International Journal of Engineering & Technology, 7(4.16) (2018) 174-176

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Classifying Anuran Call Spectrograms with Correlation Filters

Salina Abdul Samad*, Aqilah Baseri Huddin

Centre for Integrated Systems Engineering and Advanced Technologies (INTEGRA), Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia
**Corresponding author E-mail: salinasamad@ukm.edu.my*

Abstract

A method to classify anurans based on the spectrogram representations of their call vocalizations is presented. The spectrogram representations have distinctive patterns that may be used to differentiate between species. As such, they can be treated as input images to advanced correlation filters frequency employed in human biometrics applications. A type of correlation filter that has been successfully implemented for face and fingerprint biometrics is considered here. In order to obtain clear spectrograms with distinguishing features, careful selection of spectrogram parameters is performed. To demonstrate this approach, two species of anurans commonly found in Southeast Asia are classified showing that the accuracy rate is dependent on the number of call-prints used to construct the correlation filter templates.

Keywords: *Anuran Classification; Correlation Filter; Spectrogram.*

1. Introduction

Anuran species, i.e., frogs and toads, are routinely identified by their vocalization. This is usually done manually by herpetologists from field recordings. This endeavor has wide ranging applications in the field of conservation of natural habitats. For example, the anuran population is used as one of the bio-indicators in assessing the health of habitats such as wetlands and floodplains [1].

Herpetologists are able to distinguish between species of anurans because most species have their own unique sound. This has sparked interest in using automation to identify and classify anurans. This area of bioacoustics signal analysis has been mostly concentrated on using techniques similar to those used for processing speech signals. Signal segmentation of anuran sounds is routinely performed in order to isolate syllables [2]. The next step usually involved extracting the relevant features such as the popular Mel Frequency Cepstrum Coefficients (MFCC) [3], or other simpler features based on standard call variables such as duration, maximum power, maximum and minimum frequencies [4]. Similar to speech processing, once the features are identified, they are used as inputs to classifiers such as Support Vector Machines [2],[5], Nearest Neighbors [2],[6], Neural Networks [7] and many others [8]-[9].

Another approach to process anuran calls are based on spectrograms. Spectrograms are visual representations of audio signals obtained using short-time Fourier transform (STFT). Image processing techniques applied to spectrograms have been used to automatically analyzed animal calls [10]. Researchers in anuran calls have extracted features from the spectrograms such as local peaks [11] and ridges [12]. These features, or their derivatives, are then used as inputs to classifiers similar to those described above for one-dimensional bioacoustics signal processing.

This paper presents a different approach to anuran call classification. It investigates the possibility of representing anuran call spectrograms as images to a classification technique based on advanced correlation filters called the Unconstrained Minimum Average Correlation Energy (UMACE) filters. These correlation filters and their variants have been used to successfully classify biometrics images such as faces and fingerprints [13]. In order to classify anuran spectrograms, processing steps are performed to center the call in an image frame and select spectrogram parameters that highlight the salient features of the calls. A template based on repetitive spectrogram calls are constructed to represent a species which is cross-correlated with test spectrogram calls to determine the classification accuracy rate based on correlation plane parameters.

2. Constructing Spectrograms and Templates

It is important to obtain a good spectrogram representation of an anuran call. This is dependent on the parameters used for the spectrogram construction which in turn is dependent on the dataset. In this case, the recordings of the two species of anurans sampled at 44.1 kHz, are segmented into individual calls of 800-ms length using a sound editing tool.

Each segment is then filtered with a high pass filter with a cut-off frequency of 250 Hz in order to eliminate the environmental noise. A centering of the peak amplitude at 400 ms is performed before applying the short-time Fourier Transform in order to obtain the spectrogram.

Framing are performed on the calls with a chosen frame length of 256 with a 75 percent overlap with the parameters obtained after several trials where the objective is to obtain a visually clear call-prints. Windowing is performed using the Gaussian window chosen due to its superior performance in eliminating energy leakage, although the computation is more intensive compared to other windows. The window is described by

$$w(n) = e^{-\frac{1}{2}\left(\frac{2.5n}{N/2}\right)^2}, \; 0 \le |n| \le \frac{N}{2} \quad (1)$$

Each windowed frame is then transformed from the time domain signal into the frequency-domain signal by Fast Fourier Transform. The spectrogram is converted to greyscale and a simple threshold is applied to the spectrogram to obtain a cleaner image by eliminating high-frequency noise. The image is then cropped to a size of 512x512. Figure 1 shows an example of an anuran spectrogram.

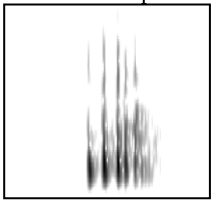

Fig 1: Anuran spectrogram example

A UMACE filter which acts as the template are synthesized in the Fourier domain using a closed form solution [13]. Several training images are used to synthesize a filter template. In this case, the training images are the anuran call spectrograms with one template representing each species.
Each image is of size 512x512 and the template is constructed from multiple calls of the training set. A separate testing set is used for the cross-correlation process with the template filter in order to determine whether the test image is from the true or false class. In this process, the filter optimizes a criterion to produce a desired correlation output plane by minimizing the average correlation energy and at the same time maximizing the correlation output at the origin. The optimization of the UMACE filter equation can be summarized as

$$\boldsymbol{U_{mace}} = \boldsymbol{D^{-1} m} \tag{2}$$

where $\boldsymbol{D}$ is a diagonal matrix with the average power spectrum of the training images placed along the diagonal elements, while $\boldsymbol{m}$ is a column vector containing the mean of the Fourier transforms of the training images.

3. Classifying Anuran Calls

To demonstrate the viability of using the proposed technique to classify calls, two species of anurans commonly found in Southeast Asia are considered. Recordings were obtained for common grass frogs (*F. limnocharis Boie*) and mangrove frogs (*F. cancrivora Gravenhorst*) and the calls were subsequently processed to obtain the spectrograms. For each species, 30 spectrograms were obtained. The templates for each class were constructed using 5, 10 and 15 call-prints. These templates were used to cross-correlate with a test spectrogram.
The template matching process for a correlation filter template T(u,v) and the input test image S(x,y) is given by

$$c(x,y) = IFFT\{FFT\big(S(x,y)\big) * T^{+}(u,v)\} \tag{3}$$

where the test image is first transformed to frequency domain and reshaped to be in the form of a vector. It is then convolved with the conjugate of the UMACE filter which is equivalent to cross correlating it with the UMACE filter. The output is transformed again to the spatial domain obtaining the correlation plane.
The resulting correlation plane produces a sharp peak at the origin while the values everywhere else are close to zero if the test image belongs to the same class as the designed filter. A simple metric called the Peak-to-Sidelobe ratio (PSR) is used to measure the sharpness of the peak where

$$PSR = \frac{peak - mean}{standard\ deviation} \tag{4}$$

where the peak is the largest value obtained from the correlation output. The mean and standard deviation are calculated from a 20x20 sidelobe region excluding a 5x5 central mask [13].
To classify the anurans into the correct classes, threshold values for each class are determined from the PSR values obtained with the cross-correlation process using the training set. Then, with the testing set, if an image has a PSR value that is greater than the threshold for the tested class, it is classified as its true class, otherwise it is classified as false. The accuracy rate is calculated, defined as the ratio of correct classification to total number of test inputs. The results are tabulated in Table 1 for different number of spectrograms per template for both species.

Table 1: Accuracy rates in percent with respect to number of spectrograms per template

Species	**5 spectrograms**	**10 spectrograms**	**15 spectrograms**
F. limnocharis Boie	10.6	23.1	63.8
F. cancrivora Gravenhorst	9.8	19.7	57.9
Average	10.2	21.4	60.9

The results demonstrate that as the number of call-prints that make-up the template increases the accuracy rate also increases. Using 5 spectrograms per template only gives an average accuracy rate of 10.2 % while using 10 spectrograms increases the accuracy rate to 21.4 %. Using 15 spectrograms per template gives the highest average accuracy rate of 60.9 %. Due to the limited dataset, this is a reasonable result and has demonstrated that it is possible to use spectrograms with correlation filters in order to distinguish between anuran species. With a larger dataset and more calls representing a species, the accuracy rate may be further increased. Another possible method to increase the accuracy rate is to utilize other types of correlation filters.

4. Conclusion

Anuran call spectrograms have been shown to be viable image inputs to correlation filters for classification. The spectrograms were obtained by selecting the spectrogram parameters in order to produce clear and distinguishing call-prints. The UMACE filter templates were constructed from multiple spectrograms to represent the species. The results of classifying two anuran species showed that the accuracy rate increases as the number of call-prints that comprises the template increases. Although the highest accuracy rate obtained is 60.9 percent, the results have demonstrated the viability of using spectrograms to distinguish between anuran species using correlation filters and point to the possibility of using other types of correlation filters in order to obtain better classification accuracy.

Acknowledgement

This research is supported by the Malaysian Ministry of Higher Education grant: FRGS/1/2016/TK04/UKM/01/1.

References

[1] Gibbs JP, Rouhani S, & Shams L (2017), Frog and toad habitat occupancy across a Polychlorinated Biphenyl (PCB) contamination gradient, *Journal of Herpetology*, 51(2), 209-214.

[2] Huang CJ, Yang YJ, Yang DX & Chen YJ (2009), Frog classification using machine learning techniques, *Expert Systems with Applications*, 36(2), 3737-3743.

[3] Gingras B. & Fitch, WT (2013), A three-parameter model for classifying anurans into four genera based on advertisement calls, *The Journal of the Acoustical Society of America*, 133(1), 547-559.

[4] Acevedo MA, Corrada-Bravo CJ, Corrada-Bravo H, Villanueva-Rivera LJ & Aide TM (2009), Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4(4), 206-214,

[5] Noda JJ, Travieso CM & Sánchez-Rodríguez D (2016), Methodology for automatic bioacoustic classification of anurans based on feature fusion. *Expert Systems with Applications*, 50, 100-106.
[6] Vaca-Castaño G. & Rodriguez D, Using syllabic Mel Cepstrum features and k-nearest neighbors to identify anurans and bird species, *Proceedings of IEEE Workshop on Signal Processing Systems (SIPS)*, (2010), 466-471.
[7] Colonna J., Peet T., Ferreira CA, Jorge AM, Gomes EF & Gama, J, Automatic classification of anuran sounds using convolutional neural networks, *Proceedings of the Ninth International C* Conference on Computer Science & Software Engineering*, (2016), 73-78.
[8] Gingras B. & Fitch WT (2013), A three-parameter model for classifying anurans into four genera based on advertisement calls. *The Journal of the Acoustical Society of America*, 133(1), 547-559.
[9] Chen WP, Chen SS, Lin CC, Chen YZ & Lin WC (2012), Automatic recognition of frog calls using a multi-stage average spectrum. *Computers & Mathematics with Applications*, 64(5),1270-1281.
[10] Dennis J, Tran HD & Li H (2011), Spectrogram image feature for sound event classification in mismatched conditions. *IEEE Signal Processing Letters*, 18(2), 130-133.
[11] Grigg G, Taylor, A, Mc Callum, H. & Watson G, Monitoring frog communities: An application of machine learning, *Proceedings of Eighth Innovative Applications of Artificial Intelligence Conference, Portland Oregon*, (1996), 1564-1569.
[12] Xie J, Towsey M., Zhang J., Dong X. and Roe P, Application of image processing techniques for frog call classification. *Proceedings of IEEE International Conference on Image Processing (ICIP)*, (2015), 4190-4194.
[13] Wang Q, Alfalou A. & Brosseau C (2017), New perspectives in face correlation research: A tutorial. *Advances in Optics and Photonics*, *9*(1), 1-78.