University Of AL-Qadisiyah

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

A Survey paper on Lossy Audio Compression Methods

Zinah S. Abduljabbar[a] , Zainab J. Ahmed[b] , Anwar H. Al-Saleh[c]

[a] Mustansiriyah University,Department of Computer Science, Baghdad, Iraq, Email: zinahsadeq@uomustansiriyah.edu.iq
[b] University of Baghdad, Department of Biology Science, Baghdad, Iraq, Email: zainabjahmed83@gmail.com
[c] Mustansiriyah University,Department of Computer Science, Baghdad, Iraq, Email: anwar.h.m@uomustansiriyah.edu.iq

ARTICLE INFO

Article history:
Received: 22 /06/2021
Rrevised form: 10 /07/2021
Accepted : 12 /10/2021
Available online: 12 /10/2021

Keywords:

compression types,

audio compression,

speech compression,

lossy type.

ABSTRACT

During the two last decades ago, audio compression becomes the topic of many types of research due to the importance of this field which reflecting on the storage capacity and the transmission requirement. The rapid development of the computer industry increases the demand for audio data with high quality and accordingly, there is great importance for the development of audio compression technologies, lossy and lossless are the two categories of compression. This paper aims to review the techniques of the lossy audio compression methods, summarize the importance and the uses of each method.

MSC. 41A25; 41A35; 41A36

DOI : https://doi.org/10.29304/jqcm.2021.13.3.853

1. Introduction

Speech is the most important means of communication between humans and for this reason, the compression of speech data is a very important topic which considered a challenge[1], especially in the rapid, growth of communication technology.

Speech compression is an operation of converting the speech signals of human into a compact representation and the product of the decoding process must the closest form of the original signal [2]. The motivation of any compression process is decreasing the data required to represent the original data and thus lead to faster data transfer and reduced storage cost.

The techniques of compressed any type of data are falling into two types, either lossless or lossy categories. The result of a lossless technique is the same signal that was entered into the encoding unit [3]. However, when the retrieved signal does not match the signal entered in this case, the loss type is present [4]. Text and images for

*Corresponding author Zinah S. Abduljabbar

Email addresses: zinahsadeq@uomustansiriyah.edu.iq

Communicated by: Dr. Rana Jumaa Surayh aljanabi.

medical and satellite type must be encoded with lossless method type, while audio, video, and some types of images may be encoded using the lossy technique. In general, the gain in compression ratio is more in the lossy type while the quality of the retrieved signal is undoubtedly better in the lossless type. Removing repetition in the input signal is the essence of any compression process, in the audio signal, the type of redundancy are perceptual irrelevancies, temporal redundancy, knowledge redundancy and statistical redundancy. Fidelity criteria of types objective and subjective are used to evaluate the lossy audio coding. Listening to the audio before and after coding is a subjective criterion, human is observing the quality by playing back the signal and check the quality when the mathematical expression is used to represent the loss of the original signal then this represents an objective criterion. Figure 1 presents the most common types of audio lossy compression methods [5].

2. *Audio lossy compression methods*

In this section, the most important and most common methods which are used to compress audio data have been highlighted. Lossy compression methods can be classified as presents in Figure (1) into four categories such as: quantization based, prediction based, frequency based compression and fractal compression.

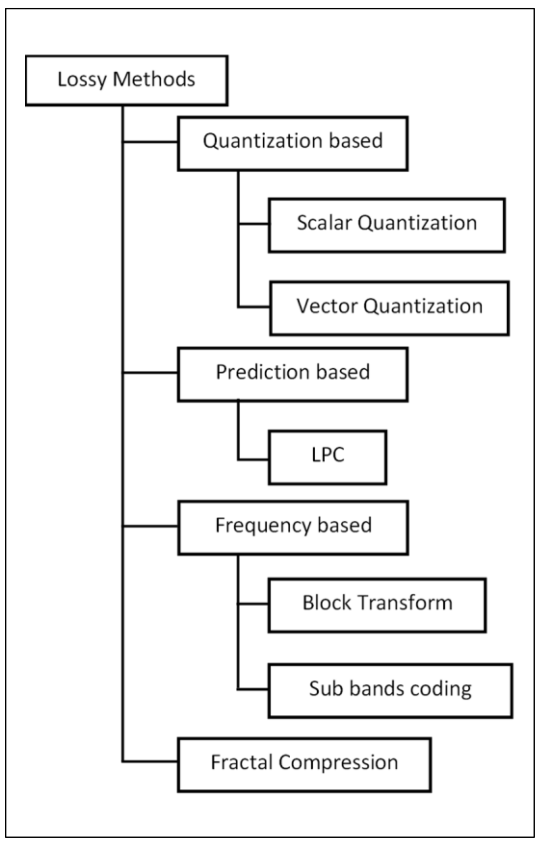

Fig1 - The most common types of audio lossy compression

2.1 Quantization

It's the most significant approach used in lossy methods which can be defined as the process of minimizing the alphabet size required to represent the desired output. A restriction of any float number types to an integer value is a quantization. The gain from quantization operation depends on the design of the quantizer which is reflected in the results. When more bits are allocated in encoding mapping, this leads to less noise and more bits are required to represent the result. On the contrary, fewer bits means more error and few bits are required to represent the result [6]. Two types of quantization are scalar quantization and vector quantization.

2.1.1 Scalar Quantization

In scalar type, the quantizer handles and quantizes each sample separately [universal rate-efficient]. Uniform and non-uniform quantization are the two categories of scalar quantization. Two differences between the types of scalar quantization, firstly the step sizes in uniform are equal while in the non-uniform type are not equal. The second variation is the quantization error, wherein the uniform type of the associated error is more than the other type [4]. Scalar quantization is often used in combination with many lossy techniques to improve the results of the compression ratio.

2.1.2 Vector Quantization

Two essential steps involved in the vector quantization (VQ) procedure: codebook generation and code vector matching. In codebook generation (training step) the similar vectors are grouped to form the clusters and every cluster is allocated to a single representative vector named a code vector, in the second step (coding or code vector matching) compressed process is performed by replacing every input vector with the closest code vector referenced by a simple index of the cluster. Index only is transmitted to the decoder and the transmitted index is used to return the required code vectors from the codebook [7]. Mohammed A. F. presented a comparison between the genetic algorithm VQ and K-means VQ compression algorithm on sound data, a good enhancement was recorded for the performance of the VQ method when mixing with a genetic algorithm where the achieved compression ratio ranges from 6 to 11 with accepted PSNR [8]. H. B. Kekre and Tanuja K. Sarode introduced the VQ technique for speech coding. Three methods were used (LBG, Keke's fast codebook generation (FCG) and Keke's Proportionate Error (KPE)). 256 was the size of the generated codebooks for six samples of various speech taking into consideration 16 and 32 the dimension of vector, 8 kHz and 16 bits speech signal was used. The results recorded indicated that, in terms of speed and performance criteria (MAE and AFCSS), the VQ of type FCG Algorithm was the best [9].

2.2 Prediction

The change of signal is slow so there is a high correlation between samples in the signal consequently, the neighboring samples have the same information with a small difference [10]. Prediction coding methods exploit the correlation to predict the next sample from the previous one [11].

2.2.1 Linear predictive coding (LPC)

This method is called parametric based speech coding because it depends on the production of the speech on the receiver side (depending on specific information related to the speech file) instead of sending the samples of the speech waveform [12]. Two components involved in this model: analysis (encoding) and synthesis (decoding). In the encoding part, the signal is breaking down into several frames and each frame is inquiring about the following factors: voiced or unvoiced frame, frame's pitch and the required parameters for building a filter in the receiving part that model the vocal track. According to the above factors, the filter is built in the synthesis part to produce the speech signal [4]. A block diagram of LPC synthesis is shown in the figure below:

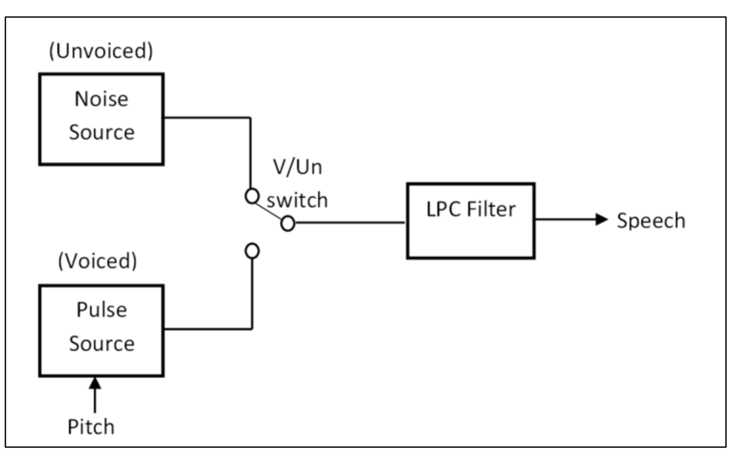

Fig2 – A block diagram of LPC synthesis

P. Venkateswaran and et al suggested an algorithm for audio compression purpose, in this algorithm two techniques were used (LPC and sub bands coding), and the results indicated reducing in the rate of data transmission from 128 Kbps to 12 Kbps with low complexity [13]. Nikhil Sharma suggested two methods, in the first method, the speech signal was encoding using LPC and synthesis filter was used to reconstruct the speech signal, while in the second method DCT was used to compress residual signal which obtained from LPC. The results indicate high quality with low bit rate when using LPC without DCT, while low bit rate with low quality for the reconstructed speech when using DCT [14]. Zinah S. and Noor Khaled suggested method for speech compression method, where the speech signal was enhanced by using normalized least mean square filter, in the encoding phase the enhanced signal was analyzed and synthesized by using LPC, the difference vector was computed by taking the difference between the synthesized vector and the input enhanced vector, then the difference vector compressed using DCT. The decoding phase decompressed the speech signal using pitch, voiced/unvoiced bits which were sent from the LPC analyzer and the difference vector was utilized to improve the quality of the reconstructed speech signal. Three samples with different sampling rate were used and the recorded results showed a high compression rate with good quality for the reconstructed speech signal [15].

2.3 Frequency domain

In the time domain, the large signal may mask the other signals and there is no doubt that the frequency range can be used as a useful tool for analyzing the speech signal. More accurate information can be obtained by performing a mathematical transformation which leads to improvement in the gain of the compression process. Block transform and Sub bands coding is used to obtain the frequency representation to the signal.

2.3.1 Discrete Cosine Transform (DCT)

This type of transformation had been originated by [Ahmed et al. 74], the characteristics of this transformation can be benefited in the compression process where the energy of the input data is concentrated in just the first few coefficients. Consequently, the first coefficients contain the most important audio information (low-frequency) while the other coefficients contain the less important audio information (high-frequency) [16]. DCT is an example of block transform and this cause artifacts [17].

The one dimension of this transform is described in the following equations:

$$Gf = \sqrt{\frac{2}{n}} Cf \sum_{t=0}^{n-1} pt \cos\left[\frac{(2t+1)f\pi}{2n}\right] \qquad (1)$$

$$Cf = \begin{cases} \frac{1}{\sqrt{2}} & f = 0 \\ 1 & f > 0 \end{cases} \quad for\ f = 0,1,\cdots,n-1 \qquad (2)$$

Where:

N is the number of audio elements

pt Audio samples block

Gf The DCT coefficients

The inverse DCT is explained in the following equation:

$$pt = \sqrt{\frac{2}{n}} \sum_{j=0}^{n-1} Cj\, Gj\ \cos\left[\frac{(2t+1)j\pi}{2n}\right]\ for\ t = 0,1,\cdots,n-1 \qquad (3)$$

Alvarado and Garcia suggested a technique for the audio signal, joint implementation of DCT used as a method to represent a sparse audio signal with the application of compressive sampling algorithm to the audio sparse representation [18].

Zinah and Dhia suggested a technique for stereo files compression, After the framing and normalizing process DCT was implemented to obtain the frequency representation then in the two channels each equal adjacent samples were isolated, Run Length Encoding (RLE) was used as entropy coding, the recorded results were promising with high quality [19].

Sankalp Shukla and et al proposed a method for audio compression. DCT was implemented on audio data and the DCT coefficients were quantized using scalar quantization then Lempel-Ziv-Welch method (LZW) was applied to remove the redundancy, the results indicate appropriate performance but may be enhanced by changing some parameters [20].

2.3.2 Discrete wavelet transforms (DWT)

Wavelet transform analyzing the input data according to scale is the essential idea behind wavelet transform. DWT considered a type of sub band coding due to split the signal, by using a band-pass filter, into several sub bands. Filter banks in DWT are utilizing to decompose the signal into approximation and detail. Blocking artifacts that are marked in DWT can be averted by using DWT [21].

M. V. Patil and et al suggested a simple technique for audio compression by applying DWT and DCT. The results indicated an improvement in compression factor and the quality of reconstruction audio when using DWT [22].

Swapnil T. Dumbre and Neeta B. Bankhele introduced a method for audio compression based on DWT, this method implemented through three stages: transformation (DWT), quantization and encoding. The wave signal decomposed using Daubechies wavelet family with different levels and the results showed good performance when level 3 was selected [23].

Zainab T. Drweesh and Loay E.George suggested coding scheme in their work where the audio signal was normalized followed by wavelet transformation (tap 9/7) then progressive hierarchal type of quantization was used to quantize the coefficients, RLE used to minimized the runs of zeros then high order shift encoding was used finally, post filtering was used to improve the quality. The results shows that the CR is increased when the passes of wavelets are increased [24].

Zainab and et al presents a technique depends on combine transforms (wavelet transform and DCT) coding scheme, to decompose the signal into low and multi high sub bands, bi-orthogonal (tab 9/7) wavelet transform was used then DCT was used to de-correlate the sub-bands then quantization (progressive hierarchical) applied followed by Run Length Encoding and finally LZW was applied to generate the compressed stream [25].

2.4 Fractal Audio Compression (FAC)

There is a redundant (self-similarity) in the audio signal. FAC can exploit these redundancies for compression purpose, the encoding unit and decoding units of fractal compression can be summarized as the figures below [5].

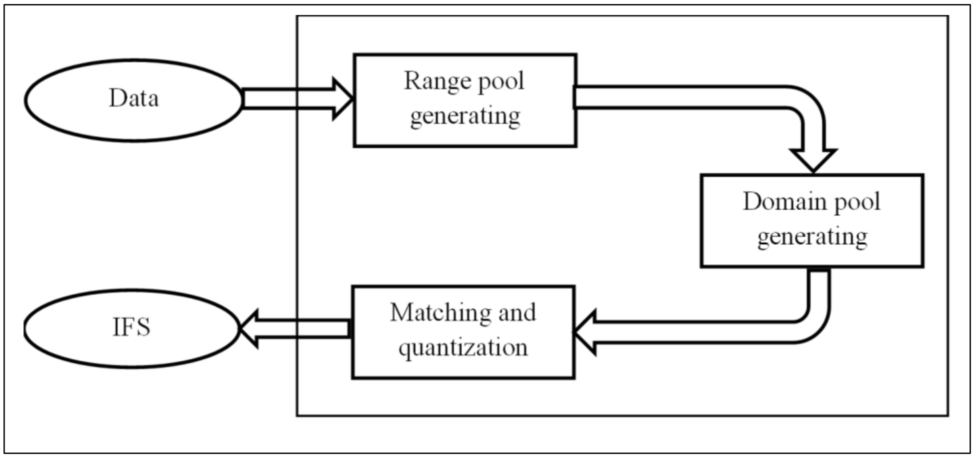

Fig3 – Encoding unit

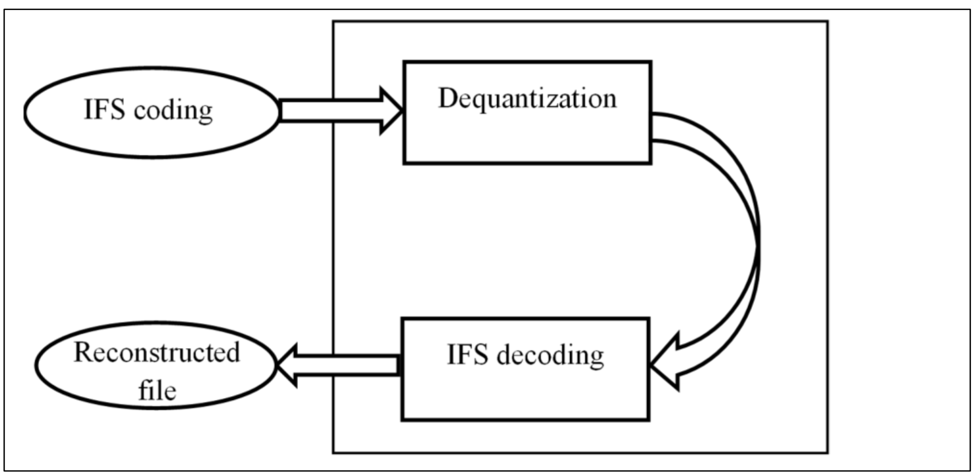

Fig4 – Decoding unit

Van and Axel were the first to discuss FAC in 1995 [26]. Wannamaker, Robert A. and Vrscay, Edward R. were Introduced their work as a novel approach, they introduced an approach for compression the audio signal by combining the wavelet transform with the FC where the fast wavelet transform applied on the signal frames and FC applied on high-frequency wavelet coefficients, the compression ratio up 6:1 with good quality for the reconstructed audio signal [27].

Zahraa A. Hasan and Loay E. George introduced a method for audio compression based on fractal coding (FC) and wavelet transform, in their paper wavelet transform was used to eliminate the drawback associated with FC when coding long-range blocks. The results indicated, there was an improvement in the value of compression ratio when the length of the block was increased and the PSNR value reached 30 dB [28].

3. Discussion

The main goal of the compression process is to reduce the space taken up and increase the data transfer and this can be achieved by applying different compression methods [29, 30] . Despite many techniques have emerged which used the previously mentioned methods during the last period in order to reach good results but no method can satisfy the requirements for all types of audio files, many factors must be taken into consideration for example when a signal has a high correlation the DCT is utilized while DWT is appropriate due the localization characteristic over the time – frequency domain [22, 31]. LPC is established on the source-filter model and despite of using LPC in compression process but the results are considerable when it used as a formant extraction tool. Promising results can be obtained by applying fractal method but the drawback of this process is the delay in reconstruction of audio file.

It is difficult to make a fair comparison between all the methods due to the lack of the same samples used in these researches, especially since they are not standardized and each research used different samples.

It is possible to include the results for the methods used by the authors in their previous work and for the same sample. An audio test sample is used to compare between two techniques of compression method which used previously in two papers [15,25], the characteristics of the audio test sample are presented in the table below:

Table 1- The characteristics of the selected sample.

Sampling rate (KHz)	Sampling resolution (bps)	Size (KB)	Audio Type
44100	8	216	Dialog (female)

The sample is a wave file format, PCM data format and mono channel. Figure (5) presents the pattern of waveform of the selected sample.

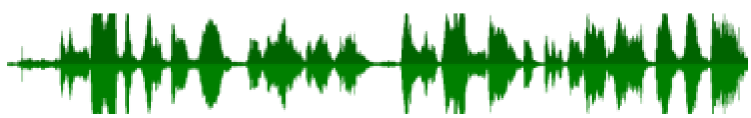

Fig5 - The waveform of the selected sample

The results presented in table below showed that the compression technique in [25] achieved better results for the selected sample.

Table 2- CR and PSNR using two methods

First method [25]		Second Method [15]	
CR	PSNR	CR	PSNR
15.230	35.423	14.107	32.754

4. Conclusions

In this paper, the general techniques of lossy audio compression are introduced. The numbers of compression techniques are expanded and some of them are in process. It's difficult to balance between the compression gain and the quality of the reconstructed signal so many techniques may be combined to implement a good system for compression and so many of previous works use a hybrid methods in order to achieve a good results. It's difficult to use specific method for all type of audio files, the results of compression affected by characteristics of the file.

Acknowledgement

The authors thanks to the "Department of Computer Science", "College of Science", "Mustansiriyah University", for supporting this work.

5. References

[1] I. Mengyi Pu, "Fundamental Data Compression", Elsevier Ltd 2005.
[2] A.M.M.A. Najih, A.R. bin Ramli, V. Prakash and A.R. Syed "Speech compression using discreet wavelet transform" *IEEE*, 2003.
[3] D. T. Yang, C.-C. Jay Kuo and C. Kyriakakis, "High Fidelity Multichannel Audio coding" *Hindawi Publishing Corporation* (2006).
[4] K. Sayood, "Introduction to Data Compression", *Elsevier* Inc (2006).
[5] H. Baker, "speeding -up fractal audio compression using double moments descriptor", thesis, 2013.
[6] B. Cavagnolo and J. Bier, "Introduction to Digital Audio Compression", *Berkeley Design Technology, Berkeley*, CA, June 2000.

[7] S.-En Qian, "Optical Satellite Data Compression and Implementation" , *SPIE*, 2013.
[8] M. A. F. Al-Husainy, "Sound Data Compression Method Using Genetic Algorithm", *Damascus Univ. Journal*, Vol. 23, No. 2, pp. 39-55, 2007.
[9] H. B. Kekre and T. K. Sarode, "Speech Data Compression using Vector Quantization", *International Journal of Computer and Information Engineering,* Vol. 2, No. 3, pp. 792-795, 2008.
[10] J. S. Chitode, "Digital Communications", *Technical Publications Pune*, 2008.
[11] H. Austerlitz, "Data Acquisition Techniques Using PCs", *Elsevier*, 2003.
[12] K. Pramila. R and G. Shital. S, "A Survey Paper on Different Speech Compression Techniques", *IJARIIE*, Vol. 2, No.5, pp.736-741, 2016.
[13] P. Venkateswaran, A. Sanyal, S. Das, R. Nandi and S. K. Sanyal, "An Efficient Time-Domain Speech Compression Algorithm Based on LPC and Sub-Band Coding Techniques", *J. of communications*, Vol. 4, NO. 6, pp.423- 428, JULY 2009.
[14] Nikhil Sharma, "Speech Compression Using Linear Predictive Coding (LPC)", *International Journal of Advanced Research in Engineering and Applied Sciences*, Vol. 1, No. 5, pp.17-28, November 2012.
[15] Z. S. Abduljabbar and N. Khaled, "Speech Enhancement Based Compression Method", *Network and Complex Systems*, Vol.8, pp. 10-16, 2018.
[16] D. Salmon. "Data Compression the Complete Reference". London: *Springer-Verlag*, (2007).
[17] A. Katharotiya, S. Patel and M. Goyani, "Comparative Analysis between DCT & DWT Techniques of Image Compression", *J. of Information Engineering and Applications*, Vol. 1, No.2, pp. 9-17, 2011.
[18] R.G. Moreno A., M.M. Garcia, "DCT-Compressive Sampling of Frequency-Sparse Audio Signals", *In proceedings of the World Congress on Engineering*, Vol. II , London, U.K, July 6 - 8, 2011.
[19] Z. S. Abduljabbar and D. A. Alzubaydi, "Audio Compression Based on Isolation of Equal Adjacent", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 5, No. 9, September 2015.
[20] S. Shukla, M. Ahirwar, R. Gupta, S. Jain and D. S. Rajput, "Audio Compression Algorithm using Discrete Cosine Transform (DCT) and Lempel-Ziv-Welch (LZW) Encoding Method", *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con)*, India, pp. 476-480, 14th -16th Feb 2019.

[21] H. M. Parmar, "Comparison of DCT and Wavelet-based Image Compression Techniques", *International Journal of Engineering Development and Research*, Vol. 2, No. 1, pp. 664-669, 2014.
[22] M. V. Patil, A. Gupta, A. Varma and S. Salil, "Audio and Speech Compression Using DCT and DWT Techniques", *International Journal of Innovative Research in Science, Engineering and Technology*, Vol. 2, No. 5, pp. 1712-1719, May 2013.
[23] S. T. Dumbre and N. B. Bankhele, "Audio Compression Using Wavelet Transform", *IOSR Journal of Engineering (IOSR JEN)*, pp. 32-36, 2019
[24] Z. T. DRWEESH and L. E. GEORGE, "Audio Compression Based on Discrete Cosine Transform, Run Length and High Order Shift Encoding", *International Journal of Engineering and Innovative Technology (IJEIT)* ,Vol. 4, No. 1,pp. 45-51, July 2014.
[25] Z. J. Ahmed , L. E. George and R. A. Hadi, "Audio compression using transforms and high order entropy encoding", *International Journal of Electrical and Computer Engineering (IJECE),* Vol. 11, No. 4, pp. 3459-3469, August 2021.
[26] A. Van de Walle, "Relating fractal image compression to transform methods" *Citeseer* , 2000.
[27] R. A. Wannamaker and E. R. Vrscay, "Fractal wavelet compression of audio signals," *Journal of the Audio Engineering Society*, vol. 45, pp. 540-553, 1997.
[28] Z. A. Hasan and L. E. George, "Hybrid audio compression using fractal coding and wavelet transform", *ARPN Journal of Engineering and Applied Sciences*, Vol. 14, No. 11, pp. 2050-2057, JUNE 2019.
[29] A. Tsegaye and G. Tariku, "Audio Compression Using DWT and RLE Techniques," *American Journal of Electrical and Electronic Engineering,* vol.7, no. 1, pp. 14-17, 2019.
[30] W. M. Abd-Elhafiez, W. Gharibi and M. Heshmat ,"An efficient color image compression technique," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 18, no. 4, pp. 2371-2377, 2020.
[31] J. James and V. J. Thomas, "Audio Compression Using DCT and DWT Techniques" *Journal of Information Engineering and Applications*, vol. 4, no. 4, pp. 119-124, 2014.
[32] S. Hussein Toman, M. Ghazi Abdul Sahib, and Z. Hussein Toman, "Content-Based Audio Retrieval by using Elitism GA-KNN Approach", JQCM, vol. 9, no. 1, pp. 153-168, Aug. 2017.
[33] Z. A. Alwan, "A New Audio Steganography Technique", JQCM, vol. 5, no. 2, pp. 89-98, Aug. 2017.
[34] M. Mahdi, A. Mohammed, and M. Jafer, "Unusual Activity Detection in Surveillance Video Scene: Review", JQCM, vol. 13, no. 3, pp. Comp Page 92 -, Sep. 2021.