

---

# Identification of Songbird Species in Field Recordings

---

**Hsiao-Yu Tung**

htung@andrew.cmu.edu  
htung

**De-An Huang**

deanh@andrew.cmu.edu  
deanh

**Xiao-Feng Xie**

xfxie@cs.cmu.edu  
xfxie

**Yurui Zhou**

yuruiz@andrew.cmu.edu  
yuruiz

**Joseph Russino**

jrussino@rec.ri.cmu.edu  
jrussino

## Abstract

### I. Introduction

Basically preprocessing and segmentation is the very first step of the bird song classification, and the methods used in preprocessing and segmentation have a big influence on classification results.

### II. Relate Work

The preprocessing and segmentation work is actually based on the method adopted in [1]. They firstly processed the audio files into grayscale image and then using standard processing techniques such as erosion, propagation and median filter. With processed image they labeled connected pixels exceeding certain spatial threshold as a segment and add on it a small area rectangle to define size and position.

### III. Methods

In preprocessing we first re-sampled the audio files to 22050 Hz guarantee a uniform audio sample format for all training and test data. For each audio file we split it into 256 segments. The the Power Spectral Density of each segments can be calculated. For each segments we use a Hanning windows with %75 overlap. Finally, the spectrogram of each audio file is treated as grayscale image for further noise reduction and segmentation.

Notice the case that in a processed grayscale image most area was occupied by the random noise. What we want is to get rid of the background noise completely and increase the contrast between real signal and the background. Given the several different algorithm tested, the best one is the so called Median clipping algorithm. The idea is simple, for each pixel we compute the median and variance of its corresponding row and column. If the pixel is above the median plus three times variance we set the pixel value to 1, otherwise the value would be set to 0. The median clipping algorithm works best because it not only removes most background noise, but also capture the sound feature clearly and precisely.

Such a algorithm, though perform well in noise removing and feature capturing, requires large amount of computation because the average and variance of each column of the row need to be calculated. Given the huge size of the processed grayscale image, we need a more efficient

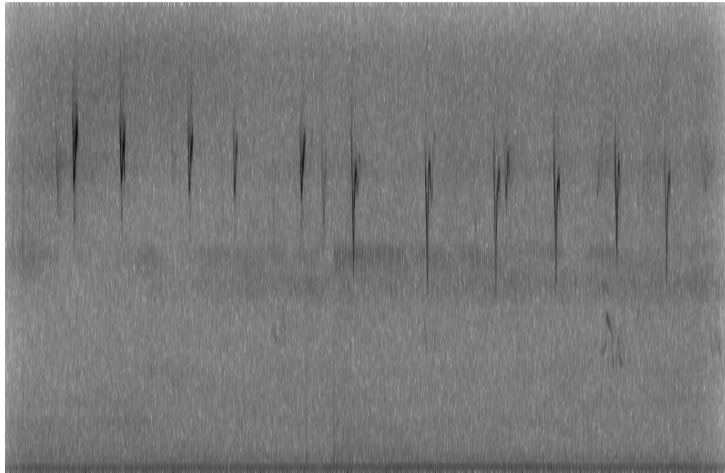


Figure 1: Spectrogram

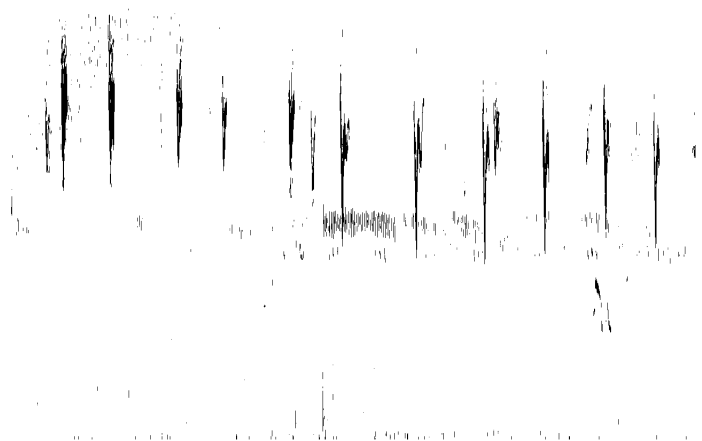


Figure 2: Median Clipped

algorithm to segment the image. Thus further experiments is needed to explore an balance point between noise reduction effectivity and efficiency.

Similar to the methods used in [1], we apply standard image processing techniques to further reduce the residual noise dot. And Finally we would use find connected pixels from the image and label the segments.

## IV. Experiments

### References

- [1] M. Lasseck. Bird song classification in field recordings: Winning solution for NIPS4B 2013 competition. In *Workshop on Neural Information Processing Scaled for Bioacoustics*, pages 176–181, 2013.

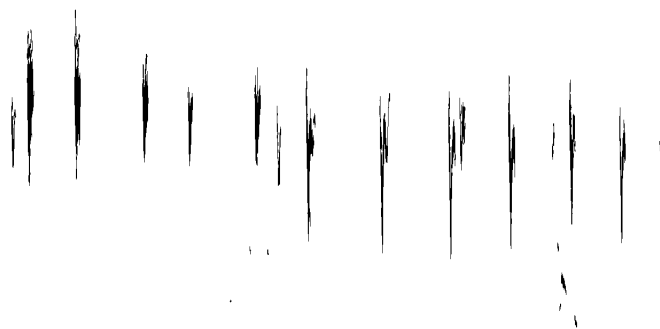


Figure 3: Eroded and Propagated

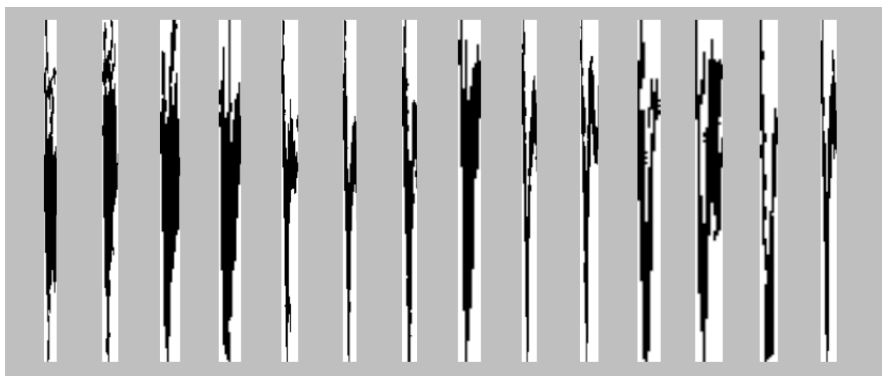


Figure 4: Segmentation