EEG Data Compression Using Tap9/7 Wavelet Transform and Double Shift-Coding

Hend A. Hadi[1(✉)] and Loay E. George[2]

[1] Ministry of Education\General Education Director of Baghdad Karkh-3, Baghdad, Iraq
hend_amir2007@yahoo.com
[2] University of Information Technology and Communication (UoITC), Baghdad, Iraq

Abstract. The main challenge of transmitting or storing Electroencephalogram (EEG) signals is the size of EEG data recordings. In this paper, an efficient and fast EEG compression system based on bi-orthogonal wavelet transform (Tap9/7) is proposed. It consists of three primary steps: (1) Tap (9/7) wavelet transform, which decomposes the EEG signal into low and multi-high subbands; (2) Progressive hierarchical quantizer, which quantizes wavelet subbands; and (3) double-shift coding, which encodes the input stream with three code-words while using the fewest total bits possible. When evaluating the efficiency of the EEG compression system, the compression ratio (CR) and mean square error (MSE) were the metrics that were utilised. The complexity of the compression system was reduced when using the Tap (9/7) wavelet transform, which also produced superior results in terms of CR and MSE compared to those obtained using the Discrete Wavelet Transform (DCT). The experimental tests are carried out with the CHB-MIT dataset, and the best compression ratio (CR = 7) is accomplished with an error level that is close to zero (MSE = 0.044).

Keywords: EEG Compression · Tap9/7 · Delta Modulation · Double-shift encoder · Histogram · Compression-Ratio · Mean Square Error

1 Introduction

Data compression techniques were applied in many fields, to efficiently transmit or store sounds, texts, images, and signals [1]. EEG signal is the recordings of the electrical activity of the brain, it is a non-surgical technique in which the electrodes are attached, in a pairwise manner, to the scalp to get the difference in voltage between specific spatial locations [2, 3]. Additionally, EEG is frequently used in a variety of disciplines, including the diagnosis of brain diseases and in brain-computer interfaces (BCI) [4]. The medical data is sensitive and required to be error-free [5]. The data can be transmitted remotely from one device to any device, such as a hospital terminal. Continuous care monitoring in the intensive care unit increases the amount of data transmitted through communication channels. This data must be transmitted quickly and without loss, so lossless compression is essential [6]. Although lossy compression provides a suitable compression rate with near-zero error, data without distortion is required for intelligent systems to accurately

detect diseases or abnormal events [7]. As a result of the importance and sensitivity of EEG data for medical professionals and researchers, a great number of studies have been carried out in this field in an effort to achieve the best possible compression results with zero or nearly zero error [2, 8].

In a lossless compression system, (Hejrat, 2017) employ inter- and intra-channel correlations. First, a differential pulse-code modulation technique is used as a preprocessing step to extract intra-channel correlation. Next, channels are grouped into distinct clusters, and arithmetic coding is utilised to determine the clusters' centroid. In the second stage, the distance between each cluster's data and its centroid is computed and then compressed using arithmetic coding. The proposed method demonstrated a higher compression rate than the presented ones [7].

(Srinivasan, 2013) proposed lossless/near-lossless compression systems using volumetric coding and image for multichannel EEG compression systems. Appropriate representations are used to utilize EEG channel correlations. Volumetric data (tensor) or image (matrix) are used to represent multichannel EEG, and then a wavelet transform is applied to those representations. The proposed compression systems are designed based on the principle of "lossy plus residual coding", consisting of a wavelet-based lossy coding layer followed by arithmetic coding on the residual. The proposed systems are applied to three EEG datasets. The proposed multichannel compression systems attained a good compression ratio compared to systems based on single-channel compression [8].

A proposal to build a lossless EEG compression system was made by Sriraam (2012), and it suggested using neural network predictions in conjunction with the correlation dimension (CD). The coefficient of determination (CD) is a measure of the correlation between EEG samples that is used to characterise irregular EEG signals. The CD value of each segment is then calculated after the input EEG signals have been segmented into one-second chunks using the segmentation tool. After that, the segments that had CD values that were the most similar were grouped together to form blocks. Because of this configuration's increased accuracy, the predictor needed fewer bits to be transmitted [9], which was a benefit to the transmission process.

(Wongsawatt, 2006) proposed a lossless multi-channel compression system by making use of the inter-correlation between EEG channels and the Karhunen-Loeve transform. In addition to this, they reduced the amount of unnecessary temporal repetition by utilising an integer time-frequency transform [10].

An effective and speedy EEG compression system that is based on the bi-orthogonal wavelet transform (Tap9/7) and double shift coding has been proposed in this body of work. An earlier study [11] suggests using DCT in conjunction with double shift coding in order to compress EEG signals. As can be seen in the section devoted to the experimental results, Tap9/7 demonstrates superior performance to DCT in terms of CR and MSE when it comes to efficiently compressing EEG data with near-zero error.

The contents of this paper are outlined below. In Sect. 2, the EEG data compression system will be presented and described. Section 3 contains a presentation of the experimental findings as well as a comparison with other relevant works. The conclusions are presented in the fourth section.

2   The Proposed Compression System

In our earlier research, we developed and validated a lossless EEG compression system using the Motor Movement Imagery dataset [12]. In another work, an EEG compression system based on DCT and Delta modulation was produced [11]. In this work, a new study on EEG data compression using biorthogonal (Tap9/7) wavelet transform is presented. The EEG compression scheme is applied through three stages; these stages are (A) Transformation, (B) Quantization, and (C) Encoding. Firstly, biorthogonal (Tap9/7) wavelet transform is applied. Secondly, the outputs are passed through a progressive hierarchical quantizer to eliminate the existing psycho-visual redundancy. Finally, the quantized values are encoded by the double-shift coding. The structure of the compression system is shown in Fig. 1.
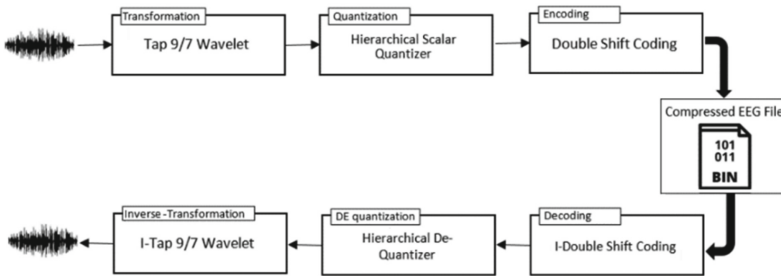

Fig. 1. (2): The proposed EEG compression system

2.1   Transformation (Bi-orthogonal Tap9/7)

The Bi-orthogonal Transform (Tap9/7) is one type of Cohen_Daubechies_Feauveau (CDF) orthogonal bi-orthogonal wavelet [13]. The Tap9/7 transform decomposes the original signal into multi-high (detailed) and low (approximation) sub-bands. For the high sub-band, no analysis is done any further, while the output low sub-band is then divided into new low and high signals [14]. Seven coefficients are included for the high pass filter, while nine coefficients are included for the low pass filter [14]. Applying the lifting and scaling steps, respectively, will result in the completion of this transformation. The lifting step is achieved using a series of phases: Split phase, Predict phase, and Update phase [15]. Tap9/7 has more complexity and accurate results than other transforms [14].

2.2   Hierarchical Quantization

Hierarchical Quantization maps the values of outcomes sub-bands from real numbers to integers [16]. Quantization is a process that involves the removal of irrelevant data as well as a decrease in the number of bits needed to accurately represent and store the values of the Tap9/7 coefficients. This reduction in the number of bits is accomplished

through the elimination of redundant information [17, 18]. Hierarchical quantization divides each sub-band coefficient by an Eq. (1)-generated quantization value (Qstp):

$$Q_{stp}(level_i) = Q_{stp}(level_{i-1}) * alpha \qquad (1)$$

where i is the ith wavelet level, and alpha <1.

2.3 Encoding (Double Shift Coding)

The double shift coder, an improved shift coder, is recommended for encoding quantized values (that uses three code words, the short code word to encode the most frequent small values, and the two long code words, to encode other less frequent large values). Double shift coding requires partitioning values into three sets. The first set encodes the most common symbols with a short code-word. The other two sets, which occur less often, are encoded with a long code and then separated into two sets using one shift bit that is either 0 or 1. These steps make up the double-shift coder [19].

Delta Modulation. In order to reduce the signal values, (DM) is computed to determine the difference between each pair of values that are adjacent to one another. (DM) is a simplified and specialised form of differential pulse code modulation (DPCM), and it is accomplished by utilising the Eqs. (2) and (3) [20]:

$$DM(i) = S(i) - S(i - 1) \quad if\ i > 0 \qquad (2)$$

where DM (i) is the ith item of delta modulation array, S is the transformed signal:

$$DM(0) = S(0) \qquad (3)$$

In order to reduce the amount of complexity in the coding process, the "mapping to positive" step is applied to the results of the "DM" step. In order to accomplish this, the positive values are mapped to even numbers, and the negative values are mapped to odd numbers. This is done so that the values can be recognised during the decompression stage [22]. The negative results of applying delta modulation can be turned into positive results using Eq. (4) [21]:

$$X_i = \begin{cases} 2X_i & if\ X_i > 0 \\ -2X_i - 1 & if\ X_i < 0 \end{cases} \qquad (4)$$

where Xi refers to the ith component of the array that is being utilised for delta modulation.

After the mapping from negative to positive has been finished, the histogram of the newly created positive array, which is referred to as the His array, is computed in order to determine the value that is the highest that is possible. In order to find two short code words in addition to the two long code words, it is essential to use the maximum value on the code optimizer.

A Coding Optimizer. Is utilised in order to establish the optimal bit length of the code words (a short code word and two longs) that will be used to encode the entire stream. The optimizer works with the histogram array and makes use of the histogram array's maximum value in order to search for the optimal bits required to encode the entire input stream with the fewest possible total bits.

Shift Encoder. Using the three code words provided by the optimizer, the input stream is finally given the opportunity to be encoded before being saved away as a binary file.

3   Experimental Results and Comparison with Related Works

The CHB-MIT dataset is utilised in the experimental tests that are carried out to assess the efficiency of the presented EEG compression system. This dataset is comprised of 23 EEG channels that were recorded from pediatric patients with epilepsy who were being treated at the "Children's Hospital in Boston, Massachusetts. [7, 22]. A total of 256 Hz were used to sample each individual channel that was included in this dataset. In this research, the utilisation of delta modulation on Tap9/7 was investigated. The results of the tests demonstrated that the performance of the system with Tap9/7 is significantly higher than that of the method that based on DCT [10].

The compression ratio (CR) is a metric for compression, and it is necessary to know how much information can be removed from the input data while data compression is taking place in order to keep only the information that is the most important and crucial to the original data. In this particular piece of research, the compression ratio, abbreviated as CR, is utilised in order to evaluate the effectiveness of the system. Compression is quantified using the CR scale. The expression that defines CR is as follows: (6) [23, 24]:

$$CR = \frac{UnCompressed\ Size}{Compressed\ Size} \tag{5}$$

Table 1 shows a selection of the results that are tested for the proposed compression system's performance in terms of CR and MSE, as well as the effect of wavelet levels and some quantization values. In accordance with Eq. (1). Figure 2 shows a curve that can be interpreted as an indicator of the overall performance of the system. When the wavelet level is set to 5, Q0 = 1, and Q1 = 5, Table 2 presents the CR and MSE values for each of the four distinct files.

On the CHB-MIT dataset, the outcomes of the proposed method were compared to the outcomes of a number of other works that are related; the outcomes of this comparison are shown in Table 3. The proposed system maintained an MSE that was very close to zero while demonstrating a higher CR than other works that are comparable (See Fig. 3).

Table 1. The impact of using wavelet levels and some tested quantization coefficients.

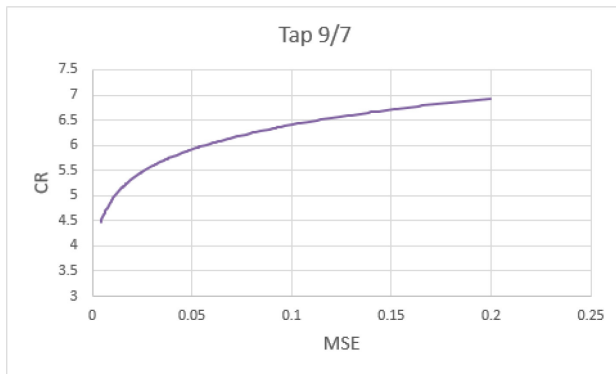| Wavelet-level	Q_1	Q_0	MSE	CR
5	1	1	0.006	4.54
4	1	1	0.005	4.55
3	1	1	0.005	4.57
2	1	1	0.004	4.58
1	1	1	0.004	4.71
2	2	2	0.017	5.23
5	3	1	0.018	5.40
4	3	1	0.018	5.43
3	3	1	0.017	5.45
6	5	3	0.061	6.08
6	5	4	0.073	6.08
5	5	3	0.063	6.11
5	5	4	0.077	6.12
7	7	6	0.143	6.59
7	7	7	0.163	6.59
8	8	7	0.179	6.79
8	8	8	0.199	6.79

Fig. 2. EEG compression system based Tap9/7 performance

Table 2. The outcomes in terms of CR and MSE of some tested files, the wavelet level = 5, Q1 = 5, and Q0 = 1.

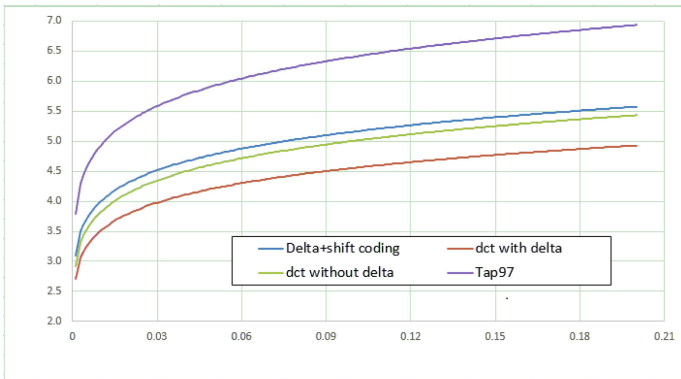| EEG-file	CR	MSE
Chb1	6.54	0.048
Chb2	6.04	0.048
Chb3	7.00	0.044
Chb5	6.05	0.048

Fig. 3. The comparison of the results obtained using Tap9/7 and previously proposed methods.

Table 3. The comparison with recent studies in terms of CR.

EEG-file	Proposed-method	[7]	[8]	[9]	[10]
Chb1	6.54	2.04	1.93	1.7	1.63
Chb2	6.04	1.61	1.68	1.67	1.38
Chb3	7.00	1.87	1.86	1.83	1.55
Chb5	6.05	1.9	1.77	1.61	1.45

4   Conclusions

A new investigation into an EEG compression method that uses Tap9/7, delta modulation and double shift coding is presented in this research. The DCT transform was outperformed by the utilisation of Tap9/7, which resulted in improved compression performance (i.e., an increase in CR and a decrease in MSE) while maintaining the quality of the EEG data. The wavelet levels as well as the quantization parameters (i.e., increasing the wavelet levels and quantization parameters increases the CR and MSE) can affect the effectiveness of the compression system. The number of total bits that should be used for encoding an entire input sequence can be calculated with the help of double

shift coding. The Chb3 file is compressed with the best possible CR = 7, achieved with an MSE of 0.044. The findings are evaluated in light of a prior study that made use of DCT and delta modulation, in addition to a few related works in terms of CR, and they are found to have a satisfactory level of performance.

An improvement can be done in the future by combining the DCT and Tap9/7 Wavelet transforms into a single compression system in order to increase the compression ratio. This would allow the system to take advantage of the benefits offered by both transforms in order to achieve a higher compression ratio. In addition, another method of coding can be utilised in order to test the functionality of the system that has been presented.

References

1. Antoniol, G., Tonella, P.: EEG data compression techniques. IEEE Trans. Biomed. Eng. **44**(2), 105–114 (1973)
2. Hadi, H.A., George, L.E.: EEG based user identification and verification using the energy of sliced dft spectra. Int. J. Sci. Res. (IJSR) **6**(9), 46–51 (2017)
3. Shaw, L., Rahman, D., Routray, A.: Highly efficient compression algorithms for multichannel EEG. IEEE Trans. Neural Syst. Rehabil. Eng. **26**(5), 957–967 (2018)
4. Dao, P.T., Li X.J., Do, H.N.: Lossy Compression Techniques for EEG Signals. In: International Conference on Advanced Technologies for Communications (ATC) (2015)
5. Higgins, G., et al.: The effects of lossy compression on diagnostically relevant seizure information in EEG signals. IEEE J. Biomed. Health Inform. **17**(1), 121–127 (2012)
6. Zhang, Z., Jung, T.-P., Makeig, S., Rao, B.D.: Compressed sensing of eeg for wireless tele-monitoring with low energy consumption and inexpensive hardware. IEEE Trans. Biomed. Eng. **60**(1), 221–224 (2012)
7. Hejrati, B., Fathi, A., Mohammadi, F.A.: Efficient lossless multi-channel eeg compression based on channel clustering. Biomed. Signal Process. Control **31**, 295–300 (2017)
8. Srinivasan, K., Dauwels, J., Reddy, M.R.: Multichannel EEG compression: wavelet-based image and volumetric coding approach. IEEE J. Biomed. Health Inform. **17**(1), 113–120 (2013)
9. Sriraam, N.: Correlation dimension based lossless compression of EEG signals. Biomed. Signal Process. Control **7**(4), 379–388 (2012)
10. Wongsawat, Y., Oraintara, S., Tanaka, T., Rao, K.R.: Lossless Multi-channel EEG Compression. ISCAS IEEE, pp. 1611–1614 (2006)
11. Hadi, H.A., George, L.E.: Comparative Study Using DCT, Delta Modulation and Double Shift coding for Compressing Electroencephalogram Data. Accepted Manuscript Iraqi J. Sci.(IJS), **63**(7) 2022
12. Hadi, H.A., George, L.E.: Lossless EEG Data Compression Using Delta Modulation and Two Types of Enhanced Adaptive Shift Coders. In: New Trends in Information and Communications Technology Applications. NTICT 2021. Communications in Computer and Information Science (2021)
13. Hadi, H.A., George, L.E.: Brainwaves for user verification using two separate sets of features based on DCT and wavelet. Int. J. Adv. Comput. Sci. Appl. **9**(1), 240–246 (2018)
14. Ahmed, Z.J., George, L.E., Hadi, R.A.: Audio compression using transforms and high order entropy encoding. Int. J. Electr. Comput. Eng. **11**(4), 3459–3469 (2021)
15. Beladgham, M., Bessaid, A., Lakhdar, A.M., Taleb-Ahmed, A.: Improving quality of medical image compression using biorthogonal CDF wavelet based on lifting scheme and SPIHT coding. Serbian J. Electr. Eng. **8**(2), 163–179 (2011)

16. Gray, R.M., Neuhoff, D.L.: Quantization. IEEE Trans. Inf. Theory **44**(6), 2325–2383 (1998)
17. Ahmed, Z.J., George, L.E.: A Comparative Study Using LZW with Wavelet or DCT for Compressing Color Images. In: 2020 International Conference on Advanced Science and Engineering (ICOASE) (2020)
18. Sagheer, A.M., Farhan, A.S., George, L.E.: Fast Intra-frame compression for video conferencing using adaptive shift coding. Int. J. Comput. Appl. **81**(18), 29–33 (2013)
19. Ahmed, Z.J., George, L.E.: Audio Compression Using Transform Coding with LZW and Double Shift Coding. In: International Conference on New Trends in Information and Communications Technology Applications(NTICT 2021) (2021)
20. George, L.E., Hassan, E.K., Mohammed, S.G., Mohammed, F.G.: Selective image encryption based on DCT, hybrid shift coding and randomly generated secret key. Iraqi J. Sci. **61**(4), 920–935 (2020)
21. Ahmed, A.H., George, L.E.: Color image compression based on wavelet, differential pulse code modulation and quadtree coding. Res. J. Appl. Sci. Eng. Technol. **14**(2), 73–79 (2017)
22. Shoeb, A.H.: Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment, (Doctoral dissertation, Massachusetts Institute of Technology) (2009)
23. Hung, N.Q.V., Jeung, H., Aberer, K.: An evaluation of model-based approaches to sensor data compression. IEEE Trans. Knowl. Data Eng. **25**(11), 2434–2447 (2012)
24. Sayood, K.: Introduction to Data Compression, Waltham, MA 02451. Elsevier, USA (2012)