IJASST International Journal of Applied Sciences and Smart Technologies

Volume 06
Issue 02

p-ISSN 2655-8564, e-ISSN 2685-943, pp. 417-428, 2024

Evaluating The Performance of DWT-DCT Feature Extraction in Guitar Chord Recognition

Linggo Sumarno

Department of Electrical Engineering, Faculty of Science and Technology, Sanata Dharma University, Yogyakarta, Indonesia
lingsum@usd.ac.id

(Received 07-10-2024; Revised 25-10-2024; Accepted 26-10-2024)

Abstract

This study presents advancements in audio signal processing techniques, specifically in enhancing the efficiency of guitar chord recognition. It is a continuation of the previous studies, which also aim at minimizing the feature extraction length with the intended performance. This study adopted two signal processing techniques that are common: Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT) for use in the feature extraction method. By conducting a systematic evaluation of two key parameters: frame blocking length and wavelet filter selection, a significant achievement could be achieved. The recognition system managed to obtain chord recognition with an accuracy of up to 91.43%, by using a feature extraction length of only three, which brought about smaller representation than the previous studies. The outcome of this study will help improve the data processing, which can be applied in real time, in this case in Field Programmable Gate Array (FPGA)-based chord recognition systems.

Keywords: chord recognition, Discrete Wavelet Transform, Discrete Cosine Transform, feature extraction

1 Introduction

The extraction of relevant information from the data is one of the crucial and critical processes in tasks like chord recognition. It is the transformation of raw information, usually a lot and difficult to work with, into a small number of useful features. These features encapsulate relevant information contained in the data and are useful in subsequent processes like classification and clustering. The feature extraction approaches

can be roughly separated into two types: spectral representation-based and non spectral representation-based.

The spectral representation-based feature extraction approach tries to represent the harmonic content of the musical data. One of the popular methods is the so-called Pitch Class Profile (PCP) [1], where it has the feature extraction length of 12, which represents the distribution of power across the pitches in a chord. Several improvements [2-4] have been developed to enhance the performance of this PCP further. However, the feature extraction length for these improvements is also 12.

On the other side, the non spectral representation-based approach tries to represent the spectral shape of the musical data. Thus, two examples of feature extraction methods based on this approach are Mel Frequency Cepstral Coefficients (MFCC) [5-6]and Discrete Sine Transform (DST-Wavelet) [7]. Recently, some studies on chord recognition in MFCC [6] and DST-Wavelet [7] showed effective recognition with a length of four in feature extraction. By using these feature extraction methods, the chord recognition system was able to achieve up to 92.14% and 92.86%, respectively.

Developments in the above chord recognition methods still leave scope for further optimization. Reducing the feature extraction length may further enhance data processing efficiency and make applications like chord recognition viable on FPGA systems [8-9]. This FPGA system will provide an added advantage of creating Application-Specific Integrated Circuits (ASICs) for electronic devices that can be customized and enabled for low power and performance.

This study introduces adopting DCT-DWT for feature extraction in a chord recognition, in order to further reduce the feature extraction length. This study particularly investigates the impact of evaluating two parameters—frame blocking length and wavelet filter selection—on attaining fewer than four feature extraction lengths while keeping the intended recognition accuracy.

2 Methodology

2.1 Recognition System Development

Fig. 1 depicts the chord recognition system built for this study, including a full block diagram of the system. It is necessary to point out that this system was implemented using Python software. The following subsections give an in-depth description of each block represented in Fig. 1.

Input and output. The input are isolated guitar chord recordings in WAV format. These recordings consist of seven major chords (C, D, E, F, G, A, B) and are sourced from a publicly available GitHub repository (https://github.com/lingsum/Chord-DST-DWT). The output indicates the corresponding chord text (C, D, E, F, G, A, or B) for each input signal.

Normalization. Normalization modifies the incoming signal data such that its peak value falls within a predetermined range, usually from 1 to -1. This modification must be made because the maximum loudness of recorded chord signals might vary greatly depending on recording conditions and how the musical equipment is used.

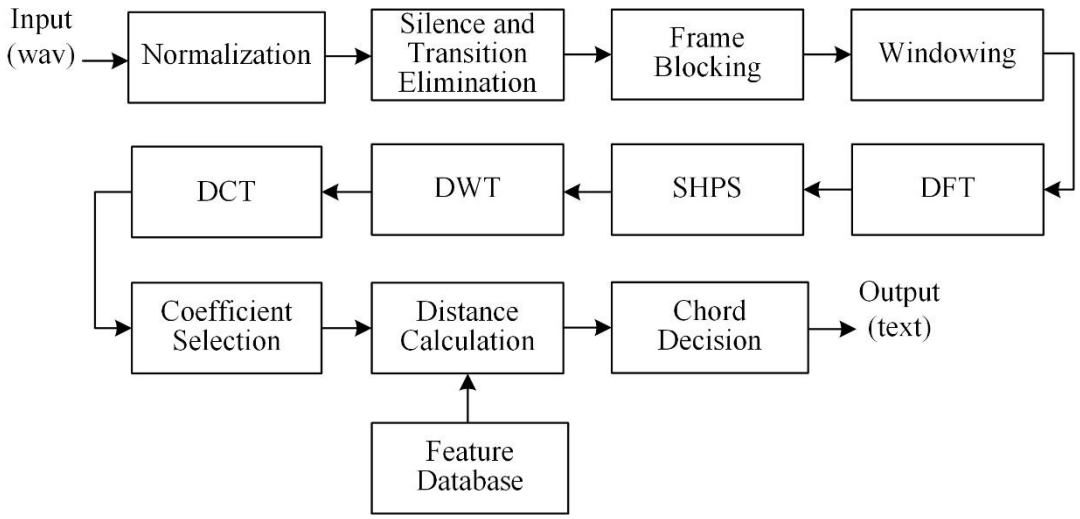

Figure 1. The developed recognition system.

Silence and transition elimination. Silence and transition elimination eliminates unnecessary portions from the signal data array. These eliminations are detailed in more depth below.

1. Silence elimination: A threshold of |0.5| was set, identified through visual examination, in order to distinguish between silence and audible segments. Data points with absolute values below this threshold are eliminated during the left-to-right scan. The scanning stops once a data point with an absolute value meeting or exceeding the threshold is found.
2. Transition elimination: The initial time of 200 milliseconds of the signal was set, identified through visual examination, in order to distinguish between transition and steady state segments. By using this initial time, the initial 200 milliseconds of the signal are eliminated.

Frame blocking. Frame blocking entails segmenting a small array of signal data [10]. This segmentation is normally carried out on the left side of the signal data. The length of this small array, also known as the frame blocking length, has a major influence on the signal's resolution resulting from the Discrete Fourier Transform (DFT) step. Either too low or too high signal resolution could have a negative impact on feature extraction discriminating capabilities, resulting in decreased identification accuracy. This study explored frame blocking lengths of 128, 256, 512, 1024, and 2048.

Windowing. Windowing diminishes the impact of a signal data array's left and right edges. If these edges are not diminished, the signal data may exhibit spectral leakage resulting from the DFT step. This leakage introduces unimportant frequencies that do not relate the chord's actual frequencies. Windowing is used to help diminish spectral leakage. This study made use of the Hamming window, a popular window in signal processing [11].

DFT. The Discrete Fourier Transform (DFT) converts a finite array of signal data into a finite array of complex values in the frequency domain. This study made use of the magnitude of the signal data's complex values. Since, the magnitude sequence is symmetric, so only the left side of the data array will be used. As a note, this study used

Fast Fourier Transform (FFT) instead of DFT. This is because FFT computation is more efficient than DFT.

SHPS. The Simplified Harmonic Product Spectrum (SHPS) is useful for reducing harmonic distortion and noise in the frequency domain. SHPS could reduce the influence of non-harmonic noise by accentuating the fundamental frequency and its harmonics while decreasing non-harmonic components. Sumarno [12] offered this technique, which is a lite version of the Harmonic Product Spectrum (HPS) introduced by Noll [13].

DWT. The Discrete Wavelet Transform (DWT) transforms a signal into a number of scales and frequency bands. Fig. 2 shows the type of DWT employed in this study. The symbols L_0 and L_1 represent low-pass (LPF) and high-pass (HPF) filters, respectively, generated from the chosen wavelet filter. The ↓2 notation indicates a down sampling factor of 2. The notation a_3 denotes the input to the DWT, while a_0, d_0, …, d_{n-2}, and d_{n-1} represent the resulting DWT coefficients at different scales and frequency bands. To maintain the signal length, periodization was applied during filtering. This study evaluated Haar, Daubechies (2-6), and Symlet (2-7) wavelet filters.

DCT. The Discrete Cosine Transform (DCT) compacts the signal's energy into a relatively small set of coefficients, a phenomenon known as energy compaction. This means that the lower-frequency components in the transformed data hold most of the important information about the signal. As a result, the DCT is especially effective for

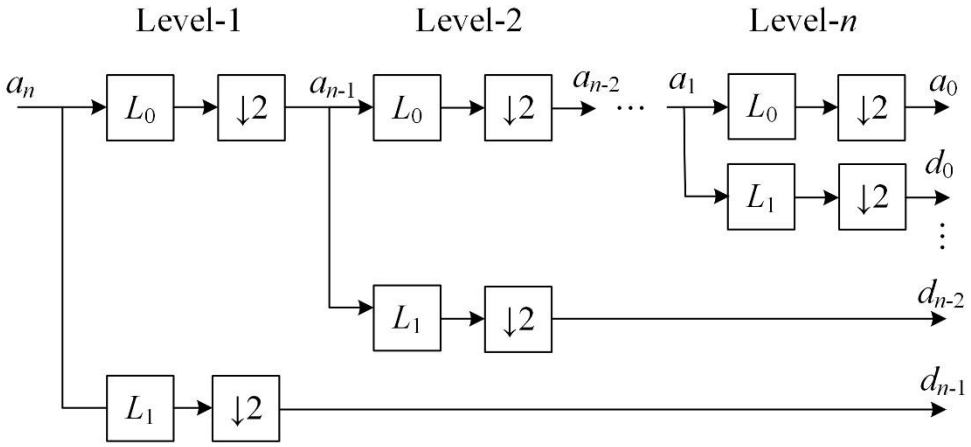

Figure 2. The type of DWT used in this study.

feature extraction, as these lower-frequency components typically carry the most discriminative information.

Coefficient Selection. Coefficient selection entails picking a subset of DCT coefficients to represent the feature extraction of the input signal. The selection process is conducted as follows.

1. Let the result of the previous DCT step be $C = \{c_0, c_1, \dots, c_{N-1}\}$ where N is the length of C.
2. Determine the number of coefficients to be selected, n, where $n \leq N - 1$.
3. The result of the coefficient selection is $S = \{c_1, c_2, \dots, c_n\}$

This study evaluated the number of coefficients to be selected (the feature extraction length) 1, 2, …, and 6.

Distance Calculation and Feature Database. Distance calculation and feature database relate to a classification method that uses template matching techniques [14-16]. The feature database contains a collection of reference feature extraction of the chords used. Distance calculation in this study used the cosine distance function, which is derived from cosine similarity. This similarity is frequently used to calculate similarity scores [17-18]. The distance calculation produces seven distance values, which indicate the comparison between the feature extraction of the input signal data and the seven feature extraction references in the feature database.

Chord decision. Chord decision entails determining the chord that corresponds to the given input signal. The smallest distance value is chosen among the seven calculated distances. The chord corresponding to this smallest distance is then assigned as the output chord.

2.2 Training and Testing

The training procedure entailed creating a thorough feature database. Firstly, features were extracted from ten training samples for each chord using the Coefficient Selection step's output in Fig. 1. Secondly, the average of these features (from ten training samples) was then calculated. Finally, these average results (from each chord) were used as feature extraction references. The final feature database had seven feature extraction

references, one for each chord. For testing purposes, other 140 samples were utilized, with 20 samples for each chord.

3 Results and Analysis

3.1 Testing Results

Tables 1 and 2 show the study's testing findings. They were acquired by systematically evaluating two parameters: frame blocking length and wavelet filter selection. Details on the variations of these two parameters can be seen in the Methodology section above.

Table 1. Testing results using Symlet 6 wavelet filter.
Results shown: Recognition accuracy (%).

Frame blocking size	Feature set size					
	1	2	**3**	4	5	6
128	14.29	26.43	37.86	57.86	62.14	82.86
256	25.71	66.43	82.14	90.00	90.71	94.29
512	28.57	39.29	**91.43**	95.00	94.29	95.00
1024	27.14	79.29	81.43	82.86	87.14	87.14
2048	21.43	67.86	74.29	80.00	84.29	84.29

Table 2. Testing results using a frame blocking size 512.
Results shown: Recognition accuracy (%).

Wavelet filter	Feature set size					
	1	2	**3**	4	5	6
Haar	28.57	41.43	74.29	75.00	66.43	75.00
Daubechies 2	25.71	62.86	82.86	87.86	87.14	91.43
Daubechies 3	28.57	79.29	82.86	90.00	88.57	88.57
Daubechies 4	28.57	75.71	85.71	80.00	83.57	90.00
Daubechies 5	25.71	63.57	80.00	81.43	82.14	82.14
Daubechies 6	16.43	41.43	72.86	78.57	80.00	80.00

Symlet 2	25.71	62.86	82.86	87.86	87.14	91.43
Symlet 3	28.57	79.29	82.86	90.00	88.57	88.57
Symlet 4	28.57	40.00	**91.43**	90.71	94.29	95.71
Symlet 5	28.57	48.57	72.14	80.71	71.43	79.29
Symlet 6	28.57	39.29	**91.43**	95.00	94.29	95.00
Symlet 7	28.57	77.14	74.29	85.71	84.29	85.71

3.2 Discussions

Table 1 shows that recognition accuracy suffers from either too short or too long frame blocking lengths. A too-short frame blocking length cannot capture enough detail in the signal data from the DFT step, leading to a too-low resolution of the signal. On the contrary, a too-long frame blocking length captures too much detail, leading to a too-high resolution of the signal. Both too low or too high a resolution of the signal has a negative impact on the discrimination level of feature extraction, thereby reducing recognition accuracy.

Table 2 shows that choosing the proper wavelet filter can increase recognition accuracy. This improvement is due to the wavelet filter's ability to properly separate low-frequency and high-frequency components through its LPF and HPF in a multi-resolution domain. This proper separation enhances the discrimination level of feature extraction, leading to increased recognition accuracy.

Tables 1 and 2 show that generally, increasing the feature extraction length from 1 to 6 improves recognition accuracy. This suggests that up to a length of 6, the extracted features still capture the essential features. These essential features help in capturing the most important and discriminative characteristics from the input data, aiding in class separation [19]. In this study, increasing the number of essential features from 1 to 6 enhances the discrimination level of feature extraction, thereby increasing recognition accuracy.

Table 3 compares the performance between the introduced feature extraction method with the other previous methods. Table 3 shows the feature extraction method in this study is the most efficient for accuracy greater than 90%. With a feature extraction

Table 3. Performance comparison between the introduced feature extraction method with the other previous methods.

Feature Extraction Methods	Feature Extraction Length	Recognition Accuracy (%)	Chord Test Set Size
Improved PCP [2]	12	95.83	192
CRP Enhanced PCP [3]	12	99.96	4608
MFCC [5]	6	91.43	140
MFCC with Kaiser windowing [6]	4	92.14	140
DST-Wavelet [7]	4	92.86	140
DWT-DCT (this study)	**3**	**91.43**	140

Note: The table shows the shortest feature extraction length required to achieve more than 90% recognition accuracy

length of three, the recognition system is able to achieve a recognition accuracy of up to 91.43%.

Notes on FPGA Implementation

There are two notes on FPGA implementation related to this study. These notes are described as follow.

1. This study uses three kinds of transformation, namely DFT, DWT, and DCT. For FPGA implementation, in order to reduce the processing time, we can use some examples of efficient architectures [20-22] for these kinds of transformations.
2. The improved data processing in this study can be applied in real-time FPGA based system by utilizing hardware-software co-design, which exploit native parallelism of the FPGA hardware [23].

4 Conclusions and Future Studies

This study explored, in detail, the recognition accuracy of a novel feature extraction method that adopted DWT-DCT. The novel method was able to achieve a recognition accuracy of 91.43% with this particular feature extraction length of only three. Such

accuracy was achieved by using a frame blocking length of 512 and also by using the Symlet 4 and 6 wavelet filters. With this adoption of DWT-DCT in the feature extraction method, it is further ensured that the relevant and discriminative features of the guitar chords are well captured, leading to better recognition performances.

This study still leaves room for improvement in terms of accuracy and efficiency. Therefore, in future studies, the focus should be directed to exploring different methods for feature extraction. In this case, the methods are expected to provide recognition accuracy that exceeds 91.43%, and the length of feature extraction is three or less.

Acknowledgements

The Institute for Research and Community Services at Sanata Dharma University has supported this study.

References

[1] T. Fujishima, “Realtime chord recognition of musical sound: a system using Common Lisp Music”, *Proceeding of the International Computer Music Conference (ICMC)*, Beijing, pp. 464–467, 1999.

[2] K. Ma, “Automatic Chord Recognition”, Department of Computer Sciences, University of Wisconsin-Madison, May 2016. [Online]. Available : http://pages.cs.wisc.edu/~kma/projects.html. [Accessed Jul. 22, 2024].

[3] P. Rajparkur, B. Girardeau, and T. Migimatsu, “A Supervised Approach to Musical Chord Recognition”, *Stanford Undergraduate Research Journal*, Vol. 15, pp. 36-40, 2015.

[4] E. Demirel, B. Bozkurt, and X. Serra, “Automatic chord-scale recognition using harmonic pitch class profiles”, *Proc. Sound Music Comput. Conf.*, pp. 72–79, 2019.

[5] L. Sumarno, “Chord Recognition using FFT Based Segment Averaging and Subsampling Feature Extraction,” *Proceeding of 8th International Conference on Information and Communication Technology (ICoICT 2020)*, pp. 465–469. 2020.

[6] L. Sumarno, "Guitar chord recognition using MFCC based feature extraction with Kaiser windowing", *Proceeding of Transdisciplinary Symposium on Engineering and Technology (TSET 2022),* Published 2024.

[7] L. Sumarno, "The performance of DST-Wavelet feature extraction for guitar chord recognition", *Proceeding of The 1st International Conference on Applied Sciences and Smart Technologies (InCASST 2023),* Published 2024.

[8] K. Vaca, M. M. Jefferies, and X. Yang, "An Open Audio Processing Platform with Zync FPGA", *Proceeding of 22nd IEEE Int. Symp. Meas. Control Robot. Robot. Benefit Humanit. ISMCR 2019,* pp. D1-2-1–D1-2-6, 2019.

[9] K. Vaca, A. Gajjar, and X. Yang, "Real-Time Automatic Music Transcription (AMT) with Zync FPGA*", Proceeding of IEEE Comput. Soc. Annu. Symp. VLSI, ISVLSI, Vol. 2019-July (2019),* pp. 378–384, 2019.

[10] O.K. Hamid, "Frame Blocking and Windowing Speech Signal", *J. Information, Commun. Intell. Syst.*, Vol. 4, No. 5, pp. 87–94, 2018.

[11] H. Rakshit, and M. A. Ullah, "A comparative work on window functions for designing efficient FIR filter", *Proceeding of 2014 9th Int. Forum Strateg. Technol. (IFOST 2014),* pp. 91–96, 2014.

[12] L. Sumarno, "Chord recognition using segment averaging feature extraction with simplified harmonic product spectrum and logarithmic scaling", *Int. J. Electr. Eng. Informatics*, Vol. 10, No. 4, pp. 753–764, 2018.

[13] A.M. Noll, "Pitch Determination of Human Speech by the Harmonic Product Spectrum, the Harmonic Sum Spectrum and a Maximum Likelihood Estimate", *Proceeding of the Symposium on Computer Processing in Communications, Vol. 19, Polytechnic Press, Brooklyn, New York,* pp. 779-797, 1970.

[14] I. Izonin, R. Tkachenko, N. Shakhovska, B. Ilchyshyn, and K.K. Singh, "A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain", *Mathematics,* Vol. 10, No. 11, 2022

[15] A.K. Jain, R.P.W. Duin, and J. Mao, "Statistical pattern recognition: A review", *IEEE Trans. Pattern Anal. Mach. Intell.* Vol. 22, No. 1, pp. 4–37, 2000.

[16] A. Massari, R.W. Clayton, and M. Kohler, “Damage detection by template matching of scattered waves”, *Bull. Seismol. Soc. Am.* Vol. 108, No. 5, pp. 2556–2564, 2018.

[17] H.U. Zhi-Qiang, Z. Jia-Qi, W. Xin, L.I.U. Zi-Wei, and L.I.U. Yong, “Improved algorithm of DTW in speech recognition”, in *Proceeding of IOP Conf. Ser. Mater. Sci. Eng.*, Vol. 563, No. 5, 2019.

[18] S. Sohangir, and D. Wang, “Improved sqrt-cosine similarity measurement”, *J. Big Data*, Vol 4, No 1**,** 2017.

[19] H.R. Shahdoosti, and F. Mirzapour, “Spectral–spatial feature extraction using orthogonal linear discriminant analysis for classification of hyperspectral data”, *Eur. J. Remote Sens.*, Vol. 50, No. 1, 2017.

[20] Y. Zhao, H. Lv, J. Li, and L. Zhu, “High performance and resource efficient FFT processor based on CORDIC algorithm”, *EURASIP J. Adv. Signal Process*, Vol. 23, 2022.

[21] M.A.M. Basiri, and P. Bharadwaja, "Efficient FPGA Implementations of Lifting based DWT using Partial Reconfiguration," *2023 36th International Conference on VLSI Design and 2023 22nd International Conference on Embedded Systems (VLSID)*, Hyderabad, India, pp. 319-324, 2023.

[22] C.A. Kumar, G.R. Poornima, R. Aruna, B.P.P. Kumar, S. Harish, & D.A.L. Vaishnavi, “Implementation of an Efficient and Reconfigurable Architecture for DCT on FPGA”, *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 12, No. 10s, pp. 597–604, 2024.

[23] I. Bravo-Munoz, J.L. Lazaro-Galilea, and A. Gardel-Vicente, “FPGA and SoC Devices Applied to New Trends in Image/Video and Signal Processing Fields”, *Electronics*, Vol. 6, No. 25, 2017.