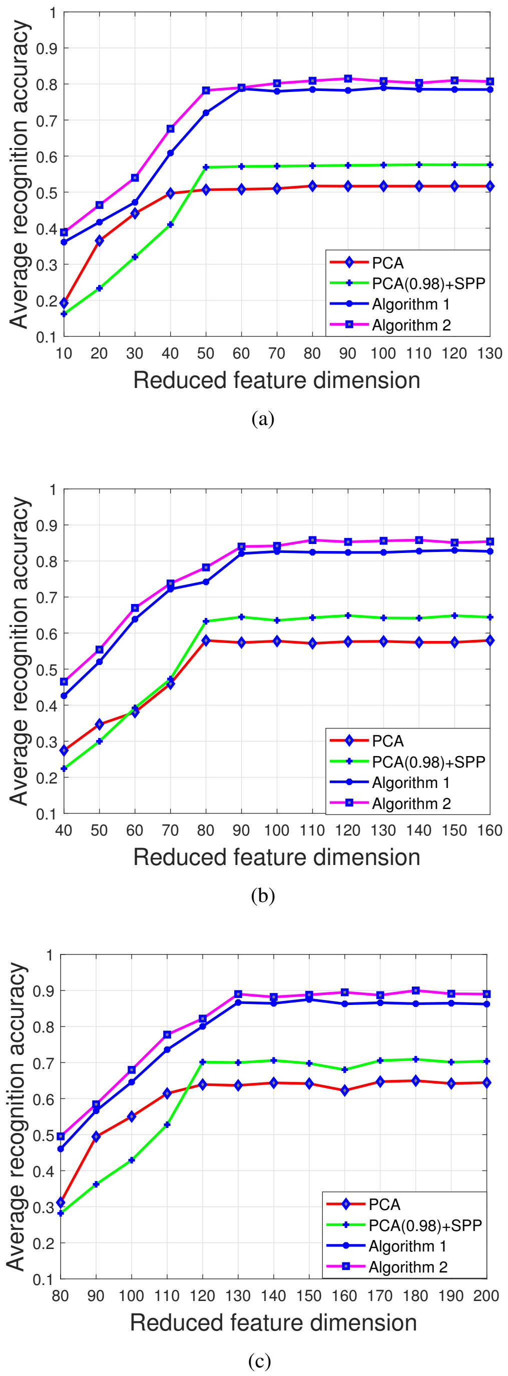


Fig. 4. (a) Recognition accuracy with different feature dimensions of L-band; (b) Recognition accuracy with different feature dimensions of X-band; (c) Recognition accuracy with different feature dimensions of S-band.

In this section, we will show the reduced feature dimension of the corresponding algorithms on the measured datasets. Figure 4 shows the average recognition accuracy of four classes with different feature dimensions base on three datasets. From the analysis, the average recognition accuracy of all methods improves with increasing feature dimensions. However, after a certain feature dimension, the average recognition accuracy performance remains almost unchanged. It can be seen that the average recognition accuracy of the four methods is very low with a lower reduced dimension. Besides, we observe that E-MMC-SPP can reach and even achieve higher recognition accuracy rates than PCA and PCA (0.98) + SPP, which indicates that E-MMC-SPP may have more capability SSS problem. Moreover, Algorithm 2 performs better than Algorithm 1 since the former can obtain more accurate projected subspace by using the structures of the coefficient matrices.

4.3 Recognition Accuracy Comparison with Different Training Sample Sizes

To verify the proposed method can solve small sample problem, the performance of the proposed method with other small sample recognition methods is compared. We have selected five comparison models. In this study, four types of DL networks are chosen, including: SAE [20], DDAEs [21], lightweight transformer network (LTN) [23], GANSO [39] and a subspace transformation ESPP [27] method.

i) Introduction to the comparative models

i) ESPP method: ESPP model is implemented through the MATLAB R2017b. The steps of the ESPP model include constructing the weight matrix, computing economic QR decomposition and the projected subspace.

ii) DL method: DL methods are implemented through the Pytorch, which includes four typical networks: LTN, SAE, DDAEs and GANSO. The feature extraction module of LTN consists of a local RNN module and a Transformer encoder modified. The classifier consists of a linear layer and a SoftMax function. The implementation of RNN is based on LSTM cells. The input sequence length is 31, and each time point is a 16-D vector. The SAE model is a stack of five auto encoders, where the number of neurons in each layer is 300, 600, 900, 2000, and 4. DDAEs are constructed with seven layers. The number of neuron from the bottom layer to the top layer are 256, 600, 400, 300, 300, 600 and 3, respectively. In addition, the sparsity target of the network is set to 0.05. GANSO consists of a generator and a discriminator. The generator consists of three linear layers and two ReLU layers. The number of linear layer is 256, 128, and 64 from the bottom to the upper layer. The discriminator consists of three linear layers and two ReLU layers. The number of linear layer is 64, 128, and 256 from the bottom to the upper layer.

In addition, if any researcher needs code, please contact the author, and it can be provided after evaluation.

ii) Experimental results using all training data

The training process of this experiment used all HRRP training samples, Among them, there are 500 samples of L-band, 800 samples of X-band and 1200 samples of S-band. The recognition rate and average recognition rate of the four classes in each band are used as evaluation indexes. Moreover, we also consider the balance between the recognition rates of various classes.

According to Tabs. 3–5, the recognition performance of different models for each class of there bands. It can be seen that the proposed model can effectively classify four classes. Comparing LTN, SAE, DDAEs and GANSO four typical DL methods, the proposed method has an average recognition rate of 6.17%, 4.88% and 7.93%, higher than the SAE model among there bands. It has an average recognition rate of 6.62%, 4.47% and 6.45%, higher than the DDAEs model. It has an average recognition rate of 2.25%, 1.45% and 1.43%, lower than the LTN model. It has an average recognition rate of 1.13%, 3.1% and 2.35%, lower than the GANSO model. Comparing with the traditional ESPP method, the model proposed in this article has an average recognition rate of 9.26%, 10.15% and 11.59% higher.

After analyzing the overall recognition performance of each method, the confusion matrix of each method is shown in Figs. 5–7. As can be seen, the accuracy of the proposed method is relatively average for all classes recognition. Among them, at S band, class4 with the highest recognition accuracy and class3 with the lowest recognition accuracy differ by only 3.43%. Other bands, the highest recognition accuracy and the lowest recognition accuracy differ is 1.99% and 1.67% respectively, which shows that the method proposed can model the characteristics of the four classes of ship in a more balanced manner. That is because the proposed model can use not only the latent physical structure information of the target shared by the training data set but also utilizes MMC to extract better separable features.

Methods	Class 1	Class 2	Class 3	Class 4	Average
Algorithm2	81.2	82.55	80.56	81.69	81.5
ESPP	70.52	76.3	78.77	63.37	72.24
LTN	80.22	90.8	76.51	89.43	84.24
SAE	70.98	82.17	78.44	69.73	75.33
DDAEs	72.62	76.84	73.5	76.56	74.88
GANSO	82.66	85.14	83.58	80.62	83

Tab. 3. Recogniton results of different methods at 125 training datasets of L-band [%].

Methods	Class 1	Class 2	Class 3	Class 4	Average
Algorithm2	86.12	85.45	84.98	86.65	85.8
ESPP	72.61	80.1	82.67	66.86	75.65
LTN	84.22	92.11	83.44	89.23	87.25
SAE	75.88	85.67	82.19	79.86	80.92
DDAEs	79.56	86.2	81.88	77.68	81.33
GANSO	87.58	92.44	89.4	86.18	88.9

Tab. 4. Recogniton results of different methods at 200 training datasets of X-band [%].

Methods	Class 1	Class 2	Class 3	Class 4	Average
Algorithm2	90.01	91.88	88.64	92.07	90.65
ESPP	75.12	81.57	84.66	74.89	79.06
LTN	91	93.87	93.14	90.31	92.08
SAE	78.59	86.74	83.77	81.78	82.72
DDAEs	82.5	88.9	84.67	80.73	84.2
GANSO	92.4	94.88	93.08	91.64	93

Tab. 5. Recogniton results of different methods at 300 training datasets of S-band [%].

Other models, more or less, have the problem of imbalance in recognition ability among various classes. For example, based on GANSO, although the average recognition rate of the four classes of ship reached 83% at S-band and 88.9% at X-band, the highest recognition accuracy and the lowest recognition accuracy differ is 4.52% and 6.26% respectively. The same problem is evident in ESPP, LTN, SAE and DDAEs.

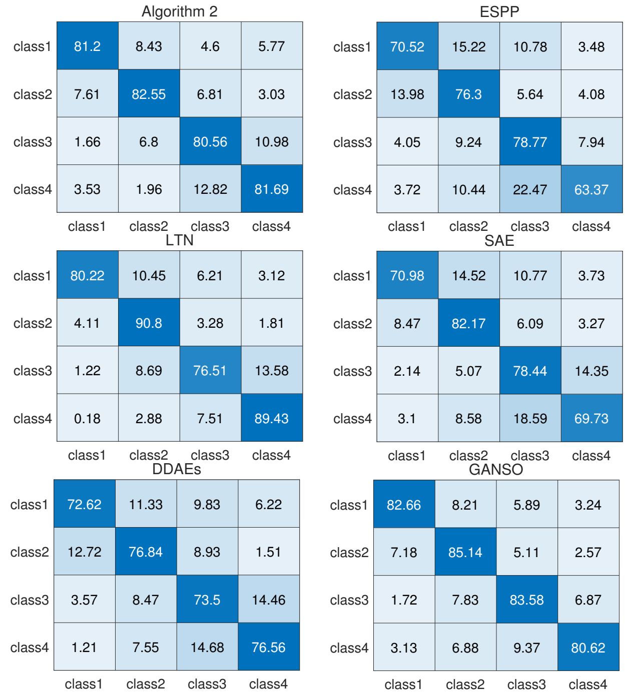


Fig. 5. Confusion matrix comparison of different models at 125 training datasets of L-band.

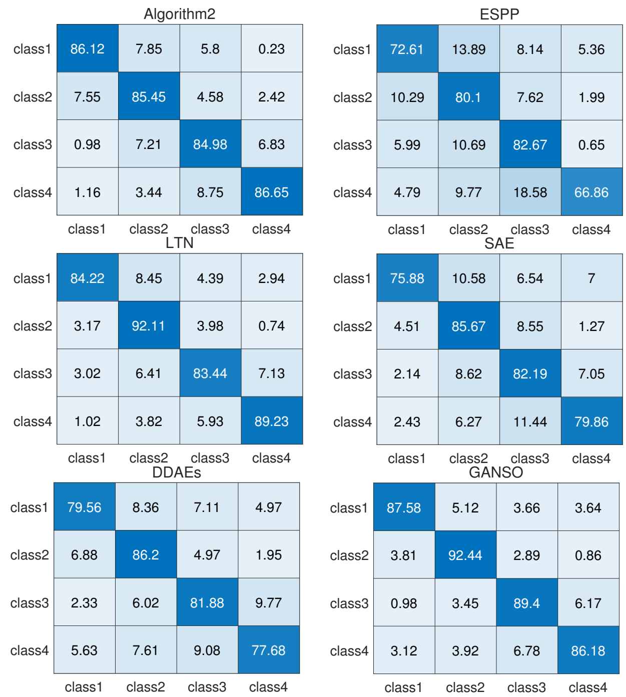


Fig. 6. Confusion matrix comparison of different models at 200 training datasets of X-band.

Algorithm2

	class1	class2	class3	class4
class1	90.01	5.88	3.47	0.64
class2	4.1	91.88	2.54	1.48
class3	1.21	4.33	88.64	5.82
class4	0.49	3.55	3.89	92.07

ESPP

	class1	class2	class3	class4
class1	75.12	11.87	9.52	3.49
class2	9.81	81.57	8.55	0.07
class3	4.97	8.24	84.66	2.13
class4	3.09	6.58	15.44	74.89

LTN

	class1	class2	class3	class4
class1	91	4.22	2.58	2.2
class2	2.14	93.87	2.23	1.76
class3	2.71	3.66	93.14	0.49
class4	0.94	3.82	4.93	90.31

SAE

	class1	class2	class3	class4
class1	78.59	12.8	5.98	2.63
class2	6.47	86.74	5.33	1.46
class3	2.7	6.66	83.77	6.87
class4	2.73	5.67	9.82	81.78

DDAEs

	class1	class2	class3	class4
class1	82.5	8.35	6.25	2.9
class2	4.05	88.9	5.37	1.68
class3	1.08	6.02	84.67	8.23
class4	2.15	6.88	10.24	80.73

GANSO

	class1	class2	class3	class4
class1	92.4	3.66	2.09	1.85
class2	1.49	94.88	2.77	0.86
class3	1.09	1.58	93.08	4.25
class4	0.78	3.55	4.03	91.64

Fig. 7. Confusion matrix comparison of different models at 300 trainingdata sets of S-band.

iii) Recognition results with different training sample sizes

The training sample size which is tabulated in Tab. 1 is set as 20%, 40%, 60%, 80%, and 100% of the original training sample size, respectively, and then we use all of the testing sample size as testing sample size to evaluate performances. Finally, the involved experimental results are shown in Fig. 8.

It can be seen from the experiments of three band datasets that under the condition of different number of training samples, the recognition accuracy of all methods increases with the increase of the number of training samples. In particular, LTN and GANSO networks are superior to the proposed E-MMC-SPP method (Algorithm 2) across all cases. But, Algorithm 2 outperforms the traditional ESPP and two types of DL methods, include SAE and DDAEs. As shown in Fig. 8, the recognition effect of SAE and DDAEs are more sensitive to the decrease of training sample size, especially when the training sample is small. This indicates that SAE and DDAEs have limited recognition ability under small sample conditions. In addition, compared with the application in air target recognition, SAE and DDAEs do not achieve the corresponding ship recognition performance with the small training samples. This may be that SAE and DDAEs networks are more suitable for anti-Gaussian white noise, and the anti-sea clutter ability is relatively weak. Meanwhile, Algorithm 2 performs better than ESPP since the former can obtain label information by using MMC.

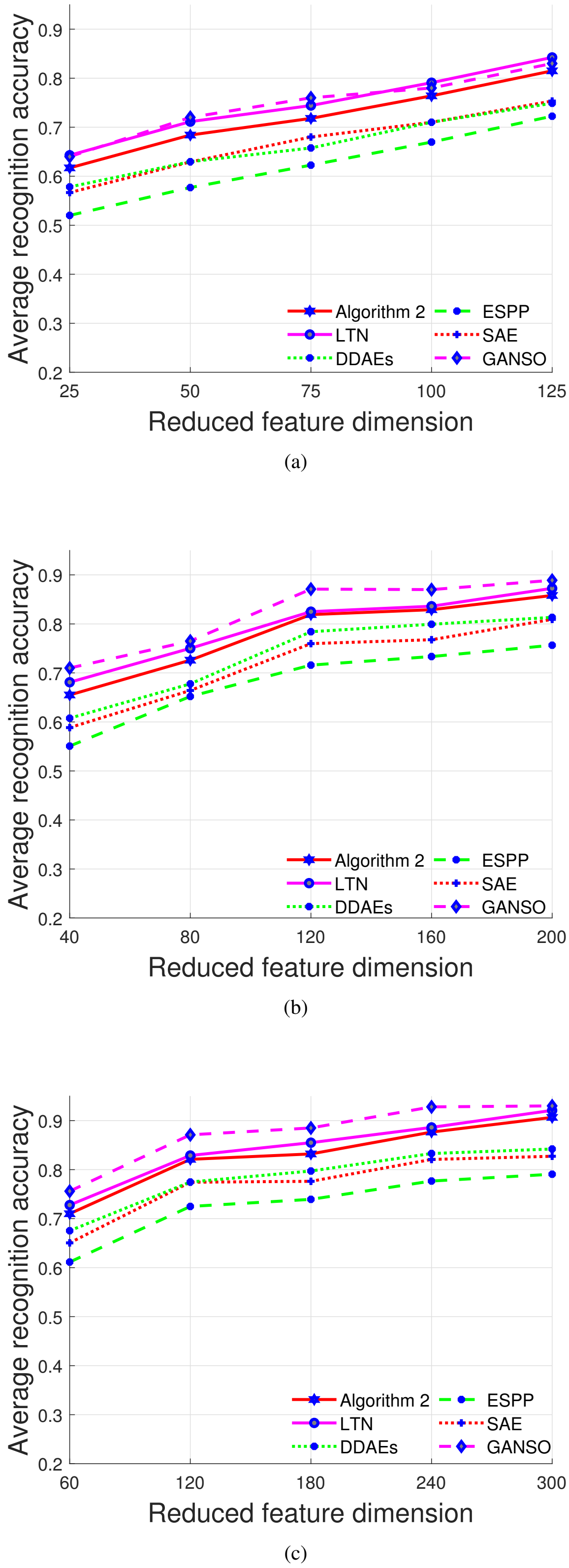


Fig. 8. (a) Recognition accuracy with different training sample size of L-band; (b) Recognition accuracy with different f training sample size of X-band; (c) Recognition accuracy with different training sample size of S-band.

5. Conclusion

In this paper, we propose a supervised exponential sparsity preserving projection (E-MMC-SPP) to enhance ship recognition based on HRRP with small training datasets. The main idea of E-MMC-SPP is that the maximum margin criterion and matrix exponential are introduced to SPP. To maximally preserve the structure of the data in the dimension-reduced space and improve the classification performance simultaneously, the merits of SPP and MMC are fused, which can maximally preserve the sparse reconstruct of data and maximize the class by separability. Meanwhile, the proposed method uses matrix exponential to ensure the positive definiteness of the coefficient matrices, which can overcome the SSS problem. In further, a numerical algorithm is presented to solve the large-scale matrix exponential eigenvalue problem which involves huge computational cost and storage requirement. The experiments are conducted on three band datasets: L, X and S. Extensive studies on the measured data demonstrate that the proposed method exhibits stable recognition performance even with small training datasets. Moreover, compared with other methods in HRRP target recognition, it shows the superiority of the proposed method.However, the comparative analysis experiments were conducted on different band data sets. In the next step, we plan to carry out relevant work on the universality of different band data and improve the recognition accuracy. In a word, the involved results could supply some reference for engineering application.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62271255), The Open Research Fund of the Key Laboratory of Radar Imaging and Microwave Photonics (Nanjing University of Aeronautics and Astronautics), Ministry of Education (NJ20220008, NJ20210001), Aeronautical Science Foundation (ASFC-201920007002), Shenzhen Science and Technology Program (JCYJ20210324134807019).

References

[1] HADAVI, M., RADMARD, M., NAYEBI, M. A new MCMC sampling based segment model for radar target recognition. *Radioengineering*, 2015, vol. 24, no. 1, p. 280–287. DOI: 10.13164/re.2015.0280

[2] ZENG, Z., SUN, J., HAN, Z., et al. Radar HRRP target recognition method based on multi-input convolutional gated recurrent unit with cascaded feature fusion. *IEEE Geoscience and Remote Sensing Letters*, 2022, vol. 19, p. 1–5. DOI: 10.1109/LGRS.2022.3192289

[3] ZHANG, K., SHUI, P. L. Estimation of complex high-resolution range profiles of ships by sparse recovery iterative minimization method. *IEEE Transactions on Aerospace and Electronic Systems*, 2021, vol. 57, no. 5, p. 3042–3056. DOI: 10.1109/TAES.2021.3068431

[4] TURK, M., PENTLAND, A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1991, vol. 3, no. 1, p. 71–86. DOI: 10.1162/jocn.1991.3.1.71

[5] BELHUMEUR, P. N., HESPANHA, J. P., KRIEGMAN, D. J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, vol. 19, no. 7, p. 711–720. DOI: 10.1109/34.598228

[6] BELKIN, M., NIYOGI, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 2003, vol. 15, no. 6, p. 1373–1396. DOI: 10.1162/089976603321780317

[7] ZHANG, Q., LU, J., LIU, T., et al. Ship HRRP target recognition based on CNN and ELM. In *Proceedings of the 2019 4th International Conference on Electromechanical Control Technology and Transportation (ICECTT)*. Guilin (China), 2019, p. 124–128. DOI: 10.1109/ICECTT.2019.00035

[8] LI, Y. G., GUO, L. B., WANG, Y., et al. A method of radar HRRP aircraft type identification and rejection based on LSTM. In *Proceedings of the 2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*. Changchun (China), 2023, p. 1115–1118. DOI: 10.1109/EEBDA56825.2023.10090595

[9] PAN, M., LIU, A., YU, Y., et al. Radar HRRP target recognition model based on a stacked CNN–Bi-RNN with attention mechanism. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, vol. 60, p. 1–14. DOI: 10.1109/TGRS.2021.3055061

[10] PAN, M., JIANG, J., KONG, Q., et al. Radar HRRP target recognition based on t-SNE segmentation and discriminant deep belief network. *IEEE Geoscience and Remote Sensing Letters*, 2017, vol. 14, no. 9, p. 1609–1613. DOI: 10.1109/LGRS.2017.2726098

[11] ZHANG, L., LI, Y., WANG, Y., et al. Polarimetric HRRP recognition based on ConvLSTM with self-attention. *IEEE Sensors Journal*, 2021, vol. 21, no. 6, p. 7884–7898. DOI: 10.1109/JSEN.2020.3044314

[12] SALAZAR, A., VERGARA, L., VIDAL, E. A proxy learning curve for the Bayes classifier. *Pattern Recognition*, 2023, vol. 136, p. 1–14. DOI: 10.1016/j.patcog.2022.109240

[13] ROWEIS, S. T., SAUL, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, vol. 290, no. 5500, p. 2323–2326. DOI: 10.1126/science.290.5500.2323

[14] QIAO, L. S., CHEN, S. C., TAN, X. Y. Sparsity preserving projections with applications to face recognition. *Pattern Recognition*, 2010, vol. 43, no. 1, p. 331–341. DOI: 10.1016/j.patcog.2009.05.005

[15] PANG, Y., WANG, S., YUAN, Y. Learning regularized LDA by clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, vol. 25, no. 12, p. 2191–2201. DOI: 10.1109/TNNLS.2014.2306844

[16] LU, J., TAN, Y. P. Regularized locality preserving projections and its extensions for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2010, vol. 40, no. 3, p. 958–963. DOI: 10.1109/TSMCB.2009.2032926

[17] LI, L., LIU, Z., LI, T. Radar high-resolution range profile feature extraction method based on multiple kernel projection subspace fusion. *IET Radar, Sonar & Navigation*, 2018, vol. 12, no. 4, p. 417–425. DOI: 10.1049/iet-rsn.2017.0391

[18] LIU, W. B., YUAN, J. W., ZHANG, G., et al. HRRP target recognition based on kernel joint discriminant analysis. *Journal of Systems Engineering and Electronics*, 2019, vol. 30, no. 4, p. 703–708. DOI: 10.21629/jsee.2019.04.08

[19] LI, C., BAO, W. M., XU, L. P., et al. Multiple kernel learning via orthogonal neighborhood preserving projection and maximum margin criterion method for synthetic aperture radar target recognition. *Optical Engineering*, 2018, vol. 57, no. 5, p. 1–10. DOI: 10.1117/1.OE.57.5.053105

[20] FENG, X., ZHANG, Y., HUO, C., et al. Small sample target recognition based on radar HRRP and sparse autoencoder. In *Proceedings of the 2023 8th International Conference on Computer and Communication Systems (ICCCS)*. Chengdu (China), 2023, p. 1043–1046. DOI: 10.1109/ICCCS.2023.000136

[21] MIAN, P., JIE, J., ZHU, L., et al. Radar HRRP recognition based on discriminant deep autoencoders with small training data size. *Electronics Letters*, 2016, vol. 52, no. 20, p. 1725–1727. DOI: 10.1049/el.2016.3060

[22] HUANG, Y., WEN, Y., SHI, L., et al. Recognition-aware HRRP generation with generative adversarial network. *IEEE Geoscience and Remote Sensing Letters*, 2022, vol. 19, p. 1–5. DOI: 10.1109/lgrs.2021.3056192

[23] YUE, Z., LU, J., WAN, L. Lightweight transformer network for ship HRRP target recognition. *Applied Sciences*, 2022, vol. 12, no. 19, p. 1–14. DOI: 10.3390/app12199728

[24] TAIPING, Z., BIN, F., YUAN YAN, T., et al. Generalized discriminant analysis: A matrix exponential approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2010, vol. 40, no. 1, p. 186–197. DOI: 10.1109/tsmcb.2009.2024759

[25] HUANG, S. C., ZHUANG, L. Exponential discriminant locality preserving projection for face recognition. *Neurocomputing*, 2016, vol. 208, p. 373–377. DOI: 10.1016/j.neucom.2016.02.063

[26] RAN, R., FANG, B., WU, X. Exponential neighborhood preserving embedding for face recognition. *IEICE Transactions on Information and Systems*, 2018, vol. E101.D, no. 5, p. 1410–1420. DOI: 10.1587/transinf.2017EDP7259

[27] WEI, W., DAI, H., LIANG, W. T. Exponential sparsity preserving projection with applications to image recognition. *Pattern Recognition*, 2020, vol. 104, p. 1–11. DOI: 10.1016/j.patcog.2020.107357

[28] LIU, J., CHERI, S. C., TAN, X. Y., et al. Efficient and robust feature extraction by maximum margin criterion. *IEEE Transactions on Neural Networks*, 2007, vol. 18, no. 6, p. 1862–1864. DOI: 10.1109/tnn.2007.900813

[29] YANG, H. C., FEI, C., WANG, X. W. Neighborhood preserving discriminant projection algorithm and its application in acoustic target recognition. *Journal of Computer Information Systems*, 2011, vol. 7, no. 12, p. 4471–4479.

[30] YANG, X.L., ZHANG, G., SONG, H. Ship recognition based on HRRP via multi-scale sparse preserving method. *Journal of Systems Engineering and Electronics*, 2024, vol. 35, no. 3, p. 599–608. DOI: 10.23919/jsee.2023.000136

[31] WU, G., FENG, T. T., ZHANG, L. J., et al. Inexact implementation using Krylov subspace methods for large scale exponential discriminant analysis with applications to high dimensionality reduction problems. *Pattern Recognition*, 2017, vol. 66, p. 328–341. DOI: 10.1016/j.patcog.2016.08.020

[32] WANG, S. J., CHEN, H. L., PENG, X. J., et al. Exponential locality preserving projections for small sample size problem. *Neurocomputing*, 2011, vol. 74, no. 17, p. 3654–3662. DOI: 10.1016/j.neucom.2011.07.007

[33] XU, B., CHEN, Y., GU, H., et al. Multi-view features-based HRRP classification via sparsity preserving projection. In *International Conference on Control, Automation and Information Sciences (ICCAIS)*. Jeju (South Korea), 2019, p. 1–6. DOI: 10.1109/ICCAIS46528.2019.9074684

[34] MIAO, X. Y., SHAN, Y. P. SAR target recognition via sparse representation of multi-view SAR images with correlation analysis. *Journal of Electromagnetic Waves and Applications*, 2019, vol. 33, no. 7, p. 897–910. DOI: 10.1080/09205071.2019.1575290

[35] LIAN, R. K.-L., SUBROTO, R. S., ANDREAN, V., et al. Matrix exponential. Chapter in *Harmonic Modeling of Voltage Source Converters using Basic Numerical Methods*, 2022, p. 73–94. DOI: 10.1002/9781119527190.ch4

[36] MIAO, C. Q., WU, W. T. On relaxed filtered Krylov subspace method for non-symmetric eigenvalue problems. *Journal of Computational and Applied Mathematics*, 2021, vol. 398, no. C, p. 1–15. DOI: 10.1016/j.cam.2021.113698

[37] XU, S. J., XUE, F. Inexact rational Krylov subspace method for eigenvalue problems. *Numerical Linear Algebra with Applications*, 2022, vol. 29, no. 5, p. 1–15. DOI: 10.1002/nla.2437

[38] CHEN, B., LIU, H. W., BAO, Z., et al. PCA and kernel PCA for radar high range resolution profiles recognition. In *IEEE Radar Conference*. Arlington (USA), 2005, p. 528–533. DOI: 10.1109/RADAR.2005.1435883

[39] SALAZAR, A., VERGARA, L., SAFONT, G. Generative adversarial networks and Markov random fields for oversampling very small training sets. *Expert Systems with Applications*, 2021, vol. 163, no. 1, p. 1–11. DOI: 10.1016/j.eswa.2020.113819

About the Authors ...

Xueling YANG was born in 1986. He received the B.S. and M.S. degrees from the College of Mathematical Sciences, Harbin Engineering University, Heilongjiang, China, in 2008 and 2011, respectively. He is currently pursuing the Ph.D. degree in Electronics and Information with the Nanjing University of Aeronautics and Astronautics, Jiangsu, China. Since 2011, he is a Senior Engineer with the Nanjing Marine Radar Institute. His current research interests are radar target recognition and radar signal processing.

Gong ZHANG was born in 1964. He received the Ph.D. degree in Electronic Engineering from Nanjing University of Aeronautics and Astronautics, Jiangsu, China, in 2002. From 1990 to 1998, he was a Member of Technical Staff at No. 724 Institute China Shipbuilding Industry Corporation (CSIC), Nanjing. Since 1998, he has been with the College of Electronics and Information Engineering, Nanjing University of Aeronautics and Astronautics, where he is currently a Professor. His research interests include radar target recognition, radar signal processing and compressive sensing. Prof. Zhang is a member of the Committee of Electromagnetic Information, Chinese Society of Astronautics (CEICSA), and a Senior Member of the Chinese Institute of Electronics (CIE).

Hu SONG was born in 1980. He received the M.S. and Ph.D degrees in Communication System, Nanjing University of Science and Technology, Nanjing, China, in 2005 and 2020, respectively. Since 2005, he is a Senior Engineer with the Nanjing Marine Radar Institute. His current research interest is radar target recognition and radar signal processing.