A Fast Baseline System for Large Scale Bird Identification

Ivan Meza[1], Adrian Espino-Gamez[2],
Frine Solano[2], and Esau Villarreal[1]

[1]Instituto de Investigaciones en Matematicas Aplicadas y en Sistemas (IIMAS)
[2]Facultad de Ingeniería (FI)
Universidad Nacional Autonoma de Mexico (UNAM)
ivanvladimir,adrian,frine,esau@turing.iimas.unam.mx

Abstract We present a description of our approach for the “Bird task Identification LifeCLEF 2015”. Our approach consists of a baseline system based on the classification of Mel-bands representations of bird singing using a random forest classification. This setting proved to be fast during testing, extraction of Mel-bands and classification was done in a couple of hours. Our best system reached a Mean Average Precision of 14.5% and Recall of 14.5%.

1 Introduction

In this work we present the description of our system submitted to the *LifeCLEF 2015 Bird task* [4] part of the *LifeCLEF 2015 Laboratory* [5]. This task is concerned with the identification of bird species based on their signing. This setting has potential applications on ecological surveillance or biodiversity conservation. This year the task was formally defined as:

> The task will be focused on bird identification based on different types of audio records over 999 species from South America centered on Brazil. Additional information includes contextual meta-data (author, date, locality name, comment, quality rates). The main originality of this data is that it was built through a citizen sciences initiative conducted by Xeno-canto, an international social network of amateur and expert ornithologists. This makes the task closer to the conditions of a real-world application: (i) audio records of the same species are coming from distinct birds living in distinct areas (ii) audio records by different users that might not used the same combination of microphones and portable recorders (iii) audio records are taken at different periods in the year and different hours of a day involving different background noise (other bird species, insect chirping, etc). [1]

At the core of our submission this year there was the goal to simplify our processing pipeline compared with our last year submission [7]. For this reason we re-write

[1] From http://www.imageclef.org/lifeclef/2015/bird (June, 2015)

our base code and our final pipeline consisted of: extracting the Mel-bands we discarded further extraction of characteristics such as MFCCs (also provided in previous challenges [**?**]). The extracted Mel-bands are reduced into a vector by extracting statistics from these and create a classifier using the resulting vectors. This pipeline would be our baseline for further improvement in our system. At this point our approach work only with audio information.

The outline of this paper is as follows. Section 2 presents the architecture of our approach.Subsection 2.1 explains the filtering stage, subsection 2.2 the extraction of Mel bands filters, subsection 2.3 the conversion of the Mel filters to vectors, subsection 2.4 present the random forest classification. Section 3 presents our results. Finally, section 4 presents some conclusions and discusses about future work.

2 Architecture of the approach

Our approach is composed of following stages:

2.1 Filtering

The original recordings were filtered using a high pass filter with a cutoff frequency of $1K$ in order to remove background noise. This cut-off frequency was empirically defined from analysing some of the bird recordings spectrograms from the training set.

2.2 Extraction of Mel bands

From the filtering recording we extract the Mel bands [3]. We extract 80 bands with a *frame size* of 1024 frames (i.e., $23ms$) and a *hop* of 512 frames (i.e., $12ms$). This corresponds to 86 frames per second for the Mel bands. For the extraction of Mel bands we limit the highest frequency to $16K$ since we notice the content of the bird singing rarely reached higher frequencies than $12k$. Finally the resulting bands were normalized by the highest energy in the whole recording.

Before extracting the band the signal was pre-emphasized by a factor of 0.95 and values smaller to 1×10^{-100} were zeroed. This configuration is typical from speech processing. We did not perform parameter optimization on the extracted Mel bands, we rather focus on the machine learning aspect of our approach.

To process the whole training setting took approximately $1hr30min$.

2.3 Conversion to vectors

The extracted Mel bands per recording are transformed into vectors by simple statistics per band. In particular, our final submission used: *mean*, *standard deviation*,*median*, *skewness*, which empirically showed to produce a fast enough system to process the whole corpus. This gave us a dimensionality of 320 dimension for the classifier to deal with.

To process the whole training of Mel-band took approximately $15min$

2.4 Classification

With a vector per recording we created a classifier using the predominant species from the training data as goal class. We focus on using the Random Forest methodology as our classifier [2]. This decision was taken from our experience talking with participants of last year challenge [6,9]. Most of our development was focus on tunning the parameters of the random forest implementation. In our development experiments using the training set we found that the performance was highly improve it by using a large amount of estimators for the random forest however this made it to take large amount of time and it did not warranty if would finish the labelling of the test data given our memory resources. The submitted runs correspond to two random forest models with 100 and 120 estimators. For our output we chose the five most probable classes from the random classification stage.

To train a model using the whole training took approximately $1hr15mn$. The whole architecture allow us to run the process in a matter of hours.

2.5 Resources

For the processing of the audio recording we used the *Essentia* library [1] and the Random Forest implementation available in the *scikit-learn* library [8]. Our code has been released under an open source license[2].

3 Experimental Results

We submitted two configurations of our system:

RF 100 Random forest using 100 estimators
RF 120 Random forest using 120 estimators

Table 1 shows the final performance in the testing set of LifeClef 2015. Additionally, we show precision, recall and f1-score metrics from our development test in Table 2. For this experiments we randomly separate the training set into two sets (80% and 20%). From these experiments we found that there 657 species that the classifier was not able to classify at all.

Table 1. Mean average precision of identification of bird species in testing.

	Without background species	background species
RF 120 (GOLEM Run 1)	17.1%	14.9%
RF 100 (GOLEM Run 2)	16.1%	13.9%

Additionally, from our development test we were able to identify species that work quite well for our system (F1-score = 1.0):

[2] `https://github.com/ivanvladimir/sonidero/tree/v0.0.1/examples/birds`

Table 2. Precision, recall and F-score for classification on development set.

	Precision	Recall	F-score
RF 120	14.5%	14.5%	12.4%
RF 100	14.4%	14.7%	12.3

- *Psarocolius viridis* (2)
- *Myiothlypis cinereicollis* (4)
- *Coccyzus euleri* (2)

However, we only identify 45 species with a score larger or equal than 0.50.

4 Conclusions and Future work

These working notes present our system proposal for the identification of bird species through singing. This proposal was built in the context of the *LifeCLEF 2015 Bird task* [4], a part of the *LifeCLEF 2015 Laboratory*[5]. This current approach is a reworking of our system from previous year. Although in its actual state our approach only corresponds to a baseline for our future work it actually means an improvement in the MAP of 4.2% with background species and 4.4% without background species from our previous approach. This taking into consideration that this year task was harder than previous years.

References

1. ESSENTIA: an Audio Analysis Library for Music Information Retrieval (2013)
2. Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)
3. Ganchev, T., Fakotakis, N., Kokkinakis, G.: Comparative evaluation of various mfcc implementations on the speaker verification task. p. 191âĂŞ194 (2005)
4. Goëau, H., Glotin, H., Vellinga, W.P., Rauber, A.: Lifeclef bird identification task 2015. In: CLEF working notes 2015 (2015)
5. Joly, A., Müller, H., Goëau, H., Glotin, H., Spampinato, C., Rauber, A., Bonnet, P., Vellinga, W.P., Fisher, B.: Lifeclef 2015: multimedia life species identification challenges. In: Proceedings of CLEF 2015 (2015)
6. Lasseck, M.: Large-scale identification of birds in audio recordings. In: Working Notes for CLEF 2014 Conference. pp. 643–653 (2014)
7. Martinez, R., Silva, L., Villarreal, T., Fuentes, G., Meza, I.: Svm candidates and sparse representation for bird identification. In: Working Notes for CLEF 2014 Conference. pp. 662–669 (2014)
8. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830 (2011)
9. Stowell, D., Plumbley, M.: Audio-only bird classification using unsupervised feature learning. In: Working Notes for CLEF 2014 Conference. pp. 673–684 (2014)