OVERLAPPING ANIMAL SOUND CLASSIFICATION USING SPARSE REPRESENTATION

*Na Lin, Haixin Sun**

Xiamen University
Key Laboratory of Underwater Acoustic
Communication and Marine Information
Technology, Ministry of Education
422 Siming South Road, Xiamen, Fujian, China

Xiao-Ping Zhang

Ryerson University
Dept. of Electrical and Computer Engineering
350 Victoria Street Toronto, Canada, M5B 2K3
xzhang@ryerson.ca

ABSTRACT

In this paper, a new method to classify the animal sound signals that are overlapped in time-frequency domain based on sparse representation is proposed. In order to obtain a discriminant sparse representation of overlapped animal sound signals, a novel dictionary atom discriminant factor is introduced. Then the proposed method generates a representation that contains crucial signal discriminant information for classification and the sparsity for sparsest representation. The experimental results show that the proposed method has a much superior performance than the conventional sparse representation based classification method for classifying the overlapped animal sound signals.

***Index Terms*—** animal sound classification, sparse representation, overlapped signal

1. INTRODUCTION

The classification of animal sounds is an important research subject for the tracking of animals for research and conservation purposes [1, 2, 3], such as marine animal songs and bird calls. Many methods [1, 2, 3, 4, 5] have been proposed for animal species classification, but little work focuses on the problem that the animal sounds are overlapped. To our knowledge, only Briggs [1, 2] and Brandes [3] mention that more work is needed to identify the multiple simultaneously vocalizing birds. It is a challange problem to reliably classify the real-world audio data collected in an acoustic monitoring scenario. And it is indeed difficult to classify the multiple simultaneously calls [1]. The bird calls and marine animal songs often have overlapping components in the time-frequency domain, especially for the audios collected by unattended omnidirectional hydrophones. Their vocalizations have significant frequency content in the range audible to a human listener and they have overlapping frequencies [2, 6]. Therefore, the classification of sounds overlapping in the time-frequency domain is an important problem to be solved.

In [2], for the classification of multiple simultaneous bird species, a time-frequency segmentation of audio is developed by using random forest classifier. Three types of features are applied to describe a segment and a multi-instance multi-label framework for supervised classification is developed. There are three stages in most classification systems: segmentation, feature construction, and supervised classification. However, in these classification systems, the performance of classification depends on the selected features. It is hard to know what features can be used to express a signal better and how to extract the features. Moreover, different selected features can result in significantly different accuracy of classification for certain acoustic signals [7], ranging from 50% to 91.53% [8, 9, 10]. So a better way is needed to classify animal sound signals.

In recent years, sparse representation has received a great deal of attention [11, 12, 13]. This is due to the fact that signals or images with high dimension can be coded by using a few representative atoms in an overcomplete dictionary. And it has been shown that sparse representation works well in signal classification [11, 13, 14, 15] due to its robustness to noise and missing data [14, 16]. So sparse representation based classification (SRC) is a good way to deal with the problem of classifying animal sound signals. But for the signals overlapped in time-frequency domain, there are same representative atoms in different classes and the overlapping part increases the probability of selecting atoms from the other classes. It will increase the classification error rate. So it is necessary to eliminate the effect of the overlapping part for the classification of overlapped signals.

In this paper, in order to remove the effect of the overlapping part, a discriminant factor is incorporated based on sparse representation to classify the animal sound signals overlapping in the time-frequency domain. We deal with the overlapping part in training stage and obtain a set of optimal atoms that have the discriminatory power for classification

*This work was supported by the National Natural Science Foundation of China (61471309, 61671394) and the Fundamental Research Funds for the Central Universities (20720170044). The authors hereby extend their hearty thanks to these organizations.

by introducing a new discriminant factor. The proposed method generates a representation that contains crucial signal discriminant information for classification and the sparsity for sparsest representation. The experimental results show that compared with sparse representation based classification, the proposed method has a superior performance for the overlapped animal sound signal classification.

The remainder of the paper is organized as follows: Section 2 presents the details of our proposed method. Section 3 discusses experimental results, and Section 4 concludes the paper with a brief summary.

2. NEW SPARSE REPRESENTATION BASED CLASSIFICATION METHOD FOR OVERLAPPED SIGNAL

In this section, the new classification method for overlapped animal sound signal based on sparse representation is formulated mathematically.

Given a test animal sound signal $\mathbf{y}_t \in \mathbb{R}^N$ with N length, the problem of sparse representation for the signal is to make sure the sparse coefficients $\mathbf{x} \in L \times 1$ are clustered around the class corresponding to the test signal $\mathbf{y}_t$. The test signal $\mathbf{y}_t$ can be expressed as a linear combination of atoms from the corresponding class.

For representation of the overlapped animal sound signal and eliminate the effect of the overlapping time-frequency part, a discriminant factor is combined to produce a discriminant representation of the overlapped animal sound signal in training stage. Suppose $\mathbf{d} \in \mathbb{R}^N$ is an atom from $\mathbf{D}_s$ which is a set containing atoms selected from dictionary $\mathbf{D}$ according to the overlapping part between class A and class B. $\mathbf{D}$ is Gabor dictionary with M atoms in $\mathbb{R}^N$ composed of sines with Gaussian envelopes for the time-frequency analysis of animal sound signal. As mentioned in [17, 18], Gabor dictionary is good to analyze the time-frequency properties of signals. $\mathbf{C}_A$ is the set of sound frames of class A with L_A atoms in $\mathbb{R}^N$ and $\mathbf{C}_B$ is the set of sound frames of class B with L_B atoms in $\mathbb{R}^N$. Take class A as an example, we define the novel atom discriminant factor as

$$P(\mathbf{d}) = \max \frac{||\mathbf{y}^T\mathbf{d}||^2_{\mathbf{y}\in\mathbf{C}_A}}{||\mathbf{y}^T\mathbf{d}||^2_{\mathbf{y}\in\mathbf{C}_B}}. \quad (1)$$

For class B, the discriminant factor is obtained by swapping $\mathbf{C}_A$ and $\mathbf{C}_B$. The discriminant factor makes sure that there are no same atoms in different classes and the optimal atoms are selected. So the effect of the overlapping part can be eliminated for the classification of overlapped animal sound signals. By incorporating the discriminant factor into sparse representation, a representation that contains crucial information for discriminitive classification and the sparsity for sparsest representation is generated.

A discriminant sparse representation is obtained by maximizing the following optimization problem,

$$J_1(\mathbf{d}, \lambda_1, \lambda_2) = P(\mathbf{d}) - \lambda_1||\mathbf{y} - \mathbf{D}\mathbf{x}_v||_2 - \lambda_2||\mathbf{x}_v||_1, \quad (2)$$

where $\mathbf{y} \in \mathbf{C}_A$. $\mathbf{x}_v$ is the vector of sparse coefficients. $\mathbf{d} \in \mathbf{D}_s \subseteq \mathbf{D}$. $\lambda_1 > 0$ and $\lambda_2 > 0$ are scalar weighting factors that balance the tradeoff among the discriminant factor, reconstruction error and sparsity. For $\mathbf{d} \in \mathbf{D}$ and $\mathbf{d} \notin \mathbf{D}_s$, it is equivalent to solving a sparse representation problem $J_2(\mathbf{d}; \lambda)$ as defined in [15] to obtain the representative atoms of training signals $\mathbf{C}_A$ and $\mathbf{C}_B$. Then instead of maximizing $J_1(\mathbf{d}, \lambda_1, \lambda_2)$, we maximize the sparse representation problem

$$J_2(\mathbf{d}; \lambda) = ||\mathbf{y} - \mathbf{D}\mathbf{x}_v||_2 + \lambda||\mathbf{x}_v||_1, \quad (3)$$

where $\lambda > 0$ is a scalar regularization parameter that balances the tradeoff between reconstruction error and sparsity.

In the test stage, suppose $\mathbf{Y}$ with L atoms in $\mathbb{R}^N$ contains the representative atoms of $\mathbf{C}_A$ and $\mathbf{C}_B$ obtained in the training stage, $\mathbf{Y} = [\mathbf{Y}_A, \mathbf{Y}_B]$. Let $\mathbf{Y}_A = [d^A_1, d^A_2, ..., d^A_{M_A}] \in \mathbb{R}^{N\times M_A}$ with M_A atoms be the matrix for $\mathbf{C}_A$. And d^A_1, d^A_2 , ..., $d^A_{M_A}$ represent the atoms in $\mathbf{Y}_A$. $\mathbf{Y}_B = [d^B_1, d^B_2, ..., d^B_{M_B}] \in \mathbb{R}^{N\times M_B}$ with M_B atoms is the matrix for $\mathbf{C}_B$. d^B_1, d^B_2, ..., $d^B_{M_B}$ are the atoms in $\mathbf{Y}_B$. $L = M_A + M_B$.

To predict the class of a test signal $\mathbf{y}_t$, the sparse coefficient is computed by optimizing the following problem according to [15],

$$\mathbf{x}_t = \arg\min_{\mathbf{x}} ||\mathbf{y}_t - \mathbf{Y}\mathbf{x}||_2 + \lambda||\mathbf{x}||_1. \quad (4)$$

In [15], the class of the test sample $\mathbf{y}_t$ is decided according to the smaller approximation error. And the errors for $\mathbf{C}_A$ and $\mathbf{C}_B$ can be computed as follows. c is A or B:

$$\mathbf{e}_c = ||\mathbf{y}_t - \mathbf{Y}_c\mathbf{x}^c_t||_2. \quad (5)$$

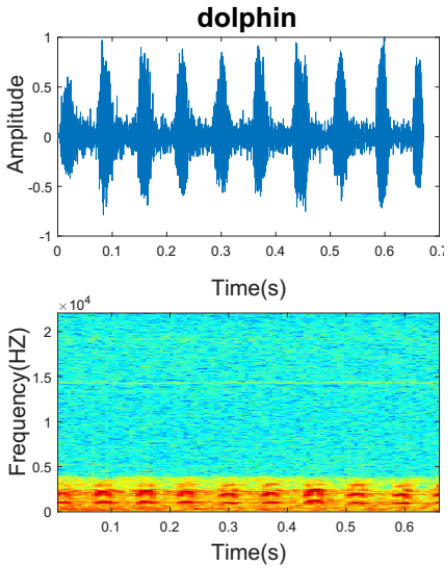
But in reality, the received signal obtained from hydrophones is a mixture of animal sounds from different classes. Then the decision rule of classification mentioned above is invalid. When the received signal is a mixture of animal sounds from different classes, the errors of the corresponding classes can be all small. It is hard to decide what class the signal is. In our method, when the errors of the classes

$$\mathbf{e}_c = ||\mathbf{y}_t - \mathbf{Y}_c\mathbf{x}^c_t||_2 < \epsilon, \quad (6)$$

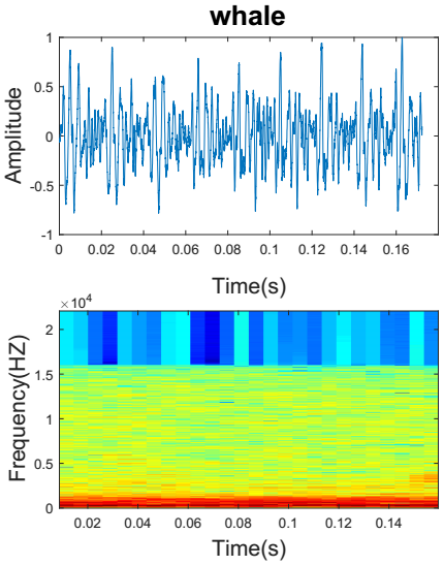
we deem the test signal as a mixture of sounds from the corresponding classes. ϵ is a scalar and an empirical threshold. According to the experiments, the errors of the two classes are nearly the same in this two-class classification scenario when the test signal is a mixture of sounds from the two different classes and the energy of different sounds are close to the same. In this case, the emprical threshold can be set as 0.5 for the test signals that are normalized.

3. EXPERIMENTAL RESULTS

The experiment of overlapped animal sound signal classification is conducted using audios from online databases in [19, 20, 21, 22], including whale sounds and dolphin sounds. These audios are segmented in the preprocessing stage and all sample sounds are nearly have a complete call. Fig.1 shows two sample sounds in our experiment. From Fig.1, we can see that dolphin sound and whale sound have overlapping frequency components.



(a) dolphin sound



(b) whale sound

Fig. 1. Sample sounds of dolphin and whale. We can see that there are overlapping frequencies between dolphin sound and whale sound.

In the experiment, there are 15 training animal sound signals and 25 test animal sound signals. Then a set of representative atoms is obtained in training stage by adopting the proposed new method for classifying the test signals. To compare with the proposed method, sparse representation based classification mentioned in [14] without introducing the discriminant factor is also performed and the classification accuracy rate of the test animal sound signals by using sparse representation based classification is 84%. While the classification accuracy rate of the test animal sound signals by adopting the proposed method is 92%.

To demonstrate the effectiveness of the proposed method, sparse coefficients are plotted when a test signal is represented as a sparse linear combination of the atoms obtained in training stage. Sparse coefficients of the two sample sounds (shown in Fig. 1) are plotted in Fig. 2 by adopting the proposed method. The horizontal axis represents the atoms obtained in training stage and the vertical axis is the coefficients of the atoms. Red represents the atoms of class whale, and blue represents the atoms of class dolphin. There are values for the atoms which are selected to represent the test signals while the coefficients of unselected atoms are zeros. From Fig. 2, we can see that most selected atoms belong to the same class as the test signals. And there is almost no selected atom from the different class. It means that the atoms obtained by adopting the proposed method is discriminatory to represent the whale sound and dolphin sound.

Fig. 3 shows the sparse coefficients obtained by using sparse representation based classification. In Fig. 3, a few coefficients are in the different class as the test signals which means that some atoms from the different class are selected to represent the test signals. It is mainly caused by the overlapping part of whale sounds and dolphin sounds. Compared with Fig. 2, it can be seen that the proposed new method is more discriminatory for the overlapped animal sound signals classification by introducing the discriminant factor.

4. CONCLUSION

In this paper, we develop a new classification strategy to classify the animal sound signals overlapping in the time-frequency domain. By introducing a novel dictionary atom discriminant factor, the optimal signal atoms are selected for classifying the overlapped animal sound signals. Then the sparse representation of overlapped animal sound signal contains crucial signal discriminant information for classification and the sparsity for sparsest representation. Experimental results show that the proposed new method is much superior for overlapped animal sound signal classification than the conventional sparse representation based classification method.

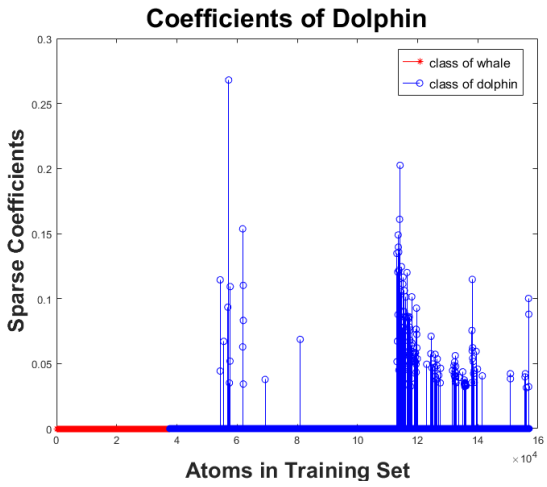


Fig. 2. Coefficients of the two sample sounds by adopting the proposed method. Red represents the atoms of class whale, and blue represents the atoms of class dolphin. Almost all the selected atoms belong to the same class as the test signals.

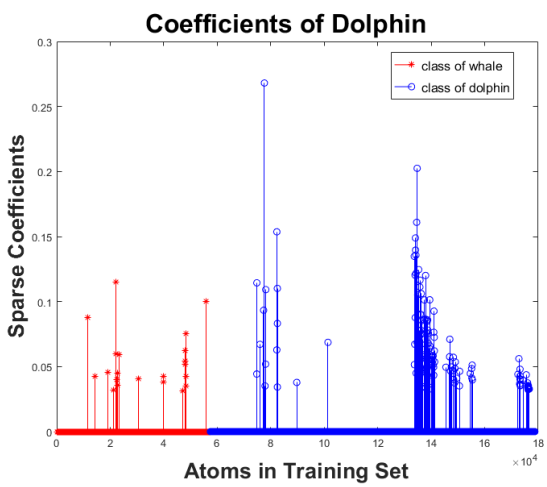


Fig. 3. Coefficients of the two sample sounds by using sparse representation based classification. Red represents the atoms of class whale, and blue represents the atoms of class dolphin. A few selected atoms are in the other class.

5. REFERENCES

[1] Forrest Briggs, Yonghong Huang, Raviv Raich, and so on, "The 9th annual mlsp competition: New methods for acoustic classification of multiple simultaneous bird species in a noisy environment," in *IEEE International Workshop on Machine Learning for Signal Processing, MLSP*. IEEE, 9 2013, pp. 1–8.

[2] Forrest Briggs, L. Balaji, N. Lawrence, X. Z. Fern, et al., "Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach," *Acoustical Society of America*, vol. 131, no. 6, pp. 4640–4650, 2012.

[3] T. Scott Brandes, "Automated sound recording and analysis techniques for bird surveys and conservation," *Bird conservation International*, vol. 18, no. 1, pp. 163–173, 1996.

[4] A.L. McIlraith and H.C. Card, "Birdsong recognition using backpropagation and multivariate statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2740–2748, 1997.

[5] Seppo Fagerlund, "Bird species recognition using support vector machines," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, pp. 038637, 2007.

[6] C. M. Binder and P. Hines, "Applying automatic aural classification to cetacean vocalizations," in *Proceedings of meetings on acoustics*. Acoustical Society of America, 2012, vol. 17.

[7] B. Wei, M. Yang, Y. Shen, et al., "Real-time classification via sparse representation in acoustic sensor networks," in *SenSys '13 Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, 2013.

[8] C. Huang, Y. Yang, D. Yang, and Y. Chen, "Frog classification using machine learning techniques," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3737–3743, 2009.

[9] A. Taylor, G. Watson, G. Grigg, and H. McCallum, "Monitoring frog communities: an application of machine learning," in *IAAI'96 Proceedings of the eighth annual conference on Innovative applications of artificial intelligence*, 1996, pp. 1564–1569.

[10] G. V. Castano and D. Rodriguez, "Using syllabic melcepstrum features and k-nearest neighbors to identify anurans and birds species," in *Signal Processing Systems (SIPS), 2010 IEEE Workshop on*, 2010, pp. 466–471.

[11] J. Wright, Y. Ma, J. Mairal, et al., "Sparse representation for computer vision and pattern recognition," in *Proceedings of the IEEE*, 2010, vol. 98, pp. 1031–1044.

[12] M. Elad, M. A. T. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," in *Proceedings of the IEEE*, 2010, vol. 98, pp. 972–982.

[13] K. Huang and S. Aviyente, "Sparse representation for signal classification," in *NIPS'06 Proceedings of the 19th International Conference on Neural Information Processing Systems*, 2007, pp. 609–616.

[14] J. Wright, A. Y. Yang, A. Ganesh, et al., "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[15] A. Shrivastava, V. M. Patel, and R. Chellappa, "Multiple kernel learning for sparse representation-based classification," *IEEE Transactions on Image Processing*, vol. 23, no. 7, pp. 3013–3024, 2014.

[16] J. K. Pillai, V. M. Patel, R. Chellappa, and N. K. Ratha, "Secure and robust iris recognition using random projections and sparse representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1877–1893, 2011.

[17] S.G. Mallat and Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[18] M. Esfahanian, *Detection and Classification of Marine Mammal Sounds*, Ph.D. thesis, Florida Atlantic University, Boca Raton, 2014.

[19] "Dolphin sounds," `http://everythingdolphins.com/dolphin-sounds.html`.

[20] "The dolphin sounds of the mediterranean sea," `http://www-3.unipv.it/cibra/edu_dolphins_uk.html`.

[21] "The whale sounds of the mediterranean sea," `http://www-3.unipv.it/cibra/edu_spermwhale_uk.html`.

[22] "Whale sounds," `http://soundbible.com/tags-whale.html`.