Applied Acoustics 159 (2020) 107057

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Integrated optimization of underwater acoustic ship-radiated noise recognition based on two-dimensional feature fusion

Xiaoquan Ke, Fei Yuan *, En Cheng

Key Laboratory of Underwater Acoustic Communication and Marine Information Technology (Xiamen University) Ministry of Education, Xiamen University, Xiamen 361005, China

ARTICLE INFO

Article history:
Received 16 February 2019
Received in revised form 10 April 2019
Accepted 19 September 2019
Available online 11 October 2019

Keywords:
Ship-radiated noise recognition
Pattern recognition
Feature fusion
Canonical correlation analysis

ABSTRACT

Feature fusion methods are introduced to ship-radiated noise recognition in this paper. Wavelet packet (WP) decomposition is used to decompose the ship-radiated noise into multiple different subbands. By considering the features extracted from the different subbands reflecting different characteristics of the ship-radiated noise, a two-dimensional feature fusion (2DFF) scheme is proposed to fuse the features extracted from the different subbands. Principal component analysis (PCA) and canonical correlation analysis (CCA) are used in the 2DFF scheme. Then, a so-called discriminative ability improving (DAI) strategy is proposed to improve the discriminative ability of the extracted features. Starting at the 2DFF, a processing chain of feature fusion and ship-radiated noise recognition is designed and jointly optimized to the task. The 2DFF scheme and DAI strategy are tested on real ship-radiated noise data recorded. Experimental results indicate that compared with the baseline, the 2DFF scheme can improve 7.25% of recognition accuracy. Experimental results also show that the DAI strategy can further improve the recognition accuracy of 13.10%.

1. Introduction

Generally, when a ship moves in the water, it produces noise, called ship-radiated noise. The ship-radiated noise, along with marine mammals' voice, natural ambient noise, constitutes the most part of acoustic sound in oceans, making underwater acoustic environment of the oceans informative and also very complicated. Generally speaking, ship-radiated noise of different classes of ships or different ships contains different acoustic characteristics, so it is possible to recognize different classes of ships by analysing the ship-radiated noise. Ship recognition based on ship-radiated noise also belong to pattern recognition problems, just like ship recognition based on images of ships. However, ship recognition based on ship-radiated noise is much more challenge than ship recognition based on the images of ships because the acoustic environment in the oceans is very complicated. The original ship-radiated noise is always seriously affected by the complicated environment during long-range transmission in the oceans.

Ship recognition based on ship-radiated noise is one of the most important and challenging subjects in underwater acoustic signal processing. Though methods of ship-radiated noise recognition have been developed for decades, performance of these methods still cannot satisfy practical demands. Typically, traditional methods of ship-radiated noise recognition are feature extraction followed by classification. Such as Jian et al. in [1] extracted line spectrum and line spectrum density features from ship-radiated noise and fed them into support vector machine (SVM) classifiers. In [2], Wei et al. introduced an approach for extracting ship-radiated noise based on $1\frac{1}{2}$D spectrum features. In [3], Mel-frequency cepstral coefficients (MFCCs) features were extracted and statistical classification of these features were based on Gaussian mixture models (GMMs). MFCCs, along with first-order differential MFCCs and second-order differential MFCCs features were also used in [4] to recognize underwater targets. Meng and Yang in [5] designed a combined feature vector of zero-crossing wavelength, peek-to-peek amplitude, and zero-crossing-wavelength difference for the recognition of ship-radiated noise. In [6], linear predictive coding (LPC) features were extracted, in [7], energy distribution features in the blocks of wavelet packet (WP) coefficients were extracted, and in [8], Hilbert spectral features were extracted for the recognition of ships.

Except the traditional feature extraction methods, more recently, to solve the feature extraction problems, many researchers also use deep neural networks to extract features. Utilizing deep neural networks to extract features requires less engineering skill, domain expertise and prior knowledge, but utilizing neural networks can also achieve a competitive performance even an

* Corresponding author.
E-mail address: yuanfei@xmu.edu.cn (F. Yuan).

https://doi.org/10.1016/j.apacoust.2019.107057
0003-682X/

outstanding performance compared with the traditional methods that extract the hand-crafted features. For example, in [9], Cao et al. used stacked autoencoder (SAE) to extract high-level features for ship recognition, and in [10], Deep Belief Nets (DBN) and a so called competitive DBN were used to extract informative features from ship-radiated noise.

The existing feature extraction methods mainly focus on extracting a single kind of specific features for ship-radiated noise recognition. However, a single kind of features usually cannot comprehensively describe different characteristics of the ship-radiated noise and that cannot satisfy other complicated situations. Therefore, in this paper, we introduce *feature fusion* methods to fully utilize different characteristics of the ship-radiated noise to meet the demands of different situations. Feature fusion has been applied to many other pattern recognition problems, such as handwritten digit recognition in [11], and face recognition in [11,12], which have yielded better recognition performance. Different features extracted from the same patterns always reflect different characteristics of these patterns. By fusing and optimizing these different features, it not only keeps the effective discriminant information of multi-feature, but also eliminates the redundant information to a certain degree. Inspired by this, we first use WP decomposition to decompose ship-radiated noise into multiple subbands and we assume that the same kinds of features but extracted from *different subbands* can also reflect *different characteristics* of the ship-radiated noise. Then, inspired by feature fusion using canonical correlation analysis (CCA) [12], in this paper, we propose a two-dimensional feature fusion (2DFF) scheme to fuse the features not only *intra* each subband but also *inter* each subband. Considering the discriminative ability of the features, we propose a so-called discriminative ability improving (DAI) strategy to improve the discriminative ability of the features during the process of the 2DFF. Utilizing Bayesian Optimization [13] and starting at the 2DFF, a processing chain of feature fusion and ship-radiated noise recognition is designed and jointly optimized to the task.

Traditionally, feature fusion methods fuse the features extracted using different feature extraction methods [11,12]. That is fusing different features from the same patterns. However, we provide the 2DFF scheme that fuses the same features but from different subbands, which can be viewed as fusing the same features but from "different patterns". The 2DFF scheme fusing the same features from different patterns is similar to the traditional feature fusion methods of fusing different features from the same patterns because both of these two methods can utilize different characteristics of the ship-radiated noise. More specially, the key idea of the 2DFF scheme is to fully utilize different characteristics of multiple subbands of the ship-radiated noise.

The rest of this paper contain 5 sections. Section 2 introduces basic methodologies of the 2DFF scheme. Section 3 will fully explain technical detail of the DAI strategy and 2DFF scheme. Section 4 demonstrates experiments and discusses experimental results. The last section is conclusion.

2. Basic methodologies of 2DFF scheme

Ship-radiated noise recognition contains two steps: feature extraction and classification. As feature fusion is introduced to ship-radiated noise recognition, process of the recognition will be enriched to: feature extraction, feature fusion and classification. Feature extraction is introduced in this section. As CCA is the core idea of the feature fusion, CCA will be introduced in detail.

2.1. Feature extraction

Many different feature extraction methods used in ship-radiated noise recognition are studied and the extracted features are grouped according to the feature extraction methods. More specially, the extracted features are *grouped* into 8 groups, including temporal features, statistical features, spectral features, cepstral features, Hilbert spectral features, wavelet features, specific features, and deep neural network features. By extracting different kinds of features, we hope that these features can comprehensively reflect different characteristics of the ship-radiated noise. Table 1 lists these 8 grouped features and dimensions (number in the angle brackets) of these features.

We preliminarily use serial feature fusion (SSF) strategy [20] to fuse all the above features and form a *feature vector*. The SSF strategy can be explained as follows:

$$\gamma = (x_1 \quad x_2) \tag{1}$$

where x_1 and x_2 are originally extracted features, and γ is serially combined feature. Obviously, if x_1 is n-dimensional and x_2 is m-dimensional, then the serially combined feature γ is $(n+m)$-dimensional. We use SSF strategy to fuse all the originally extracted features one by one and we will finally obtain a serially combined feature vector γ of 188 dimensions. The serially combined feature vector γ is formulated as follows:

$$\gamma = (x_1 \quad x_2 \quad \cdots \quad x_j \quad \cdots \quad x_{188}) \tag{2}$$

where x_j is the jth originally extracted feature. The feature vector γ is also depicted in Fig. 1.

Table 1
Feature extraction and grouping.

Feature groups	Features
Temporal ⟨12⟩	Area ⟨1⟩ [14], Line length ⟨1⟩ [14], Root mean square ⟨1⟩ [14], Zero-crossing rate ⟨1⟩ [14], Mean energy ⟨1⟩ [14], Maximum energy[1] ⟨1⟩, Total energy [2] ⟨1⟩, Entropy of energy ⟨1⟩ [15], Normalized peak number ⟨1⟩ [14], Peak variation ⟨1⟩ [14], Temporal centroid ⟨1⟩ [16], Normalized decay ⟨1⟩ [14]
Statistical ⟨2⟩	Skewness ⟨1⟩ [17], Kurtosis ⟨1⟩ [17]
Spectral ⟨39⟩	Spectral flatness ⟨1⟩ [18], Spectral roll-off ⟨1⟩ [19], Spectral entropy ⟨1⟩ [15], Spectral zone ⟨1⟩ [15], Spectral Centroid ⟨1⟩ [15], Spectral flux ⟨19⟩ [15], Chroma vector ⟨12⟩ [15], Harmonic ratio ⟨1⟩ [15], Fundamental frequency ⟨1⟩ [15]
Cepstral ⟨15⟩	MFCCs ⟨13⟩ [15], MFCC means ⟨1⟩ [15], MFCC variances ⟨1⟩ [15]
Hilbert spectral ⟨6⟩	Mean value of instantaneous amplitude ⟨1⟩ [8], Maximum value of instantaneous amplitude ⟨1⟩ [8], Mean value of the instantaneous frequency ⟨1⟩ [8], Root mean square value of the instantaneous frequency ⟨1⟩ [8], Weighted mean instantaneous frequency ⟨1⟩ [8], Weighted center frequency ⟨1⟩ [8]
Wavelet ⟨63⟩	Energy spectrum of WP transform ⟨32⟩ [7], Normalized energy spectrum of Tunable Q-factor Wavelet Transform[3] ⟨31⟩
Specific ⟨19⟩	Normalized line spectrum number ⟨5⟩ [1], Normalized line spectrum density ⟨5⟩ [1], Frequency band power ⟨5⟩ [14], LPC coefficients ⟨4⟩ [6]
Deep neural network ⟨32⟩	DBN hidden unit features[4] ⟨32⟩

[1] Maximum energy is explained in Appendix A.
[2] Total energy is explained in Appendix A.
[3] Normalized energy spectrum of Tunable Q-factor Wavelet Transform is explained in Appendix A.
[4] DBN hidden unit features is explained in Appendix A.

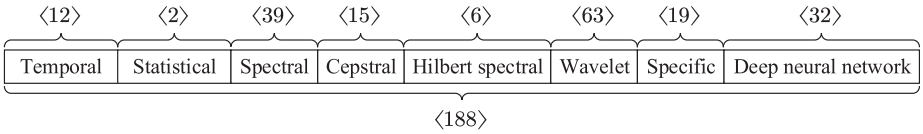


Fig. 1. Feature vector γ.

2.2. Canonical correlation analysis

In multivariate statistical analysis, correlation problem of two random vectors often needs to be studied, that is to convert the correlation research of two random vectors into that of a few pairs of variables, which are uncorrelated [11]. CCA is one of the valuable multi-data processing methods, which has been widely used to analyse the mutual relationships between two sets of variables [12]. Assume that $X \in \mathbb{R}^{p\times n}$ and $Y \in \mathbb{R}^{q\times n}$ denote two different modalities from the same patterns, where p is feature dimension of the first modality, q is feature dimension of the second modality, and n is sample number of each modality. Different modalities mean that, for the same patterns, X and Y are obtained using different feature extraction methods.

Suppose that $S_{XX} \in \mathbb{R}^{p\times p}$ and $S_{YY} \in \mathbb{R}^{q\times q}$ denote the within-sets covariance matrices of X and Y, and S_{XY} denotes the between-sets covariance matrix. The overall $(p+q)\times(p+q)$ covariance matrix S can be combined as follows:

$$S = \begin{pmatrix} cov(X) & cov(X,Y) \\ cov(Y,X) & cov(Y) \end{pmatrix} = \begin{pmatrix} D(X) & E\left(XY^T\right) \\ E\left(YX^T\right) & D(Y) \end{pmatrix} = \begin{pmatrix} S_{XX} & S_{XY} \\ S_{YX} & S_{YY} \end{pmatrix} \tag{3}$$

CCA aims to find linear combinations, $X^* = W_x^T X$ and $Y^* = W_y^T Y$ that maximize pair-wise correlations across the two feature sets:

$$\arg\max \frac{cov(X^*,Y^*)}{\sqrt{D(X^*)}\sqrt{D(Y^*)}} \tag{4}$$

where:

$$\begin{aligned} cov(X^*,Y^*) &= cov\left(W_x^T X, W_y^T Y\right) = E\left(\left(W_x^T X\right)\left(W_y^T Y\right)\right) \\ &= W_x^T E\left(XY^T\right) W_y = W_x^T S_{XY} W_y \end{aligned} \tag{5}$$

$$D(X^*) = D\left(W_x^T X\right) = W_x^T E\left(XX^T\right) W_x = W_x^T S_{XX} W_x \tag{6}$$

$$D(Y^*) = D\left(W_y^T Y\right) = W_y^T E\left(YY^T\right) W_y = W_y^T S_{YY} W_y \tag{7}$$

Eq. (4) can be transformed to:

$$\arg\max \frac{cov(X^*,Y^*)}{\sqrt{D(X^*)}\sqrt{D(Y^*)}} = \arg\max_{W_x,W_y} \frac{W_x^T S_{XY} W_y}{\sqrt{W_x^T S_{XX} W_x}\sqrt{W_y^T S_{YY} W_y}} \tag{8}$$

Conclusively, projective matrices W_x and W_y can be obtained by solving the following optimization problems:

$$\begin{cases} \arg\max_{W_x,W_y} W_x^T S_{XY} W_y \\ W_x^T S_{XX} W_x = 1 \\ W_y^T S_{YY} W_y = 1 \end{cases} \tag{9}$$

Optimization of Eq. (9) is usually performed using Lagrange multipliers [12]. Note that the linear combinations X^* and Y^* obtained by solving Eq. (9) are also known as *canonical variates*.

Following the canonical variates:

$$Z_1 = \begin{pmatrix} X^* \\ Y^* \end{pmatrix} = \begin{pmatrix} W_x^T X \\ W_y^T Y \end{pmatrix} = \begin{pmatrix} W_x & 0 \\ 0 & W_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix} \tag{10}$$

$$Z_2 = X^* + Y^* = W_x^T X + W_y^T Y = \begin{pmatrix} W_x \\ W_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix} \tag{11}$$

as the *canonical correlation discriminant features (CCDFs)* [11], used for pattern recognition.

3. Technical detail of DAI strategy and 2DFF scheme

Section 2 has introduced basic methodologies of the 2DFF scheme, this section will explain technical detail of the DAI strategy and 2DFF scheme. A detailed flow chart of this section is depicted in Fig. 2.

3.1. Database

All data of ship-radiated noise used in this paper come from a database called ShipsEar [3]. During 2012 and 2013, sounds of many different classes of ships were recorded on Spanish Atlantic coast and are included in the ShipsEar database (available at http://atlanttic.uvigo.es/underwaternoise/). The recordings were made with autonomous acoustic digitalHyd SR-1 recorders, manufactured by MarSensing Lda (Faro, Portugal). According to [3], 11 vessel types are merged into 4 experimental classes (based

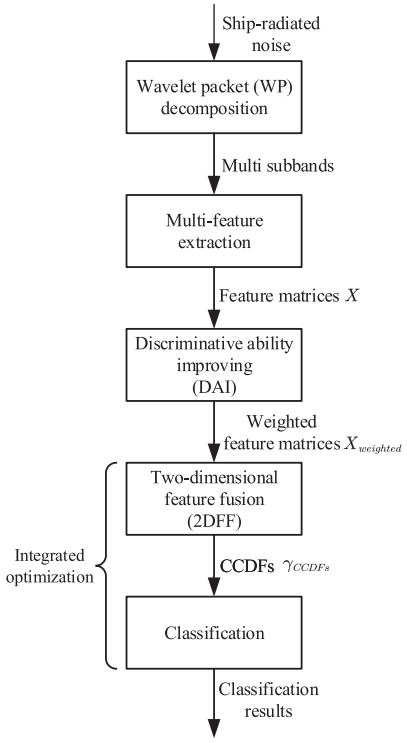


Fig. 2. The flow chart of the ship-radiated noise recognition.

Table 2
Ship grouping.

Class A	fishing boats, trawlers, mussel boats, tugboats, dredgers
Class B	motorboats, pilot boats, sailboats
Class C	passenger ferries
Class D	ocean liners and ro-ro vessels
Class E	background noise recordings

on vessel size) and 1 background noise class, as showed in Table 2. Some ship pictures and ship-radiated noise of each class are demonstrated in Fig. 3.

3.2. Multi-feature extraction

Different features extracted from the same patterns always reflects different characteristics of the patterns. In [6], WP decomposition was applied to the same patterns and a couple of subbands were decomposed from the patterns. Then LPC coefficients were extracted from these different subbands and these coefficients were assumed to reflect different characteristics of the patterns. Inspired by this, we also extract the same kinds of features from *different subbands* and we assume that these features can also reflect *different characteristics* of the ship-radiated noise.

Assume that the number of subbands decomposed by WP decomposition is I, then a *feature matrix* X extracted from a raw ship-radiated noise sample can be formulated as:

$$X = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_i \\ \vdots \\ \gamma_I \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,j} & \cdots & x_{1,188} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,j} & \cdots & x_{2,188} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i,1} & x_{i,2} & \cdots & x_{i,j} & \cdots & x_{i,188} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{I,1} & x_{I,2} & \cdots & x_{I,j} & \cdots & x_{I,188} \end{pmatrix} \tag{12}$$

where γ_i is the feature vector extracted from the ith subband, and $x_{i,j}$ is the jth feature of the ith subband. For a raw sample, the size of the feature matrix X is $I \times 188$. The feature matrix X is depicted in Fig. 4.

3.3. Discriminative ability improving strategy

For each subband, up to 188 different features are extracted, and for the feature matrix X, there are $I \times 188$ features. Therefore, it is better to consider the discriminative ability[1] of each feature of the feature matrices. In this subsection, we evaluate the discriminative ability of the extracted features and we propose the DAI strategy to improve the discriminative ability of the feature matrices. The DAI strategy will be explained as follows:

To improve the discriminative ability of X^A, we use *Fisher discriminant matrix* $\overline{D_B}$[2] to *weight* X^A:

$$X^A_{wieghted} = X^A \cdot \overline{D_B} \tag{13}$$

The same weighting operation will be applied to X^B, X^C, X^D and X^E:

$$X^B_{wieghted} = X^B \cdot \overline{D_B} \tag{14}$$

$$X^C_{wieghted} = X^C \cdot \overline{D_B} \tag{15}$$

$$X^D_{wieghted} = X^D \cdot \overline{D_B} \tag{16}$$

$$X^E_{wieghted} = X^E \cdot \overline{D_B} \tag{17}$$

After using $\overline{D_B}$ to weight the feature matrices, we think that the features with better discriminative ability will be more *highlighted*.

3.4. Two-dimensional feature fusion scheme

The process of the 2DFF scheme is clearly depicted in Fig. 5. 2DFF means that we not only fuse the features intra each subband (the first dimension), but also fuse the features inter each subband (the second dimension).

Step 1: Principal component analysis (PCA) [21] is used to reduce the feature dimension of each subband of $X_{weighted}$, the reduced dimension of each subband is d_1 ($d_1 \leqslant 188$):

$$\begin{aligned} \mathrm{PCA}(X_{weighted}) &= \begin{pmatrix} x'_{1,1} & x'_{1,2} & \cdots & x'_{1,j} & \cdots & x'_{1,d_1} \\ x'_{2,1} & x'_{2,2} & \cdots & x'_{2,j} & \cdots & x'_{2,d_1} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x'_{i,1} & x'_{i,2} & \cdots & x'_{i,j} & \cdots & x'_{i,d_1} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x'_{I,1} & x'_{I,2} & \cdots & x'_{I,j} & \cdots & x'_{I,d_1} \end{pmatrix} = \begin{pmatrix} \gamma'_1 \\ \gamma'_2 \\ \vdots \\ \gamma'_i \\ \vdots \\ \gamma'_I \end{pmatrix} \\ &= X'_{weighted}\ X_{weighted} \in \left(X^A_{weighted}, X^B_{weighted}, X^C_{weighted}, X^D_{weighted}, X^E_{weighted}\right) \end{aligned} \tag{18}$$

where $x'_{i,j}$ is the jth feature of the ith subband of $X'_{weighted}$. This step is to reduce the feature dimension and keep the effective discriminant information intra each subband.

Step 2 (1): SSF strategy is used to fuse the features of all even subbands of $X'_{weighted}$:

$$\gamma_{even} = \mathrm{SSF}_{even}\left(X'_{weighted}\right) = (\gamma'_2 \quad \gamma'_4 \quad \cdots \quad \gamma'_{2i} \quad \cdots \quad \gamma'_I), \quad i = 1, 2, \cdots, I/2 \tag{19}$$

The size of γ_{even} is $I \times d_1/2$.

(2): SSF strategy is used to fuse the features of all odd subbands of $X'_{weighted}$:

$$\gamma_{odd} = \mathrm{SSF}_{odd}\left(X'_{weighted}\right) = (\gamma'_1 \quad \gamma'_3 \quad \cdots \quad \gamma'_{2i-1} \quad \cdots \quad \gamma'_I), \quad i = 1, 2, \cdots, I/2 \tag{20}$$

The size of γ_{odd} is $I \times d_1/2$. The purpose of this step is to preliminarily fuse the information inter each band.

Step 3: PCA is used to reduce the feature dimension of γ_{even} and γ_{odd}, respectively:

$$\gamma'_{even} = \mathrm{PCA}(\gamma_{even}) \tag{21}$$

$$\gamma'_{odd} = \mathrm{PCA}(\gamma_{odd}) \tag{22}$$

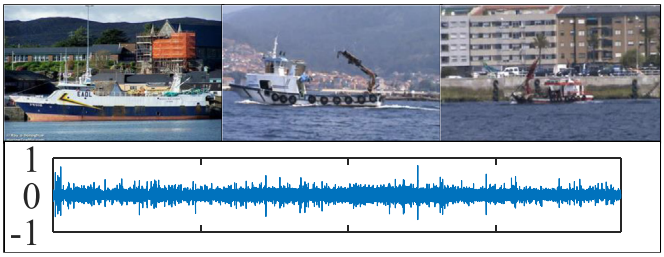
The feature dimension of γ'_{even} and γ'_{odd} are both d_2. The core idea of this step is to remove redundant information and keep the effective discriminant information inter each subband.

Step 4: CCA is eventually used to fuse γ'_{even} and γ'_{odd}:

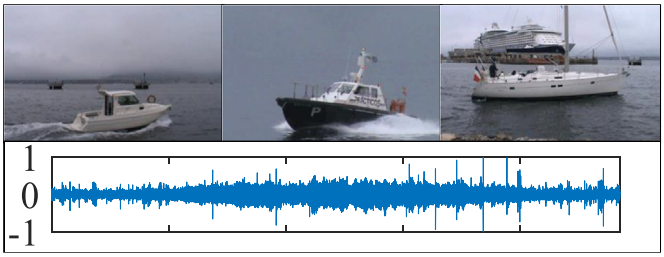
$$\gamma_{CCDFs} = \mathrm{CCA}(\gamma'_{even}, \gamma'_{odd}) \tag{23}$$

The feature dimension of γ_{CCDFs} is d_3. This step is to fuse the most discriminative information inter each subband. γ_{CCDFs} are used for ship recognition.

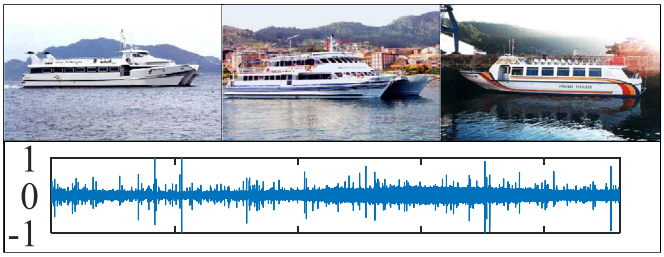
[1] The ability of distinguishing or identifying different classes of ships.
[2] Fisher discriminant matrix $\overline{D_B}$ is formulated in Appendix A.



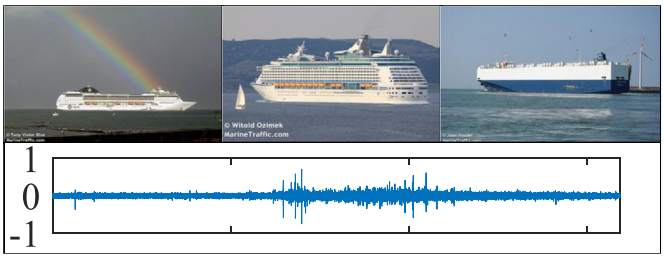
(a)



(b)



(c)



(d)

Fig. 3. Ship pictures and ship-radiated noise of each class. (a) class A. (b) class B. (c) class C. (d) class D.

3.5. Integrated optimization and classification

After the 2DFF, γ_{CCDFs} are sent to classifiers and multi-class classification errors ξ (10-fold cross-validation loss) are obtained. In this paper, k-NearestNeighbor (KNN) classifiers are used to classify γ_{CCDFs} of different classes.

During the process of the PCA, d_1 and d_2 are manually set. However, different values of d_1 and d_2 will lead to different values of d_3 and that eventually affects the classification accuracy of ship recognition. Besides this, hyperparameters of KNN also affect the classification accuracy. To figure out values selecting of d_1 and d_2, and hyperparameters setting of KNN, Bayesian Optimization [13] is used to integratedly optimize values selecting and hyperparameters setting for the best classification accuracy. Simply speaking, Bayesian Optimization aims to figure out the following problem:

$$\begin{cases} \min\ \xi \\ d_1 \in [1, 188] \\ d_2 \in [1, I \times d_1/2] \\ \text{Hyparameters of KNN} \end{cases} \tag{24}$$

The candidate hyperparameters of KNN are listed in Table 3.

After Bayesian Optimization, values of d_1 and d_2 and hyperparameters of KNN that lead to the minimum classification error will be selected.

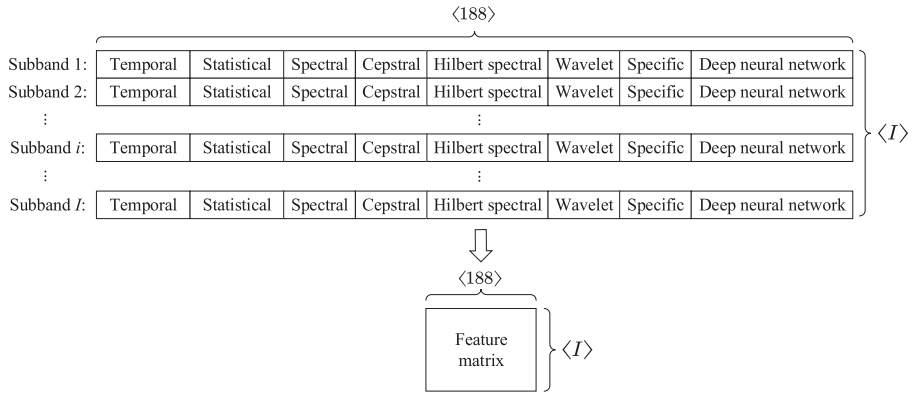


Fig. 4. Feature matrix X of a raw sample.

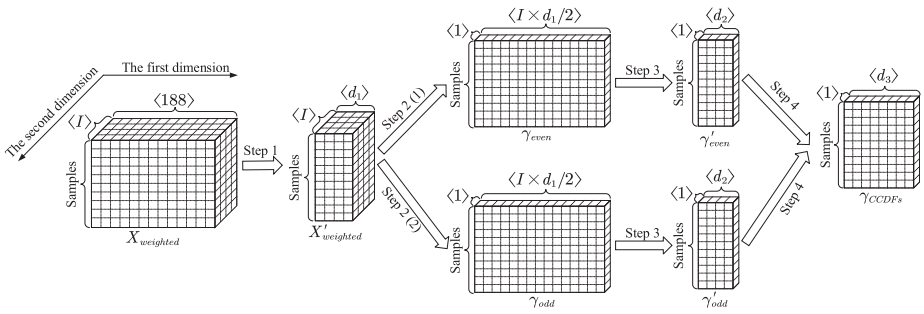


Fig. 5. Process of the 2DFF scheme.

Table 3
The candidate hyperparameters of KNN classifiers.

'Coding'	'onevsone', 'onevsall'
'Distance'	'cityblock', 'chebychev', 'correlation', 'cosine', 'euclidean', 'hamming', 'jaccard', 'mahalanobis', 'minkowski', 'seuclidean', 'spearman'
'DistanceWeight'	'equal', 'inverse', 'squaredinverse'

4. Experiments and discussions

4.1. Experiment setup

Each originally recorded signal in the database is framed using hamming window of length 2048 with 50% overlap. With sampling frequency $f_s = 52,734$ Hz, each sample lasts approximately 40 ms. Data are divided into two datasets, *dataset 1* and *dataset 2*, which are not intersected. There are 1000 samples for each class in dataset 1. Dataset 1 is used in the process of Bayesian Optimization to find the optimal values of d_1 and d_2, and the best hyperparameters of KNN. For generalization ability consideration, we only use dataset 2 to evaluate the effectiveness of the proposed 2DFF scheme.

In this paper, 7 levels of WP decomposition with fourth-order Symlet [22] wavelet is utilized to decompose each raw sample into 128 subbands. Note that fourth-order Symlet wavelet is used because it can avoid phase distortion while at the same time ensuring the orthogonality of the signal representations [6]. We depict the mean energy of each subband of each class in Fig. 6.

Fig. 6 shows that most of the energy ($> 95\%$) concentrates on 1–40 subbands, no matter which class. Therefore, in this paper, we only use 1–40 subbands. Besides this, if 1–40 subbands are used, the bandwidth analysed in this paper is:

$$B = B_{subband} \times 40 = \frac{(fs/2)}{128} \times 40 \approx 8000\ \text{Hz} \tag{25}$$

Bandwidth in Eq. (25) is equal to bandwidth analysed in [3], where the same ship-radiated noise database is adopted.

4.2. Generalization ability consideration

We extract feature matrices of each class from dataset 1 and we obtain: X^A, X^B, X^C, X^D and X^E. Then $\overline{D_B}$ of dataset 1 is calculated and we use it to weight X^A, X^B, X^C, X^D and X^E, respectively. The weighted feature matrices obtained are: $X^A_{weighted}, X^B_{weighted}, X^C_{weighted}$, $X^D_{weighted}$ and $X^E_{weighted}$. Bayesian Optimization is used to figure out Eq. (24). We set iterations of Bayesian Optimization to 1000 and the process of the Bayesian Optimization is showed in Fig. 7.

Fig. 7 shows that the minimum observed multi-class classification error is less than 0.01. Results of the Bayesian Optimization are listed in Table 4.

To evaluate the effectiveness of the Bayesian Optimization, we use hyperparameters of KNN listed in Table 4 and different values of d_1 and d_2 to conduct different classification experiments. The multi-class classification errors (mean of 10-time 10-fold cross-validation loss) are depicted in Fig. 8.

Fig. 8 shows that for a same value of d_1 (such as $d_1 = 1$), multi-class classification errors decrease with the increasing values of d_2. Fig. 8 also shows that when $d_1 = 2$, and $d_2 = 36, 37, 38, 39, 40$, multi-class classification errors achieve minimum (less than 0.01), which is corresponding to the results of Bayesian Optimization (the observed minimum classification error is 0.0094).

For generalization ability consideration, we assume that the optimal values of d_1 and d_2 obtained using Bayesian Optimization

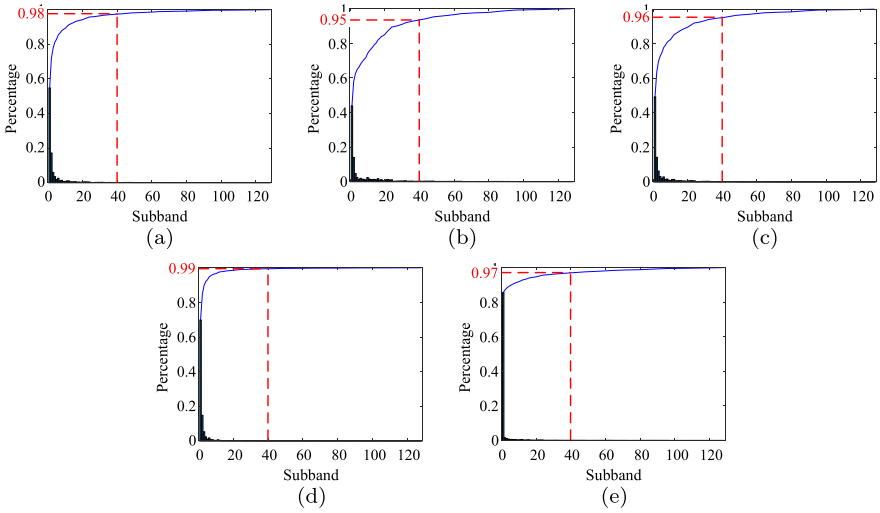


Fig. 6. The mean energy of each subband of each class. (a) class A. (b) class B. (c) class C. (d) class D. (e) class E.

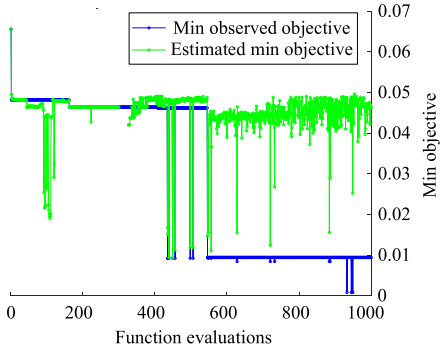


Fig. 7. The process of Bayesian Optimization.

Table 4
Results of the Bayesian Optimization.

Observed minimum classification error(10-fold cross-validation loss)	0.0094
d_1	2
d_2	37
'Coding'	'onevsone'
'Distance'	'cosine'
'DistanceWeight'	'inverse'

in dataset 1 can be transferred to dataset 2, that is, in dataset 2, when $d_1 = 2$, and $d_2 = 36, 37, 38, 39, 40$, recognition performance should be similar to that in dataset 1. In dataset 2, we conducted the same classification experiments with $d_1 = 2$, and $d_2 = 36, 37, 38, 39, 40$. In this experiment, we set the number of each class to 1000. The results are depicted in Fig. 9.

Fig. 9 shows that for different values of d_2, classification errors in dataset 2 are similar to that in dataset 1, which verifies the effec-

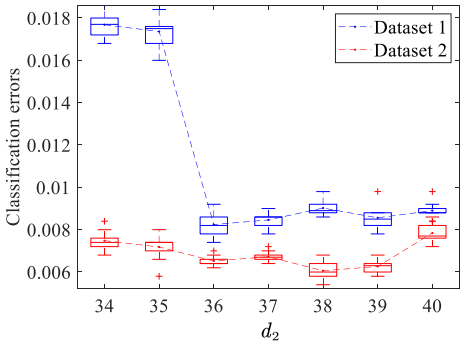


Fig. 9. Multi-class classification errors in dataset 1 and dataset 2.

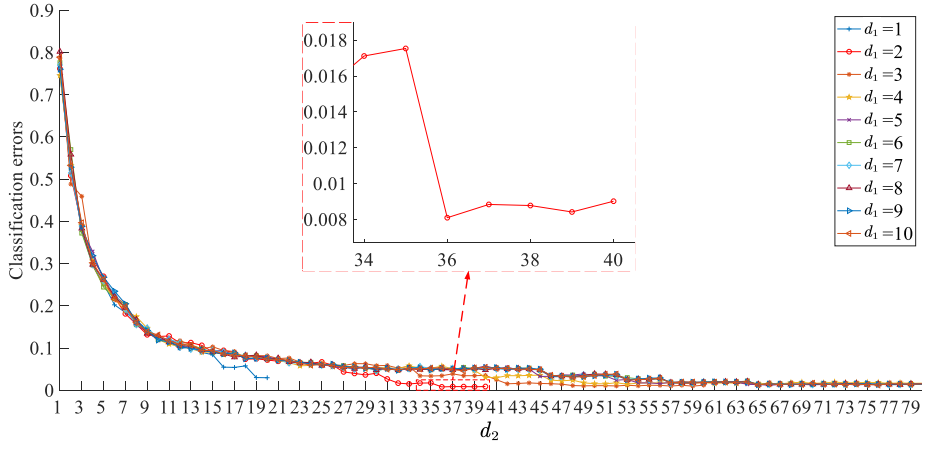


Fig. 8. Multi-class classification errors under different values of d_1 and d_2.

Table 5
Confusion matrix of unweighted feature matrices.

		Predicted Class				
		A	B	C	D	E
True Class	A	839 (83.9%)	61 (6.1%)	39 (3.9%)	61 (6.1%)	1 (0.1%)
	B	44 (4.4%)	842 (84.2%)	55 (5.5%)	58 (5.8%)	1 (0.1%)
	C	42 (4.2%)	62 (6.2%)	831 (83.1%)	65 (6.5%)	0 (0.0%)
	D	53 (5.3%)	89 (8.9%)	59 (5.9%)	799 (79.9%)	0 (0.00%)
	E	2 (0.2%)	2 (0.2%)	1 (0.1%)	1 (0.1%)	994 (99.4%)

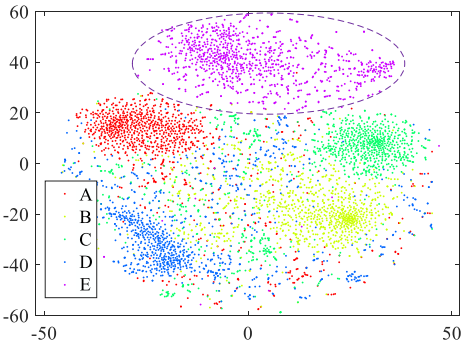


Fig. 10. Visualization results of CCDFs of unweighted feature matrices X^A, X^B, X^C, X^D and X^E.

tiveness of Bayesian Optimization. As the results of Bayesian Optimization (Table 4) have been proved to be effective in dataset 2, we will use the results to conduct classification experiments in the following ship recognition.

4.3. DAI strategy consideration

In this subsection, effectiveness of the DAI strategy proposed in subSection 3.3 is considered. The proposed DAI strategy adopts the following feature matrices: $X^A_{weighted}, X^B_{weighted}, X^C_{weighted}, X^D_{weighted}$ and $X^E_{weighted}$, called the *weighted feature matrices*.

The baseline adopts the following feature matrices: X^A, X^B, X^C, X^D and X^E, and all these feature matrices have *not* been weighted by Fisher discriminant matrix $\overline{D_B}$. We call X^A, X^B, X^C, X^D and X^E the *unweighted feature matrices*.

4.3.1. Unweighted feature matrices

The baseline adopts the following feature matrices: X^A, X^B, X^C, X^D and X^E. A confusion matrix of this case in showed in Table 5. T-Distribution Stochastic Neighbour Embedding (t-SNE) [23] is used to visualize CCDFs. The visualization technique t-SNE can map high-dimensional CCDFs into 2 dimensions and show the distributions of the high-dimensional CCDFs. The visualization results of CCDFs of this case is showed in Fig. 10.

Multi-class classification error of this case is 0.1390.

4.3.2. Weighted feature matrices

The DAI strategy adopts the following feature matrices: $X^A_{weighted}, X^B_{weighted}, X^C_{weighted}, X^D_{weighted}$ and $X^E_{weighted}$. All experimental conditions are the same as subSection 4.3.1 except the feature matrices have been weighted by Fisher discriminant matrix $\overline{D_B}$. A confusion matrix of this case is showed in Table 6. We also use t-SNE to visualize the CCDFs of this case and show the visualization results in Fig. 11.

Multi-class classification error of this case is 0.0066. Comparing Table 6 with Table 5, we can know that after being weighted by $\overline{D_B}$, feature matrices of each class can yield better discriminative ability. The multi-class classification accuracies are significantly improved after weighting feature matrices with $\overline{D_B}$. Comparing visualization results of Fig. 11 with that of Fig. 10, we can know that CCDFs become more centralized after weighting feature matrices with $\overline{D_B}$, in other words, CCDFs intra each class become more concentrated and CCDFs inter each class become more separated. After comparing Table 6 with Table 5, an interesting result can be observed: no matter using $\overline{D_B}$ to weight feature matrices or not, class E, the background noise is always truly predicted. The interesting result can also be observed from Fig. 10 that CCDFs of 4 classes (class A, class B, class C and class D) tend to mix together while CCDFs of class E are always self-contained.

The Fisher discriminant matrix $\overline{D_B}$ is showed in Fig. 12. Though different features of different subbands can reflect different characteristics of the ship-radiated noise, these features have different

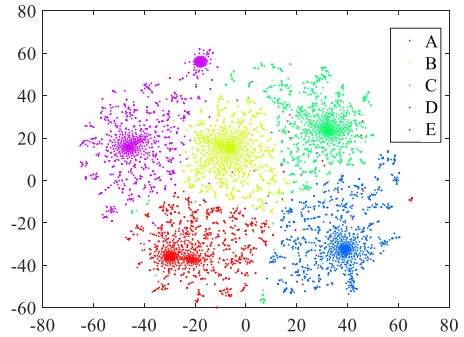


Fig. 11. Visualization results of CCDFs of weighted feature matrices $X^A_{weighted}$, $X^B_{weighted}, X^C_{weighted}, X^D_{weighted}$ and $X^E_{weighted}$.

Table 6
Confusion matrix of weighted feature matrices.

		Predicted Class				
		A	B	C	D	E
True Class	A	997 (99.7%)	0 (0.0%)	0 (0.0%)	3 (0.3%)	0 (0.0%)
	B	2 (0.2%)	989 (98.9%)	0 (0.0%)	4 (0.4%)	5 (0.5%)
	C	1 (0.1%)	3 (0.3%)	992 (99.2%)	2 (0.2%)	2 (0.2%)
	D	0 (0.0%)	0 (0.0%)	0 (0.0%)	998 (99.8%)	2 (0.2%)
	E	5 (0.5%)	2 (0.2%)	1 (0.1%)	1 (0.1%)	991 (99.1%)

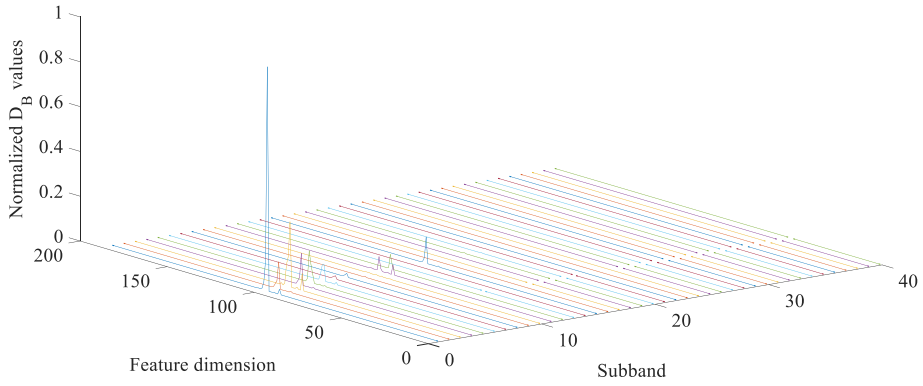


Fig. 12. Fisher discriminant matrix $\overline{D_B}$.

discriminative ability, just as shown in Fig. 12. Fig. 12 indicates that *energy spectrum of WP transform*, the 99th feature of subband 1, yields the most discriminative ability. The DAI strategy using $\overline{D_B}$ to weight feature matrices can *highlight* the features with better discriminative ability. Thus, during the process of the 2DFF, more attention will be paid to the features with better discriminative ability. More specially, the DAI strategy is similar to the "multiplicative attention" [24] mechanism which can improve the performance of natural language processing (NLP). For this, comparing Table 6 with Table 5, we can draw a conclusion that the DAI strategy can further improve the accuracy of ship recognition. Note that though feature matrices contain plenty of redundant information[3], the redundant information can be naturally removed during the process of 2DFF.

4.4. 2DFF scheme consideration

In this subsection, we are going to explain the effectiveness of the 2DFF scheme. We have assumed that the same kinds of features extracted from different subbands decomposed by WP decomposition can reflect different characteristics of the ship-radiated noise and by fusing of these features, recognition performance will be improved. The baseline of this assumption is that neither WP decomposition or feature fusion is applied, that is we directly extract feature vectors from the originally raw samples. The baseline [25] applies PCA to reduce feature dimension of the feature vectors.

4.4.1. Baseline

The baseline applies PCA to reduce feature dimension of the feature vectors extracted from the originally raw samples without WP decomposition. The flow chart of the baseline is depicted in Fig. 13.

4.4.2. Baseline without DAI strategy

In this subsection, for the baseline, we do not apply DAI strategy (DAI strategy for the baseline will be discussed in next subsection), that is we directly feed the *unweighted feature vectors* γ to PCA:

$$\gamma' = \mathrm{PCA}(\gamma) = \begin{pmatrix} x'_1 & x'_2 & \cdots & x'_j & \cdots & x'_{d'} \end{pmatrix}, \quad i = 1, 2, \cdots, d' \tag{26}$$

Then γ' of each class will be sent to KNN classifiers. For a fair comparison, we also utilized Bayesian Optimization to obtain the optimal value of d' and hyperparameters of the KNN classifiers. The results of the optimization of the baseline are listed in Table 7.

[3] Just as shown in Fig. 12, most of the features in feature matrices only yield very weak discriminative ability. These features can be viewed as redundant information.

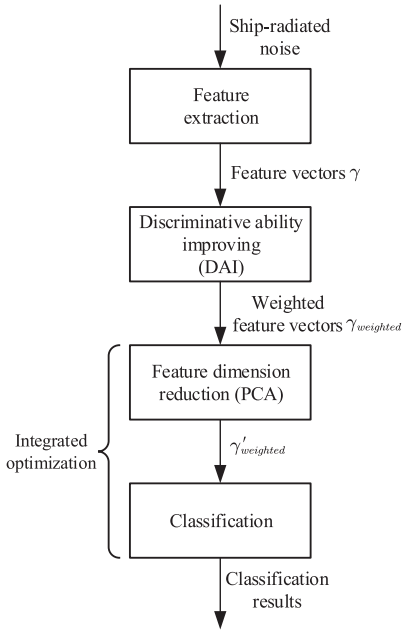


Fig. 13. The flow chart of the baseline.

Table 7
Results of Bayesian Optimization of the baseline.

Observed minimum classification error (10-fold cross-validation loss)	0.2044
$d\prime$	23
'Coding'	'onevsone'
'Distance'	'mahalanobis'
'DistanceWeight'	'equal'

Using the optimal value of d' and hyperparameters of KNN listed in Table 7, we conduct the classification experiments and experimental results are showed in Table 8.

Multi-class classification error of this case is 0.2102.

4.4.3. Baseline with DAI strategy

The DAI strategy in subSection 3.3 is also suitable for the baseline, that is we can also calculate a so called *Fisher discriminant vector* $\bar{\gamma}$.[4] The Fisher discriminant vector $\bar{\gamma}$ is plotted in Fig. 14.

[4] Fisher discriminant vector $\bar{\gamma}$ is formulated in Appendix A.

Table 8
Confusion matrix of the baseline without DAI strategy.

		Predicted Class				
		A	B	C	D	E
True Class	A	715 (71.5%)	34 (3.4%)	47 (4.7%)	126 (12.6%)	78 (7.8%)
	B	39 (3.9%)	791 (79.1%)	35 (3.5%)	66 (6.6%)	69 (6.9%)
	C	70 (7.0%)	81 (8.1%)	717 (71.7%)	75 (7.5%)	57 (5.7%)
	D	75 (7.5%)	38 (3.8%)	37 (3.7%)	789 (78.9%)	61 (6.1%)
	E	13 (1.3%)	13 (1.3%)	18 (1.8%)	19 (1.9%)	937 (93.7%)

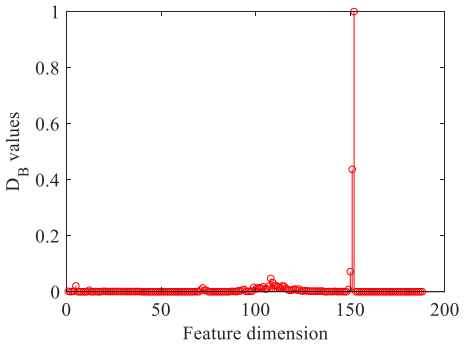


Fig. 14. Fisher vector $\bar{\gamma}$.

Fig. 14 indicates that for the originally raw samples (without WP decomposition), the feature with the most discriminative ability is the 152th feature of γ, the *frequency band power*, while applying WP decomposition, the feature with the most discriminative ability is the energy spectrum of WP transform of subband 1 (subSection 4.3.2).

We use $\bar{\gamma}$ to weight feature vectors of class A, and we obtain the *weighted feature vectors*:

$$\gamma^{A}_{weighted} = \gamma^{A} \cdot \bar{\gamma} \tag{27}$$

Similarly, the weighting operations to other classes are:

$$\gamma^{B}_{weighted} = \gamma^{B} \cdot \bar{\gamma} \tag{28}$$

$$\gamma^{C}_{weighted} = \gamma^{C} \cdot \bar{\gamma} \tag{29}$$

$$\gamma^{D}_{weighted} = \gamma^{D} \cdot \bar{\gamma} \tag{30}$$

$$\gamma^{E}_{weighted} = \gamma^{E} \cdot \bar{\gamma} \tag{31}$$

Then PCA is used to reduce the feature dimension of $\gamma_{weighted}$ of each class:

$$\begin{aligned} &\gamma'_{weighted} = \mathrm{PCA}\left(\gamma_{weighted}\right) = \left(x'_1 \quad x'_2 \quad \cdots \quad x'_j \quad \cdots \quad x'_{d'}\right), \\ &\qquad i = 1, 2, \cdots, \\ &d'\gamma_{weighted} \in \left(\gamma^{A}_{weighted}, \gamma^{B}_{weighted}, \gamma^{C}_{weighted}, \gamma^{D}_{weighted}, \gamma^{E}_{weighted}\right) \end{aligned} \tag{32}$$

Then $\gamma'_{weighted}$ of each class will be sent to KNN classifiers. The confusion matrix of this case is depicted in Table 9. In this case, multi-class classification error is 0.0850. Comparing Table 9 with Table 8, we can know that class E, the background noise always tend class E does not contain any ship-radiated acoustic characteristicsto be truly predicted with or without DAI strategy, in other words, even without DAI strategy, class E still can be truly predicted because Class E does not contain any ship-radiated acoustic characteristics (Table 6 and Table 5 in Section 4.3 have shown a similar result). These experimental results are similar to that in [3] where Santos et al. have demonstrated a experimental result that class E is always truly predicted. As the results produced by the 2DFF scheme and DAI strategy are similar to that in [3], effectiveness of the 2DFF scheme and DAI strategy is further verified.

4.4.4. Comparison of 2DFF scheme and baseline

This subsection comprehensively compares the 2DFF scheme and the baseline. 10-time 10-fold cross validation loss of corresponding experiments are listed in Table 10.

Table 10 shows that by combining the 2DFF scheme with DAI strategy, we can achieve the minimum mean multi-class classification error of 0.0067. The 2DFF scheme decomposes the originally raw samples into multiple subbands, because we want to analyse the originally raw samples with a "high-resolution" view. We assume that the same kinds of features but extracted from different resolutions (subbands) can reflect different characteristics of the ship-radiated noise. Then, using PCA and CCA, we combine and fuse the features of different resolutions (subbands) into the CCDFs. The operation of firstly "splitting into different resolutions" and then "combining these resolutions" can achieve a better ship recognition performance because the operation can fully make use of the "*potential characteristics*" of different subbands, which are invisible in the originally raw samples. The PCA and CCA naturally have the ability of keeping the effective discriminant information of multi-feature and also eliminating the redundant information, but with the aid of the DAI strategy, the ability and ship recognition performance can be further improved because the DAI strategy can additionally *highlight* the effective discriminant information of the feature matrices.

Table 9
Confusion matrix of the baseline with DAI strategy.

		Predicted Class				
		A	B	C	D	E
True Class	A	883 (83.3%)	10 (1.0%)	20 (2.0%)	49 (4.9%)	38 (3.8%)
	B	36 (3.6%)	923 (92.3%)	14 (1.4%)	13 (1.3%)	14 (1.4%)
	C	29 (2.9%)	9 (0.9%)	928 (92.8%)	18 (1.8%)	16 (1.6%)
	D	50 (5.0%)	15 (1.5%)	13 (1.3%)	884 (88.4%)	38 (3.8%)
	E	12 (1.2%)	5 (0.5%)	7 (0.7%)	19 (1.9%)	957 (95.7%)

Table 10
Comparison of 2DFF scheme and baseline.

Times	Baseline without DAI	Baseline with DAI	2DFF without DAI	2DFF with DAI
1	0.2096	0.0836	0.1370	0.0066
2	0.2124	0.0824	0.1368	0.0068
3	0.2084	0.0808	0.1374	0.0068
4	0.2100	0.0844	0.1366	0.0066
5	0.2100	0.0862	0.1374	0.0068
6	0.2126	0.0864	0.1366	0.0066
7	0.2096	0.0864	0.1382	0.0078
8	0.2124	0.0836	0.1392	0.0078
9	0.2108	0.0836	0.1388	0.0068
10	0.2100	0.0854	0.1394	0.0060
Mean	0.2102	0.0843	0.1377	*0.0067*
Variance	3.9822×10^{-6}	3.4418×10^{-6}	1.1649×10^{-6}	1.96×10^{-7}

5. Conclusion

Just like other pattern recognition problems using feature fusion methods to improve recognition performance, we also introduce feature fusion methods to ship-radiated noise recognition. We assume that the same kinds of features extracted from different subbands decomposed by WP decomposition can reflect different characteristics of the ship-radiated noise. Then, on the basis of the PCA and CCA, we propose the 2DFF scheme to fuse these features. Experimental results show that, compared with the baseline without WP decomposition, the 2DFF scheme can significantly improve the performance of ship-radiated noise recognition because the scheme can fully make use of different characteristics of different subbands of the ship-radiated noise. Last but not least, by comprehensively considering the discriminative ability of each single feature in the feature matrices, the proposed DAI strategy can further improve the accuracy of ship-radiated noise recognition.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant nos: 61871336, 61571377, and 61771412) and the Fundamental Research Funds for the Central Universities (Grant No.: 20720180068).

Appendix A

1) Maximum energy.
Maximum energy describes the maximum amplitude of the raw sample x:

$$\text{Maximum energy} = \max\left(|x_i|^2\right), \quad i = 1, 2, \cdots, N \tag{33}$$

where x_i is the ith observed value in x, N is the length of x.
2) Total energy.
Total energy is the summation of amplitude of the raw sample x:

$$\text{Total energy} = \text{sum}\left(|x_i|^2\right), \quad i = 1, 2, \cdots, N \tag{34}$$

3) Normalized energy spectrum of Tunable Q-factor Wavelet Transform.
The raw sample x is firstly decomposed by Tunable Q-factor Wavelet Transform (TQWT) [26] into multiple subbands. Then coefficients of the multiple subbands are normalized to be the normalized energy spectrum. Parameters of the TQWT are Q-factor Q, redundancy r, and level J. In this paper, Q is set to 4, r is set to 3, and J is set to 30, thus the length of the normalized energy spectrum will be 31. We randomly pick up two samples in class A and in class B and show their normalized energy spectrum in Fig. 15.
4) DBN hidden unit features.
We utilize DBN [27], a deep neural network to extract the DBN hidden unit features. We train a DBN with 4 layers of size 1536-512-128-32 to code the spectrum and reconstruct the spectrum of the raw samples. The top hidden unit values of the DBN are treated as the DBN hidden unit features [10] for ship recognition. Fig. 16 gives original spectrum of two random samples, their reconstructed spectrum, and their DBN hidden unit features.
5) Fisher discriminant matrix $\overline{D_B}$
Fisher discriminant criterion [28] is used to calculate D_B values [6] of each single feature between feature matrices of class A and class B:

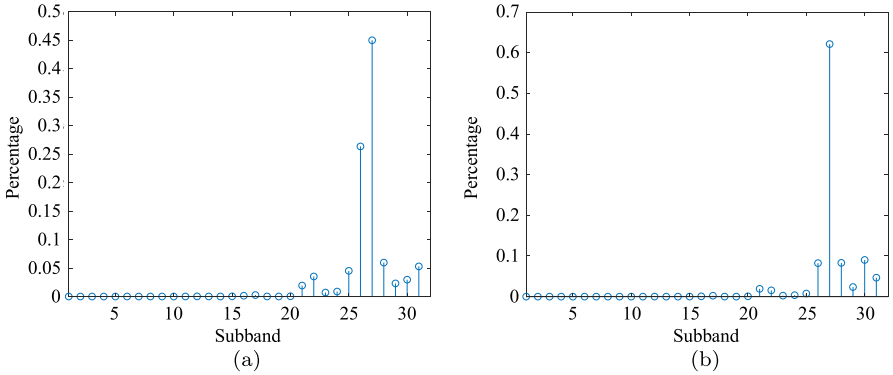


Fig. 15. Normalized energy spectrum of the TQWT of two random samples. (a) A sample in class A. (b) A sample in class B.

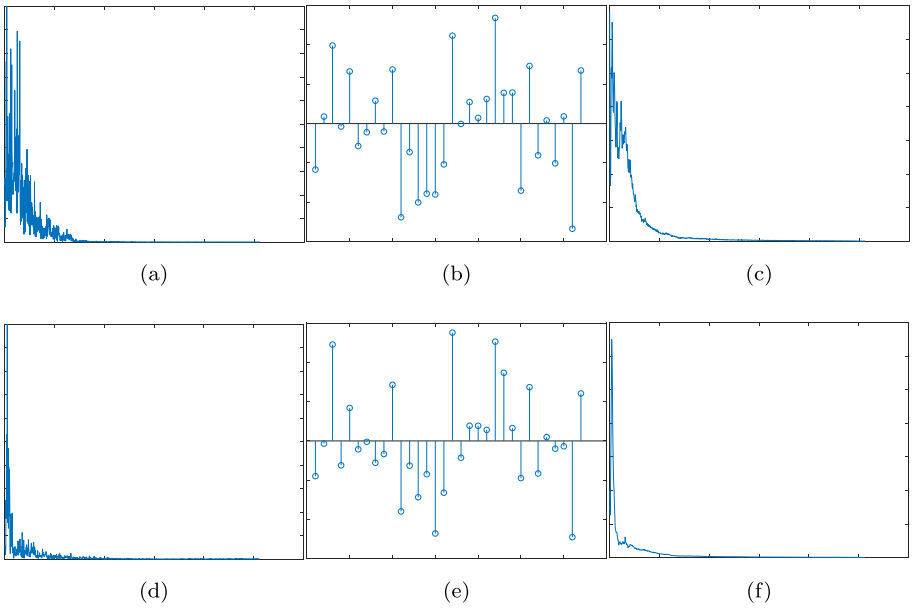


Fig. 16. DBN hidden unit features of two random samples. (a) Original spectrum of sample 1. (b) DBN hidden unit features of sample 1. (c) Reconstructed spectrum of sample 1. (d) Original spectrum of sample 2. (e) DBN hidden unit features of sample 2. (f) Reconstructed spectrum of sample 2.

$$D_B\left(X^A,X^B\right)=\begin{pmatrix}\frac{|\mu_{1,1}^A-\mu_{1,1}^B|^2}{\left(\sigma_{1,1}^A\right)^2+\left(\sigma_{1,1}^B\right)^2} & \cdots & \frac{|\mu_{1,188}^A-\mu_{1,188}^B|^2}{\left(\sigma_{1,188}^A\right)^2+\left(\sigma_{1,188}^B\right)^2}\\ \vdots & \frac{|\mu_{i,j}^A-\mu_{i,j}^B|^2}{\left(\sigma_{i,j}^A\right)^2+\left(\sigma_{i,j}^B\right)^2} & \vdots\\ \frac{|\mu_{I,1}^A-\mu_{I,1}^B|^2}{\left(\sigma_{I,1}^A\right)^2+\left(\sigma_{I,1}^B\right)^2} & \cdots & \frac{|\mu_{I,188}^A-\mu_{I,188}^B|^2}{\left(\sigma_{I,188}^A\right)^2+\left(\sigma_{I,188}^B\right)^2}\end{pmatrix} \tag{35}$$

where X^A denotes the feature matrices of class A, $\mu_{i,j}^A$ denotes the mean of the jth feature of the ith subband of X^A, $\left(\sigma_{i,j}^A\right)^2$ denotes the variance of the jth feature of the ith subband of X^A, and so dose class B. D_B values are calculated for each candidate feature in the feature matrices X to evaluate the discriminative ability. Those features for which D_B values are relatively larger than others will have a better discriminative ability. More specially, the feature $x_{i,j}$ with a larger value $\frac{|u_{i,j}^A-u_{i,j}^B|^2}{\left(\sigma_{i,j}^A\right)^2+\left(\sigma_{i,j}^B\right)^2}$ in $D_B\left(X^A,X^B\right)$ will have a better discriminative ability between class A and class B.

Similarly, D_B values are calculated for other any two combinations of class A, class B, class C, class D and class E, which are $D_B\left(X^A,X^C\right)$, $D_B\left(X^A,X^D\right)$, $D_B\left(X^A,X^E\right)$, $D_B\left(X^B,X^C\right)$, $D_B\left(X^B,X^D\right)$, $D_B\left(X^B,X^E\right)$, $D_B\left(X^C,X^D\right)$, $D_B\left(X^C,X^E\right)$ and $D_B\left(X^D,X^E\right)$. The Fisher discriminant matrix $\overline{D_B}$ is the average of these 10 values:

$$\begin{aligned}\overline{D_B}&=\left[D_B\left(X^A,X^B\right)+D_B\left(X^A,X^C\right)+\cdots+D_B\left(X^D,X^E\right)\right]/10\\&=\begin{pmatrix}\overline{D_B(1,1)} & \cdots & \overline{D_B(1,188)}\\ \vdots & \overline{D_B(i,j)} & \vdots\\ \overline{D_B(I,1)} & \cdots & \overline{D_B(I,1)}\end{pmatrix}\end{aligned} \tag{36}$$

6) Fisher discriminant vector $\bar{\gamma}$

$\bar{\gamma}$ is explained as follows:

$$\bar{\gamma}=\left[\gamma^{A,B}+\gamma^{A,C}+\cdots+\gamma^{A,D}\right]/10 \tag{37}$$

$$\gamma^{A,B}=\left(\frac{|u_1^A-u_1^B|^2}{\left(\sigma_1^A\right)^2+\left(\sigma_1^B\right)^2}\quad\cdots\quad\frac{|u_j^A-u_j^B|^2}{\left(\sigma_j^A\right)^2+\left(\sigma_j^B\right)^2}\quad\cdots\quad\frac{|u_{188}^A-u_{188}^B|^2}{\left(\sigma_{188}^A\right)^2+\left(\sigma_{188}^B\right)^2}\right),\quad j=1,2,\cdots,188 \tag{38}$$

where γ^A denotes the feature vectors of class A, μ_j^A denotes the mean of the jth feature in γ^A, $\left(\sigma_j^A\right)^2$ denotes the variance of the jth feature in γ^A, and so does class B, class C, class D, and class E.

References

[1] Jian L, Yang H, Zhong L, Ying X. Underwater target recognition based on line spectrum and support vector machine. In: International conference on mechatronics, control and electronic engineering (MCE2014). Atlantis Press; 2014. p. 79–84.
[2] Wei X. On feature extraction of ship radiated noise using 11/2 d spectrum and principal components analysis. In: Signal processing, communications and computing (ICSPCC), 2016 IEEE international conference on. IEEE; 2016. p. 1–4.
[3] Santos-Domínguez D, Torres-Guijarro S, Cardenal-López A, Pena-Gimenez A. Shipsear: an underwater vessel noise database. Appl Acoust 2016;113:64–9.
[4] Zhang L, Wu D, Han X, Zhu Z. Feature extraction of underwater target signal using mel frequency cepstrum coefficients based on acoustic vector sensor. J Sens 2016.
[5] Meng Q, Yang S. A wave structure based method for recognition of marine acoustic target signals. J Acoust Soc Am 2015;137(4):2242.
[6] Azimi-Sadjadi MR, Yao D, Huang Q, Dobeck GJ. Underwater target classification using wavelet packets and neural networks. IEEE Trans Neural Netw 2000;11 (3):784–94.
[7] Averbuch A, Zheludev V, Neittaanmäki P, Wartiainen P, Huoman K, Janson K. Acoustic detection and classification of river boats. Appl Acoust 2011;72 (1):22–34.
[8] Wang S, Zeng X. Robust underwater noise targets classification using auditory inspired time–frequency analysis. Appl Acoust 2014;78:68–76.
[9] Cao X, Zhang X, Yu Y, Niu L. Deep learning-based recognition of underwater target. In: Digital signal processing (DSP), 2016 IEEE international conference on. IEEE; 2016. p. 89–93.
[10] Yang H, Shen S, Yao X, Sheng M, Wang C. Competitive deep-belief networks for underwater acoustic target recognition. Sensors 2018;18(4):952.
[11] Sun Q-S, Zeng S-G, Liu Y, Heng P-A, Xia D-S. A new method of feature fusion and its application in image recognition. Pattern Recogn 2005;38 (12):2437–48.
[12] Haghighat M, Abdel-Mottaleb M, Alhalabi W. Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition. IEEE Trans Inf Forensics Secur 2016;11(9):1984–96.

[13] Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N. Taking the human out of the loop: a review of bayesian optimization. Proc IEEE 2016;104(1):148–75.
[14] Wulsin D, Gupta J, Mani R, Blanco J, Litt B. Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement. J Neural Eng 2011;8(3):036015.
[15] Giannakopoulos T, Pikrakis A. Introduction to audio analysis: a MATLAB approach. Academic Press; 2014.
[16] Park TH, Li Z, Biguenet J. Not just more fms: taking it to the next level. ICMC; 2008.
[17] Tuma M, Igel C, Prior M. Hydroacoustic signal classification using support vector machines. In: Signal and image processing for remote sensing. CRC Press; 2012. p. 37–56.
[18] Herre J, Allamanche E, Hellmuth O. Robust matching of audio signals using spectral flatness features. In: Applications of signal processing to audio and acoustics, 2001 IEEE workshop on the. IEEE; 2001. p. 127–30.
[19] Tuma M, Rørbech V, Prior MK, Igel C. Integrated optimization of long-range underwater signal detection, feature extraction, and classification for nuclear treaty monitoring. IEEE Trans Geosci Remote Sens 2016;54(6):3649–59.
[20] Yang J, Yang J-Y, Zhang D, Lu J-F. Feature fusion: parallel strategy vs. serial strategy. Pattern Recogn 2003;36(6):1369–81.
[21] Abdi H, Williams LJ. Principal component analysis. Wiley Interdiscip Rev 2010;2(4):433–59.
[22] Vetterli M, Kovačević J. Wavelets and subband coding, vol. 995. Prentice Hall Englewood Cliffs; 1995.
[23] Maaten Lvd, Hinton G. Visualizing data using t-sne. J Mach Learn Res 2008:2579–605.
[24] Luong M-T, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. arXiv:1508.04025.
[25] Wei X, Gang-Hu L, Wang Z. Underwater target recognition based on wavelet packet and principal component analysis. Comput Simul 2011;28:8–290.
[26] Selesnick I. Tqwt toolbox guide, electrical and computer engineering. Polytechnic Institute of New York University. Available online at: http://eeweb.poly.edu/iselesni/TQWT/index.html.
[27] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504–7.
[28] Pattern DRHP. Classification and scene analysis. New York: John Wiley and Sons; 1973.