A New Audio Watermarking System using Discrete Fourier Transform for Copyright Protection

Pranab Kumar Dhar†, Mohammad Ibrahim Khan†, and Jong-Myon Kim[1]††

†Chittagong University of Engineering and Technology, Chittagong-4349, Bangladesh
†† University of Ulsan, San 29, Mugeo 2-Dong, Nam-Ku, Ulsan, 680-749, Korea

Summary
Digital watermarking is now drawing attention as a new method of protecting multimedia content from unauthorized copying. This paper proposes a new watermarking system using discrete Fourier transform (DFT) for copyright protection of digital contents. In our proposed watermarking system, the original audio is segmented into non-overlapping frames. Watermarks are then embedded into the highest prominent peak in the magnitude spectrum of each frame. Watermarks are extracted by performing the inverse operation of watermark embedding process. Simulation results indicate that the proposed watermarking system is highly robust against various kinds of attacks such as noise addition, cropping, re-sampling, re-quantization, and MP3 compression, and achieves similarity values ranging from 13 to 20. In addition, our proposed system achieves SNR (signal-to-noise ratio) values ranging from 20 dB to 28 dB.

Key words:
Copyright protection, digital watermarking, multimedia contents, and discrete Fourier transform

1. Introduction

The recent growth in computer networks, and more specifically, the World Wide Web, copyright protection of digital audio becomes more and more important. Digital audio watermarking has drawn extensive attention for copyright protection of audio data. A digital audio watermarking is a process of embedding watermarks into audio signal to show authenticity and ownership. Audio watermarking should meet the following requirements :(a) *Imperceptibility*: the digital watermark should not affect the quality of original audio signal after it is watermarked; (b) *Robustness*: the embedded watermark data should not be removed or eliminated by unauthorized distributors using common signal processing operations and attacks; (c) *Capacity*: capacity refers to the numbers of bits that can be embedded into the audio signal within a unit of time; (d) *Security*: security implies that the watermark can only be detectable by the authorized person. All these requirements are often contradictory with each other. However, it should satisfy the important properties such as imperceptibility and robustness.

In this paper, we propose a new watermarking system using discrete Fourier transform (DFT) for audio copyright protection. The watermarks are embedded into the highest prominent peak of the magnitude spectrum of each non-overlapping frame. Experimental results indicate that the proposed watermarking system provides strong robustness against several kinds of attacks such as noise addition, cropping, re-sampling, re-quantization, and MP3 compression and achieves similarity values ranging from 13 to 20. In addition, our proposed system achieves SNR (signal-to-noise ratio) values ranging from 20 dB to 28 dB.

The rest of this paper is organized as follows. Section 2 provides a brief description of previous works related to audio watermarking. Section 3 introduces our proposed watermarking system including watermark embedding process and watermark detection process. Section 4 discusses the performance of our proposed system in terms of imperceptibility as well as robustness. Finally section 5 concludes this paper.

2. Previous Works

A significant number of watermarking techniques have been reported in recent years in order to create robust and imperceptible audio watermarks. Lie *et al.* [1] propose a method of embedding watermarks into audio signals in the time domain. The proposed algorithm exploits differential average-of-absolute-amplitude relations within each group of audio samples to represent one-bit information. It also utilizes the low-frequency amplitude modification technique to scale the amplitudes in selected sections of samples so that the time domain waveform envelope can be almost preserved. In [2], authors propose a blind audio watermarking system which embeds watermarks into audio signal in time domain. The strength of the audio signal modifications is limited by the necessity to produce an output signal for watermark detection. The watermark signal is generated using a key, and watermark insertion

[1] Corresponding author

Manuscript received June 5, 2010
Manuscript revised June 20, 2010

depends on the amplitude and frequency of audio signal that minimizes the audibility of the watermarked signal. Ling *et al.* [3] introduce a watermarking scheme based on nonuniform discrete Fourier transform (NDFT), in which the frequency points of embedding watermark are selected by the secret key. Zeng *et al.* [4] describe a blind watermarking system which embeds watermarks into DCT coefficients by utilizing quantization index modulation technique. In [5], the authors propose a watermarking system which embeds synchronization signals in time domain to resist against several attacks. Pooyan *et al.* [6] introduce an audio watermarking system which embeds watermarks in wavelet domain. The watermarked data is then encrypted and combined with a synchronization code and embedded into low frequency coefficients of the sound in wavelet domain. The magnitude of quantization step and embedding strength is adaptively determined according to the characteristics of human auditory system. Wang *et al.* [7] proposes a blind audio watermarking scheme using adaptive quantization against synchronization attack. In addition, the multiresolution characteristics of discrete wavelet transform (DWT) and the energy compression characteristics of discrete cosine transform (DCT) are combined in this scheme to improve the transparency of digital watermark. Watermark is then embedded into low frequency components by using adaptive quantization according to human auditory system. In [8], authors propose a watermarking system in cepstrum domain in which a pseudo-random sequence is used as a watermark. The watermark is then weighted in the cepstrum domain according to the distribution of cepstral coefficients and the frequency masking characteristics of human auditory system. Liu *et al.* [9] propose a blind watermarking system which takes the advantages of the attack-invariant feature of the cepstrum domain and the error-correction capability of BCH code to increase the robustness as well as imperceptibility of audio watermarking.

In Cox's method [10] watermarks are embedded into the highest m DCT coefficient of the whole sound excluding the DC component by the following equation:

$$v_i' = v_i(1+\alpha x_i) \qquad (1)$$

where, m is the length of the watermark sequence, v_i is a magnitude coefficient into which a watermark is embedded, x_i is a watermark to be inserted into v_i, α is a scaling factor, and v_i' is an adjusted magnitude coefficient. The watermark sequence is extracted by performing the inverse operation of (1) represented by the following equation:

$$x_i^* = \left(\frac{v_i^*}{v_i} - 1\right)/\alpha \qquad (2)$$

3. Proposed Watermarking System

In this section, we give an overview of our basic watermarking system which consists of watermark embedding process and watermark detection process. In this implementation, a watermark consists of a sequence of real numbers **X**= $\{x_1, x_2, x_3,..., x_n\}$. We create a watermark where each value of x_i is chosen independently according to N(0,1) where $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2.

3.1 Watermark Embedding Process

The proposed watermark embedding process is shown in Fig. 1. The embedding process is implemented in the following seven steps:

1) The original audio is first segmented into non-overlapping frames.

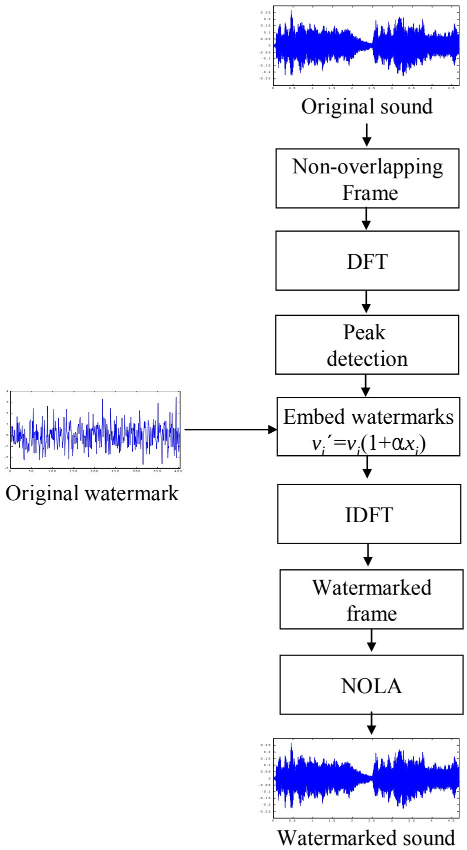

Fig. 1 Watermark embedding process

2) Calculate the magnitude and phase spectrum of each frame using discrete Fourier transform (DFT).

3) Find the most prominent peak **V** from magnitude spectrum using a peak detection algorithm.

4) Place watermarks into the highest prominent peak of the magnitude spectrum of each frame to obtain watermarked peak **V'**. This ensures that the watermark is located at the most significant perceptual components of the audio. When we insert the watermark X into V to obtain V', we specify a scaling parameter α, which determines the extent to which X alters V, shown in the equation (1) [10].
5) Insert back the modified peak into the magnitude spectrum of each non-overlapping frame.
6) Take an inverse DFT of the complex spectrum to calculate the watermarked frame.
7) Finally watermarked audio signal is computed by nonoverlap-adding (NOLA) the watermarked frames in the time domain.

3.2 Watermark Detection Process

The proposed watermark detection process is shown in Fig. 2. The detection process is implemented in the following three steps:
1) Calculate the DFT of the attacked watermark audio frame.

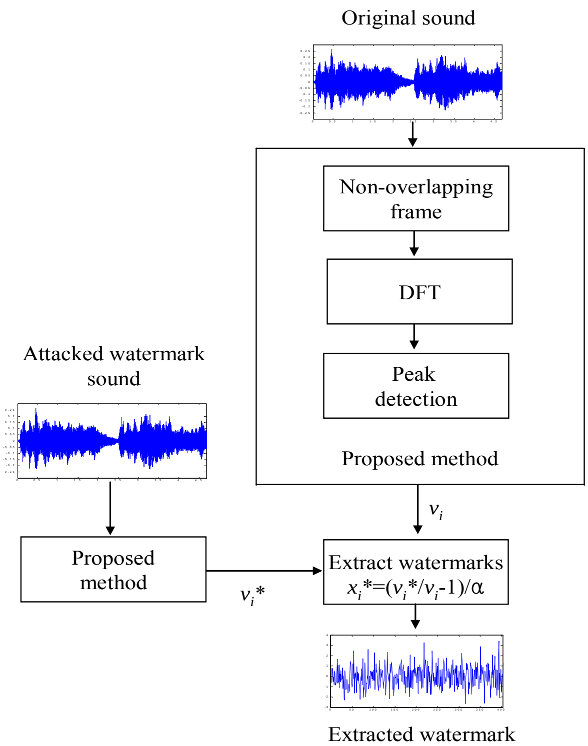

Fig. 2 Watermark detection process

2) Extract the highest prominent peak from the magnitude spectrum which is located at the same position in the embedding process above.

3) The watermark sequence $X^* = x_1^*, x_2^*, x_3^*, \ldots, x_n^*$ is then extracted by performing the inverse operation of (1) represented by the equation (2).

4. Simulation Results

In this section, we evaluate the performance of our watermarking system for four different types of 16 bit mono audio signals sampled at 44.1 kHz: (a) the song 'Let it Be,' by the Beatles; (b) the beginning of Symphony No. 5 in C Minor, Op. 67, by Ludwig van Beethoven; (c) an instrumental song 'Hey Jude' played by a Korean traditional musical instrument called the gayageum; (d) a human voice providing TOEIC (Test of English for International Communication) listening test instruction. Each audio file contains 206,336 samples (duration 4.679 sec). By considering the frame size of 512 samples, we have 403 frames for each audio sample. From each frame we detect 1 peak to embed watermark. Thus, the length of the watermark sequence is 403×1=403.
In order to evaluate the performance of the proposed watermarking system in terms of watermark detection, the correlation coefficient between the original watermark X and the extracted watermark X^* is calculated by the following similarity SIM(X, X^*) formula:

$$SIM(X, X^*) = \frac{X \cdot X^*}{\sqrt{X^* \cdot X^*}} \qquad (3)$$

It is highly unlikely that X^* will be identical to X. To decide whether X and X^* match, we determine whether the SIM(X, X^*) > T, where T is a detection threshold. In this study, the selected detection threshold (T) value is 6 [10].
Fig. 3 shows a qualitative evaluation of the original audio with a watermarked audio in which the watermarks are imperceptible using the proposed system.
In order to evaluate the quality of watermarked signal, the following signal-to-noise ratio (SNR) equation is used:

$$SNR = 10\log_{10}\frac{\sum_{n=1}^{N} S^2(n)}{\sum_{n=1}^{N}\left[S(n) - S^*(n)\right]^2} \qquad (4)$$

where S(n) and S^*(n) are original audio signal and watermarked audio signal respectively. After embedding watermark, the SNR of all selected audio signals using the proposed method are above 20 dB which ensures the imperceptibility of our proposed system.

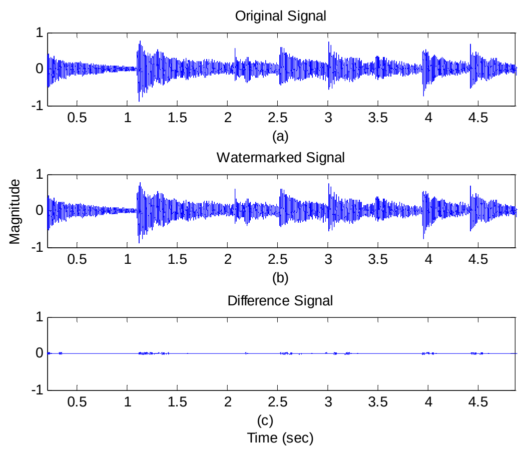

Fig.3 Imperceptibility of the watermarked audio using the proposed method: (a) original sound 'Hey Jude', (b) watermarked sound 'Hey Jude' (c) difference between original and watermarked sound

Fig. 4 shows the peak detection of the first frame of original sound 'Hey Jude'. In our proposed system, watermarks are embedded into the highest prominent peak of the magnitude spectrum of each frame which provides high robustness against different kinds of attacks as well as good SNR values for watermarked audio signals.

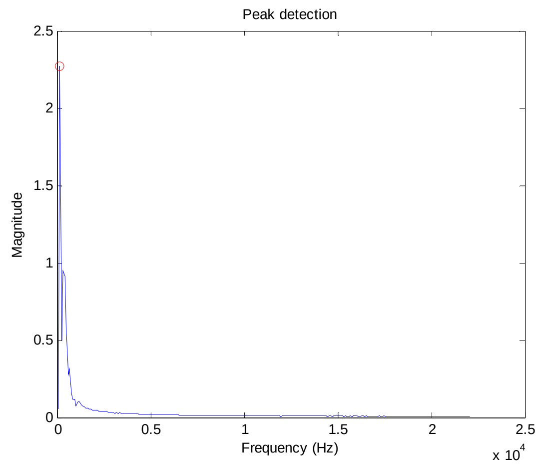

Fig .4 Peak detection in frequency spectrum

Table 1 shows the SNR results of the proposed watermarking system for different values of α. Our proposed system achieves SNR values ranging from 20 dB to 28 dB for different watermarked sounds.

Table 1: SNR results of the proposed watermarking system

Types of Signal	**SNR**		
	α=0.1	**α=0.2**	**α=0.3**
Let it be	25.654	23.176	20.202
Symphony No 5	27.395	23.652	20.649
Hey Jude	27.218	23.843	20.776
Human Voice	28.259	24.163	20.779

4.1 Imperceptibility Test

Informal listening using head set reveals that the watermark embedded into the original audio using the proposed watermarking system does not affect the quality of the sound, which ensures the imperceptibility of the embedded watermark.

4.2 Robustness Test

Table 2 shows the performance of our proposed system in terms of similarity when no attack is applied to four different types of watermarked audio signals.

Table 2: Watermark detection results of the proposed system without attacks

Types of Signal	**SIM**		
	α=0.1	**α=0.2**	**α=0.3**
Let it be	22.6345	22.7255	22.7841
Symphony No 5	22.4371	22.6285	22.6873
Hey Jude	22.5267	22.5932	22.6272
Human Voice	22.2483	22.4173	22.4756

In order to test the robustness of our proposed scheme, five different types of attacks, summarized in Table 3, were performed to the watermarked audio signal.

Table 3: Attacks used in this study for the watermarked sound

Attacks	**Description**
Noise addition	Additive white Gaussian noise (AWGN) is added with the watermarked audio signal.
Croppig	We removed 10% samples from the beginning of the watermarked signal and then replaced these samples by the original signal.
Resampling	The watermarked signal originally sampled at 44.1 kHz is resampled at 22.050 kHz, and then restored by sampling again at 44.1 kHz.
Re-quantization	The 16 bit watermarked audio signal is quantized down to 8 bits/sample and again re-quantized back to 16 bits/sample.
MP3 Compression	MPEG-1 layer 3 compression with 128 kbps is applied to the watermarked signal.

Figures 5 and 6 show the response of the watermark detector to 240 randomly generated watermarks where correct watermark is at the 120th position against Gaussian noise attack for α=0.3 using the proposed system.

Tables 4, 5, 6 and 7 show the similarity results of the proposed watermarking system in terms of robustness for different values of α against several kinds of attacks applied to four different types of watermarked audio signal 'Let it be', 'Symphony No 5', 'Hey Jude', and 'human voice' respectively. We observe that when α increases, similarity results of the proposed system also increase.

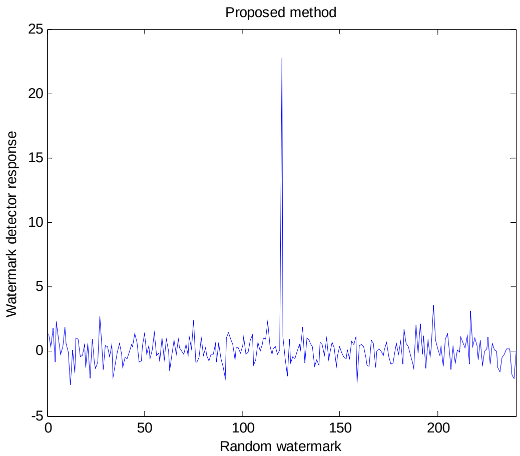

Fig. 5. Watermark detector response using the proposed method

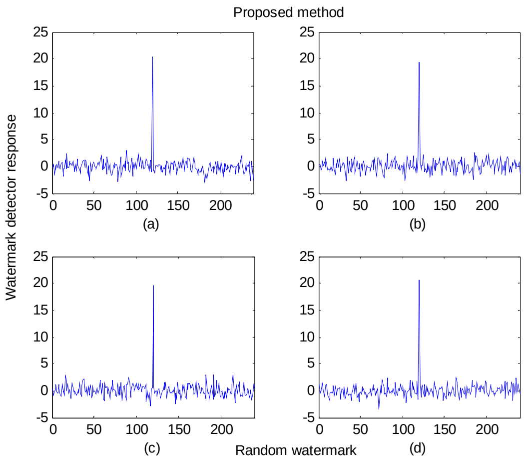

Fig. 6 Watermark detector response against Gaussian noise attack using the proposed method: (a) Let it be, (b) Symphony No. 5, (c) Hey Jude, (d) Human Voice

Table 4. Similarity results of the proposed system against different attacks for the audio signal 'Let it be'

Types of attack	**SIM**		
	α=0.1	**α=0.2**	**α=0.3**
Noise addition	19.913	20.325	20.538
Cropping	13.652	15.824	17.372
Re-sampling	20.555	21.543	21.846
Re-quantization	20.286	21.725	21.948
MP3 Compression	16.735	18.247	20.357

Table 5. Similarity results of the proposed system against different attacks for the audio signal 'Symphony No 5'

Types of attack	**SIM**		
	α=0.1	**α=0.2**	**α=0.3**
Noise addition	18.481	19.317	19.424
Cropping	17.493	18.527	19.478
Re-sampling	21.790	21.815	21.936
Re-quantization	21.921	21.957	21.983
MP3 Compression	15.352	17.528	18.674

Table 6. Similarity results of the proposed system against different attacks for the audio signal 'Hey Jude'

Types of attack	**SIM**		
	α=0.1	**α=0.2**	**α=0.3**
Noise addition	17.805	18.528	19.725
Cropping	16.737	17.419	18.348
Re-sampling	19.977	20.576	20.865
Re-quantization	17.103	18.214	18.795
MP3 Compression	16.413	18.528	19.213

Table 7. Similarity results of the proposed system against different attacks for the audio signal 'Human voice'

Types of attack	**SIM**		
	α=0.1	**α=0.2**	**α=0.3**
Noise addition	19.018	20.218	20.735
Cropping	14.468	16.583	17.672
Re-sampling	18.238	19.567	19.784
Re-quantization	19.072	20.136	20.538
MP3 Compression	15.682	17.258	18.385

4. Conclusion

In this paper, we have presented a new watermarking system using discrete Fourier transform (DFT) for copyright protection of sound contents. Experimental results indicate that our proposed watermarking system shows strong robustness against several kinds of attacks such as noise addition, cropping, re-sampling, re-quantization, and MP3 compression and achieves similarity values ranging from 13 to 20. In addition, our proposed system achieves SNR values ranging from 20 dB to 28 dB for different watermarked sounds. These results demonstrate that our proposed watermarking system can be a suitable candidate for audio copyright protection.

References

[1] W. N. Lie and L. C. Chang, “Robust and High-Quality Time-Domain Audio Watermarking Based on Low-Frequency Amplitude Modification,” IEEE Transaction on Multimedia, vol. 8, no. 1, pp. 46-59, February, 2006.

[2] P. Bassia, I. Pitas and N. Nikolaidis “Robust Audio Watermarking in the Time domain,” IEEE Transaction on Multimedia, vol. 3, no. 2, pp. 232-241, June, 2001.

[3] L. Xie, J. Zhang and H. He, “Robust Audio Watermarking Scheme Based on Nonuniform Discrete Fourier Transform,” in Proceedings of IEEE International Conference on Engineering of Intelligent System, pp. 1-5, 2006.

[4] G. Zeng and Z. Qiu, “Audio Watermarking in DCT: Embedding Strategy and Algorithm,” in Proceedings of 9th International Conference on Signal Processing (ICSP’09), pp. 2193-2196, 2008.

[5] J. Huang, Y. Wang, and Y. Q. Shi, “A Blind Audio Watermarking Algorithm with Self-Synchronization,” in

Proceedings of IEEE International Symposium on Circuits and Systems, (ISCAS'02), vol. 3, pp. 627–630, 2002.

[6] M. Pooyan and A. Delforouzi, "Adaptive and Robust Audio watermarking in Wavelet Domain," in Proceedings of International Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP 2007), vol. 2, pp. 287-290, 2007.

[7] X. Y. Wang, H. Zhao, "A novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT," IEEE Transaction on Signal Processing, vol. 54, no. 12, pp 4835-4840, December 2006.

[8] S. K. Lee and Y. S. Ho, "Digital audio watermarking in the cepstrum domain," IEEE Transactions on Consumer Electronics, vol. 46, no. 3, pp. 744–750, 2000.

[9] S. C. Liu and S. D. Lin, "BCH code based robust audio watermarking in the cepstrum domain," Journal of Information Science and Engineering, vol. 22, pp. 535–543, 2006.

[10] I. Cox, J. Killian, F. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," IEEE Transactions on Image Processing, vol. 6, no. 12, pp. 1673-1687, 1997.

Pranab Kumar Dhar received the B.S. degree in Computer Science and Engineering from Chittagong University of Engineering and Technology (CUET), Chittagong, Bangladesh in 2004. He received M.S. degree from the School of Computer Engineering and Information Technology at University of Ulsan, South Korea in 2010. Since 2005, he has been serving as a faculty member in the Department of Computer Science and Engineering at Chittagong University of Engineering and Technology (CUET), Chittagong, Bangladesh. His research interest includes Digital Watermarking, Sound Synthesis, Digital Image Processing, Multimedia Systems, and Digital Signal Processing.

Mohammad Ibrahim Khan received the B.S. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh in 1999. He received M.S. degree in Computer Science and Engineering from the same University in 2002. He received his Ph.D degree in Computer Science and Engineering from Jahangirnagar University in 2010. Since 1999, he has been serving as a faculty member in the Department of Computer Science and Engineering at Chittagong University of Engineering and Technology (CUET), Chittagong, Bangladesh. His research interest includes Digital Image Processing, Graph Theory, Cryptography, Digital Watermarking, Multimedia Systems, and Digital Signal Processing.

Jong-Myon Kim received the B.S. degree in Electrical Engineering from Myongji University, South Korea in 1995. He received M.S. degree from the School of Electrical and Computer Engineering at University of Florida, Gainesville, FL, USA in 2000. He received his Ph.D. degree from the School of Electrical and Computer Engineering at Georgia Tech, Atlanta, GA in 2005. Currently he is a faculty member in the School of Computer Engineering and Information Technology at University of Ulsan, South Korea. His research interest includes high performance, energy efficient computer architectures, application specific SOC design for embedded color image and video processing, Digital Watermarking, Sound Synthesis, Multimedia Systems, and Digital Signal Processing. Prior to joining academia, he was a processor architect in both multi-format CODEC chip design and prototyping team and reconfigurable processor team at Samsung Advanced Institute of Technology (SAIT), South Korea. He is actively participating as a member of variety of professional research societies, like IEEE, KIPS, KIISE, KSCI, and EASKO.