AUTOMATIC BIRD VOCALIZATION IDENTIFICATION BASED ON FUSION OF SPECTRAL PATTERN AND TEXTURE FEATURES

Sai-hua Zhang[a], Zhao Zhao[a,b,*], Zhi-yong Xu[a,*], Kristen Bellisario[b], Bryan C. Pijanowski[b]

[a] School of Electronic and Optical Engineering, Nanjing University of Sci. & Tech., 210094, China

[b] Department of Forestry and Natural Resources, Purdue University, West Lafayette IN47907, USA

Abstract

Automatic bird species identification from audio field recordings is studied in this paper. We first used a Gaussian mixture model (GMM) based energy detector to select representative acoustic events. Two different feature sets consisting of spectral pattern and texture features were extracted for each event. Then, a ReliefF-based feature selection algorithm was employed to select distinguishing features. Finally, classification was performed using support vector machine (SVM). The main focus of the proposed method lies in the fusion of a spectral pattern feature with several texture descriptors, which extends our previous work. Experiments used an audio dataset comprised of field recordings of 11 bird species, containing 2762 bird acoustic events and 339 detected "unknown" events (corresponding to noise or unknown species vocalizations). Experimental results demonstrate superior classification performance compared with that of the state-of-the-art method, which renders the proposed method more suitable for real-field recording analysis.

Introduction

Birds have been used widely as indicators of biodiversity because they are distributed over a wide range of natural habitat and are relatively easy to detect. As a complement to traditional human-observer-based survey methods, acoustic analysis of bird vocalizations can be used for automated species identification, leading to a promising non-intrusive method for bioacoustic monitoring.

The first stage of acoustic bird species identification is usually to segment the continuous recordings into isolated acoustic events, including manual segmentation and automated segmentation method. After segmentation, each segment is represented by a set of features that can discriminate between different classes of bird sounds. Then, a recognition algorithm is employed to identify the bird species based on the extracted features.

In our most recent study, we proposed a novel spectral pattern feature, which depicts the species-specific spectral pattern, contributing to superior robustness in real-world scenarios. More recently, texture descriptors have been investigated and successfully applied to the task of face recognition and image retrieval. It is worth remarking that fusion of different feature sets has been used for automatic acoustic classification of anurans and proved efficient in performance improvement. Based on these recent studies, we propose a new method by combining the spectral pattern feature with texture descriptors in this paper, extending our previous work. Considering the "curse of dimensionality" problem, the ReliefF algorithm is employed for feature ranking and selection, helping to obtain a distinguishing subset of the features.

Proposed method

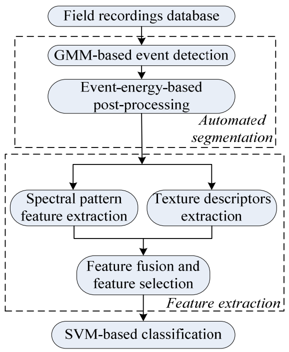
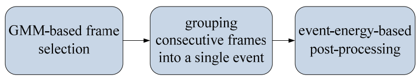


Fig. 1. Overall block diagram of the proposed method.

1. Automated segmentation



2. Feature extraction

2.1 Spectral pattern feature

• **sp**: spectral pattern(SP) feature. We employed a Mel band-pass filter bank on the spectrogram of each event. The output in each subband, i.e. a time-series containing time-varying band-limited energy, was parameterized by an autoregressive (AR) model, which resulted in a parameterized feature consisting of all model coefficients.

2.2 Texture descriptors

• **lbp**: uniform local binary pattern(ULBP)
• **lbf** : LBP histogram Fourier(LBPHF)
• **lpq**: local phase quantization(LPQ)
• **hac**: heterogeneous auto-similarities of characteristics(HASC)
• **gf**: Gabor filter(GF)

2.3 Feature selection

$$\mathbf{cfv}_i = \left[\mathbf{sp}_i^T, \mathbf{lbp}_i^T, \mathbf{lbf}_i^T, \mathbf{lpq}_i^T, \mathbf{hac}_i^T, \mathbf{gf}_i^T\right]^T, i = 1, 2, \cdots, K \quad (1)$$

where K is the total number of events.

It is worth noting that this simple concatenation leads to a high-dimensional feature vector. In this context, the ReliefF feature selection algorithm, aiming to choose a discernible subset of features to reduce the fusion feature length with the lowest information loss, can be employed to remove irrelevant and/or redundant features. With the help of ReliefF, an attribute weight was calculated for each feature, ranging from -1 to 1 with a high positive weight assigned to an important attribute. Then we sorted out N most important features as the effective subset. According to our preliminary study, we selected N=400.

3. SVM-based classification

In this work, we employed multi-class SVM based on the "one-versus-one" strategy for species classification. Besides, we used the radial basis function (RBF) as the kernel function The two parameters of the RBF kernel, gamma and cost, were set to 0.0625 and 8, respectively.

Experimental evaluations

In order to evaluate our method, the field audio recordings used were downloaded from the Xeno-canto Archive. Note that these are all real-world recordings and each recording potentially contains vocalizations of several animal species and competing noise originating from wind, rain, or anthropogenic interference. There are 11 bird species in the recordings that can be divided into five types based on sound unit shapes, including constant frequency (CF), frequency modulated whistles (FM), broadband pulses (BP), broadband with varying frequency components (BVF), and strong harmonics (SH). We provide the description of the dataset used in this study in Table 1.

Table 1. Details of species used in this work.

Bird species	Call/ Song	Sound unit shape	Number of events
Blue Jay (B-J)	Call	SH	251
Song Sparrow (S-S)	Call	SH	259
Marsh Wren (M-W)	Call	BP	249
Common Yellowthroat (C-YT)	Call	BP	256
Chipping Sparrow (C-S)	Call	FM	253
American Yellow Warbler (A-Y-W)	Call	FM	247
Great Blue Heron (G-B-H)	Call	BVF	247
American Crow (A-C)	Call	BVF	253
Cedar Waxwing (C-WW)	Call	CF	246
House Finch (H-F)	Song	SH	249
Indigo Bunting (I-BT)	Song	FM	252
Unknown	/	/	339

Results

Table 2. Comparison of various features in terms of classification accuracy (CA) equipped with the same SVM classifier.

Features	Dimension of feature vector	CA (%)
SP (**baseline method**)	320	93.7
HASC+LBP+LPQ+LBPHF+GF	475	86.6
SP+HASC+LBP+LPQ+LBPHF+GF without feature selection	795	94.0
SP+HASC+LBP+LPQ+LBPHF+GF with feature selection (**this work**)	400	96.7

Table 3. The averaged performance metrics for each species between the proposed method with the baseline method.

Classes	Recall (%)		Precision (%)		*F*-score	
	This work	Baseline method	This work	Baseline method	This work	Baseline method
B-J	**98.7**	97.9	**99.0**	97.9	**0.989**	0.979
S-S	**97.0**	92.6	**95.4**	88.4	**0.962**	0.904
M-W	**96.8**	90.1	**97.5**	94.5	**0.971**	0.922
C-YT	**95.6**	91.0	**95.1**	89.6	**0.953**	0.903
C-S	**97.7**	93.8	**97.0**	94.2	**0.973**	0.940
A-Y-W	**97.7**	93.8	**99.0**	90.7	**0.983**	0.922
G-B-H	**96.3**	91.4	**98.1**	94.8	**0.972**	0.930
A-C	**99.2**	97.9	**99.5**	96.7	**0.994**	0.973
C-WW	**96.8**	95.6	98.7	**99.0**	**0.977**	0.973
H-F	**95.1**	94.7	**95.3**	92.4	**0.951**	0.935
I-BT	**95.7**	94.0	**94.5**	93.3	**0.950**	0.936
Unknown	**95.2**	91.7	94.0	**94.4**	**0.945**	0.930

Conclusions

Aiming to improve the audio parameterization process in bird species identification tasks, we proposed an automatic acoustic classification method based on feature fusion in this work. Two different feature sets, the spectral pattern feature and texture descriptors, were extracted. As for the combination of the two sets, ReliefF-based feature selection algorithm was employed to select a distinguishing feature subset. Experimental results using real-world recordings showed that the proposed method outperformed the state-of-the-art robust approach.

Acknowledgments

This work was supported by the National Natural Science Foundations of P.R. China under Grants No. 61401203 and No. 61171167; and the State Scholarship Fund of China [grant number 201606840023].