

Audio Watermarking Combined with Compressive Sampling Based on QIM and DST-QR Techniques

Irma Safitri *, Gelar Budiman, Arfidianti Kartika Meiza Putri

*Electrical Engineering Faculty
Telkom University
Jl. Telekomunikasi 1 Dayeuhkolot
Bandung, Indonesia*

Abstract

Abuse is not only done to copy or distribute data but also to the digital copyright labels. There is a way to protect data by inserting or hiding a piece of certain information, namely a watermarking technique. In this paper, we propose audio watermarking with Quantization Index Modulation (QIM) method as an embedding process combined with Compressive Sampling (CS), Discrete Sine Transform (DST) and QR decomposition. Binary image is used as a watermark inserted in host audio. DST is used for transformation process from time domain to frequency domain, while QR is used to decompose one-dimension matrix into two-dimension matrix. Meanwhile, CS is used to obtain the compressed watermark file which is done before the embedding process. QIM method is used to embed the watermark file to the audio host file. Simulation results indicated that the proposed audio watermarking technique has good robustness against some attacks such as Low Pass Filter (LPF), resampling and linear speed change. In addition, it provides good performance in terms of imperceptibility with Signal to Noise Ratio (SNR) > 20 dB and capacity $C = 689$ bps.

Keywords: Audio watermarking, Quantization Index Modulation (QIM), Discrete Sine Transform (DST), QR, Compressive Sampling (CS)

I. INTRODUCTION

In recent years, the ease of copying and distributing digital media over the internet has increased significantly. It certainly raises the issue of copyright infringement. Therefore, a promising technique for implementing copyright protection, the watermarking technique is required. Watermarking is the process of inserting information in the form of a watermark into a host object that takes advantage of weaknesses in the Human Auditory System (HAS). Audio watermarking is a technique of inserting information in the form of digital data into a host (media host) signal in the form of an audio file. The information that is inserted can be bits, text, images, or other audio files. It has various applications, namely as content indexing, monitoring or data tracking, and data authentication [1]. A detailed survey of this technology can be found at [2].

From previous research about audio watermarking, there was several audio watermarking methods published but there were still any lack of the performance. In [3], Khademi used the Quantization Index Modulation (QIM) method for audio watermarking. He used audio signal frequency to hide the watermarking data. However, the audio watermarking system could only withstand the MP3 compression attack. In [4], Chaabane used QR technique for audio watermarking. This technique has the ability to support large amounts of information in a narrow space that requires faster insertion time.

However, the analysis did not use Time Scaling Modification (TSM) attack to test the robustness of watermark.

In [6], Yassine proposed audio watermarking for copyright protection with combination transform method of Discrete Wavelet Transform (DWT) and Discrete Sine Transform (DST), but he did not describe the method robustness against TSM and Linear Speed Change (LSC) attack. In [7], Budiman proposed Lifting Wavelet Transform (LWT) to decompose host audio before it is transformed by FFT for Spread Spectrum embedding method but the proposed method resulted in low payload of watermark.

In [8] and [9] as our previous research, there was also a method named multicarrier modulation to recover data from incomplete information. In [8], the proposed method has much lower imperceptibility level than the threshold imperceptibility scale compared to the [9]. In [9], the proposed method did not produce good bit error rate (BER) in LPF attack. However, it has a good BER with Compressive Sampling (CS). The schematic capacity can be 40 bps. In [10], Budiman conducted research on audio watermarking using Fast Fourier Transform (FFT)-DCT method. The parameters obtained are Objective Different Grade (ODG) > -1 and Signal to Noise Ratio (SNR) > 30 dB. However, by using FFT method, this research produced very low robustness against any attacks.

In this paper, we propose a Compressive Sampling (CS) procedure for audio watermarking system by using Quantization Index Modulation (QIM), Discrete Sine Transform (DST) and QR methods. The CS procedure is performed to get the results of the compression of the watermark bit to then be inserted into the audio host. In

* Corresponding Author.

Email: irmasaf@telkomuniversity.ac.id

Received: October 31, 2018 ; Revised: December 12, 2018

Accepted: February 08, 2018 ; Published: August 31, 2019

© 2019 PPET - LIPI

the DST process, the host audio signal is transformed from the time domain to the frequency domain. It is followed by the QR process to convert the signal into a square matrix, then divided into Q and R matrices. Next process is the R matrix will be used to insert the watermark bit using the QIM method. The combination of DST transformation and QR decomposition is expected to have more robust watermark when it is embedded in R matrix of QR decomposition result and also have less complexity than Singular Value (SV) decomposition.

This paper is described as follows. Part 2 introduces the basic theory of QIM, DST, QR, and CS. Section 3 presents the watermark model, including embedding and extraction processes. The evaluation is given in Section 4. The last, conclusion of this paper is written in Section 5.

II. BASIC THEORY

This section explains some methods that are used in this paper such as Compressive Sampling (CS), Quantization Index Modulation (QIM), Discrete Sine Transform (DST) and QR.

A. Compressive Sampling (CS)

Compressive Sampling (CS) is a new technique in signal processing and sensing that is currently growing rapidly [11]. The purpose of CS is to obtain less than the number of previously required samples in which signal recovery is still perfect. CS can reconstruct the signal by using a number of random measurements called sensing matrix and the signal must be random. CS uses a special measurement of multiple matrices A that satisfies the Restricted Isometry Property (RIP) criterion. CS can be formulated as (1) - (3) [12]:

$$y = A \times x \quad (1)$$

$$y = A \times \psi \times f \quad (2)$$

where y is linear vector measurement of signal ($M \times N$), A is CS measurement matrix generated randomly for insertion and decomposition only, x is coefficient transform signal, ψ is base orthonormal matrix and f is original signal. The CS reconstruction with minimum L_1 is indicated by (3) [13]:

$$\hat{x} = \underset{i \in \{1, N\}}{\operatorname{argmin}} (\|y - Ax_i\|) \quad (3)$$

where i is an iteration index in N sample of iteration until the error is converged or in minimum value.

B. Quantization Index Modulation (QIM)

Quantization Index Modulation (QIM) is a signal processing technique to quantize the signal depending on the binary watermark that will be embedded. The original data signal is the host audio and the watermark data is in the binary form. QIM is a method that has a good level of robustness. QIM formula for insertion process is written as (4) – (6) [14]:

$$F'(0) = \begin{cases} A_k, & \text{if } w = 1 \text{ and } \arg \min |F(0) - A_k| \\ B_k, & \text{if } w = 0 \text{ and } \arg \min |F(0) - B_k| \end{cases} \quad (4)$$

where:

$$A_k = \left(2k + \frac{1}{2}\right) \Delta; B_k = \left(2k - \frac{1}{2}\right) \Delta \quad (5)$$

and $k = 0, \pm 1, \pm 2, \dots$

$$\Delta = \frac{1}{2^{(nbit-1)}} \quad (6)$$

where Δ is quantization step size, $nbit$ is quantization bits, $F(0)$ is host audio, $F'(0)$ is watermarked audio and w is bit watermark. For the extraction process, the QIM extraction equation is written as (7):

$$\hat{x}(k) = \operatorname{mod} \left(\operatorname{ceil} \left(\frac{F'(0)}{\Delta} \right), 2 \right) \quad (7)$$

C. Discrete Sine Transform (DST)

Discrete Sine Transform (DST) is a technique of transforming data from time domain to frequency domain. The comparison of Discrete Sine Transform (DST) and Discrete Cosine Transform (DCT) was performed in [5]. The DST gives more insertion options from the result of the condense the signal energy into several coefficients compared to the DCT. DST also provides more security for the watermarking system and produces more excellent resilience and imperceptibility compared to the DCT. DST is identical to the spectrum shift of the Discrete Fourier Transform of a signal that is a particular permutation and modification of the original signal.

$$y(k) = \sum_{n=0}^{N-1} x(n) \sin \left(\frac{\pi kn}{N+1} \right), k = 0, \dots, N-1 \quad (8)$$

DST also has energy compaction that can be used in audio watermarking. The equation calculates the value of reverse discrete sine transformation [7]:

$$x(n) = \frac{2}{N+1} \sum_{k=0}^{N-1} y(k) \cdot \sin \left(\frac{\pi kn}{N+1} \right), n = 0, \dots, N-1 \quad (9)$$

where $y(k)$ is frequency-domain signal, $x(n)$ is time-domain signal and N is signal length.

D. QR

The QR decomposition consists of two matrices, the matrix Q and R , where Q is the matrix with the orthonormal column and R is the upper triangular matrix. QR decomposition is the most effective and widely used method to solve the total eigenvalues of common matrices. QR decomposed as a matrix A with size $M_1 \times M_2$ converted into two matrices Q and R . A QR decomposition of a square matrix A is real decomposition of A can be seen by the (10) [15]:

$$A_{M_1 \times M_2} = Q_{M_1 \times M_2} R_{M_2 \times M_1} \quad (10)$$

The role of QR decomposition and Singular Value Decomposition (SVD) is virtually identical, but the SVD takes almost ten times as much in a large number of data processing as compared to QR decomposition. Therefore, QR decomposition is much more efficient.

III. WATERMARK MODEL

In general, the design of audio watermarking system consists of two processes, embedding, and extraction. Embedding process is an inserting process of the watermark data into the audio host file. Meanwhile,

extraction process is a separation process between the watermark from the audio file. Our system must have good robustness to resist any attack. Hence, it will be tested using SNR and ODG parameters. Watermark is extracted and tested using BER parameter as the robustness performance.

A. Embedding Process

Figure 1 shows the process of inserting watermark data in an audio host file using the QIM method. The technique used is the combined technique of DST and QR. DST is only used for transformation. The audio host file is in .wav format. The steps for embedding watermark are described in the following steps:

1. Read the image file $w(m,n)$ used as a watermark.
2. Do the pre-process by changing the image size to one dimension.
3. Watermark image $w(n)$ performs a compressive sampling process that converts the sparse matrix into smaller forms.
4. The last step of CS will produce a compressed image matrix $W_c(n)$ which will be inserted into the host.
5. Read the $x(n)$ audio file that is used as the host. A segmentation process is performed which divides the host audio into multiple frames and the duration generated from the segmentation process depends on the number of frames desired.
6. Conduct DST process which is a transformation from time domain to frequency domain.
7. QR decomposition divides the one-dimensional matrix into a two-dimensional matrix, the matrix Q and R as (10). The watermark insertion process is performed in the first row of the matrix R .
8. Do the QIM process, watermark $W_c(n)$ that has been in the CS process is inserted using QIM method into the processed host.
9. Do the QR reconstruction which returns the matrix Q and R into the one-dimensional matrix.

10. Convert frequency domain to time domain with IDST. The result of this process is an audio host signal that has been embedded by the watermark $X_w(n)$.

From the above embedding steps, CS acquisition is executed in step 3 and 4. The embedding process by QIM technique is executed in step 8.

B. Extraction Process

Figure 2 shows the extraction process. The detail explanations of watermark extraction process are described in the following steps.

1. Read an audio host file that has been inserted watermark $X_w(n)$.
2. Transform the host audio by the DST technique.
3. Decompose the frequency domain signal by QR decomposition which divides two-dimensional matrices of matrix Q and matrix R . The matrix which will be extracted to get the watermark is $R_w(m,n)$ matrix.
4. The output of QR decomposition will be extracted by using QIM method that separate the audio host file with binary watermark image so as to produce $\hat{W}_c(n)$.
5. Perform the CS reconstruction on the watermark. The result of this process is $\hat{W}(n)$.
6. Finally, do the post-processing to obtain a watermark of a two-dimensional matrix or the original watermark $\hat{W}(m,n)$. In the end, it can produce the value of BER, SNR, and ODG.

IV. RESULT AND ANALYSIS

In this chapter, we use 5 types of audio hosts consisting of the host audio voice.wav, piano.wav, guitar.wav, drums.wav, and bass.wav. The watermark is in the form of binary image size 16x16 pixels as shown in Figure 3 with format of .bmp. We use Matlab as a tool for simulating the performance of the proposed audio watermarking method.

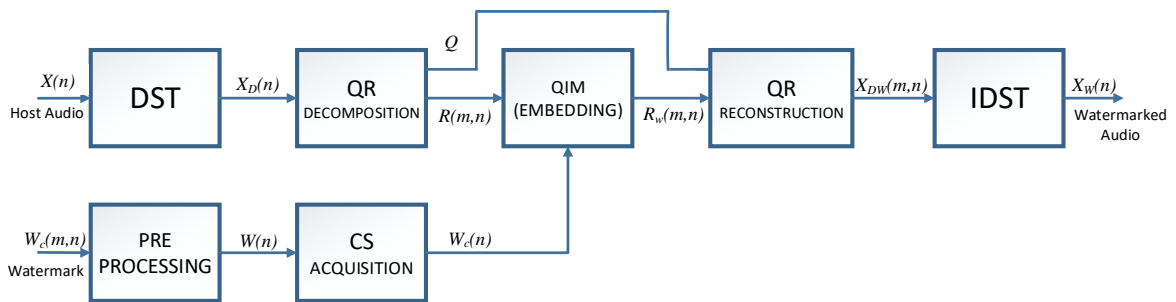


Figure 1. Flow diagram of embedding process

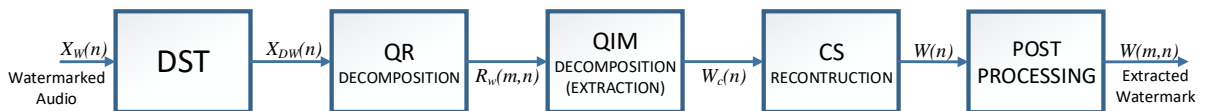


Figure 2. Flow diagram of extraction process



Figure 3. Binary image as watermark

A. Watermarking Performance without Attacks

The simulation is executed firstly with optimized compression ratio after CS acquisition which obtaining watermark compression ratio of 62.5%. After optimized result of compressed watermark is obtained, it is followed by testing the system by using non-attack parameters with the *Nframe* parameter which is to determine the number of frames; and *nbit* parameters which are to determine quantization bits in QIM. The results of this test will get the resilience value of the audio watermarking such as SNR, ODG, BER, and Capacity (C). The best parameter criterion is selected if the value of BER approaches 0, SNR value is more than 20 dB or close to 30 dB and ODG ranges from 0 to -4. Table 1 shows the best parameters selected after the tests are conducted.

B. Watermarking Performance with Attacks

In this chapter, the system will be tested using several kinds of attacks, including Low Pass Filter (LPF), Band Pass Filter (BPF), Resampling, Noise, Time Scale Modification, Linear Speed Change, Pitch Shifting, Equalizer, Echo and MP3 Compression. Figure 4 displays the watermarking robustness and imperceptibility. This figure is a result of watermarking method with MP3 attack in 128 kbps rate. Left y-axis shows the robustness and right y-axis shows the imperceptibility. We compare the watermarking performance when the quantization bit or *nbit* is 1, 3 and 5 bits. We can see that the higher the sample number per frame will be the higher the ODG and BER, or the lower the watermarking robustness.

Table 2 shows the average BER results of each audio host in various attacks. The one grey marked is the BER value which must be optimized again. Since the result of the test with the initial parameter gives a bad value of BER, optimization is required.

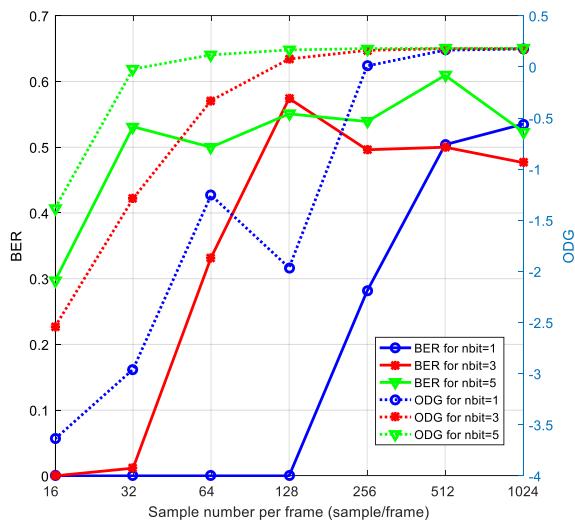


Figure 4. Watermarking robustness and imperceptibility

C. Optimization of Parameters with Attacks

This optimization process is done by taking some BER value data. Three data are taken, namely piano.wav audio host with attack BPF (100-6k) Hz which have BER = 0.26, host of guitar.wav audio with Pitch Shifting attack (1%) having BER = 0.39 and hosts bass.wav audio with Compression MP3 attacks (32k) Hz which has BER = 0.46. Table 3 shows the optimization result of the three parameters.

TABLE 1. RESULT OF BEST PARAMETERS WITHOUT ATTACKS

Nbit	NFrame	ODG	SNR	BER	C
1	512	0.16	31.26	0	98.43

TABLE 2. RESULT OF BER AVERAGE WITH VARIOUS ATTACKS

Attack	Parameter	BER				
		Voice	Piano	Guitar	Drums	Bass
LPF	3k	0	0	0	0	0
	6k	0	0	0	0	0
	9k	0	0	0	0	0
BPF	100-6k	0.41	0.26	0.31	0.57	0.6
	50-6k	0.22	0.17	0.28	0.46	0.44
	25-6k	0.15	0.08	0.3	0.42	0.43
Noise	0 dB	0.48	0.49	0.55	0.48	0.45
	10 dB	0.36	0.47	0.39	0.5	0.43
	20 dB	0.1	0.4	0.13	0.48	0.48
Resampling	22.05k	0	0	0	0	0
	11.025k	0	0	0	0	0
	16k	0	0	0	0	0
TSM	1%	0.01	0.007	0.007	0.01	0.01
	2%	0.01	0.02	0.01	0.01	0.02
	4%	0.19	0.16	0.02	0.1	0.03
LSC	1%	0	0	0	0	0
	5%	0	0	0	0	0
	10%	0	0	0	0	0
Pitch Shifting	1%	0.43	0.41	0.39	0.5	0.37
	2%	0.42	0.43	0.42	0.53	0.46
	4%	0.43	0.47	0.45	0.49	0.46
Equalizer		0.44	0.49	0.43	0.44	0.52
Echo		0.4	0.53	0.41	0.44	0.52
MP3 Compression	32k	0.04	0.51	0.07	0.31	0.46
	64k	0	0.48	0.06	0.31	0.43
	128k	0	0.46	0.01	0.3	0.43
	192k	0	0.28	0	0.14	0.52

Based on Table 3, it can be concluded that the best BER value is in the 3rd parameter, i.e., with the parameter $Nframe = 64$ and the $nbit = 1$ parameter for the MP3 compression (32k) Hz attack on the bass.wav audio host.

D. Watermarking Performance after Optimizing Robustness with Watermark

This section shows the results of testing the watermark resistance to attacks carried out on each audio host. Table 4 shows the test results of each audio host with attacks.

BER value of the extracted result will affect the watermark image quality. The higher the BER value, the poor quality of the watermark image is extracted. There is a maximum BER value on the watermark image that is still acceptable to human's eyes because visually human can still understand the contents of the watermark image even if it is damaged. The extracted watermark image and BER values are shown in Table 5. Based on the table, it can be concluded that the extracted watermark image can still be clearly seen in the BER range less than 10% using a 16x16 pixel watermark image resolution.

TABLE 3.
OPTIMIZATION RESULT OF BER ON EACH PARAMETER

Param-1	Param-2	Param-3
0.22	0.25	0.21

TABLE 4.
RESULT OF BER AND EXTRACTION PICTURE OF WATERMARK WITH OPTIMAL PARAMETER

Attack	Parameter	BER				
		Voice	Piano	Guitar	Drums	Bass
LPF	3k	0	0	0	0	0
	6k	0	0	0	0	0
	9k	0	0	0	0	0
Noise	0 dB	0.42	0.47	0.46	0.44	0.42
	10 dB	0.33	0.46	0.36	0.49	0.42
	20 dB	0.07	0.39	0.12	0.41	0.47
TSM	22.05k	0.007	0.007	0	0.01	0.01
	11.025k	0.01	0.02	0.007	0.01	0.01
	16k	0.17	0.14	0.02	0.1	0.03
Resampling	1%	0	0	0	0	0
	2%	0	0	0	0	0
	4%	0	0	0	0	0
LSC	1%	0	0	0	0	0
	5%	0	0	0	0	0
	10%	0	0	0	0	0
MP3 Compression	32k	0.02	0.47	0.06	0.26	0.45
	64k	0	0.47	0.03	0.25	0.42
	128k	0	0.45	0.01	0.27	0.42
	192k	0	0.23	0	0.13	0.52

E. Audio Watermarking Quality

There are two kinds of measuring the quality of audio watermarking. The first is measuring subjectively using Mean Opinion Score (MOS) and the second is measuring objectively using ODG and SNR. ODG is a testing parameter that is obtained from the calculation technique of Perceptual Evaluation of Audio Quality (PEAQ) which is stipulated in ITU-R BS.1387 standard (International Telecommunication Union Radio Communication Sector). SNR is a comparison between signal quality strength and noise power, while MOS is an audio quality test that is done with the help of human hearing ability. Table 6 shows the results of SNR, ODG, and MOS as the quality factors of each host audio watermarking resistance using the optimum parameters.

F. Performance Comparison

Table 7 displays the performance comparison consisting of imperceptibility, robustness, and capacity parameter performance. Our method in this paper has the highest capacity compared to the three previous methods, that is 689 bps with competitive robustness. Compared to previous method, as displayed in Table 4, our imperceptibility is also competitive than the imperceptibility in the previous paper. NA means not available, or no reporting. Our proposed method robustness against LPF and resampling are perfect. Nevertheless, the robustness of our method is worse than the previous method in MP3 attack.

TABLE 5.
VISUALIZATION OF EXTRACTED WATERMARK

BER	Extracted Watermark Image	BER	Extracted Watermark Image
0		0.23	
0.02		0.31	
0.08		0.43	
0.14		0.5	

TABLE 6.
RESULT OF SNR, ODG AND MOS WITH OPTIMAL PARAMETER

Host Audio	ODG	SNR	MOS
Voice.wav	-3.21	16.11	4.26
Piano.wav	-3.81	25.7	4.23
Gitar.wav	-3.47	12.09	4
Drums.wav	-3.91	19.99	3.73
Bass.wav	-3.79	30.74	4.03

TABLE 7.
PERFORMANCE COMPARISON

Ref.	Robustness / BER (%)				Imperceptibility/SNR (dB)	Capacity (bps)
	MP3 64 kbps	MP3 128 kbps	LPF 3k	Resampling 22.05 kHz		
[3]	0.9	0	1.8	0.4	18.7	50.0
[6]	0.1	NA	0	0	25.5	86.1
[9]	20	0	>40	0.5	9.4	40
Proposed	23	23	0	0	20.9	689

CONCLUSION

The larger the sample number per frame, the higher value of SNR, ODG, BER, and C are generated. Meanwhile, the larger *nbit* the higher value of SNR and ODG, for BER value is variable, and for the value of C does not affect anything or remain. The best parameter used is the optimal parameter with the selection of MP3 compression attack on bass.wav audio, to then displayed its watermark extraction with the output value with average of BER is 0.21, SNR value is more than 20 dB, with C is 689 bps.

Testing and analysis results with LPF (Low Pass Filter), Resampling and Linear Speed Change attacks, as well as some types of attacks on MP3 compression, are still capable of being retained by the system with the optimum parameters selected. The results of subjective evaluation of this system produce a good MOS (Mean Opinion Score) value with host audio voice.wav as the highest MOS value.

REFERENCES

- [1] F. Hartung, and M. Kutter, "Multimedia watermarking techniques," *Proc. IEEE*, vol. 87, no. 7, pp. 1079- 1107, 1999.
- [2] N. Cvejic, and T. Seppänen, in *Digital Audio Watermarking Techniques and Technologies : Applications and Benchmarks*. USA: Information Science Reference, 2008.
- [3] N. Khademi, M. A. Akhaee, S. M. Ahadi, M. Moradi, and A. Kashi, "Audio watermarking based on quantization index modulation in the frequency domain," in *Proc. 2007 IEEE Int. Conf. Signal Process. Commun.*, 2007, pp. 1127-1130.
- [4] F. Chaabane, M. Charfeddine, W. Puech, C. Ben-Amar, "A QR-code based audio watermarking technique for tracing traitors," in *Proc. European Signal Process. Conf.*, 2015, pp. 51-55.
- [5] V. K. Bhat , I. Sengupta, and A. Das, "Audio watermarking based on mean quantization in cepstrum domain," in *Proc. 2008 16th Int. Conf. Adv. Comput. Commun.*, 2008, pp. 73-77.
- [6] H. Yassine, B. Bachir, and K. Aziz, "A secure and high robust audio watermarking system for copyright protection," *Int. J. Comput. Appl.*, vol. 53, no. 17, pp. 33-39, 2012.
- [7] G. Budiman, A. B. Suksmono, D. Danudirdjo, "FFT-based audio watermarking in adaptive subband with spread spectrum framework," *2nd Advanced Research Elect. Electron. Eng. Technology*, 2017, pp. 3-8.
- [8] G. Budiman, A. B. Suksmono, D. H. Shin, "A Multicarrier modulation audio watermarking system," in *Proc. Int. Conf. Elec. Eng. Informatics*, 2015, pp. 154-160.
- [9] G. Budiman, A. B. Suksmono, D. Danudirdjo, K. Usman, D. H. Shin, "A modified multicarrier modulation binary data embedding in audio file," *Int. J. Elec. Eng. Informatics*, vol. 8, no. 4, pp. 762-773, 2016.
- [10] G. Budiman, A. B. Suksmono, D. Danudirdjo, "Fibonacci sequence based FFT and DCT performance comparison in audio watermarking," *Int. J. Eng. Tech.*, vol. 8, no. 19, pp. 1-6, 2019.
- [11] Q. Lyu, Z. Lin, et al., "A Comparison of typical L_p minimization algorithms," *Neurocomputing*, vol. 119, pp. 413-424, 2013.
- [12] A. B. Suksmono, "Progress in applied compressive sampling: a brief review on methods and devices," in *Proc. 2016 10th Int. Conf. Telecommun. Syst. Services Applicat.*, Denpasar, 2016, pp. 1-4.
- [13] H. Haider, J. A. Shah, U. Ali, "Comparative analysis of sparse signal recovery algorithms based on minimization norms" in *Proc. 2014 World Congr. Sustainable Technologies*, 2014.
- [14] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423-1443, 2001.
- [15] C. C. Tseng, S. Lee, "Closed-form design of FIR frequency selective filter using discrete sine transform," in *Proc. 2016 IEEE Asia Pacific Conf. Circuits Syst.*, 2017, pp. 591-594.