© (2014) Trans Tech Publications, Switzerland
doi:10.4028/www.scientific.net/AMM.618.362
Accepted: 2014-06-17
Online: 2014-08-18

A Method of Insect Recognition Based on Spectrogram

Qi Jia[1, a], Rui Min[2,b], Shao-guang Huang[3,c] and Ying Wei[4,d *]

[1]Shanda South Road NO.27, Jinan, Shandong province, China

[2]Shanda South Road NO.27, Jinan, Shandong province, China

[3]Shanda South Road NO.27, Jinan, Shandong province, China

[4]Shanda South Road NO.27, Jinan, Shandong province, China

* Corresponding author

[a]jiaqixdu@163.com, [b]minrui19920920@163.com, [c] shaog1314@163.com, [d] eleweiy@sdu.edu.cn

Keywords: Spectrogram, Voice activity detection, Insect recognition.

Abstract. A novel approach to insect recognition is presented in this paper. The difference between the proposed method with traditional methods is that it starts from the perspective of image and combines voice processing algorithms with image processing algorithms. The classification is done based on voice activity detection (VAD) and spectrogram. We show, by means of example that this approach can recognize different insects correctly. However, despite the potential of correct recognition, further justification of the reliability of the method need to be provided by a larger scale of experiments. Hence, some improvements will be proposed latterly.

Introduction

With the development of modern agriculture, information and intelligence contributes to agriculture greatly. In traditional agriculture, it is difficult to recognize the type of the insects in time. The common way is to spray pesticides to kill possible existing insects. One problem of this way is that it increases the economic cost. The other problem is that if there is no occurrence of insects there will be unnecessary pesticide residues which is not good for our health and breaks the ecological balance. Therefore it is attractive to find a way to recognize the occurrence and the types of the insects fast and exactly.

Traditional recognizing methods are mostly based on voice features. In [1], Mel frequency cepstrum coefficient (MFCC) was used to extract the features. Mel-frequency cepstral coefficients (MFCCs) are the coefficients that collectively represent the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. In [2], dynamic time warping (DTW) was used to do feature matching. In[3], the method combining principal component analysis with one-hidden layer time-delay neural network were proposed and the time signature of insect sounds was used as the recognition feature. Its recognition mainly focused on three kinds of insects and the recognition rate was nearly to 96%. Chesmore et al.[4] investigated techniques for automatically identifying Orthoptera (grasshoppers and crickets) with time domain signal processing and artificial neural networks. 25 species of British Orthoptera have been selected as a test set and preliminary results indicate very high classification rates.The spectrogram is an important finding in voice recognition domain. It is a kind of 3D spectrum where horizontal coordinate and vertical coordinate represent time and frequency respectively and the space value reflects the energy at the corresponding time and frequency. Because of the use of 2D space to express 3D information, the energy value is represented by a color. This means that the deep color represents strong energy of the point. The different color stripes denote the holdover time and the energy distribution of the voice [5, 6]. Little study has been done using information provided by spectrogram to recognize insects. In this paper, firstly we gained the spectrogram of each insect voice. Then we use our own approach to process the spectrogram data and obtain some projection vectors. Finally, we come to the result and analyze it. At the same time, we give a friendly user interface.

This paper is organized as follows. In section 2, the VAD and spectrogram projection is introduced. The method of gaining relative intact voice and recognizing the insects based on image projection is proposed. In section 3 the results of the projection applied to the real insect voice spectrogram is presented. In section 4 the user interface based on Matlab GUIDE and its working processes are described. Conclusions are drawn in section 5.

Methods

Before the insect sound is used for pattern recognition, it is necessary to do some preprocessing such as de-noise and activity detection. In this section, we introduce the voice activity algorithm at first and then present the proposed recognition method.

Voice Activity Detection

The VAD technique is able to find the wanted voice section from the whole voice appropriately, which means, it is easy and powerful to detect the endpoint in order to cut the valid voice rather than noise. The so-called endpoint detection is to locate the starting point and ending point in a given voice segment. Here we use the method that combines the short-term energy (STE) and zero-crossing rate (ZCR).

Short-Term Energy

Because of the short term property of speech signal, it is necessary to frame the digitized voice. It should be noted that the spectral characteristic and other physical characteristic parameters should not be changed in a frame.

The short term energy E can be calculated by Eq.1

$$E = \sum_{n=1}^{N} x^2(n) \quad . \tag{1}$$

In Eq.1, N is the length of a frame and x(n) is the signal.

The detection process is succinct. At the first stage, it compares the short term energy with the given energy threshold G1. If STE of a frame is larger than G1, we think it is a voice frame. If it reaches threshold G2 in continuous given frames, it is considered as the voice start point. Otherwise we continue to process the following frames. After confirming the start point, STE is compared with the given threshold G3 and G4. Here we adopted two thresholds to avoid misjudging on the end point when there is a short pause in voice. If the STE is larger than G3 and less than G4 in continuous given frames, the current frame is considered as voice endpoint. Otherwise we will keep processing the following frames.

Zero-Crossing Rate

In a voice frame, the number of signal waves crossing zero level is defined as Zero-Crossing Rate (ZCR). The ZCR of $x(n)$ is

$$ZCR = \sum_{n=1}^{N-1} \left| x(n) - x(n+1) \right| \quad . \tag{2}$$

To avoid silent period being too large, it is generally needed to set a threshold and make a rule that the ZCR pluses one only when the adjacent samples have opposite sign and their difference is larger than the threshold. After the start point is located by STE algorithm, the ZCR values of a certain previous frame are calculated and compared with the given threshold. If ZCR values of the following frames are all larger than the threshold, the first frame is considered as the voice start point. On account of errors produced by using STE alone, it is necessary to combine STE and ZCR together to find start point and endpoint effectively and accurately.

The Spectrogram and the Proposed Recognition Method Using Projection

The voice signal is a typical non-stationary signal and its non-stationarity comes from physical movements of vocal organ [7]. But the speed is slow compared with acoustic vibration. Hence it can be considered that voice signal is stationary in a short period of time like 10~30ms. Voice spectrogram is a 3D spectrum and shows voice spectrum variation over time. Its color depth denotes the energy intensity for an arbitrary frequency. For a given voice signal, it is convenient to gain its spectrogram by using the method of framing, windowing and FFT transform. The following Fig.1 and Fig.2 show the specific process and the spectrogram, respectively. In these figures, horizontal axis denotes time and vertical axis denotes frequency and shade stripes represent the large energy. The insect voices are obtained from data base of United States Department of Agriculture [8].

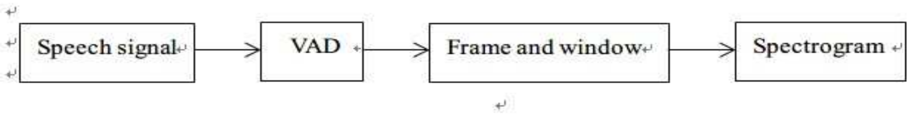

Fig.1. The process for gaining spectrogram

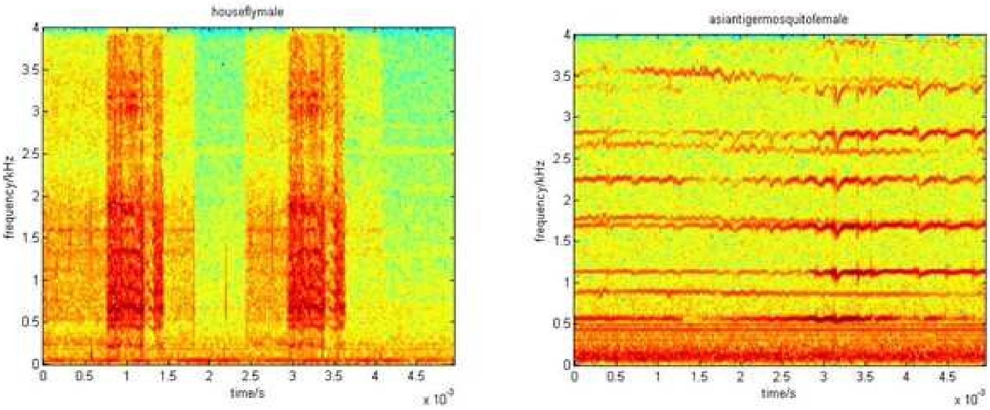

Fig.2. Spectrograms of two different insects,male housefly and female Asian tiger mosquito

The whole process of the proposed recognition method is shown in Fig. 3. After producing the spectrogram, some basic image processing methods like noise reduction, normalizing and graying [9,10] are used. Then horizontal projection of the spectrogram is calculated. For an insect, its voice has large repeatability in a certain period of time. Hence its projection curve in frequency domain just cumulates some values and the curve shape doesn’t change obliviously.

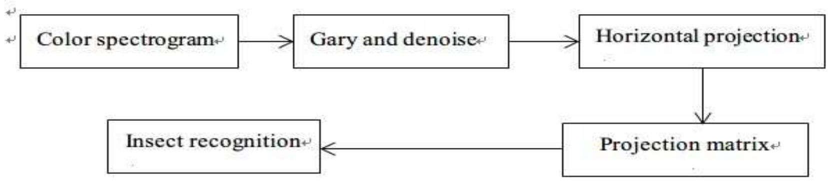

Fig.3. The basic structure of insect voice recognition

Then the values of the projection at different frequencies are arranged in the form of projection matrix. The matrix has many column vectors and each kind of insect has a projection vector. The projection vector of each insect is the template for recognition. Fig.4 shows projection data in frequency domain of two different insects.

Because the stripes in spectrogram image indicate the energy at a certain frequency, the gray image is 2D matrix whose row and column values stand for the voice energy at known frequency and frequency ranges for different insects. It is fast to acquire the horizontal projection curve with various frequencies. If $A_{M\times N}$ is the 2D matrix, then the projection can be calculated as Eq.3,

$$IPL = \sum_{n=1}^{N} A(m,:) \quad . \tag{3}$$

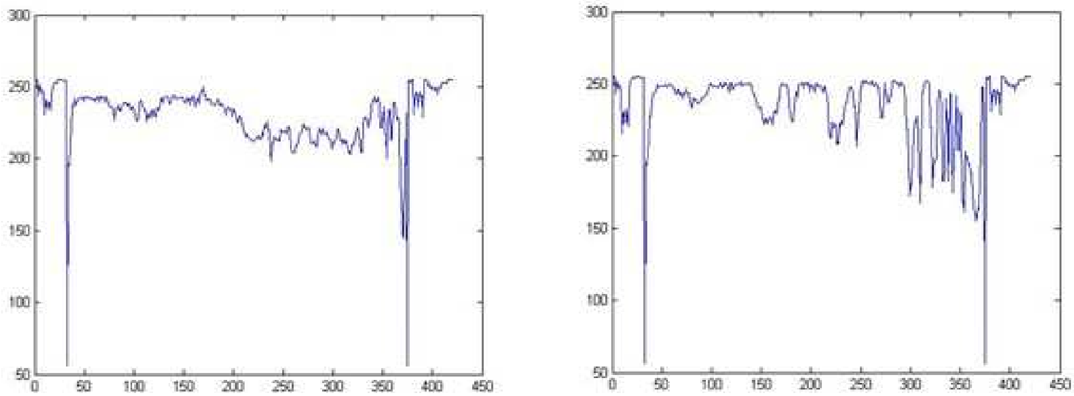

Fig.4. Projections in frequency domain of two different insects, male housefly and female Asian tiger mosquito.

In this experiment, there are ten different kinds of insects. We choose the sound of each insect that has low noise and comparatively complete sound characteristic as a template. After processing, every insect has a projection vector. These vectors are arranged as a projection matrix giving a number for each of them in order. In the test stage, the test voice goes through the same process to obtain its projection vector. The projection vector is used to compare with each vector in the projection matrix. The index of the vector in the projection matrix which is most similar with the obtained projection vector suggests the type of the insect. The similarity between the test vector and the vectors in projection IPL matrix is

$$p = \cos\langle a, b\rangle = \frac{a \bullet b}{|a| \bullet |b|} \ , \tag{4}$$

where a and b stand for the test vector and vectors in IPL matrix. The insect with the biggest value of p is the final result. Five tests have been done and the values of p are shown in Table 1. The insect sound was divided into many valid parts and every p value of each part was computed. The p value in Table 1 is the average of these p values. By checking the recognizing result, it is shown that the recognizing rate is 100% for the five tests. It should be noted that this method cannot judge the insect gender clearly.

Table1. The p values of different tests.

	Asian citrus psyllid	Mosquito	Housefly	Asian tiger mosquito	Onion housefly	result
Test 1 Asian citrus psyllid	0.9563	0.3829	0.5502	0.4062	0.5400	Asian citrus psyllid
Test 2 Mosquito	0.4602	0.9331	0.4901	0.3878	0.7015	Mosquito
Test 3 Housefly	0.5116	0.3681	0.8111	0.2661	0.4937	Housefly
Test 4 Asian tiger mosquito	0.5224	0.4714	0.4395	0.9509	0.6156	Asian tiger mosquito
Test 5 Onion housefly	0.7317	0.6374	0.5917	0.6228	0.9087	Onion housefly

User Interface

In addition to the algorithm, we have made a simple user interface based on Matlab GUIDE. The function of this interface is to make it more convenient and easier for users to recognize insect species. First, the user inputs a piece of audio of an insect into the interface. Second, the program in the system computes and displays the spectrogram of the audio and then gets the projection matrix of it. Meanwhile, the interface can show the spectrogram of the audio to the user. Third, the interface compares this matrix with matrixes for reference which we have computed before and the interface has loaded in advance and then draws a conclusion. Finally, the interface returns the recognizing result to the user and shows the name of the insect in the interface.

Conclusion

A new method for insect recognition based on voice activity detection (VAD) and spectrogram was proposed. The traditional methods mostly just use voice for recognition and the recognition rate is not very high. In this paper, the projection matrix was obtained by projecting the spectrogram in frequency domain. We calculate the correlation between projection vector of the test sound and the template projection matrix and choose the biggest value to stand for the recognizing result.

Acknowledgement

This work was supported by Application Demonstration of Agricultural Information Integrated Service Platform (grant No. 2012CX90204)

References

[1] Muda, Lindasalwa, Mumtaj Begam, and I. Elamvazuthi. "Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques." arXiv preprint arXiv:1003.4083 (2010).

[2] Bala, Anjali, Abhijeet Kumar, and Nidhika Birla. "Voice command recognition system based on MFCC and DTW." International Journal of Engineering Science and Technology 2.12 (2010): 7335-7342.

[3] Kevin M. Coggins and Jose Principe, Detection and Classification of Insect Sounds in a Grain Silo using a Neural Network

[4] Nellenbach C. Chesmore E.D., "Acoustic methods for the automated detection and identification of insects". Acta Horticulturae (562): p. 223-231, 2001.

[5] Czarnecki, K. R. Z. Y. S. Z. T. O. F., and M. A. R. E. K. Moszyński. "Using concentrated spectrogram for analysis of audio acoustic signals." Hydroacoustics 15 (2012): 27-32.

[6] Lampert, Thomas A., and Simon EM O'Keefe. "A survey of spectrogram t rack detection algorithms." Applied acoustics 71.2 (2010): 87-100.

[7] Tanyer, S. Gökhun, and Hamza Ozer. "Voice activity detection in nonstationary noise." IEEE Transactions on Speech and Audio Processing 8.4 (2000): 478-482.

[8] Information on http://www.ars.usda.gov/pandp/docs.htm?docid=10919#anastrepha%20suspensa

[9] Elad, Michael, and Michal Aharon. "Image denoising via sparse and redundant representations over learned dictionaries." Image Processing, IEEE Transactions on 15.12 (2006): 3736-3745.

[10] Chang, S. Grace, Bin Yu, and Martin Vetterli. "Adaptive wavelet thresholding for image denoising and compression." Image Processing, IEEE Transactions on 9.9 (2000): 1532-1546.

Materials, Machines and Development of Technologies for Industrial Production
10.4028/www.scientific.net/AMM.618

A Method of Insect Recognition Based on Spectrogram
10.4028/www.scientific.net/AMM.618.362