International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072

Classification of birds based on their sound patterns using GMM and SVM classifiers

Supriya[1], Prakrithi R.[2], Shwetha Bhat[3], Shivani[4], Santhosh S[5]

[1,2,3,4] Student, Dept. of CSE, SSE, Mukka, Mangalore
[5]Assistant Professor, Dept. of CSE, SSE, Mukka, Mangalore, Karnataka, India

---***---

Abstract - *This paper classifies the bird automatically considering their chirping patterns. Speech processing and Machine learning techniques are carried out to categorize the bird species. An attempt has been made to include the bird's local to South Canara region of India in order to understand the ecological interdependence. Gaussian Mixture model (GMM) and Support Vector Machine (SVM) are employed here to classify the birds into the respective classes. An overall efficiency of 95% is gained using 50 recordings of 4 bird classes using GMM and an efficiency of 86% through 50 recordings of 5 bird classes using SVM. This project is helpful for Ornithologists for identification of birds on hearing their chirps.*

Key Words: Speech processing, Machine learning, Bird species, South Canara region, Gaussian Mixture model, Support Vector Machine.

1. INTRODUCTION

Bird sounds are of great importance in ecology and in monitoring the environment. Earlier studies used small data sets and classification was done manually with minimal practical needs. Majority of the traditional studies on the analysis of birds vocalization are based on visual inspection of sound spectrogram which is a tedious task and many of the currently available methods are accurate for only a relatively small sets of bird sounds. Therefore, automating the process of identification of birds with minimal manual intervention is of great importance. Another important aspect of previously studied methods is the quality of recordings. Most of these studies are based on recordings with low or negligible noise component and rely on expensive audio equipment for this purpose. In such cases, several techniques have proved to be successful in classifying birds from audio, but cannot scale to practical scenarios without significant changes in the methodology (Somervuo et al. 2006; Juangand Chen 2007). Task specific research on classification and identification of bird sounds has been a challenging task and has only recently attracted the attention of the research community because of its wide variety of applications that are highly relevant in recent ecological scenarios. The aim of this study is to develop automated techniques for identifying the birds based on their sound patterns using Mel frequency cepstral coefficients. Possible applications of these techniques would enable people identify different bird species and use these results for further ecological or biological studies. Classification and identification of bird's sounds can be done by comparing some basic features that all birds share.

2. LITERATURE REVIEW

The songs of the birds have vocalizations which are usually very long and they have many kind of notes which are in sequence. The most commonly used technique for bird sound processing is energy based time-domain approach which is reliable for single bird's samples with low noise (Somervuo et al. 2006; Juang and Chen 2007). In the case of multiple bird's sounds in noisy environments, two dimensional time–frequency based segmentation is used (Mellinger and Bradbury 2009). The most widely used features to describe bird's sounds are linear predictive coefficients (LPC) and Mel-frequency cepstral coefficients (MFCCs) which are also prevalent in other areas of signal processing (Chen and Maher 2006; Davis and Mermelstein 1980). Recently there has been many studies undertaken which made use of multi-instance multi-label learning (MIML). These techniques were previously implemented on images, texts and audio samples where inexpensive to get labels at the bag level in order to labeling individual calls is common.

Some researchers have attempted to solve the problem of detecting bird sounds in complex environments such as the sound samples obtained from automatic recording units (Briggs et al. 2012; Bardeli et al. 2010). There are references of many papers in recent years which have addressed the fundamental problem of automatic bird species identification (ABSI) using signal processing and machine learning techniques. Majority of these papers can be analyzed based on the feature-set and the machine learning techniques employed. Somervuo et al. (2006) achieved an accuracy of about 71.3 % using Mel frequency cepstral coefficients (MFCCs). Fagerlund (2007), in another work has obtained a database of eight bird species making use of MFCC and low level signal parameters. Lopes et al. (2011) showed that the music analysis, retrieval and synthesis for audio signals (MARSYAS) feature set along with a multi-layer perceptron neural network classifier achieves the best known accuracy of three species of bird. Carlos et al. (2013) address the task of hierarchical bird species identification from audio recordings. Their work brings out the three different approaches of hierarchical classification problem, namely, the flat classification approach, local-model per parent node classifier approach and global-model of hierarchical classification approach. The first and second approach use classical Naive Bayes algorithm and the third approach uses Global Model Naive Bayes (GMNB) algorithm. Kwan et al. (2004) proposed the development of a high performance bird classification system using Hidden Markov

Model and Gaussian Mixture Model. The recognition in Hidden Markov Model uses Principal Component Analysis (PCA) and Vector Quantization (VQ). PCA is applicable for data dimension reduction. In Gaussian Mixture Model (GMM), the algorithm will have the extracted feature vectors first obtained from the sound data and then match them with trained GMM parameters where there is exactly one trained GMM for each class. The bird class is classified by comparing the preset threshold of the birds and the difference between their probabilities.

3. METHODOLOGY

The proposed approach is divided into four stages, as shown in Figure 1. Collection of bird sound data, pre-processing of bird sound recordings, feature extraction and using GMM and SVM on the features to classify the birds.

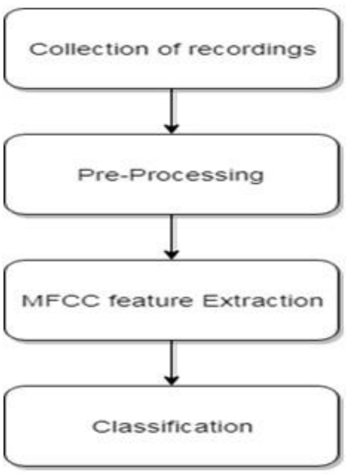

Fig -1: Stages in bird classification

3.1 Collection of recordings

The dataset used in this work is obtained from Xeno-canto where the sound recordings are labelled by users along with the region from where they have obtained the bird sound recordings. The sound recordings that we obtained were of different sampling rates and in mp3 format. So all the audio recordings were converted into wav format at the sampling rate of 8000Hz. In this work, an effort has been made to classify the birds that were local to South Canara region. Hence the bird recordings of bulbul, crow, cuckoo, myna and sparrow were taken.

3.2 Pre-processing

The recordings from the Web source were mixed with unwanted sounds. In addition to it, they were of longer duration. It is important to extract the desired bird sound from the environmental noise, or at least reduce the effect of noise. Hence pre-processing of the sound recordings is necessary to obtain the system with better accuracy. Praat tool was used for this purpose.

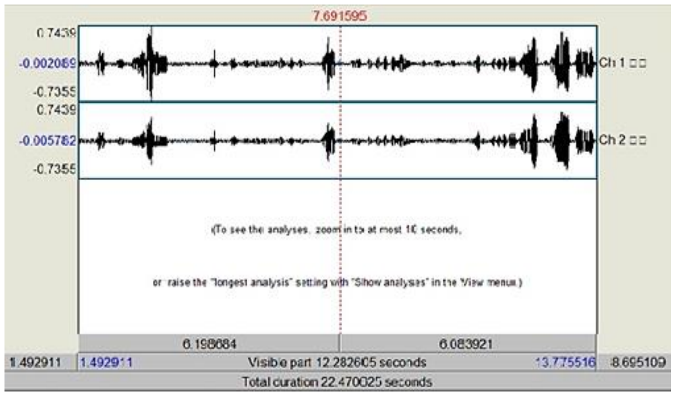

Fig -2: Bird song original waveform

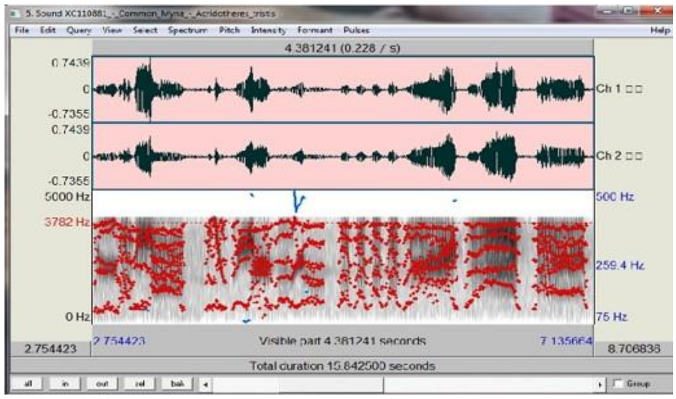

Fig -3: Bird song processed waveform

Figure 2 and 3 show the waveform of Common Myna before and after processing respectively. Praat tool is applied (as shown in Fig -2). This results in a processed waveform (as shown in Fig -3), which has better audio quality and shorter duration. This processed signal is used for feature extraction.

3.3 Feature Extraction

Feature Extraction involves the usage of MFCC for classification. Mel-frequency cepstrum (MFC) represents the power spectrum of the audio in short-term. It is based on the linear cosine transform of logarithmic power spectrum on a nonlinear mel scale of frequency. Mel-frequency cepstral coefficients (MFCCs) make up an MFC collectively. These features are unique for each individual. This property makes it eligible to classify the birds in our paper.

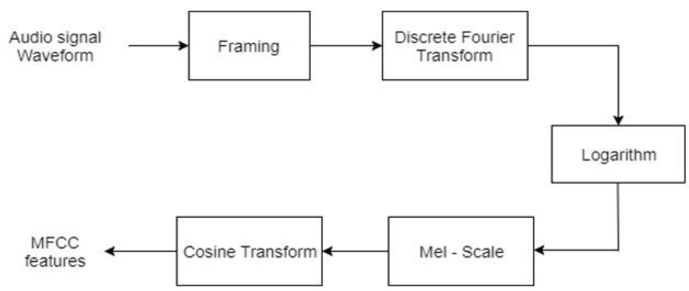

Fig -4: Stages in the extraction of MFCC

Figure 4 shows the stages in the extraction of MFCC. Initially the audio is framed into durations as small as 20ms. A Fourier Transform of the signal is taken to obtain frequency related information. Logarithm of the resulting data helps to amplify small features which would be lost otherwise. Finally Mel-Scale and Cosine transform are taken to obtain 13 MFCC features.

3.4 Classification

In this paper, we have used both GMM and SVM to classify the data.

Gaussian mixture models (GMM) are used for clustering the data. It helps to construct high performance class models for tasks like pattern recognition with the help of statistical data. A GMM is represented as a weighted sum of Gaussian component densities which is a parametric probability density function. GMM's are commonly used as an invariable model of the probability distribution. They are also used as continuous measurements in biometric systems like vocal-tract related spectral features in a speech recognition system. GMM statistical speech model is created after extracting features .When conditions of single normal distribution fails, the finite mixture models and their typical parameter estimation methods can be approximated by probability density functions(pdf). A basic distribution in which predefined distribution type is used to form a mixture. The Gaussian distribution is one of the most helpful distribution in playing an important role in statistics and various other areas of applications.

A discriminative classifier defined by separating hyper plane is called Support Vector Machine (SVM). In other words, an optimal hyper plane which categorizes new examples. SVM works on principle which is based on some previous training inputs, using supervised learning techniques to classify data. SVM classifier uses machine learning theory to boost higher accuracy and avoid the data to be over-fit automatically. Consider a set of training examples, each belongs to one or the other of two categories and training algorithm in SVM develops a model that allots new examples to one category or the other, by making it a non-probabilistic binary linear classifier. SVM model represents examples of the different categories that are separated by clear gaps as possible. Later on these examples are represented into the same space and belongs to the same category of the gap they fall. SVM is also used to test for face recognition in general pattern classification, hand writing and regression based applications. Even though having complex design and hierarchy SVM provide good results than neural networks. The main advantage of SVM is that it is easy to scale and train complex high dimensional data as compared to neural networks at the expense kernel function to guide the SVM. The SVM is used as a modeling technique for audio classification.

4. RESULTS AND DISCUSSION

In this section we discuss in detail about the results we obtained using the machine learning techniques employed. We also vary the number of recordings and show its effect on the accuracy.

Table -1, 2 and 3 shows the confusion matrix for 5 bird species with 50, 40 and 30 recordings per each class using SVM. This gives the accuracy of 86%, 85% and 83.33% respectively.

Table -1: SVM based classification with 50 recordings each.

	BULBUL	CROW	CUCKOO	MYNA	SPARROW
BULBUL	8	0	0	6	0
CROW	0	10	0	0	0
CUCKOO	0	1	9	0	0
MYNA	0	0	0	9	0
SPARROW	0	0	0	0	7

Table -2: SVM based classification with 40 recordings each.

	BULBUL	CROW	CUCKOO	MYNA	SPARROW
BULBUL	6	0	0	2	0
CROW	0	5	2	0	0
CUCKOO	0	0	6	0	0
MYNA	0	0	0	8	0
SPARROW	2	0	0	0	9

Table -3: SVM based classification with 30 recordings each.

	BULBUL	CROW	CUCKOO	MYNA	SPARROW
BULBUL	6	0	0	1	0
CROW	1	8	1	1	0
CUCKOO	0	0	4	0	0
MYNA	0	0	0	3	0
SPARROW	2	0	0	0	4

Table -4 shows the confusion matrix for 4 bird species with 40 recordings per each class using GMM. This gives an overall accuracy of 95%.

Table -4: GMM based classification with 40 recordings each.

	CROW	CUCKOO	MYNA	SPARROW
CROW	8	1	0	1
CUCKOO	0	10	0	0
MYNA	0	0	10	0
SPARROW	0	0	0	10

The results show that GMM classifies more accurately and the accuracy increases with increase in the number of recordings.

5. CONCLUSION

In this paper we presented a method of classifying real time recordings of bird sound. We have compared machine learning techniques and the number of recordings for improving the performance of the classification tasks. Performance of the proposed bird identification system still leaves scope for improvement. Based on the results of this study, bird's species may be identified with an accuracy of 95% and 86% using GMM and SVM respectively. For future works, this approach can be extended to large number of bird species. Research can be carried out to investigate the audio recordings in noisy environments. This can also be extended to study other important animal species in understanding complex and sensitive ecosystems.

REFERENCES

[1] Somervuo, P., Harma, A., & Fagerlund, S. (2006). Parametric representations of bird sounds for automatic species recognition. IEEE Transactions on Audio, Speech and Language Processing, 14, 2252–2263.

[2] Juang, C., & Chen, T. (2007). Birdsong recognition using prediction based recurrent neural fuzzy networks. Neuro computing, 71, 121–130.

[3] Mellinger, D., & Bradbury, J. W. (2007). Acoustic measurement of marine mammal sounds in noisy environments, In Proceedings of the International Conference on Underwater Acoustical Measurements: Technologies and Results.

[4] Chen, Z., & Maher, R. C. (2006). Semi-automatic classification of bird vocalizations using spectral peak tracks. The Journal of the Acoustical Society of America, 120, 2974–2984.

[5] Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, In Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing (Vol. 28, pp. 357–366).

[6] Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X. Z., Raich, R., Hadley, S., et al. (2012). Classification of multiple bird species. Journal of Acoustic Society of America, 131, 4640–4650.

[7] Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K., & Frommolt, K. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. Pattern Recognition Letters, 31, 1524–1534.

[8] Fagerlund, S. (2007). Bird species recognition using support vector machines. Journal on Advances in Signal Processing, 7, 64–71.

[9] Lopes, M. T., Koerich, A. L., Kaestner, C. A. A., Silla, Jr., C. N. (2011). Feature set comparison for automatic bird species identification, In IEEE International Conference on Systems, Man, and Cybernetics, Anchorage, Alaska.

[10] Kwan, C., Mei, G., Zhao, X., Ren, Z., Xu, R., Stanford, V., Rochet, C., Aube, J., & Ho, K. (2004). Bird classification algorithms: Theory and experimental results, In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04) (vol. 5, pp. 289–292), Montreal, Canada.