LITERATURE REVIEW

BIRD SONG IDENTIFICATION

Thanyaporn Phinthuphan

3 Sep 2018

Outline

- Bird song classification in field recordings (Mario, 2013)
- Bird identification from audio recordings (Rafael, 2013)
- Clusterized MFCC & SVM for bird song identification (Olivier, 2013)

Bird Song Classification in Field Recordings

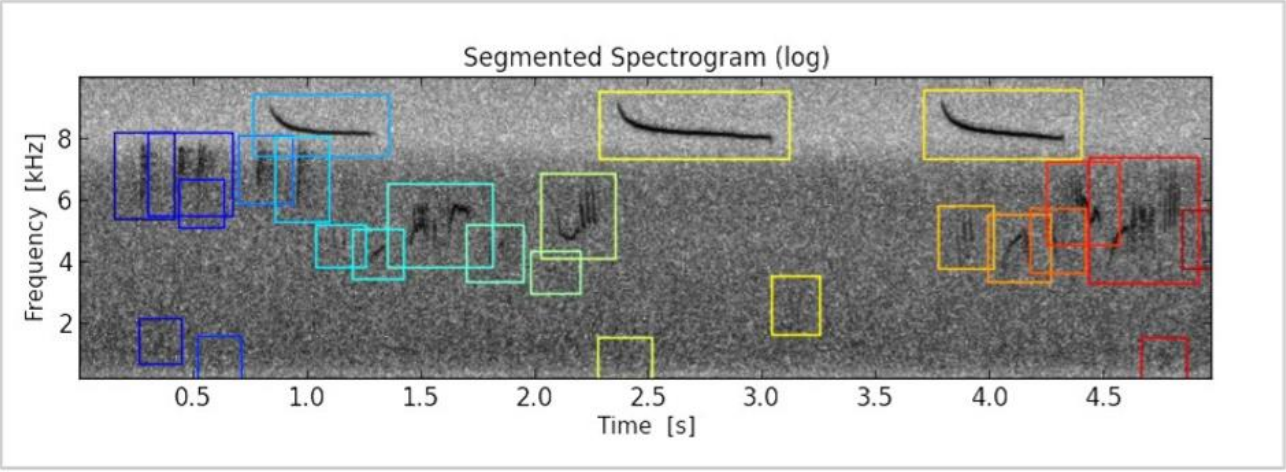
Introduction

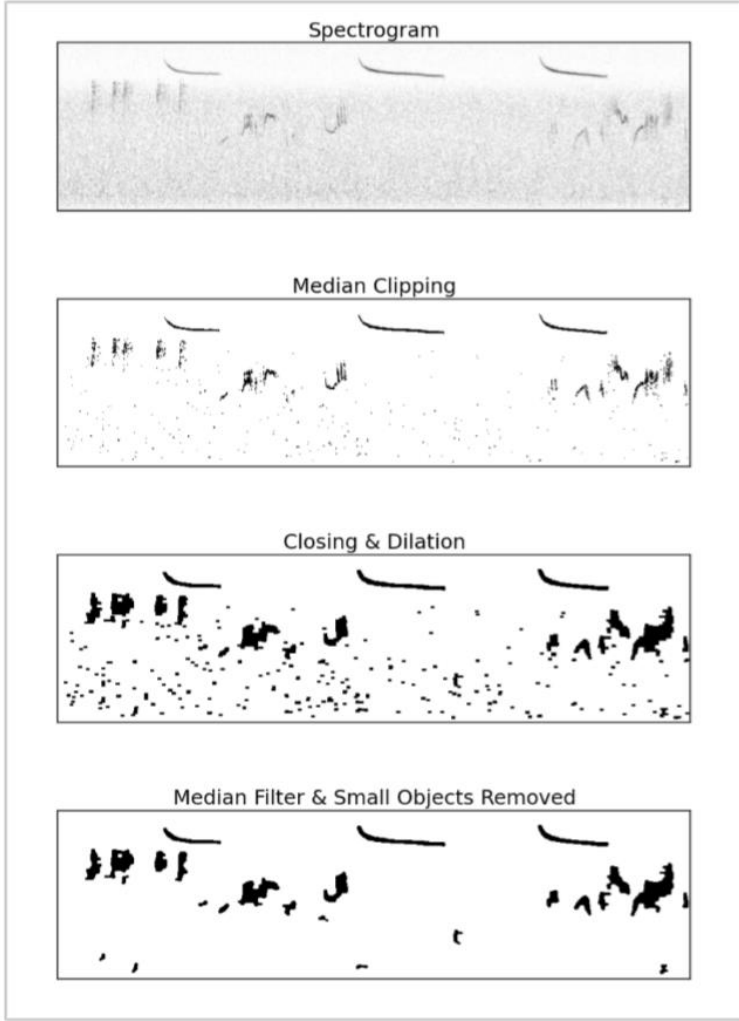
- author: Mario Lasseck
- the winning Solution for NIPS4B 2013 Competition
- starting point is the solution for the MLSP 2013 Competition
- 87 sound classes of birds (call/song)
- 687 audio file (WAV format) in the training set

Bird Song Classification in Field Recordings

Preprocessing and Segmentation

- STFT using hanning window → normalized
- reducing background noise with median clipping
- closing & dilation → segmentation (size/position)





Bird Song Classification in Field Recordings

Feature Extraction

- File-statistics

  max, min, mean, std for all values of spectrogram + 16 divided spectrogram

- Segment-statistics

  count + max, min, mean, std for weight, height, frequency position

- Segment-probabilities

  highest matching all segments using normalized cross-correlation

- 68 + 13 + 9,198 (number of segments in training) features per file

Bird Song Classification in Field Recordings

Feature Selection

- multi-label classification problem → 87 individual classification problems
- select seg-prob features from files which include same class only
- selected features for each class ~ 300 – 500 features

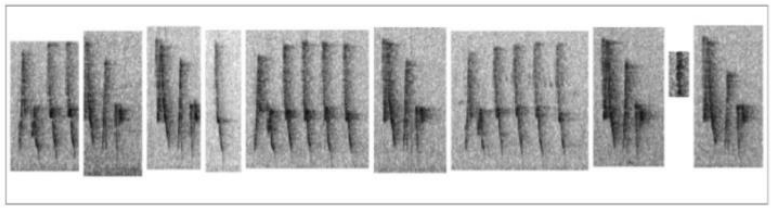
Bird Song Classification in Field Recordings

Classification

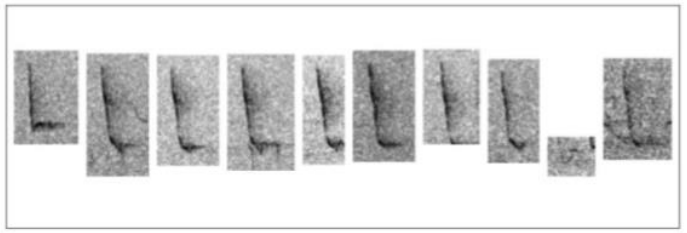
- using random forest (scikit-learn library)
- possible without file/seg-stat features and test recording segmentation
- score of 91.6% AUC
- performance for each class depends on character of importance segments

Bird Song Classification in Field Recordings

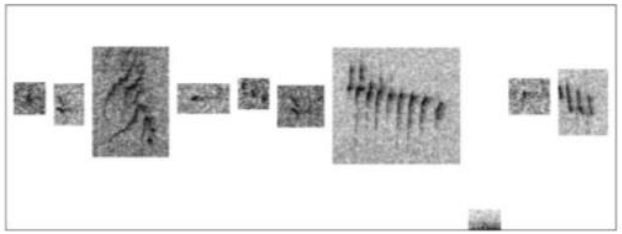
Classification



song of Cettia cetti



song of Phylloscopus collybita



call of Serinus serinus

Bird Song Classification in Field Recordings

Conclusion

- Pros

  can see the important segments for each class →

  good for visualization and manually error checking

- Cons

  too many features → may apply dimension reduction

Bird identification from audio recordings

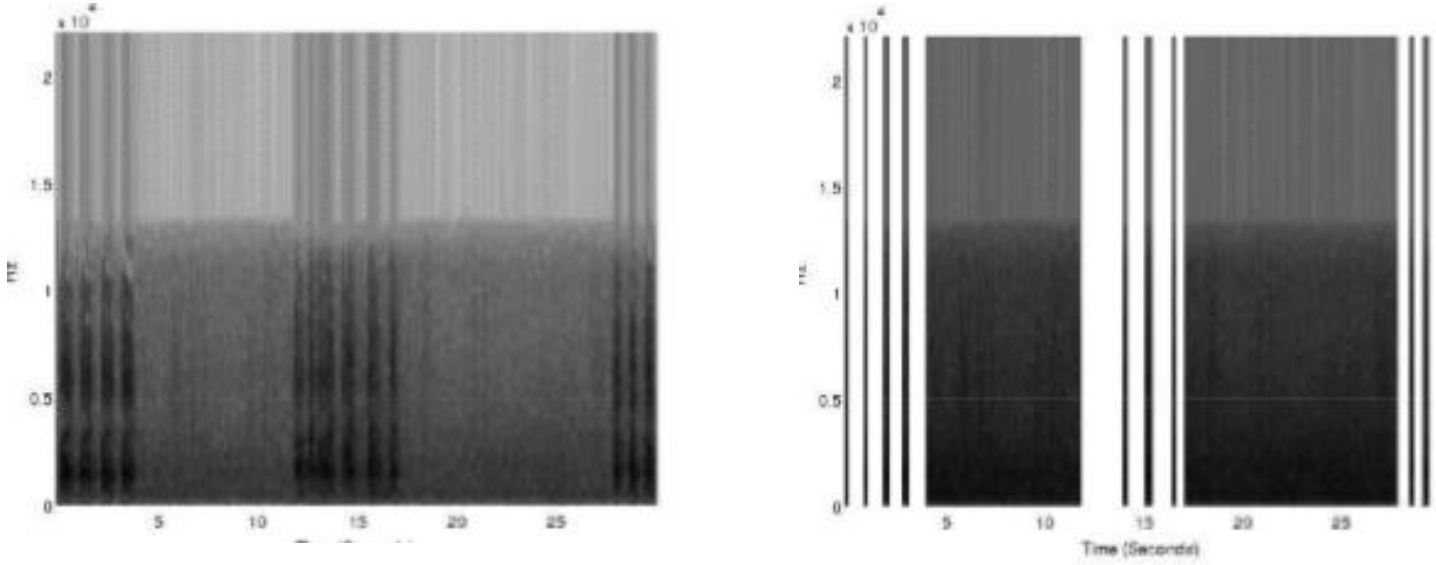
Introduction

- author: Rafael Murcia & Victor Paniagua (Spain)
- rank 1 for the ICML 2013 Bird Challenge
- train data: 35 audio recordings labeled with single species (30 sec)
- test data: 90 audio recordings (150 sec) with possibly none or multiple species

Bird identification from audio recordings

Syllable Segmentation

- signal spectrogram using Kaiser-window
- 10th-order Butterworth band-pass filter
- syllable segmentation algorithm



Bird identification from audio recordings

Feature Extraction & Dimensionality Reduction

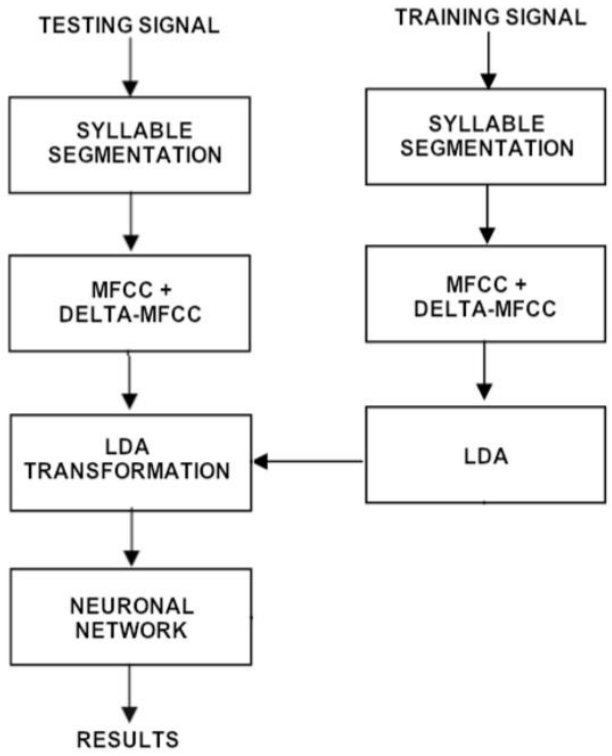
- features using the MFCCs & Delta-MFCCs
- group variables of adjacent samples into vector using sliding window

  to exploit the temporal relationship between the same class
- LDA for reducing dimension (using the train projection in test set too)

Bird identification from audio recordings

Classification

- 35 binary simultaneous classification problems
- using neural network
- *which bird, if any, sings in instant T?*
- answer using maximum score achieved during time instant
- score of 69.45% AUC



Bird identification from audio recordings

Conclusion

- Pros

  can extract some features that cannot be extracted by human

- Cons

  very hard algorithm

Clusterized MFCC & SVM for bird song identification

Introduction

- author: Oliver Dufour and team
- rank 4 for the ICML 2013 Bird Challenge

Clusterized MFCC & SVM for bird song identification

Preprocessing

- 17 MFCC feature vectors, including energy per frame
- windowing → representative of longer segments
- silence removal using clustering by average energy of frames
- 6 math features for17-MFCC → 102 features

$$f_1 = \frac{\sum_{i=1}^{n}(|v_i|)}{n}$$

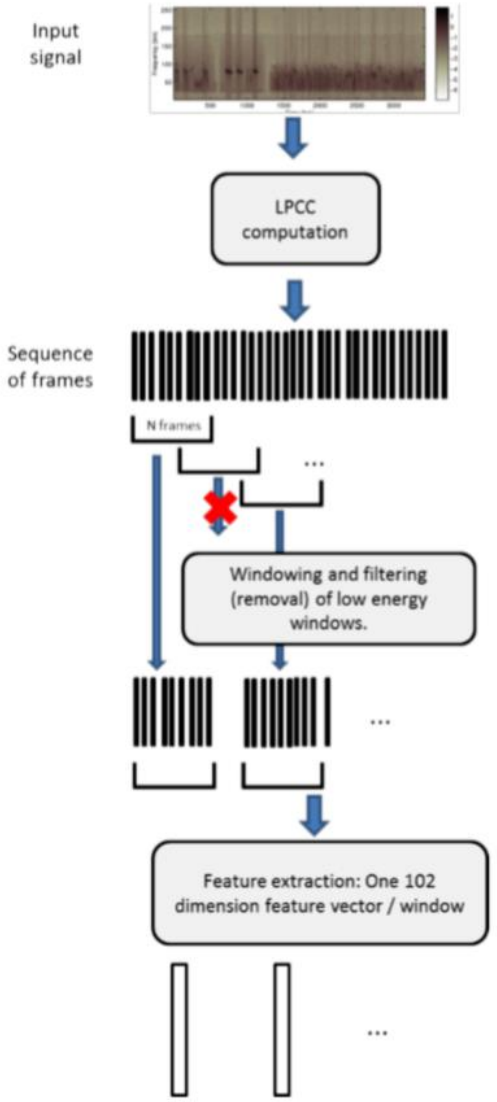
$$f_2 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(v_i - \bar{v_i})^2}$$

$$f_3 = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(d_i - \bar{d_i})^2}$$

$$f_4 = \sqrt{\frac{1}{n-3}\sum_{i=1}^{n}(D_i - \bar{D_i})^2}$$

$$f_5 = \frac{\sum_{i=1}^{n-1}|d_i|}{n-1}$$

$$f_6 = \frac{\sum_{i=1}^{n-2}|D_i|}{n-2}$$



Clusterized MFCC & SVM for bird song identification

Classification

- clustering to split call and sound for each species
- classification problem with 2K classes (K species)
- SVM in a one-versus-all fashion
- score of 64.64% AUC

Clusterized MFCC & SVM for bird song identification

Conclusion

- Pros

  unsupervised learning (clustering) to handle noise

- Cons

  too many step to implement

Q&A