Neurocomputing 155 (2015) 108–116

Contents lists available at ScienceDirect

Neurocomputing

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

Adaptive energy detection for bird sound detection in complex environments

Xiaoxia Zhang [a,b], Ying Li [b,*]

[a] Information Center of Fujian Medical University, Fuzhou 350108, China
[b] College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China

ARTICLE INFO

Article history:
Received 3 August 2013
Received in revised form
9 September 2014
Accepted 15 December 2014
Communicated by J. Zhang
Available online 29 December 2014

Keywords:
Adaptive energy detection
Noise variance estimation
Mel-scaled Wavelet packet decomposition
Sub-band Cepstral Coefficient (MWSCC)
Mel-Frequency Cepstral Coefficient (MFCC)
Bird sound recognition

ABSTRACT

A new bird sound classification approach based on adaptive energy detection was proposed to improve the recognition accuracy of bird sounds in noisy environments. In this paper, the bird sounds with background noises were divided into three linear frequency bands according to their frequency distribution in spectrogram. The noise spectrum of each band was estimated and the existent probability of the foreground bird sound for each band was computed to serve for the adaptive threshold of energy detection. These foreground bird sound signals were detected and selected via adaptive energy detection from the bird sounds with background noises. Then, the features of Mel-scaled Wavelet packet decomposition Sub-band Cepstral Coefficient (MWSCC) and Mel-Frequency Cepstral Coefficient (MFCC) were extracted from the above signals for classification by using the classifier of Support Vector Machine (SVM), respectively. Moreover, the differences of recognition performance were implemented on 30 kinds of bird sounds at different Signal-to-Noise Ratios (SNRs) under different noisy environments, before or after adaptive energy detection. The results show that MWSCC has better noise immunity function, and the recognition performance after adaptive energy detection improves more significantly, indicating that it is a very suitable approach for the bird sound recognition in complex environments.

1. Introduction

With the continuous development of modern society and economy, more and more attention is paying to the ecological environment. Environmental sound which contains a large amount of rich information is one indispensable element of ecological environment. Birds living in complex ecological environments can emit various kinds of sounds, which also contain a wealth of information closely related to the human survival environments. The relevant information of birds can be obtained by inspecting, analyzing and identifying these bird sounds. After counting the quantity and category of bird species and learning their life habits, people cannot only monitor and protect bird species, but also evaluate and predict the surrounding living or ecological environments. Therefore, the automatic classification and identification of bird species by their acoustic sound have great potential significance on the protection and sustainable development of ecological environment, which indirectly impact the development of industry and economy.

With the increasing development of science and technology, the researchers have made great achievements in the classification and recognition of bird sounds. In [1], the simple sinusoidal representations of isolated syllables were used to automatically recognize a number of bird species. In [2], a probabilistic model for audio features of statistical manifolds of histograms was proposed, and a risk-minimizing Bayes classifier was used to identify 6 species of birds, which was closely approximated by a nearest-neighbor classifier using Kullback–Leibler divergence to compare histograms of features. In [3], the automatic recognition of 14 bird species on both individual syllables and song fragments was presented; three features of sinusoidal modeling, Mel-Frequency Cepstral Coefficient (MFCC) and low-level descriptive parameters were compared; three classifiers of Dynamic Time Warping (DTW), Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM) were analyzed for the effectiveness of the classification performance.

Considering the existence of noise in real world, the researchers thought of taking advantage of the methods of noise reduction. For example, [4] presented a recognition approach which consisted of feature extraction using wavelet decomposition after reducing the noises and segmentation, and recognition of inharmonic and transient bird sounds using the unsupervised self-organizing map (SOM) and the supervised multilayer perceptron (MLP). [5] described novel algorithms for the detection of the vocalizations of two endangered bird species in natural noisy environments which contained the presence of multiple sound sources overlapping in time and

* Corresponding author. Tel.: +86 13959199804.
E-mail addresses: mismall@foxmail.com (X. Zhang), fj_liying@fzu.edu.cn (Y. Li).

http://dx.doi.org/10.1016/j.neucom.2014.12.042
0925-2312/

frequency, to apply for bird monitoring; the algorithms used noise estimation from frequency bands and spectral subtraction to make effective noise reduction. But the noise reduction methods also produce music residual noise or cause sound distortion, which may make a negative influence on the bird sound recognition performance. Therefore, we detect the bird sounds from noises rather than reduce noises, and then make recognition on the detected sounds. Such as, in [6], noise robust bird song detection based on syllable pattern and classification using HMM in noisy environments was described. [7] presented a study of automatic detection exploiting the spectral shape of sinusoidal components in the short-time spectrum, feature extraction based on tonal representation, and recognition of 95 types of tonal bird sounds in white noise and real-world waterfall noise environments respectively using Gaussian Mixture Model (GMM).

In this paper, the method of energy detection (ED) was introduced to detect the foreground bird sound of interest in complex environments owing to its simplicity. ED is most used for spectrum sensing in cognitive radio [8, 9], but its noise power is required to be known beforehand and the threshold is fixed, which is not suitable for bird sound detection. So adaptive energy detection (AED) based on noise variance estimation and adaptive threshold according to multi-band probability was proposed to apply for the bird sound detection in different bands under diverse noisy conditions. After detection, the feature of selected sounds was extracted for classification. MFCC is one of the most used features for speech and speaker recognition [10] due to its full consideration of human auditory characteristics, but its performance degrades under noisy environments. While the wavelet decomposition has noise robustness and is especially fit for analyzing the non-stationary bird sounds, seen from [4], thus the feature of Mel-scaled Wavelet packet decomposition Sub-band Cepstral Coefficient [11], marked as MWSCC, is used. Compared to other classifiers, such as GMM, HMM, K Nearest Neighbor (KNN) and Artificial Neural Network (ANN), Support Vector Machine (SVM), first proposed by Cortes and Vapnik in 1995, shows better efficiency and accuracy[12]. It has been extensively used in many fields of pattern recognition such as image analysis, text classification, speech recognition and speaker identification. However fewer works have tried to apply SVM to acoustic sound classification [13]. In this paper, we select SVM to model and recognize the various birds sounds properly.

The remaining part of this paper is organized as follows: Section 2 describes the proposed methods in detail and analyzes their characteristics. A kind of bird sound example of the method is given in Section 3. Experiments are displayed and their results are shown in Section 4. Some discussions are made in Section 5. Conclusions are drawn in Section 6 along with the future research.

2. Methods

The architecture of our acoustic sound classification for bird species can be divided into three modules: adaptive energy detection, feature extraction, and classification. The flow diagram of the proposed approach is shown in Fig. 1.

2.1. Adaptive energy detection

2.1.1. Existing energy detection scheme

We consider the observed sound signal in discrete time domain with respect to additive noise under the following two hypotheses:

$$Y(n) = \left\{ \begin{array}{ll} W(n) & H_0 \\ S(n)+W(n) & H_1 \end{array} \right\} n = 1, 2, \ldots, N \tag{1}$$

where H_0 and H_1 are respectively the "signal absent" and "signal present" hypotheses; $Y(n)$ is the observed sound signal; $W(n)$ is the background acoustic noise; $S(n)$ is the foreground uncontaminated sound signal; n is the sample index; N is the number of samples.

The accumulated energy E of the sound signal $Y(n)$ within specific time can be written as follows:

$$E = \sum_{n=1}^{N} [Y(n)]^2 \tag{2}$$

where N is the number of samples within specific time, whose value equals to the length of a frame.

Energy detection [14] is a kind of decision that decides whether the signal of interest exists by comparing the energy of the signal within specific period of time with a preset threshold. The decision rule of energy detection(ED) [15] is given by (3).

$$\frac{E}{\sigma_w^2} \left\{ \begin{array}{l} < \lambda H_0 \\ > \lambda H_1 \end{array} \right. \tag{3}$$

where σ_w^2 denotes the noise variance, λ denotes the decision threshold. We will discuss the value of noise variance and decision threshold later.

When the number N is sufficiently large, the test statistic E/σ_w^2 can be assumed to follow a Gaussian (normal) distribution [16, 17] corresponding to the hypotheses

$$H_0 : \tfrac{E}{\sigma_w^2} \sim Normal\,(N, 2N)$$

$$H_1 : \frac{E}{\sigma_w^2} \sim Normal\,(N(1+SNR), 2N(1+SNR)^2) \tag{4}$$

where the signal-to-noise ratio (SNR) is defined as $SNR = \sigma_s^2/\sigma_w^2$, σ_s^2 denotes the uncontaminated signal variance. Then, related with (4), the probability of false alarm PFA and probability of detection PD can be computed as follows:

$$PFA = P\left(\frac{E}{\sigma_w^2} > \lambda | H_0\right) = Q\left(\frac{\lambda - N}{\sqrt{2N}}\right) \tag{5}$$

$$PD = P\left(\frac{E}{\sigma_w^2} > \lambda | H_1\right) = Q\left(\frac{\lambda - N(1+SNR)}{\sqrt{2N(1+SNR)^2}}\right) \tag{6}$$

where $Q(x)$ is the standard Gaussian complementary cumulative distribution function, defined as follows:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(\frac{-t^2}{2}\right) \mathrm{d}t \tag{7}$$

and $Q^{-1}(x)$ denotes its inverse function.

Given the value of PFA or PD , the threshold λ can be derived from (5) or (6) as follows:

$$\lambda = N + \sqrt{2N} Q^{-1}(PFA) \tag{8}$$

$$\lambda = (1+SNR)(N + \sqrt{2N} Q^{-1}(PD)) \tag{9}$$

In addition, if the noise variance σ_w^2 is priori known, then we can use (3) to decide whether the foreground sound signal of interest exists. But this requires a priori knowledge of the noise variance.

In practice, the foreground bird sounds and background noises are usually mixed, which lead to the fact that the noise variance cannot be known clearly in advance. So in this paper, based on the algorithms of noise spectrum estimation [18,19] put forward by others before, we propose a method for estimating the noise variance designed for energy detection (ED) described below. However, this may cause some errors of ED during the process of noise variance estimation. In order to solve the problem, threshold adaptation is considered. Selection of the threshold level for ED is extremely important. The optimal detection threshold according to the noise variance estimation information and multi-band probability is proposed and represented below.