24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

ICSV24
23-27JULY 2017
LONDON CALLING

NOISE ROBUST FEATURE EXTRACTION FOR UNDERWATER TARGETS CLASSIFICATION VIA LABEL CONSISTENT DICTIONARY LEARNING

Lu Wang, Lin Chen and Xiangyang Zeng

Northwestern Polytechnical University, School of Marine Science and Technology, Xi'an, China

Email: wanglu@nwpu.edu.cn

Underwater targets classification using traditional features suffers significant performance loss under the mismatched noisy condition when the training data and the testing data are measured at totally different noise levels. We consider to denoise the measured signals by the sparse representation. However, the real acoustic signals for underwater targets are too complicated to be represented by a fixed simple dictionary, we suggest to learn the representative dictionary from the data. To simultaneously learn the dictionary and discriminative sparse features, the underlying structure that sparse patterns are similar for sample data from the same target and differ a lot among different targets is exploited by using the label consistent dictionary learning method. The resulted sparse coefficients are then used as the input features of the Support Vector Machine (SVM) for target classification. Experimental results using real underwater acoustic dataset demonstrate that the proposed noise robust feature outperforms the traditional ones under the mismatched noisy condition.

Keywords: Underwater targets classification; noise robust feature; dictionary learning; sparse representation.

1. Introduction

Underwater target classification has received much attention recently in underwater acoustic signal processing community. Research in this field focuses on the underwater target feature extraction and classifier optimization. Traditional features for underwater targets are extracted from the signal spectrum, such as the zero-crossing rate, the frequency bandwidth, the spectral centroid, line spectrum and the energy of the spectrum in the low frequency range [1, 2]. Those features always represent certain targets physical information, for example, the number of the blades, the rotating velocity of the propellers, and so on, hence are typical and useful for underwater target classification. However, the spectrum based features are vulnerable to noise and the time variant channel distortion. Another prevailing feature for acoustic signal stemming from the automatic speech recognizers is the mel-frequency cepstral coefficients (MFCC)[3, 4, 5, 6]. It is found that MFCC outperforms the above mentioned physical-related features [7, 8] in classification due to its powerful imitation of the mask effect and the frequency responses of human ears by a set of auditory based Mel filters. What's more, since the discrete cosine transformation is applied on the cepstrum in MFCCs, it has the advantage of high frequency channel distortion suppression. However, a high level background noise would severely depress the discriminating ability of both the traditional physical-related features and the

MFCCs. It is probably the case when the training and testing data are measured at totally different noise levels, since the underwater environment is very complicated and time varying. Under such a background noise mismatched condition, the correct rate of classification would drop dramatically.

Sparse representation (SR) finds a parsimonious representation of the signal as a linear combination of a few atoms from an over-complete dictionary, which has an advantage of effective signal denoising. The sparse coefficients extracted by SR are representative and noise robust features which outperform the traditional ones and the MFCC in noisy environments[9]. However, typical real acoustic signals for underwater targets are always too complicated to be represented parsimoniously by a fixed simple dictionary, we consider to learn the representative dictionary and the corresponding sparse features of the targets from the data. Furthermore, it is observed that sample data share similar sparse supports for the same target and exhibit totally different ones for different targets. Such an observation can be properly modeled by block sparse structure with each nonzero block being representative to one target. Those block structures when applied are capable to reduce the within-class covariance and promote the indiscriminateness of sparse features. Since multiple measurements with labelled samples are available, the discriminative sparse features and the representative dictionary can be simultaneously learned by a dictionary learning method, named the Label consistent K-SVD (LC-KSVD) [10]. The low-frequency band of the spectra of the real data can be used as the inputs of the LC-KSVD to extract the sparse features with block sparsity. Experimental results using real underwater target dataset demonstrate that the proposed noise robust feature outperforms the traditional ones under the mismatched noisy condition.

2. Sparse Feature and Dictionary Learning by LC-KSVD

The spectra of most underwater targets have energy concentrated in the low-frequency range, which are typically composed by both continuous and line spectra reflecting important target information. However, they can be easily contaminated by noise. We consider instead to extract the noise robust sparse features. Unfortunately, as illustrated by the spectrograms of three different underwater targets in Fig. 1, not every target has a sparse spectrum. We have to learn a representative dictionary from the data itself in order to use the sparse representation. What's more, note from Fig. 1 that each target has its unique spectrum range and the spectra of samples measured from the same target show a great similarity. This fact motivates us to learn the sparse features and the representative dictionary simultaneously by LC-KSVD method, since the a block sparse structure is encouraged in LC-KSVD to model and describe the targets' specifications where sparse representations of samples from the same target share similar supports. As we will shown later in the experiments part, such underlying structure when applied can

Note that if the spectra are directly used as the inputs to the LC-KSVD, an over-completed dictionary with huge dimensions will be learned, which will be computational complicated and time-consuming. We suggest to apply the Principal Component Analysis (PCA) [11] to reduce spectra to a smaller dimension before being used to extract the sparse feature by the LC-KSVD. In this way, a dictionary with reasonable size will be learned efficiently. When the PCA is applied, the degree of cumulative contribution against the dimension of the spectra can be calculated which is given in Fig. 2. As seen from Fig. 2, the first 140-dimension of the principal components represents 90.05% of the feature energy. Therefore, the first 100-dimension of the principal components are preserved as the inputs to LC-KSVD for dictionary learning.

Let $\boldsymbol{Y} = \left[\boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_N}\right]$ be a set of preserved principal components after PCA. According to LC-KSVD, an objective function is constructed as follows:

$$\begin{aligned} \langle \boldsymbol{D}, \boldsymbol{W}, \boldsymbol{A}, \boldsymbol{X} \rangle = \arg \min_{\boldsymbol{D}, \boldsymbol{W}, \boldsymbol{A}, \boldsymbol{X}} \left\| \boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X} \right\|_2^2 \\ + \alpha \left\| \boldsymbol{Q} - \boldsymbol{A}\boldsymbol{X} \right\|_2^2 + \beta \left\| \boldsymbol{H} - \boldsymbol{W}\boldsymbol{X} \right\|_2^2 \quad s.t. \forall i, \left\| \boldsymbol{x_i} \right\|_0 \leq \boldsymbol{T} \end{aligned} \tag{1}$$

where $\boldsymbol{D}$ and $\boldsymbol{X}$ are the representative dictionary and the corresponding sparse feature, respectively;

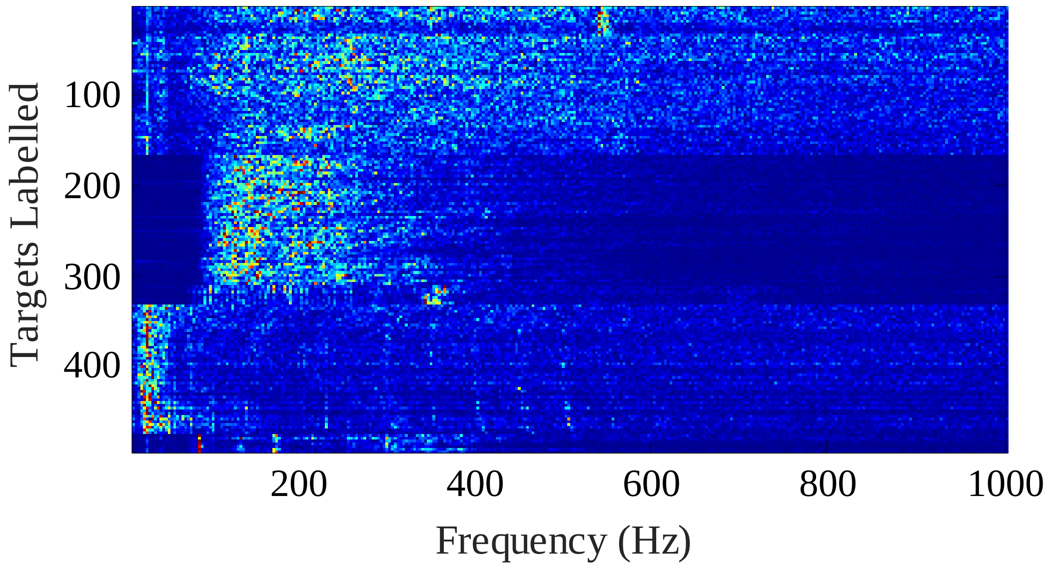

Figure 1: The spectrograms of 3 kinds of underwater targets. Samples indexed from 1 to 180 belong to first class; Samples indexed from 181 to 320 are measured from class 2 and the rest comes from class 3.

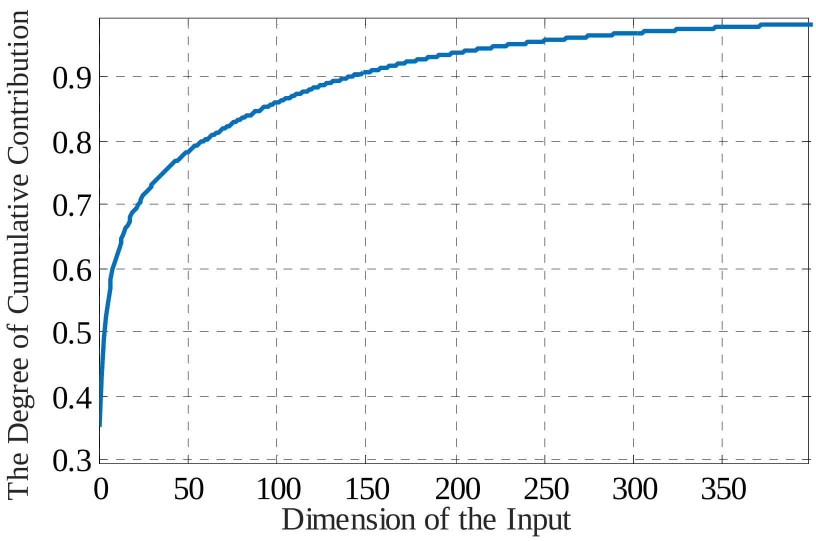

Figure 2: The degree of cumulative contribution against the dimension of the spectra

$\|\boldsymbol{Y}-\boldsymbol{DX}\|_2^2$ calculates the data fitting error; Block modeling error of the sparse feature is given by $\|\boldsymbol{Q}-\boldsymbol{AX}\|_2^2$ with $\boldsymbol{Q}=[\boldsymbol{q_1}\cdots\boldsymbol{q_N}]$, $\boldsymbol{q_i}=[q_i^1\cdots q_i^k]^T=[0\dots1,1\dots0]^T\in R^k$ and $\boldsymbol{A}$ being a linear transformation matrix; label consistent regularization is give by $\|\boldsymbol{H}-\boldsymbol{WX}\|_2^2$ with $\boldsymbol{H}=[\boldsymbol{h_1}\cdots\boldsymbol{h_N}]$ the class labels of the input signals; and α and β control the relative contributions of the corresponding terms. The sparsity is imposed by the term $\|\boldsymbol{x_i}\|_0\leq\boldsymbol{T}$, where $\|.\|_0$ is the l_0-norm and T is the maximum sparsity allowed (corresponding to the number of the nonzeros in each $\boldsymbol{x_i}$). Further more, the block sparsity is imposed by the term $\|\boldsymbol{Q}-\boldsymbol{AX}\|_2^2$, which restricts the sparse features to be discriminative by setting elements in $\boldsymbol{q_i}$ to be one in a block-wise manner. By minimizing the above objective function in Eq.(1), the learned sparse representation dictionary $\boldsymbol{D}$ and the sparse coefficient matrix $\boldsymbol{X}$ with block structure can be learned iteratively. The detailed learning procedures can be found in [10].

3. Numerical Experiments

Real data experiments with 3 different underwater moving targets are conducted to test the proposed sparse features with block structure. For each class, 600 clips each with a duration of 1s are collected. The dataset is split into six folds for cross validation (CV). Five folds are used for training (1,500 clips) and one for testing (300 clips). We add additional white Gaussian noise into the testing data to mimic the noisy mismatched conditions with the signal to noise ratio (SNR) of the testing data from -5 dB to 15 dB, where we treat that the original data set is noiseless. We calculate the spectrum for each clip of the data. And PCA is applied on the spectra in low frequency range from 10 Hz to 1 KHz to reduce the feature to a relatively low dimension of 140. The preserved principal components are used as the final input of the LC-KSVD to learn the representative dictionary. To predict the label of the testing data, the learned dictionary is used to extract the sparse features firstly and then the support vector machine (SVM)[12] is applied for classification. The traditional features of MFCC and the spectra in low frequency range from 10 Hz to 1 KHz are also tested as a comparison. MFCCs are extracted form the spectra with 256 ms Hamming window and then passed by a voicebox of 26 Mel-filters. The first 13 dimensions of the filter are preserved as the final MFCC feature, since it is found to yield the best classification results. The allowed maximum sparsity in LC-KSVD is set to be 12 in our experiments, and the regulation parameters α and β are set empirically as 1.

The block structure of the extracted sparse feature is indicated in Fig. 3, where most of the nonzero coefficients of the sparse features for different targets distributed in different blocks and features of the same target always share similar support pattern both for training and testing data. Fig. 4 shows the final correct recognition rates using different features against the levels of the SNRs of the testing data. As the SNR increases, all the correct recognition rates increase. Among which, our proposed features yield highest the correct recognition rate especially when the SNR is from -3 to 9 dB.

4. CONCLUSION

In this paper, we propose to use the LC-KSVD method to learn the discriminative sparse features and the representative dictionary for underwater targets classification under mismatched noisy conditions. PCA is firstly applied to the targets spectra in low frequency band to reduce the dimensionality. And the preserves principle components are used as the rough feature to learn the dictionary and the block sparse discriminative features. We demonstrate by numerical experiments with real underwater dataset that the sparse features with block structure learned by LC-KSVD are effective for underwater target classification under the mismatched noisy conditions.

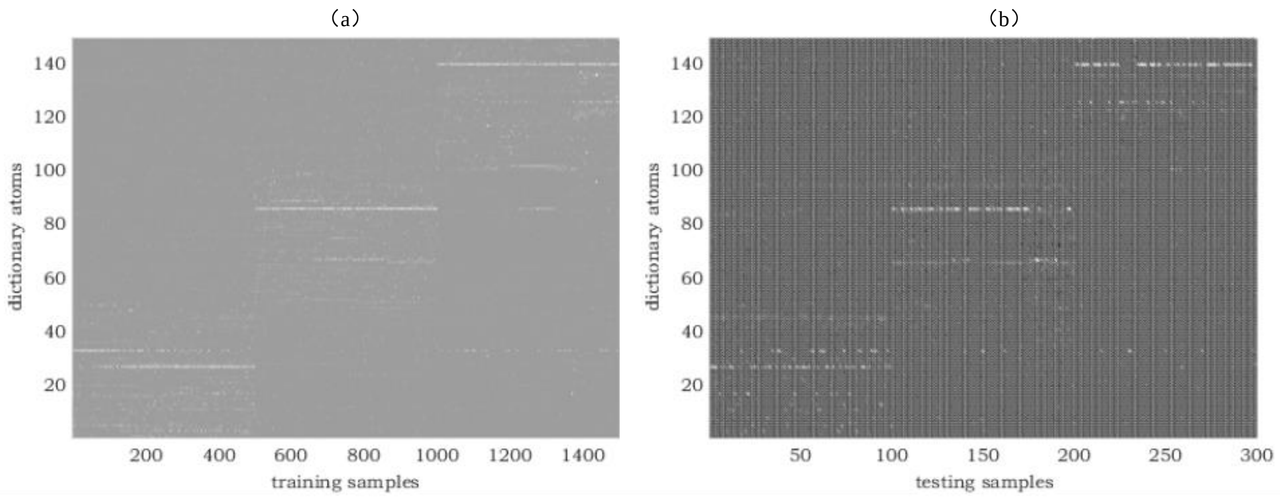

Figure 3: Illustration of the sparse codes obtained by LC-KSVD algorithm. (a) training signals; (b) testing signals. For training data in (a), samples indexed from 1 to 500 belong to first class; samples indexed from 500 to 1000 are measured from class 2 and the rest comes from class 3. For testing data in (b), samples indexed from 1 to 100 belong to first class; samples indexed from 100 to 200 are measured from class 2 and the rest comes from class 3.

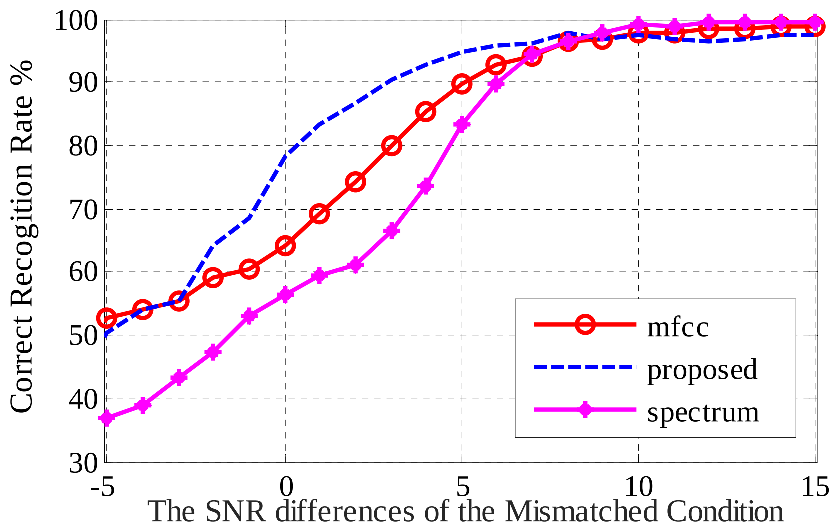

Figure 4: Classification rate using three different features against the levels of SNR differences between the training and testing data.

Acknowledgment

This work was supported by National Natural Science Foundation of China under Grant 61501375 and the Fundamental Research Funds for Central Universities under Grant 3102016ZY006.

REFERENCES

1. Jian, L., Yang, H., Zhong, L. and Ying, X. Underwater target recognition based on line spectrum and support vector machine, *Proceedings of International Conference on Mechatronics, Control and Electronic Engineering (MCE2014)*, Atlantis Press, (2014).

2. Chung, K. W., Sutin, A., Sedunov, A. and Bruno, M. DEMON acoustic ship signature measurements in an urban harbor, *Advances in Acoustics and Vibration*, (2011).

3. Mak, B. A mathematical relationship between full-band and multiband mel-frequency cepstral coefficients, *IEEE Signal Processing Letters*, **9**(8), 241–244, (2002).

4. Zhao, X. and Wang, D. Analyzing noise robustness of MFCC and GFCC features in speaker identification, *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, BC, 7204–7208, (2013).

5. Chia, O., Hariharan, M. and Yaacob, S. Classification of Speech Dysfluencies with MFCC and LPCC Features,*Expert Systems with Applications*, **39**(2), 2157–2165,(2012).

6. Hung, W. W. and Wang, H. C. On the use of weighted filter bank analysis for the derivation of robust mfccs, *IEEE Signal Processing Letters*, **8**(3), 70–73, (2001).

7. Chandran, C., Kamal, S., Mujeeb, A. and Supriya, M. Novel class detection of underwater targets using self-organizing neural networks, *2015 IEEE Underwater Technology (UT)*, 1–5, (2015).

8. Cipli, G., Sattar, F. and Driessen, P. F. Multi-class acoustic event classification of hydrophone data, *Proceedings of 2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, Victoria, BC, 473–478, (2015).

9. Shen, P., Tamura, S. and Hayamizu, S. Multistream sparse representation features for noise robust audio-visual speech recognition, *Acoustical Science and Technology*, **35**(1), 17–27, (2014).

10. Jiang, Z., Lin, Z. and Davis, L. S. Learning a discriminative dictionary for sparse coding via label consistent K-SVD, *Proceedings of CVPR 2011*, Providence, RI, 1697–1704, (2011).

11. Aj, D., Leonardis. A. and Bischof, H. Weighted and robust learning of subspace representations. *Image & Vision Computing*, **40**(5), 1556–1569, (2007).

12. Chang, C. C. and Lin, C. J. LIBSVM: A library for support vector machines, *ACM Trans. Intell. Syst. Technol.*, **2**(3), 1–27, (2011).