34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Losless EEG Signal Compression

Arda Gumusalan, Ziya Arnavut*, and Huseyin Kocak

Abstract—**In this work, we study the lossless compression of EEG (electroencephalograph) signals using linear prediction and arithmetic coder. We show that, when we separate the less significant bits of each signal, linear prediction techniques yield better prediction, and with a structured arithmetic coder not only our technique achieves better compression rates than other techniques reported previously, but also our technique is much faster than the others.**

I. INTRODUCTION

There are several compression techniques applied to EEG signals. These techniques can be classified as multi-channel compression techniques or single-channel compression techniques. In this work, we investigate single-channel compression of EEG signals. In single-channel, the data collected from different channels is considered separately. Some of the commonly used compression techniques for EEG signals are: repetition count , dictionary-based compression, linear prediction, Burrow-Wheeler Transformation, Huffman coding [1], adaptive linear predictive compression [2], recurrent neural network predictors [3], context-based bias cancelation [4], [5], chaos-based modeling compression [6], power spectral density [7], and context-based error modeling using neural networks [8]. The common compression programs that we use in daily life are also suitable for EEG compression such as gzip and bzip2. Also, some of the techniques used for image compression can also be implemented for EEG signal data [9].

In this work, 21 different data sets obtained from piglets are used. The piglets were one week old and weighed about 3 to 4 kilograms. These data were provided by Prof Dr. Nitish Thakor from John Hopkins University, Baltimore for Prof. Memon [4], whom we obtained the data from.

Therefore, we used same data sets as the ones from the influential study on EEG signal compression by Memon et al. [4] for a fair comparison our results with some of the previous work.

Arda Gumusalan was an undergraduate student at SUNY Fredonia, Fredonia, NY 14063, USA, (e-mail: agumusalan@fredonia.edu) currently a graduate student at George Mason University.
*Ziya Arnavut is with SUNY Fredonia, Fredonia, NY 14063, USA, (phone:7166733864; fax:7166733804; e-mail: arnavut@fredonia.edu)
Huseyin Kocak is with University of Miami, Miami, FL 33124, USA, (e-mail: hk@cs.miami.edu)

This paper is organized as follows: In section 2 we explain the proposed technique. In section 3, we present the experimental results and compare them with the results obtained previously. Finally, in section 4, we conclude our work.

II. PROPOSED TECHNIQUE

There are three stages in our proposed EEG compression technique. These stages are: (1) the separation of the least significant bits from the data; (2) the linear prediction; and (3) the arithmetic coder. These stages are shown in Figure 2.

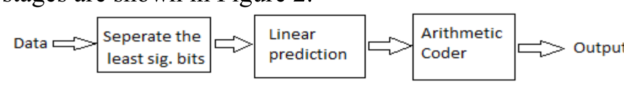

Figure 1 – The pipeline representation of the stages

A. The Least-Significant-Bit Technique

In our experiments, 21 different EEG signal files were used. Each data file is 12000 bytes in size and digitized as 16 bits per sample but each EEG signals used in this project consist of 10-bit data samples.

In the least-significant-bit technique, the left (higher) 8 bits and the right (lower) 2 bits are separated from each EEG signal. The lowest 2 bits and the highest 8 bits are stored in separate files. The lowest 2 bits are stored in raw data format which are 1500 bytes in size.

$$6000 * 2 / 8 = 1500$$

In the real time implementation of this technique, there is a buffer for higher 8 bit values and another buffer for lower 2 bit values. In order to obtain and compare the compression results, each buffer is stored in a separate file for experimental purposes. All the other stages are applied to the file which contains the higher 8 bit values. Reducing 10 bit values to 8 bit values reduces the variance in the distribution of the values. Figure 2 shows the histogram for the eegl12.dat test sample, whereas Figure 3 shows the histogram for the eegl12.dat.upper (The highest 8 bits of the eegl12.dat test sample.) In Figure 2, it is clearly seen that the data values range roughly between -450 and 450. On the other hand, Figure 3 shows the range of the higher 8 bit of the eegl12.dat sample. The new values range roughly between -120 and 120.

After separating off the least 2 significant bits from each data sample, the new distribution of the data can range between -128 and 127 (-2^7, 2^7-1) whereas the original data values can range between -512 and 511 (-2^{10}, $2^{10}-1$). This reduction in range results in a smaller standard deviation. In the eegl12.dat, for example, the standard deviation decreases from 138.1361 to 34.52201. Figure 4 illustrates the new signal obtained from the upper 8 bits of eegl12.dat test sample.

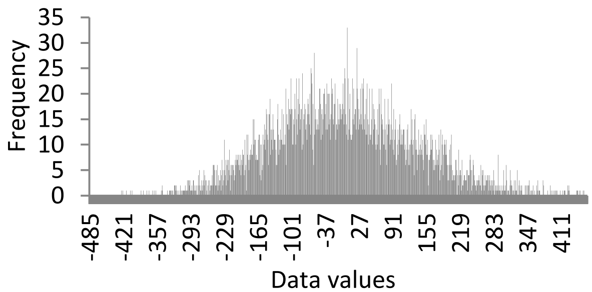

Figure 2: Histogram of eegl12.dat

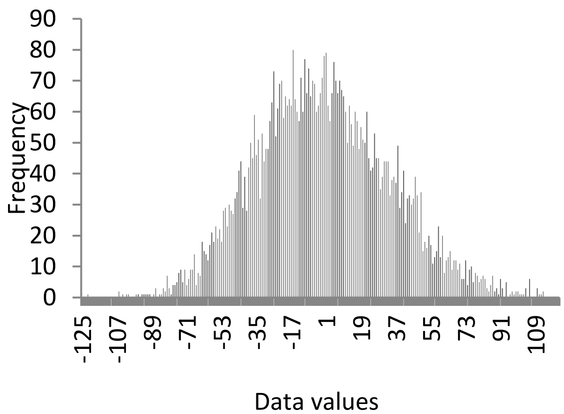

Figure 3-Histogram of eegl12.dat.upper

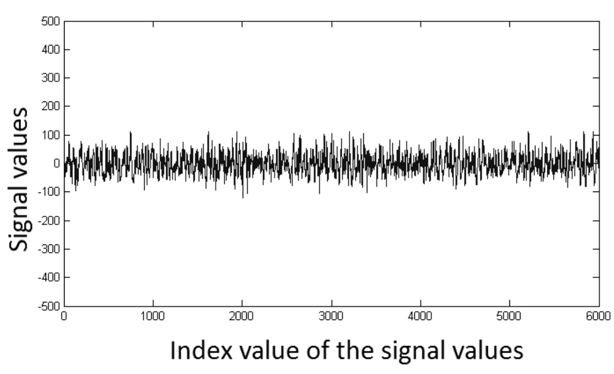

Figure 4- eegl12.dat.upper

The least-significant-bit technique might be implemented differently; for example, separating the right 3 bits from the left 7 bits will lower the standard deviation even more. However, for each file, storing 3 bits as raw data then will require 2250 bytes, which may cancel out the gain from lowering the standard deviation. Experimental results showed that separating the right 2 bits from the left 8 bits gives the best compression results.

B. Linear Prediction Stage

Prediction-based compression techniques are well known and commonly used technique for data compression [10]. There have been previous studies on compressing EEG signals using prediction-based compression [11, 12].

With the predictive approach, the value of the current sample is predicted according to the previous values that have already been transmitted. In order to reconstruct the transmitted data, only the transmission of the error values is needed. More precisely, if x_i is the actual value of the sample, $\hat{x}_i$ is the predicted value, and $\hat{e}_i$ is the error value, then

$$x_i = \hat{x}_i + \hat{e}_i$$
$$\hat{e}_i = x_i - \hat{x}_i$$

Various models of linear prediction are available and have been applied to different types of data. It is common to treat EEG signals as time series; in this case, Autoregression (AR) is a commonly used linear prediction method for time series, and so is applied to EEG signals. AR is based on the formula

$$x_k = \sum_{i=1}^{p} a_i\, x_{k-i} + e_k = \hat{x} + e_k$$

a_i is an autoregression coefficient, x_k is the value series under investigation, p is the order of the AR process, e_k is the prediction error.

The idea behind the AR model is that the current term of the series can be estimated by using a linear weighted sum of the previous terms, where the weights are the autoregression coefficients. Memon et al. [4] obtained the best results with six order autoregressive model. Taking a relatively larger order makes the prediction more accurate but may decreases the compression ratio. Previous studies have determined that using six order autoregressive model gives the best compression [13].

In our work, we used the *maximum entropy* method to calculate the AR coefficient values for each data file, which is based on the formula

$$e_k^q = x_k - q\left(\sum_{i=1}^{p} a_i\, x_{k-i}\right) = x_k - \hat{x}_k^q$$

where *q()* is a quantization function that rounds its variables to the nearest integer.

In order to obtain the original data, the receiver has to know the coefficients and the error values. Upon receiving the transmitted error and coefficients values, the receiver can produce the original signal values. Figure 5 shows the histogram of the error values after remapping the resulting values to positive integers.

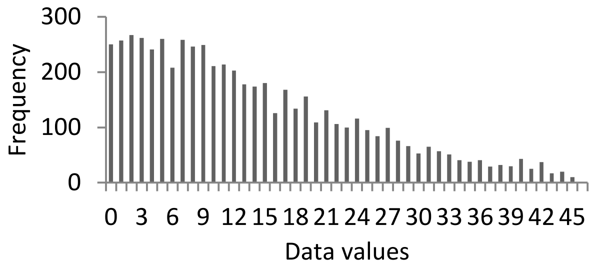

Figure 5- Histogram of upper byte of the EEG signals, after AR modeling and remapping.

The distribution of the error values for higher 8-bit ranges roughly between 0 and 100 in this specific example (only the frequencies for 0 to 45 are shown on the graph for a clear view). Note that all the error values are either 0 or a positive number. This is because, after AR modeling, we map the error values to a positive 8-bit integer value, so the error values can range between 0 and 256 (2^8). However, experimental results showed that the range is usually between 0 and 100. The new standard deviation of this data distribution is 13.128. As it can be seen, AR modeling lowers the standard deviation of the data significantly.

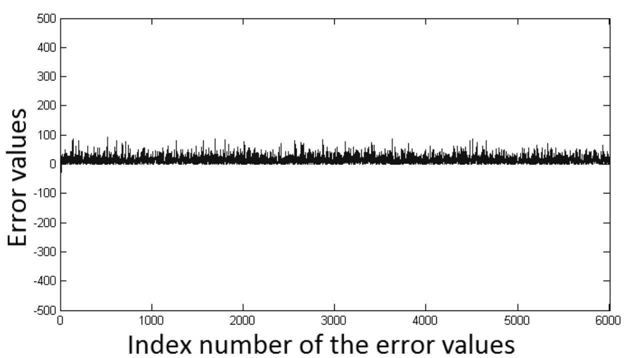

Figure 6- Error values, after AR modeling and remapping.

The separation of the left 2 bits of the data from the right 8 bits (see section 2.1) helps us keep the error value range less than 256. The experimental results show that, applying AR modeling to the data without the separation of the left 2 bits, and the right 8 bits, results in higher error values.

C. Arithmetic Coder Stage

Arithmetic coding is the final stage in our proposed technique. Arithmetic coding is a form of entropy coding which uses variable length codes. As a result, this coding procedure maps source symbols to a variable number of bits. Arithmetic coding is commonly used in lossless data compression due to its speed, low storage requirements, and efficiency. Huffman coding is another common entropy coder. However, in our experiments, we achieved better compression ratios with arithmetic coding. Several new implementation of arithmetic coding which incorporates improvements over a widely-used earlier version has been developed. During 1994-1996, while Peter Fenwick was studying block sorting compression technique [15], he developed a structured Arithmetic Coder that includes a run length coder for 0s and 1s. As shown in Figure 5, we observed that after autoregression applied to EEG data, the first 10 values are more dominant than the others. Hence, with this observation, in our study, we utilized a version of Fenwick's structured arithmetic coder, which yield the best results.

III. EXPERIMENTAL RESULTS

In our experiments, we used the same data files that were used in previous studies of prediction-based EEG signal compression by Memon et al. [4], and Arnavut and Koçak [1]. Figure 7 compares our experimental results with theirs.

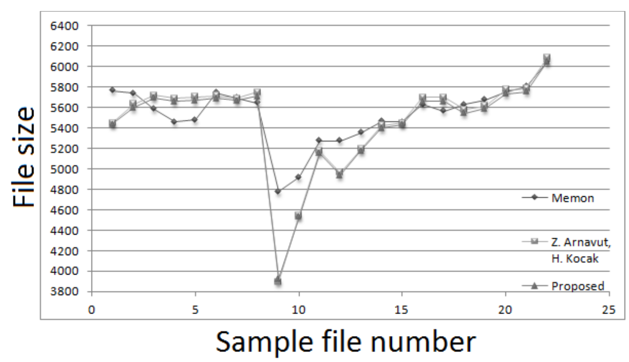

Figure 7- Comparison of experimental results

Memon, et al obtained their results by applying an order six AR modeling, and then context-modeling with a bias correction on error values [4]. They finally used a context-model binary arithmetic coder on the resulting error values. On the other hand, Arnavut and Koçak applied an order six AR modeling followed by Burrows-Wheeler Transformation with Inversion Coder (BWIC) technique [14] on the error values, generated by the AR model.

On these 21 sets of test data, our proposed method gives about 1.72% better compression than the method proposed in [4], on the 21 test data files. Moreover, our proposed method is faster than the method suggested in [4], since we utilize a structured arithmetic coder; while in [4] authors use bias cancelation and a context-model binary arithmetic coder. We also experimented with bias-cancelation technique. However, bias-cancelation did not yield any improvement in compression, when it is used with the structured arithmetic coder.

When we compare our results with the method proposed by Arnavut and Koçak [1], our proposed technique yields 0.48% better compression, on the same set of data files. However, our proposed technique runs in O(n) times, where BWIC runs in O(n log n) time, since BWIC uses BWT and Inversion coder. Both transformations require O(n log n) time [14], where n is the number of elements. Hence, our proposed technique is much faster and yields better results than the previously proposed linear predictive techniques for EEG signal compression.

To demonstrate practical relevance of our proposed algorithm, we implemented the techniques proposed in [4] and [14] and test them on the same UNIX server, using the same 21 data files. To get a more accurate reading of compression times, all three techniques were run 10 times each, on each file. Based on our experiments, we conclude that our proposed technique compresses 32% faster than the technique proposed in [4], while it compresses 64% faster than the technique proposed in [14].

IV. CONCULSION

Linear-prediction-based compression of EEG signals is a topic that has been employed in previous studies. In this work, we showed that by separating the least significant bits and then utilizing autoregression, and later compressing the residuals with a structured arithmetic coder gives better compression results than the results reported previously. We also showed that there is no need to use bias-cancelation, or the BWIC, when a structured arithmetic coder is employed after AR modeling for EEG signal compression.

Moreover, we showed that our technique is much faster than the other techniques reported previously.

REFERENCES

[1] Z. Arnavut, H. Koçak; , "Lossless EEG signal compression," *Fifth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control*, ICSCCW 2009, pp.1-4, 2-4 Sept. 2009.

[2] G. Antoniol and P. Tonella, "EEG data compression techniques," *IEEE Trans. Biomed*. Eng., vol. 44, pp. 105-114, 1997.

[3] F. Bartolini, V. Capellin, S. Nerozzi, and A. Mecocci, "Recurrent neural network predictors for EEG signal compression," in *Proc. Int. Conf. Acoust. Speech Signal Process*., 1995, pp. 3395-3398.

[4] N. D. Memon, X. Kong, and J. Cinkler, Context based lossless and near-lossless compression of EEG signals. *IEEE Trans. Inf. Technol. Biomed*. 3(3):231-238, 1999.

[5] N. D. Memon, X .Kong, and J. Cinkler, "Lossless and Near-Lossless Compression of EEG Signals", *Proceedings of IEEE Signals, Systems and Computers,* vol. 2, pp. 14321436, November 2-5, 1997.

[6] V. Kavitha and D. N. Dutt, "A chaos-based model for low complexity predictive coding scheme for compression and transmission of electroencephalogram data", *Medical and Biological Engineering and Computing*, vol 37, pp.316-321, 1999.

[7] T. Madan, R. Agarwal and M.N. S. Swamy, "Compression of long-term EEG using power spectral density", *26th Annual International Conference of the IEEE EMBS*. vol. 1, Sept. 2004, pp. 180-183.

[8] N. Sriraam and C. Eswaram, "Context-Based Error Modeling for Lossless Compression of EEG Signals Using Neural Networks", *Journal of Medical Systems*, vol. 30, pp. 439448, 2006.

[9] N. D. Memon and X. Wu, "Recent developments in lossless image compression," *The Computer Journal*, vol. 40, nos 2/3, pp. 127-136, 1997.

[10] R. Logeswaran and C. Eswaran, "Performance survey of several lossless compression algorithms for telemetry applications," *Int. J. Comput. Appl*., vol.23, no.1, pp. 1-9, 2001.

[11] N. Sriaam and C. Eswaran, "Performance evaluation of two stage-stage lossless compression of EEG signals," *Int. J. Signal Process*., vol. 1, no.2, pp. 94-97, 2004.

[12] F. Bartolini, V. Capellin, S. Nerozzi, and A. Mecocci, "Recurrent neural network predictors for EEG signal compression," in *Proc. Int. Conf. Acoust. Speech Signal Processing*, 1995, pp. 3395-3398.

[13] X. Kong, V. Goel and N. Thakor, "Quantification of injury-related EEG signal changes using Itakura distance measure," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, 1995, pp. 2947-2950.

[14] Z. Arnavut, "Inversion coding," *The Computer Journal*, vol.47, no. 1, pp.46-47, January 2004.

[15] P. Fenwick, "The Burrows-Wheeler Transform for Block Sorting Text Compression -- Principles and Improvements", *The Computer Journal*, Vol. 39, No 9, pp 731-740, 1996.