Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.199

IJCSMC, Vol. 9, Issue. 2, February 2020, pg.117 – 123

Creating Human Speech Identifier using WPT

Dr. Amjad Hindi; Dr. Majed Omar Dwairi; Prof. Ziad Alqadi
Albalqa Applied University, Faculty of Engineering Technology, Jordan, Amman
amjadhindi@bau.edu.jo; majeddw@gmail.com; natalia_maw@yahoo.com

Abstract: Human speech signal is widely used in various vital applications such as computer security system, speech signal usually has big size, which make it difficult to identify the speech by direct matching, sample by sample, so the process of using an efficient and accurate method of creating speech print is an important task in speech identification process.

In this research paper we will introduce a wavelet packet decomposition, the generated wavelet packet tree will be analyzed in order to create a unique features (identifier) to each speech signal, we will show how wavelet packet decomposition is flexible in creating speeches identifiers by providing a variety of selections, each features selection will lead to generate a unique and small in size identifier, which can be used later on in any application requiring human speech recognition.

Keywords: Speech, features vector, identifier, WPT, approximation, detail, entropy, STD.

1- Introduction

Speech digital signal is one of the most widely type of digital signals [1], [2], [3], [4], it is defined as the expression of thoughts and feelings by articulating sounds. Speech is the most natural, intuitive and preferred means of communication by human beings [5], [6], [11]. The perceptual variability of speech exists in the form of various languages, dialects, accents, while the vocabulary of speech is growing day by day. More intricate variability at the speech signal level exists in the form of varying amplitude, duration, pitch, and timbre and speaker variability. Speech and text analysis have wide applications in the present world. They have different representations, and many dissimilarities and challenges are encountered in their analysis [7], [8]. Speeches are recorded as a 16 bit signal at a sampling frequency of for example 16 kHz, which means there are 16000 samples for each second of the signal and each sample has a resolution of 16 bits per sample. Sampling frequency of a speech signal determines the resolution of the audio samples, higher the sampling rate, higher is the resolution of the signal. It can be seen from figures 1 and 2 that speech can be represented as a variation of amplitude with time. The amplitude is normalized such that the maximum value is *1*[9], [10], [11].

Speech signal usually recorded based on a sampling rate or frequency, this will lead to a big number of samples and a big memory space needed to store the speech digital file [12]. Using this digital file in direct matching will require a big time making the recognition system inefficient, table 1 shows some recorded speeches and there sizes, these speeches will be used later on in the experimental part.

Table 1: Recorded speeches

Speech number	Spoken speech	Samples	Size(byte)
1	Please open the window	122610	980880
2	Albalqa applied university	154829	1238632
3	Faculty of engineering technology	173071	1384568
4	Amman is the capital city of Jordan	183544	1468352
5	Aqaba is a beautiful city, it is located on the red sea	255037	2040296
6	Speech signal features	125834	1006672
7	Wavelet packet tree	124801	998408
8	Good morning	86627	693016
9	Have a nice day	94378	755024
10	My name is Ziad Alqadi	144478	1155824

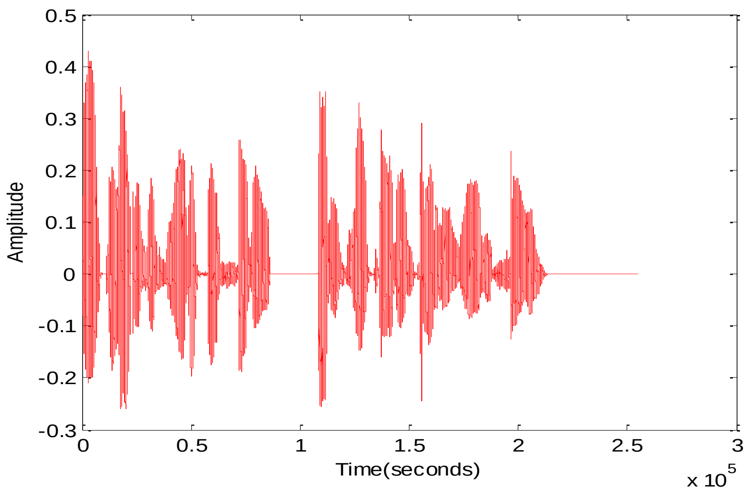

Figure 1: Speech signal wave (speech 5) (time domain)

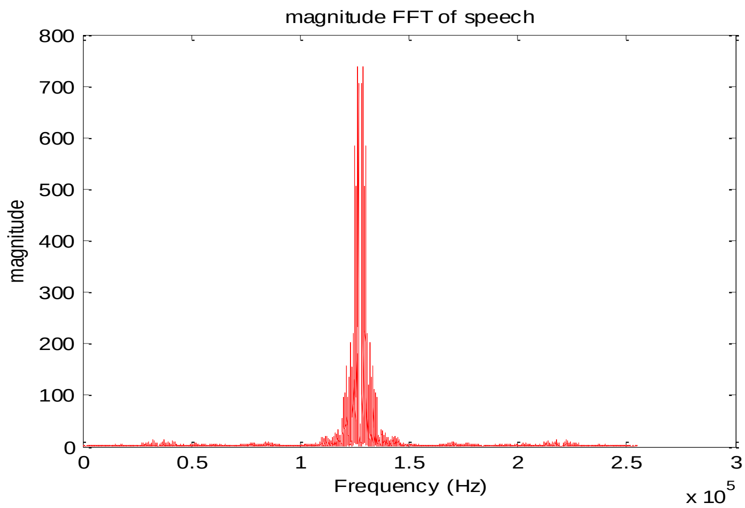

Figure 2: Speech signal wave (speech 5) (frequency domain)

2- Speech Signal Decomposition using WPT

Creating speech identifier is an important task in building any speech recognition system, many methods were introduced to create a speech signal features, some of these methods were based on calculating local binary pattern (LBP)[13], [14], some of them were based on calculating the coeffecient of linear prediction coding(LPC)[15], [16], [17], [18], some of them were based on using k-means method of clustering[19], [20], [21], [22] . These methods are effeciently used in creating speech signal features, but here in our research paper we will show how WPT is efficient and how it provides a user of varieties in selecting the speech signal features [31], [32].

Wavelet packet tree (WPT) [23], [24], [25] is one of the popular transformations used to decompose a digital signal into partitions called packet, the digital signal will be divided into approximation and details as shown in figure 3 and the number of approximations and details will depend on the selected level of decomposition as shown in figures 3 and 4[26], [29], [30]:

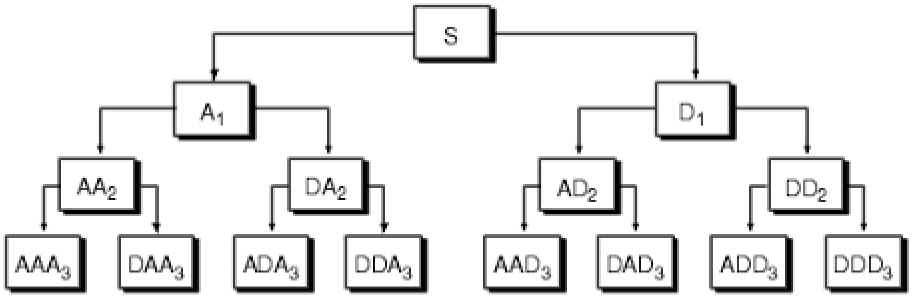

Figure 3: Signal decomposition

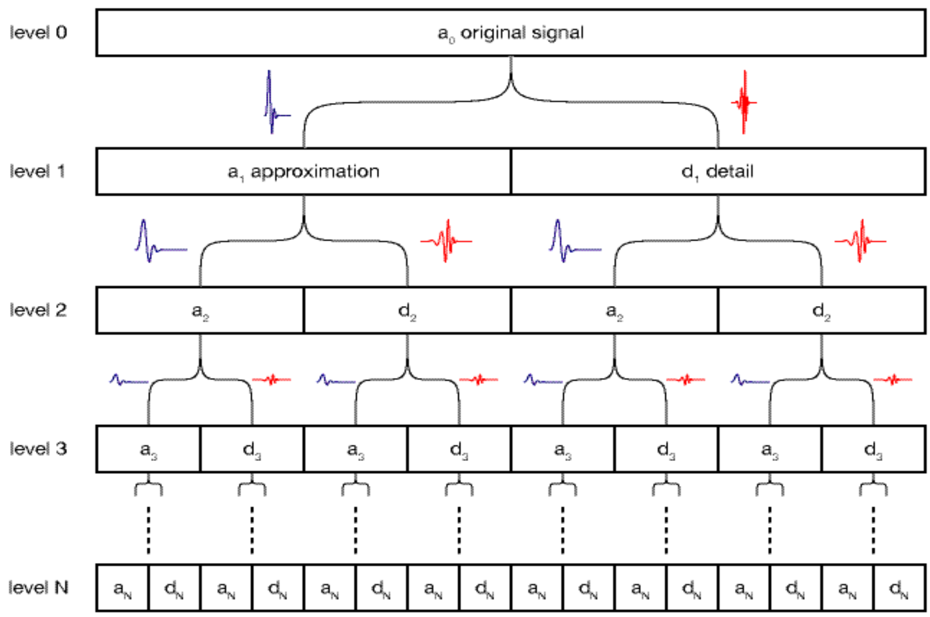

Figure 4: Levels of signal decomposition

The speech signal can be decomposed using WPT by selecting the level of composition and calculating the approximations and details at each level.

The approximation or the Haar scaling (low pass [27] , [28]) function is by formula 1:

$$s_{j+1,i} = \frac{even_{j,i} + odd_{j,i}}{2} \quad (1)$$

While the details or Haar wavelet (High pass) function is calculated by formula 2

$$d_{j+1,i} = \frac{even_{j,i} - odd_{j,i}}{2} \quad (2)$$

Table 2 shows an example of how to decompose a digital signal of 8 discrete values into approximations and details using 3 levels of decomposition:

Figure 5 shows some of the approximation results of a speech signal 1.

Table 2: Worked example

Signal	**-3**	**-1**	**0**	**4**	**5**	**8**	**2**	**0**
Level 1	**Approximation**				**Detail**			
	-2	**2**	**6.5**	**1**	**-1**	**-2**	**-1.5**	**1**
Level 2	**Approximation**		**Detail**		**Approximation**		**Detail**	
	0	**3.75**	**-2**	**2.75**	**-1.5**	**-0.25**	**0.5**	**-1.25**
Level 3	**Approximation**	**Detail**	**Approximation**	**Detail**	**Approximation**	**Detail**	**Approximation**	**Detail**
	1.875	**-1.875**	**0.375**	**-2.375**	**-0.875**	**-0.625**	**-0.375**	**0.875**

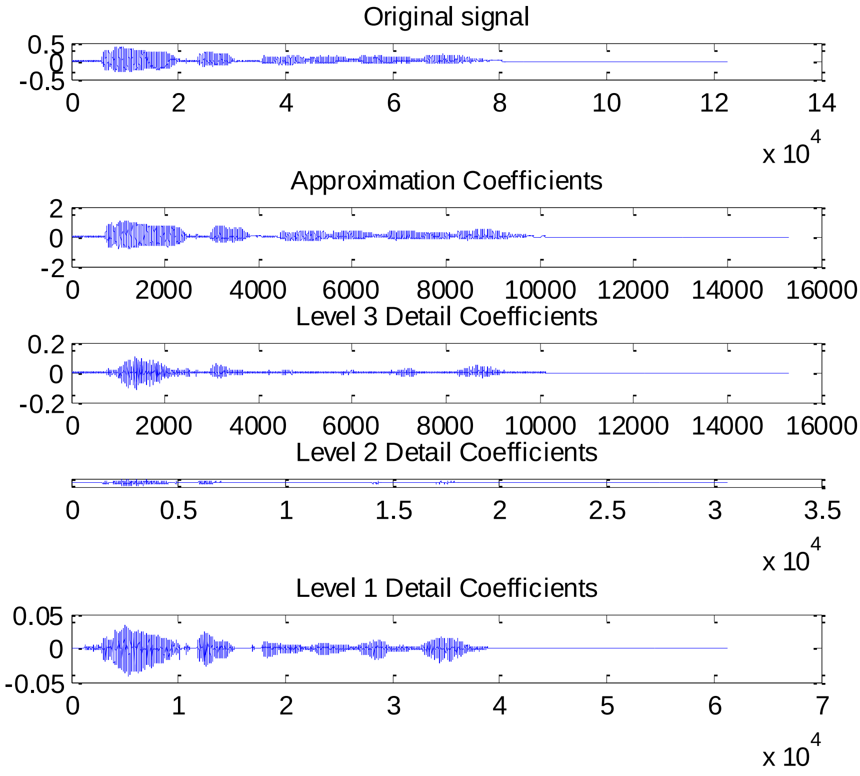

Figure 5: Speech 1 decomposition.

3- *Implementation and Experimental Results*

The recorded speech signals shown in table 1 were decomposed using WPT, for each approximation we calculated the entropy, table 3 shows the results of this implementation.

Table 3: Calculated entropy

Speech number	Packets(Approximation)			
	A10	A20	A30	A40
1	**1501.8**	**1154.0**	**807.1**	**469.1**
2	**3427.2**	**2519.9**	**1627.6**	**805.0**

3	**2162.1**	**1708.6**	**1254.5**	**812.1**
4	**2543.6**	**1935.7**	**1335.6**	**800.0**
5	**2780.6**	**2202.8**	**1627.4**	**1068.9**
6	**1798.0**	**1349.3**	**905.5**	**474.7**
7	**1063.9**	**862.7**	**660.7**	**459.0**
8	**1427.1**	**1079.3**	**733.6**	**399.1**
9	**771.6716**	**625.7456**	**480.4965**	**337.8110**
10	**2195.5**	**1721.5**	**1249.6**	**804.3**

From table 3we can see that entropy values for various packets of the decomposed signal are unique, thus they can be easily used to for a speech signal identifier, which can be stored and used later on in a recognition system. WPT method provides a valuable various set of information, each of these sets can be separately used as an identifier for a speech signal, these sets includes a unique values for each speech signal, tables 4-7 shows the values obtained for these set using the same recorded speech signals.

Table 4: Average of the speech packets(approximations)

Speech number	A10	A20	A30	A40
1	**0.0002409**	**0.0003407**	**0.0004819**	**0.0006814**
2	**0.0005**	**0.0007**	**0.0010**	**0.0014**
3	**0.0001231**	**0.0001741**	**0.0002462**	**0.0003482**
4	**0.0004**	**0.0005**	**0.0007**	**0.0010**
5	**0.0002017**	**0.0002852**	**0.0004034**	**0.0005705**
6	**0.0001614**	**0.0002282**	**0.0003227**	**0.0004564**
7	**0.0002112**	**0.0002987**	**0.0004225**	**0.0005974**
8	**0.0002894**	**0.0004093**	**0.0005788**	**0.0008185**
9	**0.0000580**	**0.0000820**	**0.0001160**	**0.0001640**
10	**0.0000622**	**0.0000879**	**0.0001243**	**0.0001758**

Table 5: Standard deviation of the speech packets(approximations)

Speech number	A10	A20	A30	A40
1	**0.0903**	**0.1275**	**0.1794**	**0.2495**
2	**0.1295**	**0.1825**	**0.2553**	**0.3479**
3	**0.0866**	**0.1222**	**0.1716**	**0.2372**
4	**0.0973**	**0.1371**	**0.1916**	**0.2599**
5	**0.0802**	**0.1131**	**0.1587**	**0.2185**
6	**0.1004**	**0.1417**	**0.1993**	**0.2777**
7	**0.0678**	**0.0957**	**0.1344**	**0.1863**
8	**0.1075**	**0.1518**	**0.2137**	**0.2969**
9	**0.0662**	**0.0933**	**0.1311**	**0.1819**
10	**0.0968**	**0.1365**	**0.1909**	**0.2600**

Table 6: Minimum values of the speech packets(approximations)

Speech number	A10	A20	A30	A40
1	**-0.4231**	**-0.5966**	**-0.8294**	**-1.0567**
2	**-0.5542**	**-0.7732**	**-1.0680**	**-1.4612**
3	**-0.3526**	**-0.4962**	**-0.6853**	**-0.9530**
4	**-0.6888**	**-0.9308**	**-1.2513**	**-1.4109**
5	**-0.3675**	**-0.5173**	**-0.7137**	**-0.9674**
6	**-0.4346**	**-0.6096**	**-0.8172**	**-1.0961**
7	**-0.2300**	**-0.3232**	**-0.4495**	**-0.5937**
8	**-0.5906**	**-0.8295**	**-1.1461**	**-1.5191**
9	**-0.2910**	**-0.4104**	**-0.5749**	**-0.7797**
10	**-0.3518**	**-0.4955**	**-0.6660**	**-0.9112**

Table 7: Maximum values of the speech packets(approximations)

Speech number	A10	A20	A30	A40
1	**0.5391**	**0.7531**	**1.0290**	**1.3663**
2	**0.7765**	**1.0890**	**1.4403**	**1.7790**

3	**0.4816**	**0.6789**	**0.8769**	**1.1257**
4	**0.8369**	**1.1772**	**1.6199**	**1.8889**
5	**0.6064**	**0.8350**	**1.1187**	**1.3788**
6	**0.5099**	**0.7048**	**0.9936**	**1.2360**
7	**0.3908**	**0.5449**	**0.7054**	**0.9018**
8	**0.5730**	**0.8063**	**1.1218**	**1.4967**
9	**0.3755**	**0.5166**	**0.7089**	**0.9708**
10	**0.6059**	**0.8304**	**1.0893**	**1.1710**

Conclusion

WPT method of speech signal decomposition was proposed and implemented. From the obtained experimental results it was shown that it is easily to use the signal packets(approximations) to create a speech signal identifier, these packets contains a valuable information which can be used by the user to form a digital signal features.

It was shown that WPT method is a flexible by providing a variety of forming the signal features.

Packet entropy, packet standard deviation, packet average, packet minimum and maximum values can be used separately to generate a speech signal identifier.

References

[1]. Ziad A. Alqadi, Majed O. Al-Dwairi, Amjad A. Abu Jazar and Rushdi Abu Zneit, Optimized True- RGB color Image Processing, World Applied Sciences Journal 8 (10): 1175-1182, ISSN 1818-4952, 2010.

[2]. A. A. Moustafa, Z. A. Alqadi, Color Image Reconstruction Using A New R'G'I Model, journal of Computer Science, Vol.5, No. 4, pp. 250-254, 2009.

[3]. Jamil Al Azzeh, Hussein Alhatamleh, Ziad A. Alqadi, Mohammad Khalil Abuzalata, Creating a Color Map to be used to Convert a Gray Image to Color Image; International Journal of Computer Applications , November 2016,Volume 153,Issue 2.

[4]. AlQaisi Aws and AlTarawneh Mokhled and Alqadi Ziad A. and Sharadqah Ahmad A, Analysis of Color Image Features Extraction using Texture Methods, TELKOMNIKA, volume 17, number 3, pages1220—1225, year 2019.

[5]. K Matrouk, A Al-Hasanat, H Alasha'ary, Ziad Al-Qadi, H Al-Shalabi, Speech fingerprint to identify isolated word person, World Applied Sciences Journal, vol. 31, issue 10, pp. 1767-1771, 2014.

[6]. Saleh Khawatreh, Belal Ayyoub, Ashraf Abu-Ein, Ziad Alqadi, A Novel Methodology to Extract Voice Signal Features, International Journal of Computer Applications, vol. 975, pp. 8887, 2018.

[7]. Jihad Nadir, Ashraf Abu Ein, Ziad Alqadi, A Technique to Encrypt-decrypt Stereo Wave File, International Journal of Computer and Information Technology, vol. 5, issue 5, pp. 465-470, 2016.

[8]. Majed O. Al-Dwairi, Amjad Y. Hendi, Mohamed S. Soliman, Ziad A.A. Alqadi, A new method for voice signal features creation, International Journal of Electrical and Computer Engineering (IJECE), vol. 9, issue 5, pp. 4092-4098, 2019.

[9]. Ashraf Abu-Ein, Ziad AA Alqadi, Jihad Nader, A TECHNIQUE OF HIDING SECRETE TEXT IN WAVE FILE, International Journal of Computer Applications, 2016.

[10]. Ayman Al-Rawashdeh, Ziad Al-Qadi, Using wave equation to extract digital signal features, Engineering, Technology & Applied Science Research, vol. 8. Issue 4, pp. 1356-1359, 2018.

[11]. Ziad Alqadi, Analysis of stream cipher security algorithm, Journal of Information and Computing Science, vol. 2, issue 4, pp. 288-298, 2007.

[12]. Jihad Nadir, Ashraf Abu Ein, Ziad Alqadi, A Technique to Encrypt-decrypt Stereo Wave File, International Journal of Computer and Information Technology, vol. 5, issue 5, pp. 465-470, 2016.

[13]. Aws Al-Qaisi, Saleh A Khawatreh, Ahmad A Sharadqah, Ziad A Alqadi, Wave File Features Extraction Using Reduced LBP, International Journal of Electrical and Computer Engineering, vol. 8, issue 5, pp. 2780, 2018.

[14]. Ziad Alqad, Prof. Yousf Eltous , Dr. Ghazi M. Qaryouti, Prof. Mohammad Abuzalata, Analysis of Digital Signal Features Extraction Based on LBP Operator, International Journal of Advanced Research in Computer and Communication Engineering, vol. 9, issue 1, pp. 1-7, 2020.

[15]. Ziad A. AlQadi Amjad Y. Hindi, Majed O. Dwairi, PROCEDURES FOR SPEECH RECOGNITION USING LPC AND ANN, International Journal of Engineering Technology Research & Management, vol. 4, issue 2, pp. 48-55, 2020.

[16]. Dr. Ghazi M. Qaryouti Dr. Amjad Hindi, Prof. Yousif Eltous, Prof. Mohammad Abuzalata, Prof. Ziad Alqadi, USING FIR FILTER COEFFICIENTS TO CREATE COLOR IMAGE FEATURES, International Journal of Engineering Technology Research & Management, vol. 4, issue 2, pp. 6-14, 2020.

[17]. Dr. Amjad Hindi, Dr. Ghazi M. Qaryouti, Prof. Yousif Eltous, Prof. Mohammad Abuzalata, Prof. Ziad Alqadi, Color Image Compression using Linear Prediction Coding, International Journal of Computer Science and Mobile Computing, vol. 9, issue 2, pp. 13 – 20, 2020.

[18]. Prof. Yousif Eltous Dr. Amjad Hindi, Prof. Ziad Alqadi, Dr. Ghazi M. Qaryouti, Prof. Mohammad Abuzalata, Using FIR Coefficients to Form a Voiceprint, International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, vol. 8, issue 1, pp. 1-6, 2020.

[19]. Prof. Ziad Alqadi Dr. Ghazi M. Qaryouti, Prof. Mohammad Abuzalata, Prof. Yousf Eltous, Comparative Study of Voice Signal Features Extraction Methods, IOSR Journal of Computer Engineering (IOSR-JCE), vol. 22, issue 1, pp. 58-66, 2020.

[20]. Prof. Yousf Eltous Prof. Ziad Alqadi, Dr. Ghazi M. Qaryouti, Prof. Mohammad Abuzalata, Enhancing Color Image Clustering using K-Means Method, International Journal of Advanced Research in Computer and Communication Engineering, vol. 9, issue 1, pp. 78-84, 2020.

[21]. Yousf Eltous Ziad A. AlQadi, Ghazi M. Qaryouti, Mohammad Abuzalata, ANALYSIS OF DIGITAL SIGNAL FEATURES EXTRACTION BASED ON KMEANS CLUSTERING, International Journal of Engineering Technology Research & Management, vol. 4, issue 1, pp. 66-75, 2020.

[22]. Prof. Yousif Eltous, Dr. Ghazi M. Qaryouti, Prof. Mohammad Abuzalata, Prof. Ziad Alqadi, Evaluation of Fuzzy and C_mean Clustering Methods used to Generate Voiceprint, IJCSMC, vol. 9, issue 1, pp. 75 -83, 2020.

[23]. S. Choi, Y. Shin, H.-K. Park, Selection of wavelet packet measures for insufficiency murmur identification, Expert Syst. Appl. 38 (2011) 4264–4271.

[24]. C. Ahlstrom, P. Hult, P. Rask, J.-E. Karlsson, E. Nylander, U. Dahlström, et al, Feature extraction for systolic heart murmur classification, Ann. Biomed. Eng.34 (2006) 1666–1677.

[25]. Y. Chen, S. Wang, C.-H. Shen, F.K. Choy, Matrix decomposition based feature extraction for murmur classification, Med. Eng. Phys. 34 (2012) 756–761.

[26]. Ziad A. AlQadi Amjad Y. Hindi, Majed O. Dwairi, Analysis of Digital Signals using Wavelet Packet Tree, IJCSMC, vol. 9, issue 2, pp. 96 – 103, 2020.

[27]. Ziad Alqadi, Bilal Zahran, Jihad Nader, Estimation and Tuning of FIR Lowpass Digital Filter Parameters, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 7, issue 2, pp. 18-23, 2017.

[28]. Jamil Azzeh, Bilal Zahran, Ziad Alqadi, Salt and Pepper Noise: Effects and Removal, JOIV: International Journal on Informatics Visualization, vol. 2, issue 4, pp. 252-256, 2018.

[29]. [29] Jamil Al-Azzeh, Ziad Alqadi, Mohammed Abuzalata, Performance Analysis of Artificial Neural Networks used for Color Image Recognition and Retrieving, International Journal of Computer Science and Mobile Computing, vol. 8, issue 2, pp. 20-33, 2019.

[30]. Ahmad Sharadqh Naseem Asad, Ismail Shayeb, Qazem Jaber, Belal Ayyoub, Ziad Alqadi, Creating a Stable and Fixed Features Array for Digital Color Image, IJCSMC, vol. 8, issue 8, pp. 50-56, 2019.

[31]. Prof. Mohammed Abu Zalata Dr. Ghazi. M. Qaryouti , Dr.Saleh Khawatreh, Prof. Ziad A.A. Alqadi, Optimal Color Image Recognition System (OCIRS), International Journal of Advanced Computer Science and Technology., vol. 7, issue 1, pp. 91-99, 2017.

[32]. Aws AlQaisi, Mokhled AlTarawneh, Ziad A. Alqadi, Ahmad A. Sharadqah, Analysis of Color Image Features Extraction using Texture Methods, TELKOMNIKA, vol. 17, issue 3, pp. 1220~1225, 2019.