Novel Compression Techniques for Time Series Signals

Oinam Suchitra Devi, Hemanth Kumar P., S Basavaraj Patil

Abstract*—A time series signal can be defined as a sequence of data items which is measured through repeated measurements over uniform time intervals. Time series analysis comprises techniques for analyzing time series data in order to obtain meaningful statistics and other characteristics of the data transmission time. Compression is the techniques of reduction in size of data in order to save space or transmission time. Wavelet compression technique is a form of data compression well defined for image compression. The design of time series signal compression techniques involves trade-offs among various factors which includes the degree of compressing the data, the amount of distortion introduced and the computational resources required to compress and decompress the time series data. This paper analyzes different wavelet compression techniques like Wavelet Decomposition, Wavelet Packet, Decimated Discrete Wavelet, Fixed encoding, Huffman encoding and Novel Encoding Compression technique. Analyzing this paper discuss about novel approach for compressing time series signal. There exist several measures to know the quality of the reconstructed time series signal after compression of signal data. The most popularly used measured parameters are Percentage Root mean square Difference (PRD), Peak Signal to Noise Ratio (PSNR) and Maximal Absolute Difference (MAD) etc. From the results it is observed that Novel Compression Encoding technique gives better performance in compression of time series signal as it achieve high PSNR with better quality of compression, smaller PRD and MAD with less distortion compare to other compression techniques.

***Index Terms*—Decimated Discrete Wavelet, Fixed encoding, Huffman encoding, Wavelet Decomposition, Wavelet Packet**

I. INTRODUCTION

Time series signals data like ECG, EEG or stock market need to be stored each and every second as the data generated is enormous and the importance of time series compression is well justified from the necessity of reducing the storage space, transmission time, bandwidth and the quantity of information. Biological signals like ECG, EEG has an important role in diagnosis of human health. As most of the hospitals around the world are implanting the use of the electronic patient record (EPR), reducing storage requirements for clinical examinations (like ECG,EEG) is essential to include the results of these examinations within the EPR without the saturation of the storage system. The key concept is to preserve the diagnostic quality of the raw signal. The main goal of any compression technique is to achieve maximum data reduction while preserving the raw or original morphology reconstruction.

Manuscript received June, 2013.
Oinam Suchitra Devi,
Hemanth Kumar P.,
S Basavaraj Patil

In many applications, simply reducing the size of the data is not sufficient, some additional scalable and embedded properties are also needed. Compression reduces the amount of data needed to represent a particular amount of information of the data. If certain data contain repeated information or information that is not useful then that information is usually called as redundant data. The need for time series compression is that data contain large amounts of information that requires much storage space, long transmission times and large transmission bandwidths. Therefore it is advantageous to compress the data by storing only the essential information required to reconstruct the signal data. In order to compress, redundancies of the data must be exploited.

Wavelet techniques can be used to divide the information of a data into approximation and detail sub-signals. If these details of the signal data are very small then they can be set to zero without significantly changing the signal. The value below which the details of the data are considered as small enough to be set to zero is called as the threshold. If the number of zeros are greater then the compression achieve by the data is greater. The fundamental goal of data compression is to reduce the amount of data required to represent and this is done by transforming and removing data redundancies.

Two major categories of compression techniques can be classified. Lossless compression techniques in which the raw signal is reconstructed perfectly and theoretical limit on maximal compression ratios. The techniques of the lossless compression can obtain an exact reconstruction of the raw signal, but they do not achieve low data rates. Lossless compression performs compression without any loss of information from the raw or original one so it can be useful mainly for medical records like ECG and EEG. Most common and widely used lossless compression techniques are Huffman encoding, LZW and arithmetic coding methods whereas the lossy compression method provide larger compression ratios but when the previous original information is reconstructed, some information is lost obviously. Lossy methods do not obtain an exact reconstruction of the signal but higher compression ratios can be obtained. Lossy compression techniques are widely used in video conferencing and broadcast television, etc where certain information lost is tolerable for the high rates of compression.

The commonly used data compression methods are lossy in nature. These fall into two methods (i) Direct methods, in which actual data samples are analyzed (time domain). Direct compression methods are Amplitude-Zone-Time Epoch Coding (AZTEC) method, the coordinate reduction time coding system(CORTES) (ii) Transformational methods, in which first apply a transform to the signal data and perform spectral and energy distribution and analysis of signals.

Some of the transformations used in the method are Fourier transform, Walsh Transform, Karhunen-Loeve Transform (KLT), Discrete cosine transform(DCT), and Wavelet Transform (WT) etc.

II. DATA DESCRIPTION

Table I shows the description of data set. Data set is a time series signal which is equivalent to an Electrocardiogram (ECG) signal. The data signal must be filtered or smoothed with an n-point smoother, Savitzky-Golay FIR filter. In below table y, y_1 and y_2 represents the Savitzky-Golay Filtering. Savitzky-Golay Filtering sgolayfilt (k,f,x) smoothes the signal x using a Savitzky-Golay smoothing filter. The polynomial of order k should be less than the frame size f and f should be odd number. The length of the input x must be greater than or equal to the frame size f. If x is a matrix then filtering is done on the columns of x**.** If the polynomials order k equals f-1 then no smoothing will occur. The data set is the time series signal that is the collection of observations of well-defined data obtained through repeated measurements over a uniform time interval.

Table I. Data descriptions

Data	Value	Min	Max
x	500x1 double	-0.8390	1
y	500x1 double	-0.7912	0.9583
y_1	500x1 double	-0.5712	0.7734
y_2	500x1 double	-0.7529	0.9249

III. EXISTING COMPRESSION TECHNIQUE

A. Wavelet Decomposition

Wavelet Decomposition is the projection of the signal data on the set of wavelet basis vectors. Each wavelet coefficient can be computed as the dot product of the signal data with its corresponding basis vector. The signal data can be fully recovered from the wavelet decomposition so it is lossless compression. A wavelet can be defined as a wave like oscillation with amplitude that begins at zero and increases its value, and then decreases back to zero. Wavelets are signal generally have irregular shapes and have an ideal property for compact signal representation that is orthogonality. This property ensures that the signal data is not over represented. A signal data can be decomposed into many shifted and scaled representations of the original raw wavelet. Wavelet transform can be used to decompose a signal data into its component wavelets. Once it is decompose the coefficients of the wavelets can be decimated to remove some of the details. Wavelets have the great advantage of being able to separate the fine details in a signal time series. Very small wavelets can be used to isolate very fine details in a signal data, while very large wavelets can identify coarse details In order to represent complex signals efficiently, a basis function should be localized in time as well as frequency domains. The wavelet function is localized in both time and frequency domain, and it is a function of variable parameters. The wavelet decomposes the signal data, and it generates four different horizontal and vertical frequencies outputs. These outputs are approximation, horizontal detail, vertical detail and diagonal detail. The approximation contains low frequency horizontal and vertical components of the data. The decomposition procedure of the wavelet is repeated on the approximation sub-band to generate the next level of the wavelet decomposition, and so on.

B. Wavelet Packet

Wavelet Packet Compression is a compression of wavelet transform where the discrete-time or sample signal is passed through more filters than the discrete wavelet transform (DWT). In the Discrete Wavelet Transform, each level is calculated by passing the previous wavelet approximation coefficients through discrete-time low and high pass quadrature mirror filters. But in the Wavelet Packet decomposition, both the detail and approximation coefficients are decomposed to create the full binary tree. For n levels of decomposition the wavelet packet decomposition produces 2^n different sets of coefficients as opposed to (3n+1) sets for the Discrete Wavelet Transforms. Due to the downsampling process the overall number of coefficients is still the same and there is no redundancy. From the point of view of compression, the standard wavelet transform may not produce the best result of compression since it is limited to wavelet bases that increase by a power of two towards the low frequencies.Wavelet packet transform(WPT) is applied to low pass results as well as high pass results . Wavelet packets are more able to represent the high frequency information. Wavelet packets represent a generalization of multiresolution wavelet decomposition. In the wavelet packets decomposition (WPD), the recursive procedure is applied to the coarse scale approximation along with horizontal detail, vertical detail and diagonal detail, which leads to a complete binary tree.

C. Decimated Discrete Wavelet

In Decimated Discrete Wavelet techniques, it uses only the fixed values for wavelet scales based on powers of two. Wavelet positions are fixed, none overlapping and also form a set of wavelet basis vectors of length N. Discrete Wavelet Transform (DWT) is based on sub-band coding and which is found to yield a fast computation of Wavelet Transform. It is easy to implement and it reduces the computation time and resources required. In Discrete Wavelet Transform (DWT), a time-scale representation of the digital signal is obtained using digital filtering techniques. The signal data to be analyzed is passed through filters with different cut off frequencies at different scales. Filters are most widely used signal processing functions. Wavelets can be realized by iteration of filters with rescaling. The resolution of the signal is a measured of the amount of detail information in the signal and is determined by the filtering operations, and the scale is determined by up sampling and down sampling operations. The filtering and decimation process of the wavelet transform is continued until the desired level is reached. The maximum number of levels is depends on the length of the signal. The Discrete Wavelet Transform of the original data is then obtained by concatenating all the coefficients, starting from the last level of decomposition. In the reconstruction of the raw data from the wavelet coefficients basically, the reconstruction is the reverse process of decomposition. Decimated Discrete Wavelet Transform compression is lossy in nature.

In lossy compression, the raw data cannot be exactly reconstructed from the compressed data. The reason is that, much of the details in signal data can be discarded without greatly changing the appearance of the signal.

D. Fixed Encoding

In Coefficients Thresholding Methods there are two parameters, the first one is related to the threshold value and the second one is the number of classes for quantization. Thresholding is the processing tool which is commonly used in wavelet multi-resolution analysis. The wavelet Thresholding method was mainly developed to remove noise and outliers. The basic ideas presented in true compression techniques are used which cascade in a single step, Coefficient Thresholding i.e. it may be global or by level and is encoding by quantization. Global Thresholding of coefficients is used in the compression method. A fixed length encoding is an encoding process such as ASCII and is convenient because the boundaries between characters are easily determined and the pattern used for each character is completely fixed value for example 'a' is always exactly 97. The standard ASCII uses 8 bits amount of space in character encoding to store each character. Common characters don't get any special treatment and they require the same 8 bits that are used for much rarer characters also such as 'ü'. For example a file of 1000 characters encoded using the ASCII scheme will take 1000 bytes that is 8000 bits. Fixed length encoding uses the same number of bits for each and every symbol and k-bit code supports 2^k different symbols.

E. Huffman Encoding

The Huffman encoding algorithm is an optimal signal data compression algorithm when only the frequency of individual letters is used to compress the signal data. The idea behind the Huffman encoding algorithm is that if we have some letters that are more frequent than others, it makes sense to use fewer bits to encode those letters than to encode the less frequent letters. Huffman coding algorithm is an entropy encoding and is used for lossless data compression. The term encoding refers to the use of a variable length code table for encoding a source symbol such as a character in a file where the variable length code table has been derived in a particular way. It is based on the estimated probability of occurrence for each possible value of the source symbol. The Huffman encoding algorithm takes advantage of the disparity between frequencies. Huffman encoding algorithm is a variable length encoding such that some characters may only require 2 or 3 bits and other characters may require 7, 10, or 12 bits. The savings from not having to use all the full 8 bits for the most common characters that makes up for having to use more than 8 bits for the rare characters. The overall effect of the encoding is that the file almost always requires less data space. The algorithm works by creating a binary tree. These can be stored in a regular array and the size of which depends on the number of symbols n. A node can be either an internal node or a leaf node. Initially, all nodes are leaf nodes, which contain the symbol itself, the weight of the symbol and optionally, a link to a parent node which makes it easy to read the code starting from a leaf node. Internal nodes contain the symbol weight, links to two child nodes and the optional link to a parent node.

IV. NOVEL APPROACH

A. *Novel compression encoding technique*

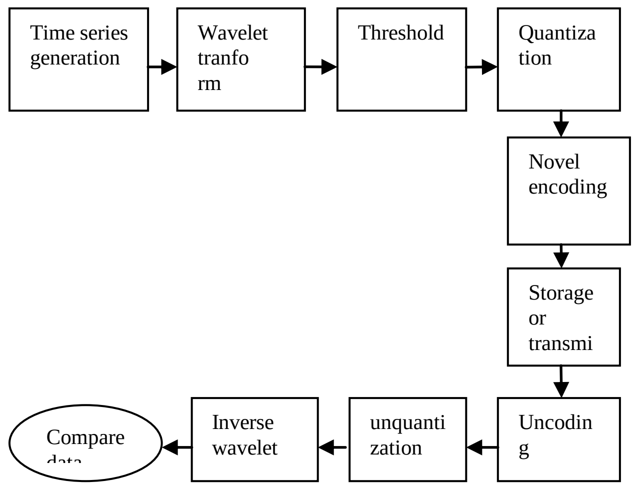

Fig 1.Block diagram of Novel compression encoding

In signal data Compression, we addressed the aspects specifically related to compression using wavelets. However, in addition to the algorithms related to wavelets like DWT and IDWT, it is necessary to use other ingredients concerning the quantization process and the coding method in order to deal with true compression. The effect of quantization on the image is corresponded to a matrix of integers ranging between 0 and 255. Through quantization we can decrease the number of colors. This quantization leads to a compression of the data. Indeed, with a fixed length binary code, 8 bits per pixel are needed to code 256 colors and 4 bits per pixel to code 16 colors. The data obtained after quantization is of good quality. However, within the framework of true compression, quantization is not used on the original data, but on its wavelet decomposition. The basic ideas presented here are used by three methods which cascade in a single step, coefficient Thresholding (global or by level), and encoding by quantization. Fixed or Huffman or novel coding can be used for the quantization depending on the method.

Novel encoding algorithm is an optimal data compression algorithm when only the frequency of individual letters is used to compress the data. The idea behind the novel encoding algorithm is that some letters that are more frequent than the others used fewers bits to encode the data. Novel encoding algorithm is a variable length encoding such that some characters may require 2 or 3 bits and some other characters may required 7,10 or 12 bits. The savings from not having to use all the full 8bits for the most common characters that makes up for having to use more than 8 bits for the rare characters. The overall effect of the encoding is that the file almost always requires less data space. Novel encoding algorithm uses different number of bits to encode different characters. The encoded symbols and its corresponding bit strings are represented as a novel tree and the novel tree is used for compressing as well as decompressing. The algorithm works by creating a binary tree. These can be stored in a regular array and the size of which depends on the number of the symbols.

V. METHODOLOGY

Compression of time series signal is implemented in these steps:

Step 1) Time series signal like EEG,ECG or stock market etc is generated and stored in the Database.

Step 2) Apply Compression

i) Wavelet Decomposition
ii) Wavelet Packet
iii) Decimated Discrete Wavelet compression
iv) Fixed encoding
v) Huffman encoding
vi) Novel encoding

Steps for data compression

A) Decompose

Choose a wavelet and a level L. Compute the wavelet decomposition method of the signal at level L.

B) Threshold detail co-efficient

A threshold level is selected for each level from 1 to L.

C) Quantization

Fixed or Huffman or novel coding can be used for quantization process depending on the method. For Wavelet Decomposition, Wavelet Packet and Decimated Discrete Wavelet, no quantization step in compression.

D) Reconstruct

Compute wavelet reconstruction using the original approximation coefficient of level L and modified the detail coefficient of level from 1 to L.

Step 4) Compare the result in various comparing measured parameters that show how much reconstructed signal is similar to the raw one.

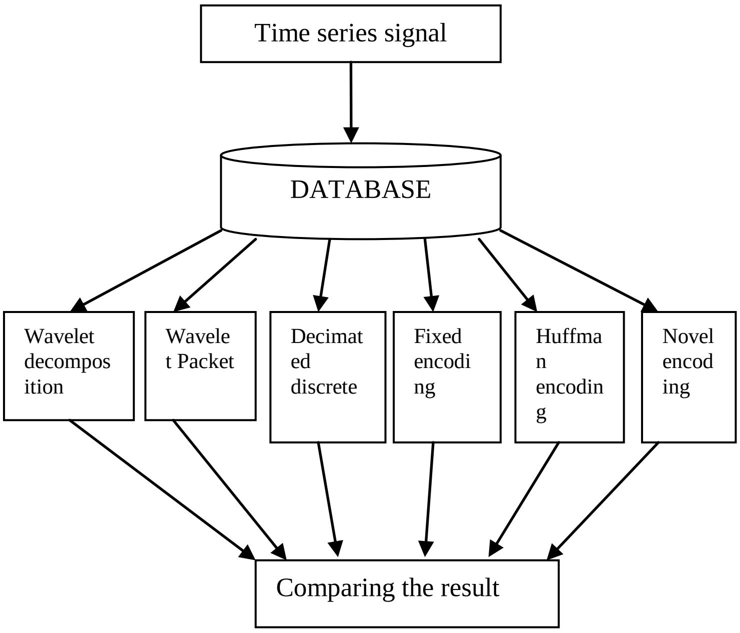

Fig 2. Block diagram of the Time series compression and its comparison.

VI. UANTITATIVE AND PERCEPTUAL QUALITY MEASURED PARAMETER

There are several measures commonly used to evaluate the perceptual quality of the data. MSE is the Mean Square Error and it can be defined as the mean squared error between the compressed and the original data. It measures the average of the squares of the errors and the error is the amount by which the value implied by the estimator differs from the quantity to be estimated. The difference is occurred due to randomness or as the estimator doesn't account for information that could produce a more accurate estimate.

MSE is given by:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m=1}\sum_{j=0}^{n=1}[X(i,j) - Xc(i,j)]^2 \qquad (1)$$

Where X(i,j) and Xc(i,j) are the original data and its corresponding compressed data respectively. The lower the value of MSE, the lower is the error of the compression.

PSNR is the Peak Signal to Noise Ratio and it is a measure of the peak error and is usually expressed in decibels. PSNR is the signal to noise ratio used to measure the quality of reconstruction of lossy compression. The signal in this case is the original time series signal and the noise is the error introduced due to signal data compression. The higher PSNR generally indicates that the reconstruction is of good quality. PSNR is defined by:

$$PSNR = 10 \log_{10}\left(\frac{255^{\wedge}2}{MSE}\right) \qquad (2)$$

The higher the PSNR, the better is the quality of the compressed or reconstructed signal data. Typical values for lossy compression of a signal data are in between 30 and 50 dB and when the PSNR is greater than 40 dB, then the two data are indistinguishable.

Also there exists another measure to know the quality of the reconstructed time series signal after compression. The most popularly used measured one is Percentage Root mean square Difference (PRD**).** PRD represents a numerical measure of the root mean square (rms) error. This parameter is a quality measurement that can mask the real performance of an algorithm as the PRD depends a lot on the mean value of the raw signal. It is given by

$$PRD = 100 \times \sqrt{\frac{\sum_{i=1}^{N}(x(i) - \widetilde{x(i)})^{\wedge}2}{\sum_{i=1}^{N} x(i)^2}} \qquad (3)$$

Where x(i) and $\widetilde{x(i)}$ are the ith sample of the original and reconstructed time series signal of length N respectively.

MAD represents Maximal absolute difference and is the maximum absolute difference that measures between the original data and the compressed or decompressed data. As PRD does not show exactly how much the signal is distorted in different time position, but shows only the cumulative distortion. Maximal Absolute Difference is used to measure the difference between reconstructed and raw signal.

VII. RESULT AND DISCUSSION

The result of the time series compression by six different methods is given below in Table II. In this paper we discussed several measured parameters that show how much reconstructed signal is similar to the raw one and the most used are Percentage Root mean square Difference (PRD), Peak Signal to Noise Ratio (PSNR), Mean square Error (MSE) and Maximal absolute difference (MAD) etc.

The lower the value of MSE, the lower is the error of the compression. From the table II, we observed that novel encoding compression has lower MSE value compared to the other methods so, it has lower error in the compression.

Another measures which is commonly used to evaluate the perceptual quality is Peak Signal to Noise Ratio (PSNR). The higher PSNR generally indicates that the reconstruction is of higher quality. From the Table II, we observed that novel encoding compression has higher PSNR value so it has higher quality of the compressed or reconstructed signal compared to the other methods.

The lower the PRD value the better will be the quality of the reconstructed signal. And we observed that the novel encoding compression has less PRD value compared to the other method so it has better in quality of compression.

Higher the value of the Maximal absolute difference (MAD), the more is the distortion of the original and compressed signal so for the better in compression MAD value should be the lesser in value. In the Table II, we observed that MAD value of the novel encoding compression had lesser in value compared to the other so it has better in compression.

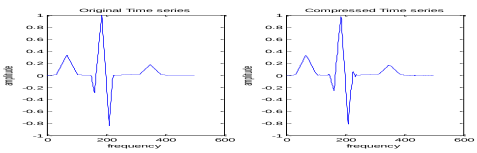

Fig 3. Time Series Compression by Wavelet Decomposition

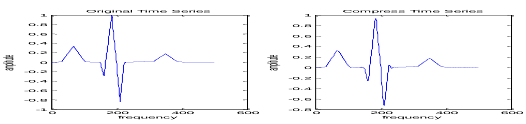

Fig 4. Time Series Compression by Wavelet Packet

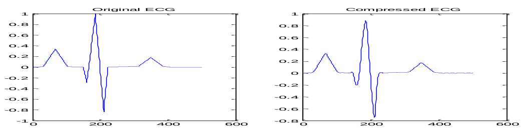

Fig 5. Time Series Compression by Decimated discrete

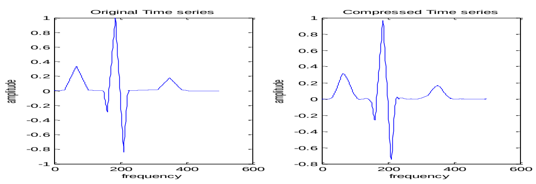

Fig 6. Time Series Compression by Fixed encoding

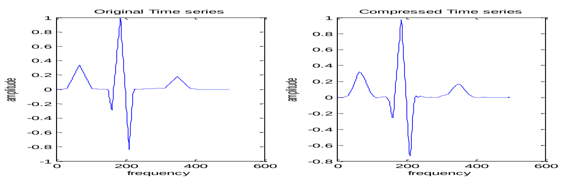

Fig 7. Time Series Compression Huffman encoding

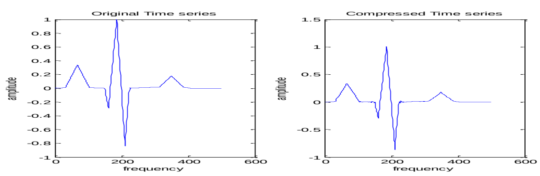

Fig 6 : Time Series Compression by Novel encoding

Table II. MSE, PSNR, PRD and MAD of compressed data

	MSE	PSNR	PRD	MAD
Wavelet decomposition	3.7814e-04	82.354	0.019	0.134
Wavelet packet	1.8575e-04	85.441	0.013	0.123
Decimated discrete	1.5114e-04	86.337	0.012	0.108
fixed encoding	1.1025e-04	87.706	0.010	0.096
Huffman encoding	6.2752e-05	90.154	0.007	0.067
Novel encoding	3.4015e-05	92.814	0.005	0.036

VIII. CONCLUSION

In this paper, wavelet decomposition and wavelet packet compression, decimated discrete wavelet compression, Fixed encoding, Huffman encoding and novel encoding Compression is implemented. From the overall detail of the quantitative and perceptual quality measures of the compression of the time series, we observed that novel encoding compression is better in compression as it has high PSNR with low PRD and minimum MAD compare to the other compression method. So, novel encoding compression has better in compression .Compressing time series signal can be fundamental not only for the obvious storage size reduction but also for improving its performance. This is because of the less data need to be read or written on disk stored in a block device. The work presented in the paper may be helpful for the design of efficient ECG compressor and EEG compression in many hospitals.

REFERENCES

1. Goudarzi, M.M, and Moradi , H. M., 2005, “Electrocardiogramsignal compression Using multiwavelet transform,” Transactions on Engineering, Computing and Technology .
2. Benzid, R., Marir, F., and Bouguechal, N. E., 2006, “Qualitycontrolled compression method using wavelet transform for electrocardiogram signals,” International Journal of Biomedical Science.
3. Benzid, R., Marir, F., Benyoucef, M., and Arar, D., 2003, “Fixed percentage of wavelet coefficients to be zeroed for ECG compression “Electronics Letters Vol 39 No 11.
4. Bell,T.C.,Cleary,J.G.,W.,Written,I.H.,Text Compression,(Prentice Hall,Englewood Cliffs(1990)).
5. Donoho,D.L.,Vetterli,M.,Devore,RA.,Daubechies,I.,”Data Compression and Harmonic Analysis”, IEEE Trans . Inf. Theory.
6. H.C.Liu and G.L. Zick ,”Automatic determination of scene changesnin MPEG compressed video,”in Proc.ISCAS –IEEE Int .Symp.Circuits and System,1995.
7. J. M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients .IEEE Trans. Signal Processing.
8. D. Sinha and A. Tewfik. Low bit rate transparent audio compression using adapted wavelets. IEEE Trans. Signal Processing,1993.
9. R. Baraniuk. Optimal tree approximation using wavelets. In A. J. Aldroubi and M. Unser, editors, Wavelet Applications in Signal Processing, volume VII, pages 196–207. SPIE, 1999.
10. Blelloch, E., 2002. Introduction to Data Compression, Computer Science Department, Carnegie Mellon University.
11. Phillips, Dwayne, "LZW Data Compression," The Computer Applications Journal, Circuit Cellar Ink, vol. 27, June/July 1992.
12. Al-Nashash, H. A. M., 1994, "ECG data compression using adaptive Fourier coefficients estimation", Med. Eng. Phys., Vol. 16.

AUTHORS PROFILE

Oinam Suchitra Devi, presently pursuing M.Tech in Digital Electronics and Communication at AMC Engineering College, Bangalore, India. She was a student of Manipur Institute of Technology (MIT) during her B.E. She got 2nd rank in BE in the stream of Electronics & communication from Manipur University, Canchipur. She has already contributed two international journal and two national conference papers. Her Area of research interest includes Signal processing, Wavelets, Advanced digital image processing, Neural Network and Wireless Communication.

Hemanth Kumar P. currently pursuing Ph.D. in the stream of Computer Science Engineering from VTU Resource Research Centre, Belgaum, Karnataka had received his Master's degree in Digital Communication from National Institute of Technology (MANIT) Bhopal, Madhya Pradesh, India in the year 2010 and Bachelor's Degree from Global Academy of Technology, Bangalore in 2008.Currently working as Assistant Professor in AMC Engineering College, Bangalore in Department of Electronics & Communication Engineering. His area of interest includes Soft Computing Techniques, Data Mining, Pattern Classification, Communication etc. He has have three years of Teaching Experience and guided six M.Tech projects successfully. He has published more than 10 International Journals.

Dr. S.basavaraj p**atil,** is Professor & Head, CSE Deptt., BTL Institute of Technology and he is also Principal Consultant & Founder, Predictive Research. Started career as Faculty Member in Vijayanagar Engineering College, Bellary (1995-1997). Then carried out Ph.D. Research work on Neural Network based Techniques for Pattern Recognition in the area Computer Science & Engineering. Along with that took initiatives to bring several central government projects (AICTE sponsored) and executed successfully. Also consulted many software companies (Aris Global, Manthan Systems, etc.,) in the area of data mining and pattern recognition. Later joined Industry(2003-2004) and worked in diversified domains
in the Pharma, Retail and Finance in the predictive analytics space in different capacities such as Technical Architect, Technology Consultant, Manager and Assistant Vice President (2004-2010). The last one was Assistant Vice President, Equity Quantitative Research with HSBC for around five years. He was responsible for building the revenue generating global quantitative research team at
HSBC, Bangalore, that spread across different geographical locations such as London, Hongkong and Bangalore from scratch. For smaller stints he also worked at Canary wharf, HSBC Headquarters, London. His areas of research interests are Neural Networks and Genetic Algorithms. He founded Predictive Research in 2011. Now the Company has domestic and international client base and staff strength of 12 people. He is presently mentoring two PhD and one MSc (Engg by Research) Students. He has published around 12 International, 8 National (in referred journals and conferences) and 14 Commercial Research Papers (with HSBC Global Research). His total experience is 19 years in Teaching and Research (10 years in Teaching and 9 years in
Industry)