Comparative Study of Lossless ECG Signal Compression Techniques for Wireless Networks

Rupali V Tornekar[1], Suhas S Gajre[2]

[1]Vishwakarma Institute of Technology, Pune, India
[2]Shri Guru Gobind Singhaji Institute of Engineering & Technology, Nanded, India

Abstract

Electrocardiogram (ECG) is an important physiological signal which represents electrical activity of Heart. ECG plays important role in diagnosis of cardiovascular diseases. Telemedicine, telemonitoring requires huge amount of data to be stored for analysis and diagnosis purpose. Wireless sensor nodes consume lot of energy in data transmission. So Data Compression is needed for reducing storage space, transmission rate and effective utilization of bandwidth. This paper includes comparative study of various lossless compression methods for ECG signals in terms of compression ratio and execution time. It is found that minimum variance Huffman coding is best suited for ECG signal compression. Implementation is done in MATLAB software and database used is MITBIH Arrhythmia. 50% storage space can be saved with Minimum variance Huffman code with computational complexity of $NLog_2N$*. Bandwidth is effectively utilized and buffer design complexity is also reduced.*

1. Introduction

Cardiovascular disease is the leading cause of death in the world. Recent development in telecommunication and health care services is Telemedicine [7]. Through this advanced technique, people get provisional or first aid help in medicine and diagnosis. In rural communities where there is unavailability of medical experts, people can take advantage of telemedicine facility to save their time and money. In particular to heart disease analysis, signal monitoring and diagnosis, huge amount of ECG data is required to store which takes lot of storage space. In telemedicine, recorded data of patients is transferred over long distances for medical expert's opinion. Compression plays an important role here. Lossless compression of biomedical data is always preferred because of important clinical information irrespective of low compression ratio achieved.

Wireless sensor nodes are powered by limited energy supply. Hence energy consumption must be reduced. Data transmission consumes the largest amount of energy in a sensor node. Thus, one method to reduce the energy used is by compressing the data before transmitting it [1][6].

In the past, various algorithms for ECG signal compression are presented in the literature. They are based on time domain, transform domain, parametric based methods. Current research of ECG compression uses wavelet transform mainly because single transform resolves issues of time frequency localization, denoising, compression, feature extraction[2][3][4]. Even if lossy compression methods are preferred since it gives high compression ratio, still lossless compression is preferred in biomedical field.

It is mentioned in the literature that the Huffman algorithm is able to reduce the data size by 43% on average [6]. It is noticed that minimum variance Huffman codes are not focused in ECG signal lossless compression methods [5][8]. Since minimum variance codes are always preferable for effective utilization of channel, effectiveness of Minimum variance Huffman codes is studied.

This paper includes Comparative Analysis of Relative Effectiveness of Entropy Based Lossless ECG Signal Compression Techniques for wireless sensor node application. MITBIH Arrythmia database is used for experimentation. Database contains 47 patients ECG recordings of 10 second duration. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. Result shows Minimum variance huffman coding is better entropy based lossless compression method for ECG.

2. Lossless compression methods

ECG signals are compressed using various lossless compression methods like Shannon Fano, Normal Huffman, Minimum variance Huffman, Lempel Ziv and Welch (LZW) Dictionary technique and examined for their relative effectiveness. Comparison of these methods in terms of compression ratio achieved and execution time required is done. Lead II MITBIH Arrhythmia

 ISSN: 2325-887X DOI:10.22489/CinC.2017.095-236

database of 48 patients with sampling frequency 360 Hz and 11 bit resolution is used.

This paper deals with entropy coding algorithms like Shannon Fanon coding, Huffman's coding, minimum variance Huffman coding, LZW dictionary technique and run length coding.

2.1 Shannon Fano coding

Shannon says for any lossless compression method, we can not reduce number of bits than entropy value. Calculate probability of each unique sample present in the ECG signal. Arrange the sample values in descending order of probability. Split the samples into two groups such that difference of probabilities of two sets is minimized. Assign zero to one set and one to another set. Continue till there is no further splitting of two sets is possible.

2.2 Huffman coding

This is the most popular technique for removing coding redundancy which was developed by David Huffman. Huffman code creates variable length codes that have unique prefix attributes which means that they can be correctly decoded despite being variable in length. Satisfactory compression ratios are achieved at the cost of large amount of memory must be allocated both on the compression and the decompression side.

The Huffman codes are based on two observations regarding optimum prefix codes:

1-In an optimum code, symbols that occur more frequently (have higher probability of occurrence) will have shorter code words than symbols that occur less frequently.

2-In an optimum code, the two symbols that occur least frequently will have highest and equal codeword length with a bit difference of one. Calculate probability of each unique sample present in the ECG signal. Arrange the sample values in descending order of probability. Combine the lowest two probabilities and rearrange in descending order until only two values remain whose addition equals one. Assign zero to the higher probability & one to the lower. Back trace and assign binary bits as mentioned above. The final code obtained for the samples forms the Huffman code.

2.3 Minimum variance Huffman code

In minimum variance Huffman code, variability in codeword lengths is reduced. So difference in maximum and minimum codeword lengths is reduced. It is useful in effective utilization of bandwidth and buffer design complexity is also reduced. In this technique, Combined probability should be kept as high in the list as possible. Rest all process is same as Normal Huffman coding. Since the height of the tree is reduced, computational complexity for binary tree travel is reduced. To obtain minimum variance Huffman code, we always put the combined letter as high in the list as possible. Available transmission rates are generally fixed. Since bit generation rate varies, as per the transmission capacity, channel is either having deficit or excess. In case of excess of bit generation rate, buffer is to be designed to smooth out the operation. Greater the variance in the codewords, more difficult buffer design problem becomes.

$$\text{Variance} = \sum_{i=1}^{N} p_i (n_i - L)^2 \text{ ------------- (1)}$$

Where N- number of unique symbols
P_i- probability
n_i-individual codeword length
L- average codeword length

2.4 lempel Ziv Welch(LZW) coding

It is a dynamic dictionary technique in which index number of the dictionary element acts as encoder output. Every new dictionary element starts with last letter of previous dictionary element. As it has to handle the entire dictionary, the compression procedure takes quite a long time. The compression bandwidth achieved is not the most suitable. Time complexity of algorithm is O (N)

2.5 Runlength Encoding

Run value represents number of times a symbol occurs in consecutive way. This technique is efficient for data having is high local correlation. Simplicity in implementation and high speed makes it useful. The decompression procedure takes too much time which results in extremely low decompression bandwidths. Time complexity of algorithm is O (N)

3. Experimental results

Implementation is done in MATLAB software. ECG database used for experimentation is MITBIH Arrhythmia. The compression algorithm is tested on Lead II recordings of 10 sec of 47 patients with sampling frequency 360 Hz and 11 bit resolution.

Table 1. Comparative analysis in terms of Compression ratio

Re cord	**Shan non**	**Huff man**	**Min Var.Huffma n**	**L ZW**	**Runle ngth**

100	1.85	1.85	**1.85**	1.63	0.59
101	1.70	1.70	**1.70**	1.50	0.59
103	1.61	1.61	**1.61**	1.43	0.58
105	1.57	1.57	**1.57**	1.38	0.57
106	1.48	1.48	**1.48**	1.25	0.54
107	1.23	1.23	**1.23**	1.08	0.53
108	1.65	1.65	**1.65**	1.41	0.56
109	1.35	1.35	**1.35**	1.15	0.55
111	1.60	1.60	**1.60**	1.36	0.56

Table 2. Comparative analysis in terms of execution time (sec)

Record	**Shannon**	**Huffman**	**Min Var.Huffman**	**LZW**	**Runlength**
100	1.85	1.85	**1.85**	1.63	0.59
101	1.70	1.70	**1.70**	1.50	0.59
103	1.61	1.61	**1.61**	1.43	0.58
105	1.57	1.57	**1.57**	1.38	0.57
106	1.48	1.48	**1.48**	1.25	0.54
107	1.23	1.23	**1.23**	1.08	0.53
108	1.65	1.65	**1.65**	1.41	0.56
109	1.35	1.35	**1.35**	1.15	0.55
111	1.59	1.60	**1.60**	1.36	0.56

Variance gives information of variability in codeword length. Huffman and Minimum variance Huffman may give same efficiency but since variance value in minimum variance is less, they are preferred from data transmission point of view for effective utilization of bandwidth. In Many applications, available transmission rate is always fixed. 10 sec recording of ECG contains 3600 samples per second. We might ask for transmission capacity of 10000 bits per second. This means that during each second the channel expects to receive 8000 bits/sec no more and no less. As the bit generation rate will vary around 8000 bits per second, the output of the source coder is generally fed to buffer. Due to low variance, either in case of underutilization or overloading of channel, complexity is less. Implementation is done on computer with intel I3 processor with 3.6GHz speed and 4GB of RAM.

Table 3. Comparative analysis in terms of variance

Record	**Shannon**	**Huffman**	**Min Variance Huffman**
100	2.46	2.47	**2.45**
101	1.96	1.93	**1.92**
103	2.32	2.38	**2.32**
105	2.28	2.28	**2.15**
106	2.68	2.42	**2.41**
107	1.95	2.00	**1.95**
108	1.99	1.80	**1.80**
109	1.44	1.44	**1.44**
111	2.29	2.07	**2.07**

Table 1, 2, 3 shows comparative results of following methods in terms of compression ratio, execution time and variance value.

4. Conclusion:

Minimum Variance Huffman codes are best suited for ECG data compression. Compression achieved with Minimum variance Huffman code is 50% with computational complexity of $NLog_2N$. Bandwidth is effectively utilized and buffer design complexity is also reduced. Run length coding technique is not suited for ECG signal compression since it is found that run length coding is much suitable when data is geometrically distributed. Further scope for this work done is in implantable medical devices like pacemaker where storage space, power are stringent requirements. In recent era, ECG recordings may be used for person identification where compression plays important role. Compression methods are also being used in digital imaging and communications (DICOM), Health level 7 (HL7) standards to store medical images and health records and transfer data through wireless sensor nodes.

5. References:

[1] Priyanka, Indu Saini. "Analysis ECG Data Compression Techniques-A Survey Approach" *International Journal of Emerging Technology and Advanced Engineering* 3.2 (2013): 544-548.

[2] Gupta, Komal, and Ram Lautan Verma "Minimum Entropy based lossless image compression using predictive coding and integer wavelet transform" *International Journal of Engineering Science and Innovative Technology (IJESIT) Volume* 2 (2013).

[3] Miaou, Shaou-Gang, and Shu-Nien Chao. "Wavelet-based lossy-to-lossless ECG compression in a unified vector quantization framework" *Biomedical Engineering, IEEE Transactions on* 52.3 (2005): 539-543.

[4] Gurkan, Hakan. "Compression of ECG signals using variable-length classified vector sets and wavelet transforms." *EURASIP Journal on Advances in Signal Processing* 2012.1 (2012): 1-17.

[5] Fira, Catalina Monica, and Liviu Goras. "An ECG signals compression method and its validation using NNs." *Biomedical Engineering, IEEE Transactions on* 55.4 (2008): 1319-1326.

[6] Jambek, Asral Bahari, and Nor Alina Khairi. "Performance Comparison of Huffman and LZW Compression for wireless sensor node application", American Journal of Applied Sciences 11 (1): pp119-126, 2014

[7] Deepu, Chacko John, Chun-Huat Heng, and Yong Lian. "A Hybrid Data Compression Scheme for Power Reduction in Wireless Sensors for IoT." IEEE Transactions on Biomedical Circuits and Systems 11.2 (2017): 245-254.

[8] Abhinandini, U., et al. "Comparative Study On ECG Data Compression Methods." International Journal of Innovative Research and Development, ISSN 2278–0211 2.11 (2013).

Address for correspondence.

Rupali V Tornekar
Lalit E 505, Nanded City, Sinhgad road, pune -41
E-mail (rupali.tornekar@gmail.com).