Applied Mechanics and Materials
ISSN: 1662-7482, Vol. 855, pp 75-81
doi:10.4028/www.scientific.net/AMM.855.75

Submitted: 2016-06-23
Revised: 2016-08-15
Accepted: 2016-08-23
Online: 2016-10-25

Nondestructive Determination of Maturity of the Monthong Durian by Mel-Frequency Cepstral Coefficients (MFCCs) and Neural Network

Peerapol Khunarsa[1,a*], Julalug Mahawan[2,b], Pisit Nakjai[1,c] and Nerissa Onkhum[1,d]

[1]Computer Science Program, Faculty of Science and Technology, Uttaradit Rajabhat University, Uttaradit, 53000 THIALAND.

[2]Information Technology Program, Faculty of Science and Technology, Uttaradit Rajabhat University, Uttaradit, 53000 THIALAND.

[a]peerapol@uru.ac.th, [b]Julalug@uru.ac.th, [c]mynameisbee@uru.ac.th, [d]nerissaonkhum@uru.ac.th

Keywords: Mel Frequency Cepstral Coefficient (MFCC), Neural Network, Signal processing, Pattern recognition.

Abstract. The challenging for buyers around the globe to identify good quality of Durian. For several kinds of Durian, it may be difficult for buyers to determine the Durian quality by appearance. The ability to select only good quality Durian without cutting or cleaving is useful because buyers will not waste money ordering undesirable Durian.

This paper proposes a nondestructive technique to determine the stages of maturity of durian fruits. The presented methodology utilizes the concept of pattern matching. We used the local knocking equipment to knock the durian for knocked-sound. After that the knocked-sound was analyzed and generated to Mel-frequency cepstral coefficients (MFCCs) that is used to train data for the classifier. Feed-Forward Neural Network was used for the classifier and can effectively classification the stages of maturity of durian fruits with accuracy rate more than 82%.

Introduction

In Asia, Thailand, Malaysia, Indonesia, India, Vietnam and the Philippines are the main growers of Monthong durians. Thailand exports Monthong durians to countries such as the U.S.A., China, Hong Kong and Singapore. The skin of a durian is thick and consists of hundreds of hard spikes, as shown in Fig. 1. It is difficult to determine durian ripeness by observing the skin. Buyers depend on vendors' recommendations and sometimes they may not get value for money. Consumers have felt disappointed because cutting Monthong durians too early results in missing out on the delicious taste and texture. Therefore, it will impact people worldwide if we can develop a practical automated method that can determine the ripeness of Monthong durians in advance without the need to damage them.

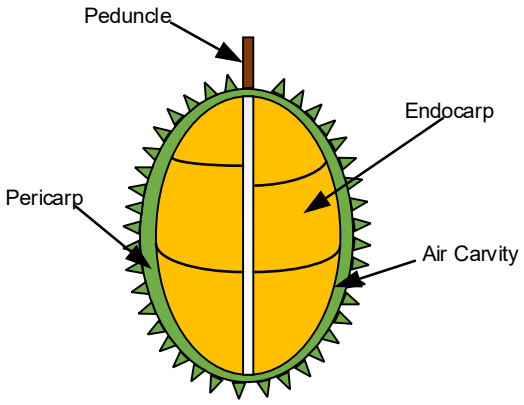

Fig. 1 Physical appearance of a durian fruit.

Fig. 2 Photograph of a durian

Fig. 3 Local knocking equipment

A major challenge facing most exporters of durian fruits is that the fruits must reach a certain minimum degree of ripeness or maturity when it arrives at the destination. However, at present, there is no specific means to guarantee that the exported durian fruits would reach the proper stage of maturity upon arrival in the importing countries. In addition, if immature or "too young" durian fruits are exported, then they would never ripen even after it arrives at their destination and even after a long time. Therefore, it is crucial that durian fruits are exported when their proper maturity stage is reached; otherwise, the importing customers could refuse to take delivery of the entire shipment of the fruits.

Several techniques have been adopted by durian growers to estimate the maturity of durian fruits, most of which rely on the physical appearances of the fruits, e.g., size, shape, and color. The conventional techniques for estimating the stage of maturity of durian fruits are by trial and error whereby the experienced durian growers count the number of days after anthesis (DAA) as well as adopt the practice of acoustic tapping. The acoustic tapping is a technique whereby the growers hold a durian fruit close to their ear before gently tapping it with a flat object, e.g., a gardening knife, couple of times to listen to the nature of the reverberated sound. A firm sound indicates that the durian fruit is "too young" while a loosening sound indicates that the fruit is mature. This type of determination technique for maturity of the fruit relies on the size of the air cavity between the endocarp (i.e., the flesh of the durian fruit) and the pericarp (i.e., the fruit's outer thorny peel) of a durian fruit, the size of which grows larger with degree of ripeness. Nevertheless, the reliability of the tapping technique is greatly subject to the growers' experience.

In addition to the use of physical appearances to determine the stage of ripeness, the stage of maturity can also be determined by the chemical composition of the durian fruits, i.e., dry weight, ethylene, and sugar content. Even though highly accurate, the destruction of the durian fruit is required for measurement of the chemical properties.

However, very little research has focused on applying signal-processing techniques to animals or agricultural produce. Traditionally, signal processing has been applied to other fields such as gender identification, emotion classification and speech recognition. In the gender identification field, Mel-frequency cepstral coefficients (MFCCs) and pitch frequencies are acoustic features that have been used to identify a speaker's gender [1,2,3,4,5]. The duration of speech segments has been studied to determine a speaker's gender [6]. With regard to emotion, features such as pitch frequencies, spectral features, energy features and their augmentations have been investigated to distinguish five emotional states [7]. Speech recognition technology has progressed and Mel-frequency cepstral coefficients (MFCCs) are acoustic features that have been widely used in speech recognition [8-10]. The Hidden Markov Model (HMM) is an efficient method employed in speech recognition [11, 12] to model signal phenomena. Furthermore, HMM multiple speech classifiers have been studied to improve a voice-controlled robot [13]. There has been limited research that investigates animal sounds. Some studies consider animal sounds among other classes of sounds [14]. Ways of distinguishing the sounds of birds, cats, cows and dogs have been examined [15]. Recently, methods of recognizing dog sounds have been developed [16]. A developed animal voice recognition system has used zero-cross-rate (ZCR) to find dog voice boundaries and used Mel-frequency cepstral coefficients (MFCCs) to recognize animal voice.

Selecting fruit using digitized signals is both interesting and practical. Countless consumers have been disappointed by the quality of the fruit that they have purchased from supermarkets and market stalls. When looking at the exterior color of fruit like Monthong durians, it can be very difficult to determine whether the fruit selected is of acceptable quality. Therefore, for several kinds of fruit, image-processing technology cannot efficiently classify fruit quality. It is hoped that in the future, with the help of portable computers and mobile devices, buyers will be able to select good quality fruit more accurately and fruit retailers will be able to verify the quality of their produce. In addition, the fruit industry will have machines with the capability to automatically classify large quantities of fruit not only by size but also by quality.

This paper proposes a nondestructive technique to determine the stages of maturity of durian fruits. The presented methodology utilizes the concept of pattern matching. We used the local knocking equipment to knock the durian for knocked-sound. A durian ripeness knocked-sound recognition method that used Mel-frequency cepstral coefficients (MFCCs) and Neural Network to solve this problem.

Experimental evaluation

Data Collection. For the evaluation and development of unripe and ripe Monthong durians knocked-sound recognition system, the following 100 unripe and 100 ripe Monthong durians were captured to generate a data set. The data set contains 20-30 knocked-sound for each durian. Sample files were coded in stereo of frequency 44.2 kHz with 128/s bit rate. The files were converted in mono and down sampled to 11 kHz.

Feature extraction. Feature extraction is the process of computing a compact numerical representation that can be used to characterize a segment of audio. The design of descriptive feature for a specific application is the main challenge in building pattern recognition systems. The present work uses Mel Frequency Cepstral Coefficients analysis that is based on Fast Fourier transform (FFT).

The use of Mel Frequency Cepstral Coefficients can be considered as one of the standard method for feature extraction [17]. The use of about 20 MFCC coefficients is common in ASR, although 10-12 coefficients are often considered to be sufficient for coding speech [18]. The most notable downside of using MFCC is its sensitivity to noise due to its dependence on the spectral form. Methods that utilize information in the periodicity of speech signals could be used to overcome this problem, although speech also contains aperiodic content [19].

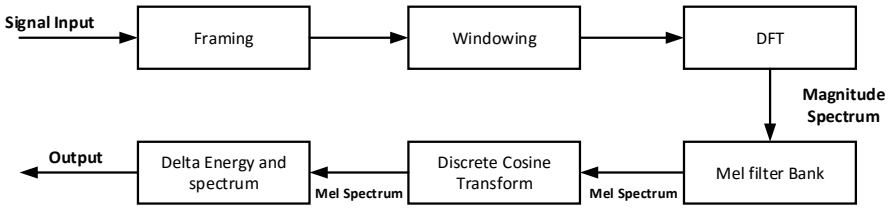

Fig. 4 MFCC Block Diagram

Figure 4 shows the process of creating MFCC features. The first step is to divide the speech signal into frames, usually by applying a windowing function at fixed intervals. The aim here is to model small (typically 20ms) sections of the signal that are statistically stationary. The window function, typically a Hamming window, removes edge effects. We generate a cepstral feature vector for each frame.

The next step is to take the Discrete Fourier Transform (DFT) of each frame. We then retain only the logarithm of the amplitude spectrum. We discard phase information because perceptual studies have shown that the amplitude of the spectrum is much more important than the phase. We take the logarithm of the amplitude spectrum because the perceived loudness of a signal has been found to be approximately logarithmic.

The next step is to smooth the spectrum and emphasize perceptually meaningful frequencies. This is achieved by collecting the 256 spectral components into 40 frequency bins as shown in Figure 4. Although one would expect these bins to be equally spaced in frequency, it has been found that for speech, the lower frequencies are perceptually more important than the higher frequencies. Therefore, the bin spacing follows the so-called 'Mel' frequency scale.

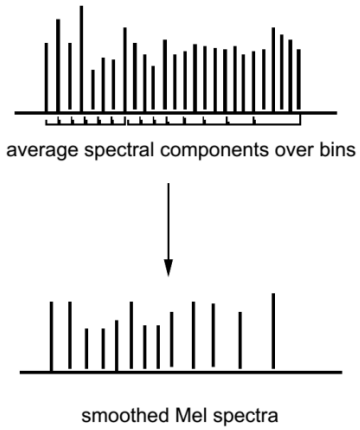

Fig. 5 Mel scaling and smoothing of the log amplitude spectrum. Spectral components are averaged over Mel-spaced bins to produce a smoothed spectrum.

The Mel scale is based on a mapping between actual frequency and perceived pitch as apparently the human auditory system does not perceive pitch in a linear manner. The mapping is approximately linear below lkllz and logarithmic above. Figure 5 shows the Mel function.

The components of the Mel-spectral vectors calculated for each frame are highly correlated. Speech features are typically modeled by mixtures of Gaussian densities. Therefore, in order to reduce the number of parameters in the system, the last step of MFCC feature construction it to apply a transform to the Mel-spectral vectors which decorrelates their components. Theoretically, the Karhunen-Loeve (KL) transform (or equivalently Principal Components Analysis (PCA)) achieves this. In the speech community, the KL transform is approximated by the Discrete Cosine Transform (DCT). Using this transform, 13 (or so) cepstral features are obtained for each frame.

Data Classification. Neural networks have many similarities with Markov models. Both are statistical models which are represented as graphs. Where Markov models use probabilities for state transitions, neural networks use connection strengths and functions. A key difference is that neural networks are fundamentally parallel while Markov chains are serial. Frequencies in speech, occur in parallel, while syllable series and words are essentially serial. This means that both techniques are very powerful in a different context.

As in the neural network, the challenge is to set the appropriate weights of the connection, the Markov model challenge is finding the appropriate transition and observation probabilities. In many speech recognition systems, both techniques are implemented together and work in a symbiotic relationship. Neural networks perform very well at learning phoneme probability from highly parallel audio input, while Markov models can use the phoneme observation probabilities that neural networks provide to produce the likeliest phoneme sequence or word. This is at the core of a hybrid approach to natural language understanding.

13-dimension acoustic features, consisting of 12 Mel-frequency cepstral coefficients (MFCCs) with energy as well as their 1st- and 2nd-order derivatives, are extracted and used to distinguish unripe and ripe Monthong durians. To recognize durian ripeness from the extracted acoustic features, acoustic models of unripe and ripe Monthong durians, sequences of acoustic models for unripe and ripe Monthong durians. After that, Feed-Forward neural network with a 30 Hidden are used for Classification technique. All classifiers were implemented by using the Pattern Recognition Matlab 2015B. The performances were evaluated by 5-fold cross-validation technique.

Result and discussion

In each experiment, we performed 50 runs of the 5-fold cross-validation to obtain statistically reliable results. The mean recognition rate was calculated based on the error average for one run on test set. Table 1 shows the test classification performance of different Feature Extraction. Experiments by using Feed Forward Neuron network with 30 hidden neuron show the best performance of 82.3% by using Mel-frequency cepstral coefficients (MFCCs).

Table 1 Average accuracy of all classes from each feature extraction.

Feature Extraction	Ripe Accuracy(%)	Unripe Accuracy(%)	Avarage Accuracy(%)
Mel-frequency cepstral coefficients (MFCCs)	83.1	81.5	82.3
Spectral rolloff	62.0	60.2	60.8
Autocorrelation	80.1	80.0	80.0
Brightness	71.1	62.5	64.8
lowenergy	52.3	51.1	51.7
Linear predictive coding (LPC)	68.9	66.1	67.1
Regularity	46.7	53.6	51.4
Root Mean Square	58.1	60.2	59.5
Roughness	54.9	59.9	57.8

After that, we want to find out the Feature Extraction technique that gives the best accuracy rate in classification for the data set. Feature Extraction technique include Spectral roll-off, Autocorrelation, Brightness, low energy, Linear predictive coding (LPC), Regularity, Root Mean Square and Roughness. After the experiments, it was found that Mel-frequency cepstral coefficients (MFCCs) show the best accuracy rate in classification.

Conclusion

This paper proposes a nondestructive technique to determine the stages of maturity of durian fruits. The presented methodology utilizes the concept of pattern matching. We used the local knocking equipment to knock the durian for knocked-sound. After that the knocked-sound was analyzed and generated to Mel-frequency cepstral coefficients (MFCCs) that is used to train data for the classifier. Feed-Forward Neural Network was used for the classifier and can effectively classification the stages of maturity of durian fruits with accuracy rate more than 82%.

References

[1] R. Phoophuangpairoj, S. Phongsuphap, and S. Tangwongsan, "Gender identification from Thai speech signal using a neural network," Lecture Notes in Computer Science, Vol. 5863, 2009, pp. 676-684.

[2] H. Ting, Y. Yingchun, and W. Zhaohui, "Combining MFCC and pitch to enhance the performance of the gender recognition," in Proceedings of International Conference on Signal Processing, 2006, pp. 16-20.

[3] S. M. R. Azghadi, M. R. Bonyadi, and H. Sliahhosseini, "Gender classification based on feedforward backpropagation neural network," IFIP International Federation for Information Processing, Vol. 247, 2007, pp. 299-304.

[4] M. H. James and J. C. Michael, "The role of F0 and formant frequencies in distinguishing the voices of men and women, attention," Perception and Psychophysics, Vol. 71, 2009, pp. 1150-1166.

[5] C. R. Pernet and P. Belin, "The role of pitch and timbre in voice gender categorization," Frontiers in Psychology, Vol. 3, Article 23, 2012, pp. 1-11.

[6] M. Sigmund, "Gender distinction using short segments of speech signal," International Journal of Computer Science and Network Security, Vol. 8, 2008, pp. 159-162.

[7] D. Ververidis and C. Kotropoulos, "Automatic speech classification to five emotional states based on gender information," in Proceedings of the European Signal Processing Conference, Vol. 1, 2004, pp. 341-344.

[8] A. Deemagarn and A. Kawtrakul, "Thai connected digit speech recognition using Hidden Markov Models," in Proceedings of the 9th International Conference on Speech and Computer, 2004, pp. 731-735.

[9] L. Fuhai, M. Jinwen, and D. Huang, "MFCC and SVM based recognition of Chinese vowels," Lecture Notes in Computer Science, Vol. 3802, 2005, pp. 812-819.

[10] S. Tangwongsan and R. Phoophuangpairoj, "Boosting Thai syllable speech recognition using acoustic models combination," in Proceedings of International Conference on Computer and Electrical Engineering, 2008, pp. 568-572.

[11] S. Tangruamsub, P. Punyabukkana, and A. Suchato, "Thai speech keyword spotting using heterogeneous acoustic modeling," in Proceedings of IEEE International Conference on Research, Innovation and Vision for the Future, 2007, pp. 253-260.

[12] S. Tangwongsan, P. Po-Aramsri, and R. Phoophuangpairoj, "Highly efficient and effective techniques for Thai syllable speech recognition," Lecture Notes in Computer Science, Vol. 3321, 2004, pp. 259-270.

[13] R. Phoophuangpairoj, "Using multiple HMM recognizers and the maximum method to improve voice-controlled robots," in Proceedings of International Conference on Intelligent Signal Processing and Communication Systems, 2011, pp. 1-6.

[14] G. Guo and S. Z. Li, "Content-based audio classification and retrieval by support vector machines," IEEE Transactions on Neural Networks, Vol. 14, 2003, pp. 209- 215.

[15] D. Mitrovic, M. Zeppelzauer, and C. Breiteneder, "Discrimination and retrieval of animal sounds," in Proceedings of the 12th International Multi-Media Modelling Conference, 2006, pp. 339-343.

[16] C. Y. Yeo, S. A. R. Al-Haddad, and C. K. Ng, "Animal voice recognition for identification (ID) detection system," in Proceedings of the 7th IEEE International Colloquium on Signal Processing and Its Applications, 2011, pp. 198-201.

[17] Motlíček P.: Feature Extraction in Speech Coding and Recognition, Report, Portland, to research, data, and theory. Belmont, CA: Thomson/Wadsworth, 2003 US, Oregon Graduate Institute of Science and Technology, pp. 1-50, 2002

[18] Hagen A., Connors D.A. & Pellm B.L.: The Analysis and Design of Architecture Systems for Speech Recognition on Modern HandheldComputing Devices. Proceedings of the 1st IEEE/ACM/IFIP international conference on hardware/software design and system synthesis, pp. 65-70, 2003

[19] Ishizuka K.& Nakatani T.: A feature extraction method using subband based periodicity and aperiodicity decomposition with noise robust frontend processing for automatic speech recognition. Speech Communication, Vol. 48, Issue 11, pp. 1447-1457, 2006