Multiple Hybrid Compression Techniques for Electroencephalography Data

Mohamed Adel[1], Mohamed El-Naggar[2], M. Saeed Darweesh[1,3], Hassan Mostafa[2,3]
[1]Institute of Aviation Engineering and Technology, Giza, Egypt
[2]Faculty of Engineering, Cairo University, Giza, Egypt
[3]Nanotechnology Department, Zewail City of Science and Technology, Egypt

***Abstract*—The large data size of Electroencephalography (EEG) is a result of long-time recording, the large number of electrodes, and a high sampling rate together. Therefore, the required bandwidth and the storage space are larger for efficient data transmission and storing. So, for higher efficiency transmission with less bandwidth and storage space, EEG data compression is a very important issue. This paper introduces two efficient algorithms for EEG compression. In the first algorithm, the EEG data is transformed through Discrete Wavelet Transform (DWT). Then it passes through Set Partitioning in Hierarchical Trees (SPIHT) compression algorithm. While in the second algorithm the data is segmented into N segments and these segments are transformed using Discrete Cosine Transform (DCT) then encoded using Uniform Quantized Huffman (UQH) scheme. Finally, the Lempel Ziv Welch (LZW) is used as a second lossless encoding algorithm for making a heavy compression. The system performance is evaluated in terms of the total time for compression and reconstruction, compression ratio, and root mean square error. The proposed hybrid technique DCT/UQH/LZW achieves 95% compression compared to 59% by DCT/RLE with the same similarity. Furthermore, it reduces 50% less root mean square error.**

Keywords—Electroencephalography, Data Compression, Lossy, Lossless, Hybrid Techniques.

I. INTRODUCTION

Recently, transmission of electrophysiological monitoring signal such EEG signal is the most important problem in the medical applications. Because of the large amount of data resulting from a group of sensors placed over or inside the human body and long hours recording the data, that makes real-time monitoring is a challenge. Therefore, data compression techniques are required for efficient communication purposes [1]. Data compression techniques can be classified into Lossy and Lossless techniques. In lossy compression, the original data can be completely reconstructed without any losing data that is in Lossless compression. While in lossy compression, some of the data can be loosed which causes a non-complete recovery [2].

One of the difficulties of EEG signal that it has high randomness [3]. Therefore, there are limitations on compression ratio (CR) in the lossless techniques Thus, lossy compression techniques are used with an acceptable level of distortion to moderate the CR, and it's a tradeoff between choosing reasonable distortion with acceptable CR.

The EEG data compression is discussed by several literature works. The presented work in [1] considered a compression algorithm for ECG data composed from DCT and RLE. While [4] considered the use of the DCT algorithm for lossy EEG data compression, it is not achievable to get high CR by using the DCT only. Authors in [5] considered run length encoding. Finally, the work in [2] is proposing the hybrid system DCT with Uniform Quantized Huffman. High CR can be achieved in this work, but it consumes a long-time for compression and reconstruction processes. In this paper, we propose a hybrid compression technique using DCT follows by uniform quantized Huffman coding then Lempel Ziv coding to introduce high compression ratio. The performance is evaluated with respect to some metrics as CR, RMSE, SSIM, and the total time consumed for compression and reconstruction processes.

The paper is covering the following: Section II gives a description of EEG compression techniques; DCT, SPIHT, UQH, and LZW. The implementation of the proposed system and performance metrices are introduced in Section III. The simulation results are presented in Section IV. Finally, the whole work is concluded in section V.

II. DATA COMPRESSION TECHNIQUES

This section gives an overview of the data compression techniques that are introduced. The data compression techniques are divided to two approaches: lossless and lossy.

A. Discrete Cosine Transform (DCT)

Fourier transform techniques are responsible for decomposition of the signal into its frequency components. A backward decomposition operation is responsible for retrieving the original signal from its component. So, the Fourier transforms can be called as an invertible transform. The Fourier based transforms are used for regular time-invariant signals. To get the time-specific information, we must apply the Fourier-based transforms over a pre-determined temporal window [6]. The DCT is lossy compression technique and one of the versions of the windowed Fourier transform. It provides great properties in the field of data compression. The DCT also represent the

978-1-5386-8167-1/18/$31.00 ©2018 IEEE

input signal with a minimal number of parameters [2]. Let Y(u) is the output, the output is a set of n coefficients of DCT and f(x) is the input which is a set of n data values. For n real numbers, the one-dimensional DCT is expressed as follows [7]:

$$y(u) = \sqrt{\frac{2}{n}}\,\alpha(u)\sum_{x=0}^{n-1} f(x)\cos\left(\frac{\pi(2x+1)}{2n}\right) \qquad (1)$$

Where

$$\alpha(u) = \begin{cases} \frac{1}{\sqrt{2}}, & u = 0 \\ 1 & u > 0. \end{cases} \qquad (2)$$

Y(0) is the DC coefficient and contain the mean value of the original signal. The rest of the coefficients are the AC coefficients.

B. Set Partitioning in Hierarchical Trees (SPIHT)

SPIHT depends on the significant coefficients only. The coefficients are considered significant only if they fulfill the condition $|C_{i,j}| \geq 2^n$ so basically the coefficients are divided into partitioning subsets T_m and by applying the previous condition on them we get two options, first one "no" indicating that all coefficients in subset T_m are insignificant. While "yes" means that the subset is significant and should be divided into subsets $T_{m,l}$ and these subsets should be subjected to the same condition and this test continues for all significant sets to identify all significant coefficients [8]. The objective here is to provide two types of sets, one contains insignificant coefficients and this one should have most of the elements and another one contains the significant coefficients and this one with only one significant element [8]. The following equation (3) simply indicates whether the set is significant or not where $S_n(T)$ is the sets of T coordinates and $|C_{i,j}|$ is the coefficient located in certain (i, j) position.

$$S_n(T) = \begin{cases} 1, & max_{(i,j)\epsilon T}\{|C_{i,j}|\} \geq 2^n \\ 0, & otherwise \end{cases} \qquad (3)$$

C. Uniform Quantized Huffman (UQH)

Huffman coding is an important class of lossless algorithm as it has a prefix code where the length of the code is proportional related to the entropy of the code [9]. Although classical Huffman algorithm is optimal for unrelated symbols with a known input probability distribution therefore, it is suitable for finite number of levels, while it cannot be applied for discrete-time continuous-amplitude signals (EEG samples), because of it has an infinite and uncountable set of numbers, so that the probability distribution of the infinite levels cannot be estimated therefore, we propose what we called "Quantized Huffman Algorithm" [2]. Each EEG sample value is mapped to a discrete level (to represent this level with a sequence of bits) in a process called quantization. Quantization is generally irreversible and results in loss of information, because of this process introduces distortion into the quantized signal that cannot be eliminated. The fundamental trade-off in this choice is the resulting signal quality versus the amount of data needed to represent each sample. Also, it is natural to expect that the amount of distortion introduced depend on the quantizer. The quantizer used here in this paper is uniform quantizer which the transition and reconstruction levels are all equally spaced.

D. Lempel Ziv Welch (LZW)

LZW is a lossless data compression algorithm that can be applied to any discrete source. In this algorithm a defined rule is used to purse strings of symbols into substrings or words from a finite alphabet A. These substrings or words must not exceed Length of string (L_s). Also, the substrings need to be mapped over the same alphabet A sequentially into uniquely decipherable code words of fixed Length of code (L_c) [10].

III. IMPLEMENTATION OF THE PROPOSED SYSTEMS

A. DWT and SPIHT Technique.

As shown in Fig. 1, the EEG signal is transformed through the DWT [6] and the confidence resulted from this block is the input to the SPIHT block which determine the significant coefficients and the insignificant ones and sort them in sets called list of insignificant sets, list of insignificant pixels and list of significant pixels according to a controlled threshold which adjusts to yield a specific the accuracy of the output signal. This threshold is chosen in this paper to be 2^{20}. The accuracy and compression ratio are inversely related. In the decoder side the inverse process is done as the ISPIHT rearrange the coefficients again to be the input of the IDWT which transform the coefficients to get the EEG signal.

B. DCT, UQH and LZW Technique.

As shown in Fig. 2, the input EEG signal is transformed through the DCT block into the frequency domain and the high frequency components in the coefficients considered being redundant data, so a level of the threshold is set to get rid of these redundancies and this threshold is the controller of the compression ratio and the resulted accuracy. The output of the DCT is the input to the UQH to transform it into a binary form with lower size. Then the binary data enters the LZW block, it firstly initializes dictionary contacting strings that have the long one and hence find the longest string S that matches the current input in the dictionary, S is removed from the input and the dictionary input for S is emitted, finally S and the next symbol is added to the input to the dictionary. These steps are repeated to all inputs.

IV. SIMULATION RESULT

Computer specifications that used in the simulation of the EEG compression are Intel(R) Core (TM) i7-3770 CPU @ 3.4GHz

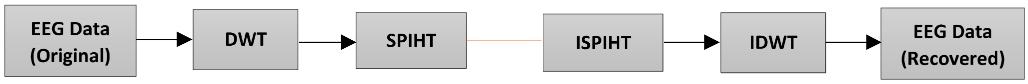

Fig. 1. Block diagram of DWT and SPIHT

Fig. 2. Block diagram of DCT, UQH and LZW

and internal memory (RAM) 8.00 GB. The five-performance metrics which are used in this paper are presented as follows:

1) *Compression Ratio* [1].
2) *Normalized Root Mean Square Error (NRMSE)* [2]
3) *Compression and Recovery Time* [1].
4) *Structural Similarity Index (SSIM)* [1].
5) *Figure of Merit (FOM):* FOM is a scoring parameter which gives a recommendation for one system from the proposed systems as in (4).

$$FOM = \frac{NRMSE \times Time}{CR \times SSIM} \tag{4}$$

A. DWT and SPIHT Technique

Fig. 3 shows that by varying the CR, RMSE is varying and It is obvious from Fig. 3 that the SPIHT algorithm achieves a compression ratio around 90.5% at NRMSE 0.016.

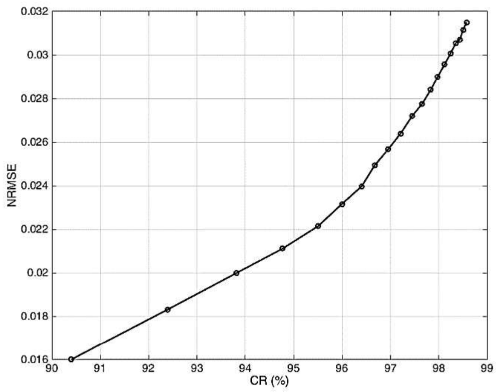

Fig. 3. Compression ratio versus NRMSE

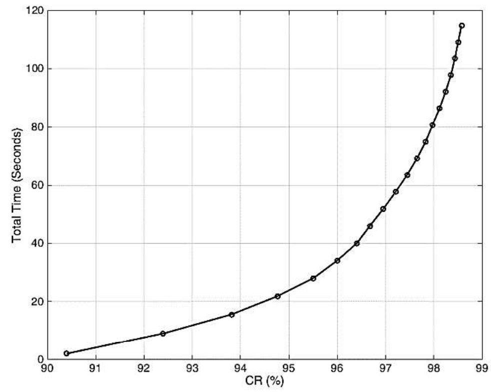

Fig. 4. Total time versus CR

While Fig. 4 shows the total time with compression ratio. We notice from this figure that, to get a high the compression ratio then a long-time is needed but this will increase the NRMSE.

B. DCT, UQH and LZW Technique.

The quantization level and the threshold are the main two parameters that control the compression ratio. So, the maximum compression ratio 0.95 at the minimum quantization level 21 and maximum threshold number 0.5.

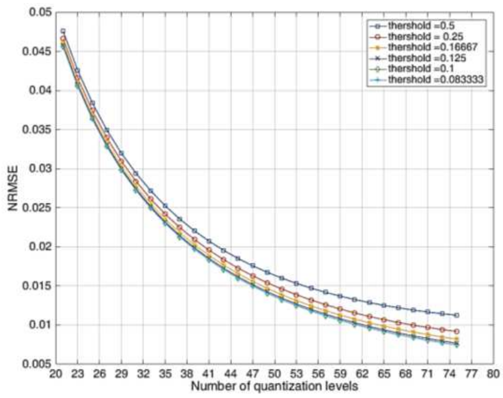

Fig. 5. Normalized RMSE versus different quantization levels

NRMSE deals with the number of quantization levels and the threshold value. As shown in Fig. 5, the minimum value of NRMSE occurs at the maximum quantization level with the minimum threshold. The total time is direct proportionally with the complexity of the algorithm. The minimum value of the total time (43.26 seconds) occurs at the maximum CR at maximum threshold 0.5 with the minimum quantization level 21 as shown and Fig. 6. From Table 1, it is demonstrated that, the (DCT/UQH/LZW) technique can achieve compression ratio greater than the compression ratio of (DCT/SPIHT) with more similarity but the time consumption is bigger for the (DCT/UQH/LZW).

As shown in Fig. 7, the DWT/SPIHT algorithm is significantly superior in FOM parameter due to its significant time response. Even though its compression ratio and the similarity are not the best among other algorithms but is the optimal algorithm as it provides a relatively good compression ratio and similarity with the minimum time. On the other hand, the DCT/UQH/LZW algorithm gives a good performance because of it depends on Huffman/LZW compression algorithm which is based on the stream of unrelated symbols with a known input probability distribution. But the DCT/UQH/LZW algorithm drawback is that it takes a huge time due to its quantization levels so its FOM is not the best among other techniques.

Table 1: Comparison with previous work

	[1]	[2]	[11]	[12]	DWT/ SPIHT	DCT/ UQH/ LZW
RMSE	14.73	14.72	12.07	30.85	22.003	7.4
Compression	59%	90%	70%	83.1%	90.5%	95%
Similarity	0.97	0.97	-	-	0.95	0.97
Time	23.91	13.81	0.08	Real-time	2.014	43.26
FOM	0.647	0.424	-	-	0.0542	0.365

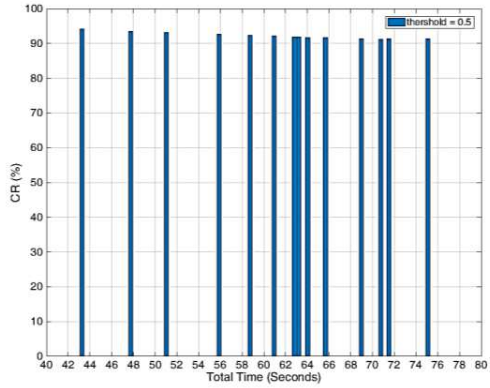

Fig. 6. Compression ratio versus total time at the maximum threshold

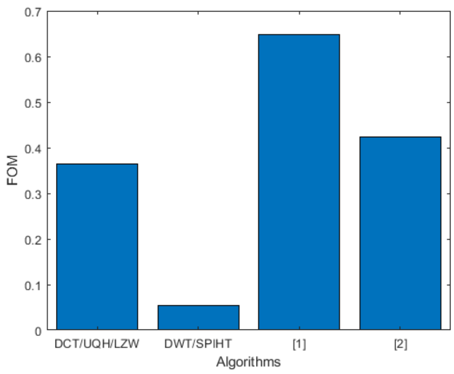

Fig. 7. Figure of Merit score graph

V. CONCLUSION

In this paper, two proposed algorithms are presented which are SPIHT which is a technique based on the wavelet transformation. This algorithm proposed a significant improvement in the total time of compression/decompression among the other presented techniques despite the similarity that considered to be the lowest one among other techniques. While the other proposed algorithm is a hybrid compression algorithm between DCT, UQH and LZW technique, where the EEG signal is segmented into N segments then the DCT compress each segment. Then uses UQH to compress the signal into binary code then the binary coded in the codebook using LZW algorithm. This hybrid algorithm achieved a significant result in the compression ratio and RMSE among the other techniques. But on the other hand, the total time is long compared to other techniques.

ACKNOWLEDGMENT

This research was partially funded by ONE Lab at Zewail City of Science and Technology, Egypt and Cairo University, Egypt.

REFERENCES

[1] M. Alsenwi, T. Ismail and H. Mostafa, "Performance analysis of hybrid lossy/lossless compression techniques for EEG data," *in 28th International Conference on Microelectronics (ICM)*, pp. 1-4, 2016.

[2] R. M. Elaskary, M. Saeed, T. Ismail, H. Mostafa and S. Gabran, "Hybrid DCT/Quantized Huffman compression for electroencephalography data," *in Japan-Africa Conference on Electronics, Communications and Computers (JAC-ECC)*, pp. 111-114, 2017.

[3] D. Birvinskas, I. Jusas, and Damasevicius, "Fast DCT algorithms for EEG data compression in embedded systems," *in Computer Science and Systems*, vol. 12, no. 1, pp. 49–62, 2015.

[4] L. J. Hadjileontiadis, "Biosignals and compression standards," *in MHealth.* Springer, pp. 277–292, 2006.

[5] S. Akhter and M. Haque, "ECG compression using run length encoding," *in 18th European Signal Processing Conference*, pp. 1645–1649, 2010.

[6] M. Ashraf, H. Mostafa, and A. Eladawy, "A Low-Power Area-Efficient Design and Comparative Analysis for High-Resolution Neural Data Compression," in *IEEE International Conference on Microelectronics (ICM 2016)*, Egypt, pp. 217-220, 2016.

[7] M. Alsenwi, M. Saeed, T. Ismail, H. Mostafa and S. Gabran, "Hybrid Compression Technique with Data Segmentation for Electroencephalography Data," *in 29th International Conference on Microelectronics (ICM 2017)*, Lebanon, pp. 235-238, 2017.

[8] A. Said, W.A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *in IEEE Transaction of Circuit Systems*, vol. 6, no. 3, pp. 243–250, 1996.

[9] R. M. Elaskary, M. Saeed, T. Ismail, "Non-Uniform Quantized Huffman Compression Technique for EEG Data," in *13th International Computer Engineering Conference (ICENCO)*, pp. 219-223, Cairo, December 2017

[10] Ziv, Jacob; Lempel, Abraham, "A Universal Algorithm for Sequential Data Compression," *in IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 337-343, 1977.

[11] Hamza Djelouat, Hamza Baali, Abbes Amira, and Faycal Bensaali, "An Adaptive Joint Sparsity Recovery for Compressive Sensing Based EEG System," *Wireless Communications and Mobile Computing*, Article ID 9823684, 2017. https://doi.org/10.1155/2017/9823684.

[12] L. Shaw, D. Rahman and A. Routray, "Highly Efficient Compression Algorithms for Multichannel EEG," *in IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 5, pp. 957-968, May 2018.