Comparative Study of Voice Signal Features Extraction Methods

Dr. Ghazi M. Qaryouti, Prof. Mohammad Abuzalata, Prof. Yousf Eltous, Prof. Ziad Alqadi

Albalqa Applied University, Faculty of engineering technology, Jordan, Amman

Abstract: Creating voice signal features is an important and vital task, these features can be used in any security system to identify human and to identify the spoken word (phrase) by this human.

In this research paper we will investigate three methods, each of them can be used to create a voice signal features, we will adopt and modify LBP method to be capable to create a features for a voice signal. The three methods will be programmed and implemented, the obtained experimental results will studied and analyzed in order to give some judgment about these method.

Key words: Crest factor, dynamic range, Mu, sigma, LBP, CSLBP, MLBP, K-means, throughput, speed up.

Date of Submission: 15-01-2020 Date of Acceptance: 03-02-2020

I. Introduction

A voice signal is a representation of sound, typically using a level of electrical voltage for analogue signals, and a series of binary numbers for digital signals. Voice signals have frequencies in the audio frequency range of roughly 20 to 20,000 Hz, which corresponds to the lower and upper limits of human hearing [1].

The analogue voice signal can be converted to digital signal by applying the operations shown if figure 1(sampling, quantization and encoding):



Figure 1: Converting analogue voice to digital)

The number of samples generated in a unit of time is called sampling rate (frequency FS) as shown in figure 2:



Each voice signal is characterized by the amplitude and frequency, and it can be presented in a time domain or in a frequency domain as shown in figure 3:



Figure 3: Voice signal

Voiceprint refers to the acoustic frequency spectrum that carries the speech information in a human voice. Like fingerprints, it has unique biometric signatures, is individual-specific, and can function as an identification method. The acoustical signal is a one-dimensional continuous signal. Comparison of two recorded speech by means of spectrogram or voice prints for the purpose of identification is called as Voice fingerprinting. Forensic voice analysis has been used in a wide range of criminal cases such as murder, rape, drug dealing, bomb threats and terrorism [2], [3].

Voice biometrics works by digitizing a profile of a person's speech to produce a stored model voice print, or template. Biometric technology reduces each spoken word to segments composed of several dominant frequencies called formants. The tones collectively identify the speaker's unique voice print.

Will Voice Biometrics Be Accurate Enough? To begin with, it is important to recognize that no biometric is 100% accurate. For example, a 2014 study on iris recognition determined system accuracy could be between 90 and 99%, a broad range. Voice biometric accuracy is also in this range for a variety of reasons.

Voiceprint generation method (feature extraction) must process the digital wave file and must create unique features, which can be simply used to identify the voice [5], [6].

Here in this research we will study and analyse some method used to extract voice features in order to define the advantages and disadvantages of each method and to give some recommendation of how to create a voice signal features [7-14].

II. Features Extraction Methods

In this section we will some of the most papular and widely used methods, which were produced for digital data features extraction.

2-1 Statistical methods

Here in this method we will use some voice statistical parameters, crest factor, dynamic range, Mu, and sigma [4].

Crest factor:

Crest or peak factor (CF) of a voice signal is the dB difference between the peaks and the RMS value of the signal (see figure 4). The RMS (Root Mean Square) is defined as the "heating value" of the signal – the voltage that would generate the same heat as a DC (Direct Current) signal, over the same time [5].



Figure 4: Crest (peak) factor calculation

Dynamic range

Dynamic range (DR) is used to describe the ratio between the smallest and the largest possible values of a changeable quantity, frequently encountered in a recorded voice signal. DR is another way of starting the maximum signal-to-noise ratio. For sound signal, DR is the ratio of the loudest signal to that of the quietest signal in a system as expressed in decibels (dB).

Mu

Mu is the estimate mean of the sound signal values.

Sigma

Sigma is the standard deviation of the normal distribution of the data in the voice signal.

2-2 Local binary pattern method

Local binary pattern (LBP) method was introduced to manipulate color images [15], [16], [17], and it is based on LBP [18], [19], [20] operator calculation for each pixel as shown in figure 5:



Figure 5: LBP operator calculation (example for one pixel)

Based on LBP methods many methods such as central symmetric LBP (CSLBP) were introduced to form a feature array for digital image, figure 6 shows how to use CSLBP to calculate the operator value, here using this method we can generate a feature array of 16 elements.



Figure 6: CSLBP calculation example (for one pixel)

Here in this research paper we will adopt the idea of CSLBP method by introducing modified LBP (MLBP) used to create a feature array of 8 elements for any wave file, MLBP can be calculated for each voice signal amplitude value, figure 7 shows how MLBP operates:



Figure 7: MLBP operations

2-3 K-means clustering method

K-means clustering is used to group an input data signal values into groups, and here we can take the centroids of the clusters, or the number of points in each cluster, or the within clusters sums as a signal features [24-29].

To increase the efficiency of k-means method we propose using the LBP histogram of the voice signal, instead of using the whole wave file as an input data set for the clustering method(later in the experimental part the time required to calculate LBP histogram will be included in the clustering time (features extraction time)). To calculate voice signal LBP histogram [21-24], we select a window of 9 values and do the comparisons with value located in the centre, the results of comparisons generate a binary number, the binary number then must be converted to a decimal value, which refers to the index of LBP histogram, this index must be incremented by 1, theses operation must be performed for each value in the wave file, figure 8 shows how to calculate an index in LBP histogram.

A(i-4)	0.1831		1
A(i-3	-0.0610		0
A(i-2)	0.1526		0
A(i-1)	0.0615	Comparisons	0
A(i)	0.1671		-
A(i+1)	0.0610		0
A(i+2)	0.2610		1
A(i+3)	-0.0305		0
A(i+4)	-0.1221		0
-			
-			
Binary = 00100	001 =		
Decimal = 33			

Figure 8: Calculating voice signal histogram index

III. **Implementation And Experimental Results**

The three mentioned methods were programmed using matlab environment, the written programs was implemented using various voce signal.

We select 6 persons and 5 spoken words for each person, then the wave file were used in the following experiments, table 1 shows the used wave files:

radic 1. Osed wave mes									
Spoken word	1	2	3	4	5				
Person									
1	Bye-bye-1	Okay-1	Yes-1	No-1	Zero-1				
2	Bye-bye-2	Okay-2	Yes-2	No-2	Zero-2				
3	Bye-bye-3	Okay-3	Yes-3	No-3	Zero-3				
4	Bye-bye-4	Okay-4	Yes-4	No-4	Zero-4				
5	Bye-bye-5	Okay-5	Yes-5	No-5	Zero-5				
6	Bye-bye-6	Okay-6	Yes-6	No-6	Zero-6				

Table 1. Used wave files

Experiment 1: Statistical method implementation

Table 2 shows the results of implementing this method using the first word for the 6 persons, while table 3 shows the results of implementing this method using the 5 spoken words for the first person, here we change the amplitude and the sampling rate, to show theses changes affect the generated features.

Experiment 2: LBP method implementation

Table 4 shows the results of implementing this method using the first word for the 6 persons, while table 5 shows the results of implementing this method using the 5 spoken words for the first person, here we change the amplitude and the sampling rate, to show theses changes affect the generated features.

Experiment 1: K-means clustering method implementation

Table 6 shows the results of implementing this method using the first word for the 6 persons, while table 7 shows the results of implementing this method using the 5 spoken words for the first person, here we change the amplitude and the sampling rate, to show theses changes affect the generated features.

		Tabl	le 2: Experim	ent 1 results	1		
Voic	e	Bye-bye-1	Bye-bye-2	Bye-bye-3	Bye-bye-4	Bye-bye-5	Bye-bye-6
Features of original wave	Crest fac- tor(db)	14.0451	13.178	16.6458	16.1097	14.8419	17.6223
file	Dynamic range(db)	83.9406	81.434	81.4391	83.2393	82.4679	79.9092
	Sigma	0.19849	0.21933	0.14713	0.1565	0.1811	0.13149
	Mu	-2.5899e-	-5.0772e-	-2.2025e-	-4.8363e-	-9.4536e-	-0.00013186
		005	006	005	008	006	
Features of	Crest factor	14.0451	13.178	16.6458	16.1097	14.8419	17.6223
wave file with dividing FS by	Dynamic range	83.9406	81.434	81.4391	83.2393	82.4679	79.9092
2	Sigma	0.19849	0.21933	0.14713	0.1565	0.1811	0.13149
	Mu	-2.5899e-	-5.0772e-	-2.2025e-	-4.8363e-	-9.4536e-	-0.00013186
		005	006	005	008	006	
Features of	Crest factor	14.0451	13.178	16.6458	16.1097	14.8419	17.6223
wave file with	Dynamic	83.9406	81.434	81.4391	83.2393	82.4679	79.9092

Comparative Study of Voice Signal Features Extraction Methods

doubling ampli-	range						
tude	Sigma	0.19849	0.21933	0.14713	0.1565	0.1811	0.13149
	Mu	-2.5899e-	-5.0772e-	-2.2025e-	-4.8363e-	-9.4536e-	-0.00013186
		005	006	005	008	006	
Size(sar	nple)	36787	36787	223607	44145	64745	39730
Extraction	time(s)	0.0320	0.0360	0.0520	0.0340	0.0350	0.0340
	Average size				74300		
Avera	age extraction tin	me			0.0372		
Throughp	out (sample per s	econd)			1997300		

Table 2: Experiment 1 results 2

V	oice	Bye-bye-1	Okay-1	Yes-1	No-1	Zero-1
Features of	Crest factor(db)	14.0451	14.4317	18.7455	14.4441	15.4313
original wave	Dynamic	83.9406	84.6284	85.4112	84.8095	60.7982
file	range(db)					
	Sigma	0.19849	0.18986	0.11554	0.18958	0.1692
	Mu	-2.5899e-	-1.8806e-	-2.1345e-	-4.6484e-	0.0022494
		005	005	005	006	
Features of	Crest factor	14.0451	14.4317	18.7455	14.4441	15.4313
wave file with	Dynamic range	83.9406	84.6284	85.4112	84.8095	60.7982
dividing FS by	Sigma	0.19849	0.18986	0.11554	0.18958	0.1692
2	Mu	-2.5899e-	-1.8806e-	-2.1345e-	-4.6484e-	0.0022494
		005	005	005	006	
Features of	Crest factor	14.0451	14.4317	18.7455	14.4441	15.4313
wave file with	Dynamic range	83.9406	84.6284	85.4112	84.8095	60.7982
doubling ampli-	Sigma	0.19849	0.18986	0.11554	0.18958	0.1692
tude	Mu	-2.5899e-	-1.8806e-	-2.1345e-	-4.6484e-	0.0022494
		005	005	005	006	

Table 4: Experiment 2 results 1

Voice	Bye-bye-1	Bye-bye-2	Bye-bye-3	Bye-bye-4	Bye-bye-5	Bye-bye-6
Features of	14792	13690	83288	16855	23748	11450
original wave	706	1126	8354	1367	2797	1809
file	1470	1614	9609	1389	2197	4891
Features of	776	1161	7437	1216	2554	523
wave file with	648	944	6540	1033	2343	454
dividing FS by	1587	1741	9925	1528	2396	4959
2	728	1143	7862	1417	2790	1883
	16074	15362	90586	19334	25914	13755
Features of	14792	13690	83288	16855	23748	11450
wave file with	706	1126	8354	1367	2797	1809
doubling am-	1470	1614	9609	1389	2197	4891
plitude	776	1161	7437	1216	2554	523
Features of	648	944	6540	1033	2343	454
original wave	1587	1741	9925	1528	2396	4959
file	728	1143	7862	1417	2790	1883
	16074	15362	90586	19334	25914	13755
Features of	14792	13690	83288	16855	23748	11450
wave file with	706	1126	8354	1367	2797	1809
dividing FS by	1470	1614	9609	1389	2197	4891
2	776	1161	7437	1216	2554	523
	648	944	6540	1033	2343	454
	1587	1741	9925	1528	2396	4959
	728	1143	7862	1417	2790	1883
	16074	15362	90586	19334	25914	13755
Size(sample)	36787	36787	223607	44145	64745	39730
Extraction	0.0020	0.0010	0.0090	0.0020	0.0020	0.0020
time(s)						
	Average size			743	00	
Av	verage extraction tim	me		0.00	30	
	Throughput			24767	7000	

Table 5: Experiment 2 results 2

Voice	Bye-bye-1	Okay-1	Yes-1	No-1	Zero-1
Features of origi-	14792	11927	6643	7971	3262
nal wave file	706	1209	2448	864	722
Features of wave	1470	1955	2067	2817	255
file with dividing	776	488	4647	390	785
FS by 2	648	522	4601	388	405

Comparative Study of Voice Signal Features Extraction Methods

	1587	2138	2341	2996	535
	728	1305	2561	991	751
	16074	12823	8530	8593	5823
Features of wave	14792	11927	6643	7971	3262
file with doubling	706	1209	2448	864	722
amplitude	1470	1955	2067	2817	255
Features of origi-	776	488	4647	390	785
nal wave file	648	522	4601	388	405
	1587	2138	2341	2996	535
	728	1305	2561	991	751
	16074	12823	8530	8593	5823
Features of wave	14792	11927	6643	7971	3262
file with dividing	706	1209	2448	864	722
FS by 2	1470	1955	2067	2817	255
	776	488	4647	390	785
	648	522	4601	388	405
	1587	2138	2341	2996	535
	728	1305	2561	991	751
	16074	12823	8530	8593	5823

Table 6: Experiment 3 results 1

Voice	Bye-bye-1	Bye-bye-2	Bye-bye-3	Bye-bye-4	Bye-bye-5	Bye-bye-6
Features of original wave file	13.5	22.1	13.8	24	67	19.8
-	94	150.9	108.5	225	941	284.9
	809.8	779.9	464.4	852	3704	1140.9
	9896.5	8800	5137.6	11785	13238	5772.3
Features of wave file with divid-	13.5	22.1	13.8	24	67	19.8
ing FS by 2	94	150.9	108.5	225	941	284.9
	809.8	779.9	464.4	852	3704	1140.9
	9896.5	8800	5137.6	11785	13238	5772.3
Features of wave file with dou-	13.5	22.1	13.8	24	67	19.8
bling amplitude	94	150.9	108.5	225	941	284.9
	809.8	779.9	464.4	852	3704	1140.9
	9896.5	8800	5137.6	11785	13238	5772.3
Size(sample)	36787	36787	223607	44145	64745	39730
Extraction time(s)	0.1810	0.1000	0.1030	0.1000	0.1010	0.0960
Average size				74300		
Average extraction tir	ne			0.1135		
Throughput (sample per se	econd)			654630		

Table 7: Experiment 3 results 2

Voice	Bye-bye-1	Okay-1	Yes-1	No-1	Zero-1
Features of origi-	13.5	9.9	32.2	20.7	8.7
nal wave file	94	132.5	334	403	53.3
	809.8	656.1	1165.3	2028	163.6
	9896.5	7086	4143.5	4855	2026.3
Features of wave	13.5	9.9	32.2	20.7	8.7
file with dividing	94	132.5	334	403	53.3
FS by 2	809.8	656.1	1165.3	2028	163.6
	9896.5	7086	4143.5	4855	2026.3
Features of wave	13.5	9.9	32.2	20.7	8.7
file with doubling	94	132.5	334	403	53.3
amplitude	809.8	656.1	1165.3	2028	163.6
Features of origi-	9896.5	7086	4143.5	4855	2026.3
nal wave file					

From the obtained results shown in tables 2 through 7 we can raise the following facts:

- \checkmark All the introduced method gave unique features for each wave file.
- \checkmark Each method can be used to form a features which can be used a wave file to identify a person.
- ✓ Each method can be used to form a features which can be used a wave file to identify a spoken word by a person.
- \checkmark Changing the voice amplitude does not affect the features.
- \checkmark Changing the voice sampling rate does not affect the features.
- ✓ The more efficient method is LBP methods, it provides a highest throughput and the lowset average features extraction time as shown in tables 8 and 9

Table 8: Efficiency measures								
Method	Average extraction time	Throughput						
Statistical	0.0372	1997300						
LBP	0.0030	24767000						
K-means	0.1135	654630						

K-means	1135		654630					
Table 9: Speedup								
Method	Statistical	LBP		K-means				

Statistical	1	0.0806	3.0511
LBP	12.4000	1	37.8333
K-means	0.3278	0.0264	1

IV. Conclusion

Three methods of voice features extraction were presented, programmed and implemented. The obtained experimental results showed the accuracy of using each method to create unique features for each wave signal; these features can be used in any security system to identify word/person.

LBP method of features extraction has the highest efficiency, and it has a significant high speed up comparing with the other 2 method.

References

- R. Szabo, A. Gontean, I. Lie, "Sound Based Coin Recognition and Clapper", 16th International Conference on Soft Computing (MENDEL), pp. 509–516, June 23-25, 2010.
- [2]. A. K. Paul, D. Das, M. M. Kamal, "Bangla Speech Recognition System Using LPC and ANN", Seventh International Conference on Advances in Pattern Recognition (ICAPR), Kolkata, India, pp. 171–174, February 4-9, 2009.
- [3]. R. Szabo, A. Gontean, "Human Voice Signal Synthesis and Coding", IFAC Proceedings Volumes, Vol. 46, No. 28, pp. 336-341, 2013.
- [4]. K. M. Matrouk, A. Alhasanat, H. Alasha'ary, Z. Alqadi, H. M. AlShalabi, "Speech Fingerprint to Identify Isolated Word-Person", World Applied Sciences Journal, Vol. 31, No. 10, pp. 1767-1771, 2014.
- [5]. Jihad Nader Ismail Shayeb, Ziad Alqadi, Analysis of digital voice features extraction methods, International Journal of Educational Research and Development, v. 1, issue 4, pp. 49-55, 2019.
- [6]. 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), v. 9, issue 5, pp. 4092-4098, 2019.
- [7]. Ziad A. AlQadi Amjad Y. Hindi, Majed O. Dwairi, A Novel Technique for Data Steganography, Engineering, Technology & Applied Science Research, v. 9, issue 6, pp. 4942-4945, 2019.
- [8]. 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, v. 8, issue 5, pp. 2780, 2018.
- [9]. Ayman Al-Rawashdeh, Ziad Al-Qadi, Using wave equation to extract digital signal features, Engineering, Technology & Applied Science Research, v. 8, issue 4, pp. 1356-1359, 2018.
- [10]. Saleh Khawatreh, Belal Ayyoub, Ashraf Abu-Ein, Ziad Alqadi, A Novel Methodology to Extract Voice Signal Features, International Journal of Computer Applications, v. 975, pp. 8887, 2018.
- [11]. Ziad Alqadi, Bilal Zahran, Jihad Nader, Estimation and Tuning of FIR Low pass Digital Filter Parameters, international Journal of Advanced Research in Computer Science and Software Engineering, v. 7, issue 2, pp. 18-23, 2017.
- [12]. Jihad Nader Ahmad Sharadqh , Ziad Al-Qadi, Bilal Zahran, Experimental Investigation of Wave File Compression-Decompression, International Journal of Computer Science and Information Security, v. 14, issue 10, pp. 774-780, 2016.
- [13]. Ashraf Abu-Ein, Ziad AA Alqadi, Jihad Nader, A TECHNIQUE OF HIDING SECRETE TEXT IN WAVE FILE, International Journal of Computer Applications, 2016.
- [14]. Jihad Nadir, Ashraf Abu Ein, Ziad Alqadi, A Technique to Encrypt-decrypt Stereo Wave File, International Journal of Computer and Information Technology, v. 5, issue 5, pp. 465-470, 2016.
- [15]. Majed O Al-Dwairi, Ziad A Alqadi, Amjad A Abujazar, Rushdi Abu Zneit, Optimized true-color image processing, World Applied Sciences Journal, v. 8, issue 10, pp. 1175-1182, 2010.
- [16]. AL-Azzeh Jamil, Bilal Zahran, Ziad Alqadi, Belal Ayyoub, Mazen Abu-Zaher, A Novel Zero-Error Method to Create a Secret Tag for an Image, Journal of Theoretical and Applied Information Technology, 2018.
- [17]. Jihad Nader, Ziad A. A. Alqadi, Bilal Zahran, Analysis of Color Image Filtering Methods, International Journal of Computer Applications, v. 174, issue 8, pp. 12-17, 2017.
- [18]. J Al-Azzeh M Abuzalata, Ziad Alqadi, Modified Inverse LSB Method for Highly Secure Message Hiding, International Journal of Computer Science and Mobile Computing, v. 8, issue 2, pp. 93-103, 2019.
- [19]. Qazem Jaber Ziad Alqadi, Jamil azza, Statistical analysis of methods used to enhance color image histogram, XX International scientific and technical conference, 2017.
- [20]. ZIAD ALQADI, A MODIFIED LBP METHOD TO EXTRACT FEATURES FROM COLOR IMAGES, Journal of Theoretical and Applied Information Technology, v. 96, issue 10, pp. 3014-3024, 2018.
- [21]. Bilal Zahran Belal Ayyoub, Jihad Nader, Ziad Al-Qadi, Suggested Method to Create Color Image Features Victor, Journal of Engineering and Applied Sciences, v. 14, issue 1, pp. 2203-2207, 2019.
- [22]. A Waheeb, Ziad AlQadi, Gray image reconstruction, Eur. J. Sci. Res, v. 27, pp. 167-173, 2007.
- [23]. 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, v. 975, pp. 8887, 2016.
- [24]. Qazem Jaber Ziad Alqadi, Jamil azza, Statistical analysis of methods used to enhance color image histogram, XX International scientific and technical conference, 2017.
- [25]. Ahmad Sharadqh Naseem Asad, Ismail Shayeb, Qazem Jaber, Belal Ayyoub, Ziad Alqadi, Creating a Stable and Fixed Features Array for Digital Color Image, IJCSMC, v. 8, issue 8, pp. 50-56, 2019.

ľ

- [26]. Ahmad Sharadqh Jamil Al-Azzeh, Rashad Rasras, Ziad Alqadi, Belal Ayyoub, Adaptation of matlab K-means clustering function to create Color Image Features, International Journal of Research in Advanced Engineering and Technology, v. 5, issue 2, pp. 10-18, 2019.
- [27]. Ghazi M. Qaryouti Ziad A. AlQadi, Yousf Eltous, Mohammad Abuzalata, Detecting and Counting Minutiae in Human Fingerprint, OSJ, v. 5, issue 1, pp. 1-5, 2020.
- [28]. Yousf Eltous Ziad A. AlQadi, Ghazi M. Qaryouti, Mohammad Abuzalata, ANALYSIS OF DIGITAL SIGNAL FEATURES EX-TRACTION BASED ON KMEANS CLUSTERING, International Journal of Engineering Technology Research & Management, v. 4, issue 1, pp. 66-75, 2020.
- [29]. 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, v. 9, issue 1, pp. 75-83, 2020.

Dr. Ghazi M. Qaryouti, et.al, Comparative Study of Voice Signal Features Extraction Methods" *IOSR Journal of Computer Engineering (IOSR-JCE)*, 22.1 (2020), pp. 58-66.

DOI: 10.9790/0661-2201015866