Circuits Syst Signal Process
DOI 10.1007/s00034-014-9764-y

Hybrid Fractal-Wavelet Method for Multi-Channel EEG Signal Compression

Jamal Saeedi · Karim Faez · Mohammad Hassan Moradi

Received: 9 July 2013 / Revised: 12 February 2014 / Accepted: 14 February 2014

Abstract In this paper, a hybrid method is proposed for multi-channel electroencephalograms (EEG) signal compression. This new method takes advantage of two different compression techniques: fractal and wavelet-based coding. First, an effective decorrelation is performed through the principal component analysis of different channels to efficiently compress the multi-channel EEG data. Then, the decorrelated EEG signal is decomposed using wavelet packet transform (WPT). Finally, fractal encoding is applied to the low frequency coefficients of WPT, and a modified wavelet-based coding is used for coding the remaining high frequency coefficients. This new method provides improved compression results as compared to the wavelet and fractal compression methods.

Keywords Electroencephalograms · Compression · Fractal coding · Wavelet-based coding · Multi-objective genetic optimization

1 Introduction

Electroencephalograms (EEG) are electrical signals recorded along the brain surface, which are generated by firing of neurons within the brain [18]. In clinical analysis, i.e., to diagnose the disease (such as epilepsy and sleep disorders), EEG signals are

J. Saeedi (✉) · K. Faez
Electrical Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), 424, Hafez Ave., Tehran, Iran
e-mail: jamal.saeedi@yahoo.com; kfaez@aut.ac.ir

M. H. Moradi
Biomedical Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), 424, Hafez Ave., Tehran, Iran
e-mail: mhmoradi@aut.ac.ir

Published online: 13 March 2014

recorded over a short period, typically 20–40 min. However, EEG signals are sometimes recorded over extended periods (several days, weeks, or even months), e.g., in telemedicine for neurological patients, where EEG is continuously recorded outside the hospital. Long-term EEG recording results in huge EEG datasets. Therefore, compression the EEG signals is needed before storage or transmission. EEG compression is challenging, specifically when the number of EEG channels is large, and the sampling rate is high (to capture spikes and high-frequency fluctuations in the EEG signal). In order to compress EEG signals, several types of redundancies must be taken into account. Antoniol and Tonella [2] presented the survey on EEG lossless compression algorithms using predictive coding, transform coding, vector quantization together with the entropy coding and compared with some well-known lossless compression algorithms. EEG signals are often modeled as auto-regressive (AR) processes. Various AR predictors have been developed: linear AR predictor [2], least squares [2], and adaptive neural network predictors [22].

Before compressing signals, it could be fruitful to transform them in another domain. Various common transformations have been used for EEG compression, including discrete cosine transform [2], sub-band transform, wavelet transform, wavelet packet transform, and integer lifting wavelet transform [2,21]. Some of the well-known 2D compression methods are also used for EEG signal compression: Higgins et.al have used JPEG2000 algorithm [11], and Mitra and Sarbadhikari have used iterative function system (fractal) for EEG compression [16]. The idea behind most hybrid fractal-wavelet coders is to apply a discrete wavelet transform (DWT) to the signal and then use fractal methods in the wavelet domain [25]. It is well-known that the wavelet energy concentration is located primarily in the low-pass filter values, and this makes the approximation sub-band extremely favorable to the application of fractal techniques.

In order to solve the problem concerning the expensive fractal encoding time, without losing the quality of the EEG signal, this work involves a hybrid coder. We present a new, fast, and efficient EEG compression algorithm that applies the speed of the wavelet transform to the quality of the fractal compression. In our proposed method, PCA is applied to decorrelate multi-channel EEG signal. The decorrelated EEG signal looks like a low-pass signal in most channels as compared to the original multi-channel EEG signal. Wavelet-based coding is a good choice to compress a low-pass signal. In addition, there are many self-similarities in EEG signal, which make it appropriate for fractal compression. Therefore, a hybrid coder using fractal and wavelet can be a good combination to get better results.

The remainder of this paper is organized as follows. The proposed method will be described in Sect. 2. Then, EEG compression using fractal coding is reviewed in Sect. 3. Section 4 gives the description of the modified wavelet-based coding. Section 5 illustrates experimental results and compares to other methods. Finally, conclusions are given in Sect. 6.

2 Hybrid Fractal-Wavelet Compression

Data compression is the process of detecting and eliminating redundancies in a given dataset. EEG signals are simply measured from different electrode positions on human scalp. Therefore, the neighboring channels of EEG signals usually have a high degree of similarity in their structures. In order to efficiently compress the multi-channel data, this inter-channel redundancy must be exploited. In this study, the effective decorrelation is performed through the PCA of different channels [8]. PCA has been widely used as feature extraction in pattern recognition. The main concept of PCA is to project the original feature vector onto principal component axes. These axes are orthogonal and correspond to the directions of greatest variance in the original feature space. Hence, projecting input vectors onto this principal subspace allows reducing the redundancy in the original feature space as well as the dimension of input vectors. In order to achieve to these goals, PCA computes new variables called principal components, which are obtained as linear combinations of the original variables. The first principal component is required to have the largest possible variance. The second component is computed under the constraint of being orthogonal to the first component and to have the largest possible inertia. The other components are computed likewise. The values of these new variables for the observations are called factor scores; these factors scores can be interpreted geometrically as the projections of the observations onto the principal components.

In PCA, the components are obtained from the singular value decomposition of the data $\mathbf{X}$ [12]:

$$\mathbf{X} = \mathbf{P\Delta Q^T}, \tag{1}$$

where $\mathbf{P}$ is the $\mathbf{I} \times \mathbf{L}$ matrix of left singular vectors, $\mathbf{Q}$ is the $\mathbf{J} \times \mathbf{L}$ matrix of right singular vectors, and $\mathbf{\Delta}$ is the diagonal matrix of singular values. Note that $\mathbf{\Delta}^2$ is equal to $\mathbf{\Lambda}$, which is the diagonal matrix of the (non-zero) eigenvalues of $\mathbf{X^TX}$ and $\mathbf{XX^T}$.

The I $\times$ L matrix of factor scores, denoted F, is obtained as

$$\mathbf{F} = \mathbf{P\Delta} \tag{2}$$

The matrix $\mathbf{Q}$ gives the coefficients of the linear combinations used to compute the factors scores. This matrix can also be interpreted as a projection matrix, because multiplying $\mathbf{X}$ by $\mathbf{Q}$ gives the values of the projections of the observations on the principal components. This can be shown by combining (1) and (2) as

$$\mathbf{F} = \mathbf{P\Delta} = \mathbf{P\Delta QQ^T} = \mathbf{XQ} \tag{3}$$

In this context, the matrix $\mathbf{Q}$ is interpreted as a matrix of direction cosines (because $\mathbf{Q}$ is orthonormal). The matrix $\mathbf{Q}$ is also called a loading matrix. In this context, the matrix $\mathbf{X}$ can be interpreted as the product of the factors score matrix by the loading matrix as

$$\mathbf{X} = \mathbf{FQ^T} \quad \textbf{with } \mathbf{F^TF} = \mathbf{\Delta}^2 \quad \textbf{and } \mathbf{QQ^T} = \mathbf{I} \tag{4}$$

The proposed compression algorithm uses decorrelated multi-channel EEG signal to achieve more compression ratio (CR) compared to single-channel method. It can be seen from Fig. 1 that the decorrelated EEG signal has only low frequency (LF)

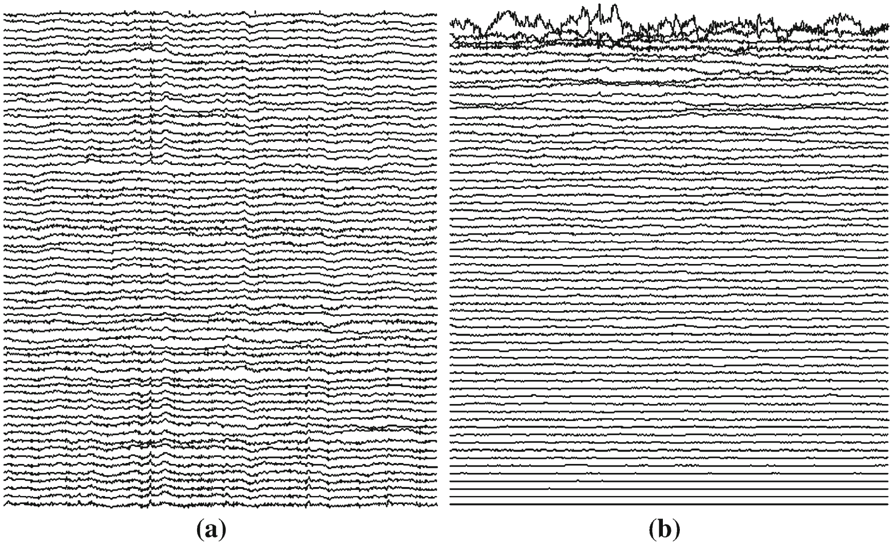

Fig. 1 Applying PCA to multi-channel EEG signal, **a** multi-channel EEG, and **b** decorrelated multi-channel EEG

information in most channels as compared to the original multi-channel EEG signal. In one hand, wavelet-based coding can be used to efficiently compress the high frequency (HF) information of the decorrelated EEG signal. However, the LF coefficients in wavelet-based coding cannot be efficiently coded. In other hand, the LF structures in EEG signal have many self-similar regions, which can be efficiently coded by fractal compression. Therefore, a hybrid coder using fractal and wavelet is used to take advantage of each method.

After applying PCA, the decorrelated EEG signal is decomposed using WPT. Then, fractal encoding is applied to the LF coefficients of WPT, and a modified wavelet-based coding is used for coding the remaining HF coefficients. The schematics of the proposed hybrid coder and decoder can be seen in Fig. 2. For further data compaction, a proper bit encoding format can be used to quantize and transmit the data at low bit rates. A low bit rate representation can be achieved by using an entropy coder like Huffman coding or arithmetic coding.

3 EEG Signal Compression Using Fractal

The theory of compression using iterative function system (IFS) and Collage theorem was first proposed by Barnsley [3]. Since then, this technique has been used successfully in image compression by several researchers [24]. The detailed mathematical descriptions of the IFS theory, collage theorem, and other relevant results are available in [3,5]. Only the important features of coding through IFS are given below. Let I be a given signal that belongs to the set X. Our purpose is to find a set F of affine contractive maps for which the given signal I is an approximate fixed point. F is

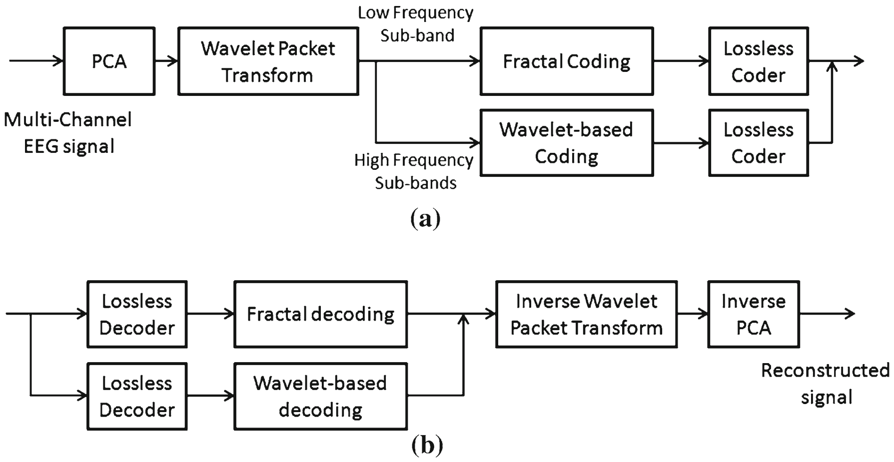

Fig. 2 Block diagram of the proposed hybrid compression. **a** Coder. **b** Decoder

constructed in such a way that the distance between the given signal and the fixed point (attractor) of F is very small. The attractor "A" of the set of maps F is defined as follows:

$$\lim_{N\to\infty} F^N(J) = A, \qquad \forall J \in X \tag{5}$$

and $F(A) = A$, where $F^N(J)$ is defined as

$$F^N(J) = F\left[F^{N-1}(J)\right]$$

with

$$F^1(J) = F(J), \quad \forall J \in X$$

In addition, the set of maps F is defined as follows:

$$d\left[F(J_1), F(J_2)\right] \le s.d(J_1, J_2), \quad \forall J_1, J_2 \in X \text{ and } 0 \le s < 1 \tag{6}$$

Here, d is called the distance measure, and s is called the contractivity factor of F: Let

$$d\left[I, F(I)\right] \le \varepsilon \tag{7}$$

where ε is a small positive quantity. Now, by the collage theorem [3], it can be shown that

$$d\left[I, A\right] \le \frac{\varepsilon}{1-s} \tag{8}$$

where A is the attractor of F.

From (8) it is clear that, after a sufficiently large number of iterations N, the set of affine contractive maps F produces a set that belongs to X and is very close to the given original signal I. Here, (X, F) is called iterative function system, and F is called the set of fractal codes for the given signal I.

3.1 Generation of Fractal Codes

Let I be a given signal having w points. The signal is first partitioned into n non-overlapping segments having b number of points and represented by $R = \{R_1, R_2, \ldots, R_n\}$. Each R_i is called as a range segment. Note that $n = w/b$. Let D be the collection of all possible segments having $2b$ number of points. Let $D = \{D_1, D_2, \ldots, D_m\}$. Each D_j is called as a domain segment with $m = (w-2b)$. Let $F_j = \{f : D_j \rightarrow \Re^2;$ f is an affine contractive map$\}$. Now, for a given range segment R_i, let, $f_{i|j} \in F_j$ be such that

$$d\left(R_i, f_{i|j}\left(D_j\right)\right) \leq d\left(R_i, f\left(D_j\right)\right), \quad \forall f \in F_j, \forall j \tag{9}$$

Here d measures the distance between two sets of points.

$$d\left(R_i, f_{i|k}\left(D_k\right)\right) = \min_j \; d\left(R_i, f_{i|j}\left(D_j\right)\right) \tag{10}$$

Our aim is to find $f_{i|k}\left(D_k\right)$ for each $i \in \{1, 2, \ldots, n\}$. In other words, for every range segment R_i, one needs to find an appropriately matched domain segment D_k, as well as an appropriate transformation $f_{i|*}$. The set of transformations $F = \{f_{1|.}, f_{2|.}, \ldots, f_{n|.}\}$ accordingly obtained is called the fractal code of the given signal I.

To find the best matched domain segment, as well as the best matched transformation, we have to search all possible domain segments as well as all possible transformations with the help of (6). The affine contractive transformation $f_{i|*}$ is constructed using the fact that the points of the range segment are a scaled and shifted version of the points of domain segment. Thus, the affine transformation has two parts. The first part indicates which point of the range segment corresponds to which point of the domain segment. The second part is to find the scaling and shift parameters.

Fractal EEG compression first introduced by Mitra and Sarbadhikari [16], including of the representation of signal blocks through the contractive transformation coefficients. In Mitra's approach, the EEG signal is partitioned into non-overlapped segments named range-blocks $\{R_i\}$ with b points. Each R_i is compared to all down-sampled and affine transformed $2b$ blocks (named domain blocks, D_j) of the same signal. Generally, the decoding process only needs the blocks positions and the affine transform coefficients of the best matching. The affine transformations consist of the contrast scaling s, luminance shift o, and the isometry (eight possible isometric transformations [16] as shown in Table 1). The matching process consists of minimizing of the following:

$$d\left(R_i, D_j\right) = \sum_{p=1}^{b} \left(R_i\left(p\right) - \left(s_i \times D_j\left(p\right) + o_i\right)\right)^2 \tag{11}$$

Table 1 The list of isometric transformations on the domain segments

Identity	Second half-reflection
$l_1\left(D_j(p)\right) = D_j(p)$	$l_2\left(D_j(p)\right) = \begin{cases} D_j(p) & p \le \frac{b}{2} \\ D_j\left(\frac{3b}{2}+1-p\right) & p > \frac{b}{2} \end{cases}$
First half-reflection	First–second and third–fourth quarter reflection
$l_3\left(D_j(p)\right) = \begin{cases} D_j\left(\frac{b}{2}+1-p\right) & p \le \frac{b}{2} \\ D_j(p) & p > \frac{b}{2} \end{cases}$	$l_4\left(D_j(p)\right) = \begin{cases} D_j\left(\frac{b}{2}+1-p\right) & p \le \frac{b}{2} \\ D_j\left(\frac{3b}{2}+1-p\right) & p > \frac{b}{2} \end{cases}$
First half-reflection and second half-swapping	First half-swapping and second half-reflection
$l_5\left(D_j(p)\right) = \begin{cases} D_j(b+1-p) & p \le \frac{b}{2} \\ D_j\left(p-\frac{b}{2}\right) & p > \frac{b}{2} \end{cases}$	$l_6\left(D_j(p)\right) = \begin{cases} D_j\left(p+\frac{b}{2}\right) & p \le \frac{b}{2} \\ D_j(b+1-p) & p > \frac{b}{2} \end{cases}$
Reflection about mid-point	Second–third quarter reflection
$l_7\left(D_j(p)\right) = D_j(b+1-p)$	$l_8\left(D_j(p)\right) = \begin{cases} D_j(b-p) & \frac{b}{4} < p < \frac{3b}{4} \\ D_j(p) & \text{else} \end{cases}$

where R_i is the range block, D_j is the domain block, o_i is the luminance shift, s_i is the contrast scaling, and b is the number of points.

Using the least square analysis, the analytic solution for (11) is given by

$$\begin{aligned} \frac{\partial d}{\partial s_i} = 0,\ \frac{\partial d}{\partial o_i} = 0 \ \Rightarrow s_i \ &= \frac{b\left(\sum_{p=}^{b} R_i(p)D_j(p)\right) - \left(\sum_{p=1}^{b} D_j(p)\right)\left(\sum_{p=1}^{b} R_i(p)\right)}{b\left(\sum_{p=}^{b} D_j^2(p)\right) - \left(\sum_{p=1}^{b} D_j(p)\right)^2} \\ \Rightarrow o_i \ &= \sum_{p=1}^{b} R_j(p)/b - s_i \sum_{p=1}^{b} D_j(p)/b \end{aligned} \tag{12}$$

The whole above process of fractal coding is described in Fig. 3. In the encoding process, for each range segment the matched domain segment and the matched transformation have been obtained. Thus, we have to store (a) location of the domain segment; (b) orientation (isometry) of the domain segment; (c) the scaling factor; and (d) the shift factor.

3.2 Decoding

The decoding scheme simply consists of iterating the fractal code F on any initial signal I_0, until convergence to a stable decoded signal is obtained. The transformation of a signal under the fractal code is done sequentially. The transformation is applied on the kth domain segment (D_k) of the current signal to produce the ith range segment (R_i) of the next signal. In our case, all the EEG signals have been reconstructed from the respective fractal codes, starting from an arbitrary signal having all the amplitude values zero. In addition, the process of iterating the fractal codes has been stopped after 20 repetitions, because a stable decoded EEG signal is obtained.

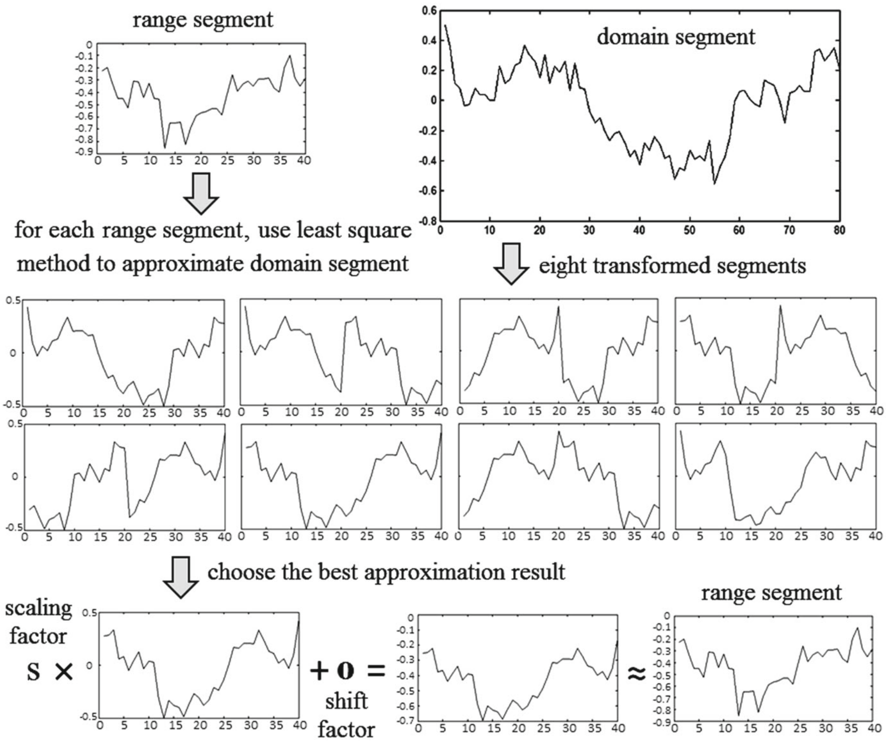

Fig. 3 The schematic diagram of the fractal encoding process

4 Modified Wavelet-Based Compression

The idea behind signal compression using wavelets is primarily linked to the relative sparseness of the wavelet domain representation for the signal. Wavelets concentrate signal information into a few neighboring coefficients [14]. Data compression is then achieved by treating small valued coefficients as insignificant data and thus discarding them. EEG signal is a non-stationary random process due to the time varying nature of the human brain system. Several transitory drifts and abrupt changes describe non-stationary signals. The localization feature of wavelets, together with its time-frequency resolution properties, makes them appropriate for coding EEG signals. The process of compressing a signal using wavelets involves a number of different stages as follows: (1) choosing optimal wavelets for EEG signal, (2) decomposition level in wavelet transforms, (3) Threshold criteria for the truncation of coefficients, (4) efficiently representing zero valued coefficients, and (5) Quantizing and digitally encoding the coefficients.

Wavelets work by decomposing a signal into different resolutions or frequency bands, and this task is carried out by choosing the wavelet function and computing the DWT [14]. Signal compression is based on the concept that selecting a small number of approximation coefficients (at a suitably chosen level), and some of the

Table 2 The process of encoding truncated wavelet coefficients

Wavelet coefficients after thresholding									
0 0 0 0 0	0 0	0 0	0 0	12312	0 0 0	32393	0	2187	0
Coded vector									
11 + 65,408	12,312	3 + 65,408	32,393	1 + 65,408	2,187	…			

First row: truncated wavelet coefficients, and *second row*: coded vector, which is contained of the number of consecutive zeros plus a constant value (65,408) and non-zero coefficients

detail coefficients can accurately represent regular signal components. After calculating the wavelet transform of the EEG signal, compression involves truncating wavelet coefficients below a threshold. Two different approaches are available for calculating thresholds. The first, known as “global thresholding,” involves taking the wavelet expansion of the signal and keeping the largest absolute value coefficients. Thus, only a single parameter needs to be selected. The second approach known as “by level thresholding” consists of applying visually determined level-dependent thresholds to each decomposition level in the wavelet transform [14]. After zeroing wavelet coefficients with negligible values based on either calculating threshold values or simply selecting a truncation percentage, the transform vector needs to be compressed. One way of representing the high-magnitude coefficients is to encode consecutive zero valued coefficients [14], with two values. One value is to indicate a sequence of zeros in the wavelet transforms vector, and the second value is representing the number of consecutive zeros. Due to the sparsity of the wavelet representation of the EEG signal, this encoding method leads to higher CRs than storing the non-zero coefficients along with their respective positions in the wavelet transform vector [7].

In this paper, a modified run-length coding approach is used, which is used only one value for representing the number of consecutive zeros as it can be seen in Table 2. In other words, only the number of consecutive zeros plus a constant value (bigger than maximum values of wavelet coefficients) is coded. In the decoding process, if a value in the coded vector is bigger than a constant value, it demonstrates consecutive zeros in the wavelet sub-bands. If we store the EEG data with 16 bits, we have 2^{16} =65,536 distinct values. For the proposed run-length coding, we need a space to consider the amount of consecutive zeros. In our case, because we have the maximum of 128 samples in each sub-band of WPT, then the maximum of 128 is needed to demonstrate the amount of consecutive zeros. Therefore, in the proposed run-length coding, we can consider 65,408 as the maximum value of wavelet coefficients (every values above this point is equal to 65,408), and the consecutive zeros are added to this value. In the decoding process, if a value in the coded vector is bigger than 65,408, it demonstrates consecutive zeros in the wavelet sub-bands.

4.1 Wavelet Packet Transform

Wavelets are mathematical functions that cut up data or function into different frequency components and then study each component with a resolution matched to its

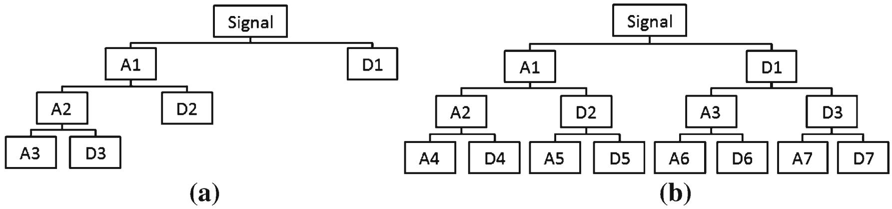

Fig. 4 **a** Wavelet decomposition tree, and **b** wavelet packet decomposition tree

scale [4]. Wavelets, which are oscillatory functions of zero mean and of finite energy, can be used to obtain a time-frequency representation of a process.

Due to decomposition of only the approximation component at each level using the dyadic filter bank, in a regular wavelet analysis, the results of frequency resolution in higher-level DWT decompositions (e.g., A1 and D1 in Fig. 4a) are less desirable. It may cause problems while applying DWT in certain applications, which the important information is, located in higher frequency components. The frequency resolution of the decomposition filter may not be fine enough to extract necessary information from the decomposed component of the signal. The necessary frequency resolution can be achieved by implementing a WPT to decompose a signal further. The wavelet packet method is a generalization of wavelet decomposition that offers a richer range of possibilities for signal analysis. In wavelet analysis, a signal is split into an approximation and a detail. The approximation is then itself split into a second-level approximation and detail, and the process is repeated. For n-level decomposition, there are $n + 1$ possible ways to decompose or encode the signal. In wavelet packet analysis, the details as well as the approximations can be split. This yields more than different ways to encode the signal. For instance, wavelet packet analysis allows the signal S to be represented as $A1 + A6 + D6 + D3$ in Fig. 4b. This is an example of a representation that is not possible with ordinary wavelet analysis. The wavelet decomposition tree is a part of this complete binary tree. As mentioned above, the wavelet packet analysis is similar to the DWT with the only difference that in addition to the decomposition of the wavelet approximation component at each level, the wavelet detail component is also decomposed to obtain its own approximation and detail components as shown in Fig. 4b.

Each component in this wavelet packet tree can be viewed as a filtered component with a bandwidth of a filter decreasing with increasing level of decomposition, and the whole tree can be viewed as a filter bank. At the top of the tree, the time resolution of the WP components is good but at an expense of poor frequency resolution, whereas at the bottom with the use of wavelet packet analysis, the frequency resolution of the decomposed component with high frequency content can be increased. As a result, the wavelet packet analysis provides better control of frequency resolution for the decomposition of the signal [1]. A wavelet packet is represented as a function, ψ, where 'i' is the modulation, 'j' is the dilation, and 'k' is the translation parameter [17].

$$\psi^i_{j,k}(t) = 2^{-j/2}\, \psi^i (2^{-j}\, t\, - k) \tag{13}$$

where $i = 1, 2 \ldots j.n$ and 'n' is the level of decomposition in wavelet packet tree. The wavelet is obtained by the following recursive relationships:

$$\psi^{2i}(t) = \frac{1}{\sqrt{2}} \sum_{k=-\infty}^{\infty} h(k)\psi^{i}\left(\frac{t}{2} - k\right) \tag{14}$$

$$\psi^{2i+1}(t) = \frac{1}{\sqrt{2}} \sum_{k=-\infty}^{\infty} g(k)\psi^{i}\left(\frac{t}{2} - k\right) \tag{15}$$

where $\psi^{i}(t)$ is called as a mother wavelet, and the discrete filters $h(k)$ and $g(k)$ are quadrature mirror filters associated with the scaling and the mother wavelet functions. These two filters, $h(k)$ and $g(k)$, are also called group conjugated orthogonal filters [6].

The wavelet packet coefficients C corresponding to the signal $f(t)$ can be obtained as

$$C^{i}_{j,k} = \int_{-\infty}^{\infty} f(t)\psi^{i}_{j,k}(t)\mathrm{d}t \tag{16}$$

Provided the wavelet coefficients satisfy the orthogonality condition. The wavelet packet component of the signal at a particular node can be obtained as

$$f^{i}_{j}(t) = \sum_{k=-\infty}^{\infty} C^{i}_{j,k}\, \psi^{i}_{j,k}(t) \tag{17}$$

After performing wavelet packet decomposition up to jth level, the original signal can be represented as a summation of all wavelet packet components at jth level as shown in equation:

$$f(t) = \sum_{\mathrm{i}=1}^{2\,i} f^{i}_{j}(t) \tag{18}$$

In wavelet packet, standard structure composed of low and high pass filters is used in perfect reconstruction filter bank [23]. In this study, WPT is used to decompose the EEG signal for compression.

4.2 Thresholds Selection Based on Multi-Objective Genetic Algorithm

The global- and level-dependent thresholds for truncating wavelet coefficients do not produce optimum compression results based on CR and quality of decompressed signal. In this study, a new algorithm based on multi-objective genetic optimization (MOGO) is used to optimally estimate the thresholds for truncating wavelet coefficients in different WPT sub-bands. The performance of compression algorithms usually is computed using percent root-mean square difference (PRD) and CR. If No is the size of the original file in bits, and Nc is the size of the compressed signal, then CR = No/Nc, and PRD is defined as follows

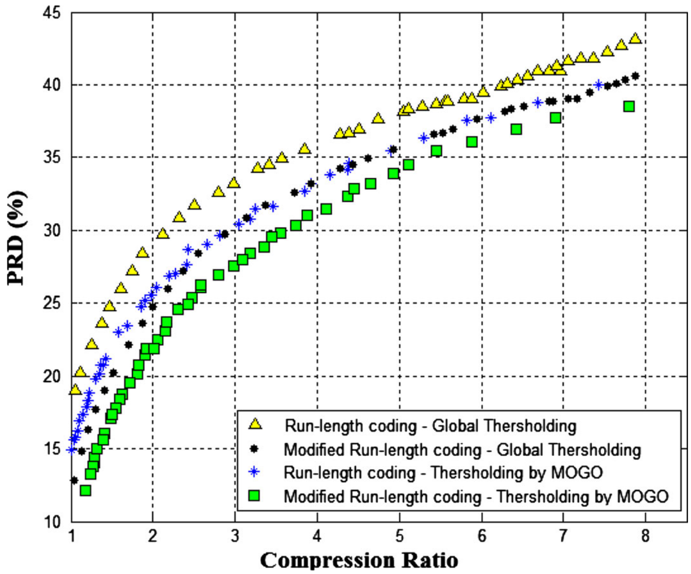

Fig. 5 PRD variations with compression ratio for different wavelet-based coding

$$\mathrm{PRD} = \sqrt{\sum_{i=1}^{N} (x[i] - \hat{x}[i])^2 \Big/ \sum_{i=1}^{N} x^2[i]} \times 100 \tag{19}$$

where x is the original EEG signal, $\hat{x}$ is the reconstructed EEG signal, and N is the number of samples.

In this study, the CR and PRD are chosen for the optimization problem. In other words, the optimization is a minimization problem, and the goal is to estimate the optimal thresholds for truncating wavelet coefficients in each WPT sub-bands of the EEG signals, which simultaneously minimize the two objectives including of the inverse of CR, and PRD expressed in (19). For this purpose, the MOGO is used for optimization problem [15], which is implemented in Matlab optimization toolbox. Here, the performances of different wavelet-based coding are compared and illustrated in Fig. 5. The results are obtained for 1,024 EEG samples. As can be seen from Fig. 5, the proposed wavelet-based compression with modified run-length coding and thresholding using MOGO clearly outperforms other wavelet-based compression schemes.

4.3 Multi-Objective Genetic Algorithm

In single–objective genetic algorithm (GA) [9], a candidate solution for a specific problem is called an individual or a chromosome and consists of a linear list of genes. Each individual stands for a point in the search space and therefore a possible solution to the problem. A population consists of a finite number of individuals. Each individual is decided by an evaluating system to obtain its fitness value. Based on this fitness value and undergoing genetic operators, a new population is iteratively generated with each successive population referred to as a generation.

Three basic genetic operators are sequentially applied to each individual with certain probabilities during each generation, i.e., selection, crossover (recombination), and mutation. First, a number of best individuals are picked based on a user defined fitness function. The remaining individuals are discarded. Next, a number of individuals are selected and paired with each other. Each individual pair generates one offspring by partially exchanging their genes around one or more randomly selected crossing points. At the end, a certain number of individuals are selected, and the mutation operations are applied (i.e., a randomly selected gene of an individual abruptly changes its value). The GA is called a population-based technique, because instead of operating on a single potential solution, it uses a population of potential solutions. The larger the population, the greater the diversity of the members of the population, and the larger the area searched by the population. The basic steps of the classical GA as follows: (a) Generate an initial population; (b) Evaluate fitness of individuals in the population; Repeat (c) Select parents from the population; (d) Recombine parents to produce children; (e) Mutate children; (f) Evaluate fitness of the children; (g) Replace some or all of the population by children; Until a satisfactory solution has been found.

A multi-objective decision problem is defined as follows: Given an n-dimensional decision variable vector $x=\{x_1,\ldots,x_n\}$ in the solution space X, find a vector x that minimizes a given set of K objective functions $z(x^*) = \{z_1(x^*),\ldots, z_K(x^*)\}$. The solution space X is generally restricted by a series of constraints, such as $g_j(x^*) = b_j$ for $j = 1, \ldots, m$.

If all objective functions are for minimization, a feasible solutions x is said to dominate another feasible solution $y(x > y)$, if and only if, $z_i(x) \leq z_i(y)$ for $i = 1, \ldots, K$ and $z_j(x) < z_j(y)$ for least one objective function j. A solution is said to be Pareto optimal if it is not dominated by any other solution in the solution space. A Pareto optimal solution cannot be improved with respect to any objective without worsening at least one other objective. The set of all feasible non-dominated solutions in X is referred to as the Pareto optimal set, and for a given Pareto optimal set, the corresponding objective function values in the objective space are called the Pareto front. For many problems, the number of Pareto optimal solutions is enormous (maybe infinite). The ultimate goal of a multi-objective optimization algorithm is to identify solutions in the Pareto optimal set.

Being a population-based approach, GA are well suited to solve multi-objective optimization problems. The ability of GA to simultaneously search different regions of a solution space makes it possible to find a diverse set of solutions for difficult problems with non-convex, discontinuous, and multi-modal solutions spaces. In this study, we have used a classical approach to solve multi-objective optimization problem. This method assigns a weight w_i to each normalized objective function $z'_i(x)$ so that the problem is converted to a single objective problem with a scalar objective function as follows:

$$\min z = w_1.z_1'(x) + w_2.z_2'(x) + \cdots + \mathrm{w_k}.z_k'(x) \tag{20}$$

where $z\prime_i(x)$ is the normalized objective function, and $\sum w_i = 1$. This approach is called a priori approach, since the user is expected to provide the weights. Solving a problem with the objective function (1) for a given weight vector $w = \{w_1, \ldots, w_k\}$ yields a single solution, and if multiple solutions are desired, the problem should be solved multiple times with different weight combinations.

The main advantage of the weighted sum approach is a straightforward implementation. Since a single objective is used in fitness assignment, a single objective GA can be used with minimum modifications. In addition, this approach is computationally very efficient. For our optimization problem, there are two objective functions, PRD and CR, which should be minimized. For the multi-objective GA discussed here, the weights are $\{0.45, 0.55\}$, the population size is set to 16, and the generation size is set to 20 as a good compromise between accuracy and complexity.

5 Experimental Results

We have analyzed 64-channel EEG from Physiobank EEG motor movement/imagery dataset [10,19]. This dataset consists of over 1,500 1- and 2-min EEG recordings, obtained from 109 volunteers. The EEG was recorded at a sampling rate of 160 Hz, and we considered EEG segments that are 1,024 samples long. The compression performance is measured using two parameters: CR and PRD expressed in (3). Here, four methods are considered for comparison: fractal coding expressed in Sect. 2, wavelet-based coding using global thresholding, the proposed modified wavelet-based coding expressed in Sect. 3, and the proposed hybrid methods.

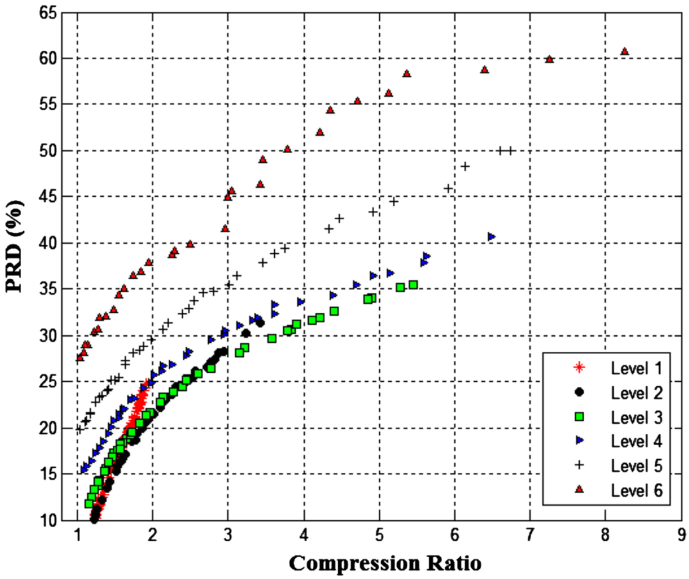

Fig. 6 PRD variations with compression ratio for different decomposition levels in the modified wavelet coding

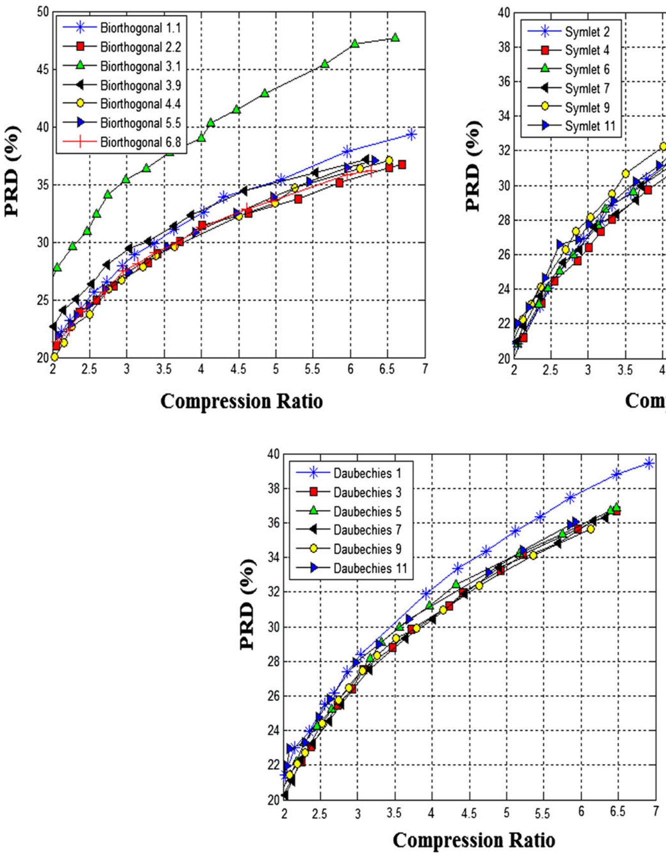

Fig. 7 PRD variations with compression ratio for different mother wavelet functions in the modified wavelet coding

According to the results of this study for the fractal coding, the CR is related to the range segment size. In order to evaluate the fractal coding for EEG signal compression, different sizes are used for the range segment, which result in different CRs and corresponding PRDs. For wavelet-based coding, three decomposition levels are considered for WPT. Hence, seven HF sub-bands of EEG signal are used for coding as described in Sect. 3, and LF sub-band is not coded. For wavelet-based coding using global thresholding, a fixed threshold is used for truncating HF coefficients in all sub-bands of WPT, in which changing the threshold yielding different CRs and corresponding PRDs. For the modified wavelet-based coding, the thresholds for truncating HF coefficients in each sub-band of WPT are obtained using the proposed method based on MOGO as described in Sect. 3.

As we have discussed in Sect. 3, there are two parameters in wavelet-based compression that should be selected: optimal wavelet type for EEG signal, and decomposition level in wavelet transform. Figure 6 shows experimental results of using different decomposition levels in the modified wavelet-based coding. As it can be seen from

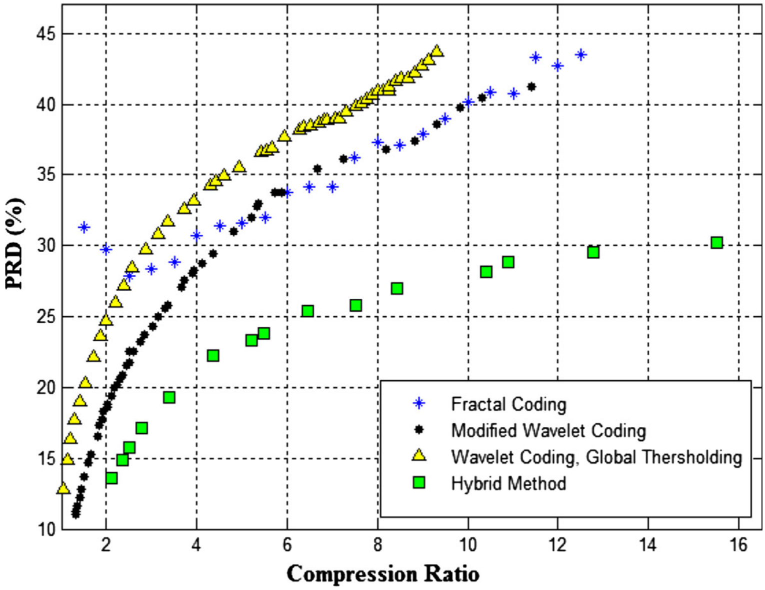

Fig. 8 PRD variations with compression ratio for different compression methods

Fig. 6, more decomposition levels lead to higher CRs, but it introduces distortion into the processed signal (larger PRD). Here, three decomposition levels are considered in wavelet-based compression.

Wavelets with more vanishing moments provide better reconstruction quality, as they introduce less distortion into the processed signal and concentrate more signal energy in a few neighboring coefficients [13]. However, the computational complexity of the DWT increases with the number of vanishing moments, and hence, for real time applications, it is not practical to use wavelets with an arbitrarily high number of vanishing moments. The modified wavelet-based coding is evaluated with using different mother wavelet functions shown in Fig. 7. The results are shown that wavelets with more vanishing moments provide better reconstruction quality. From the results, we can find out that the modified wavelet coding with ‘symlet 4’ mother wavelet function gives slightly better results.

In the proposed hybrid method, the fractal compression is used for coding the LF coefficients, and the modified wavelet-based coding is applied for the HF coefficients. It should be mentioned that performance evaluation between different methods was performed on the 64×1024 EEG samples (64 channels). The results of different algorithms are summarized in Fig. 8. As can be seen from Fig. 8, the hybrid compression scheme clearly outperforms other compression schemes, especially at large CRs. In addition, two examples of reconstructed EEG signals using different compression methods are shown in Figs. 9 and 10.

Since some phenomena of EEG signals, such as spikes and high frequency fluctuations are very important features for the subsequent processing and applications, such as recognition and seizure prediction, another experiment has been performed to

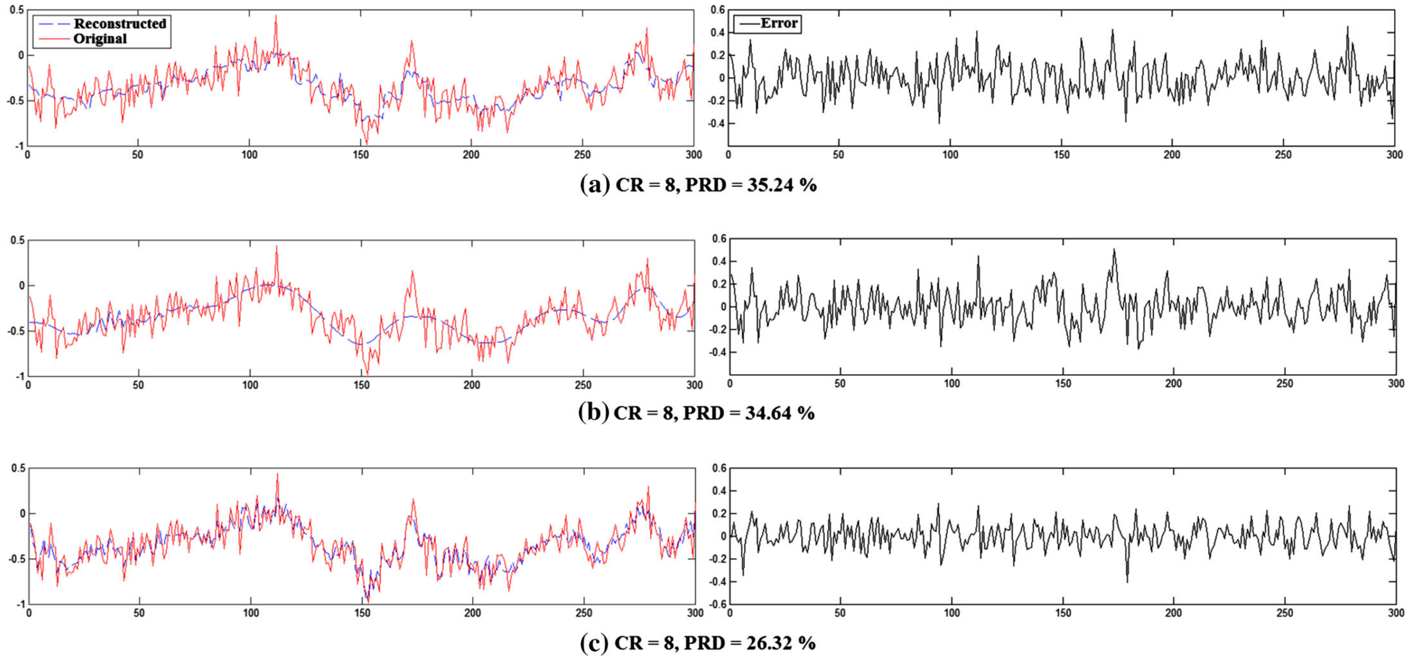

Fig. 9 Reconstructed and error signals for an example of EEG signal using different compression methods (CR = 8). **a** Fractal coding, **b** modified wavelet coding, and **c** the proposed hybrid methods

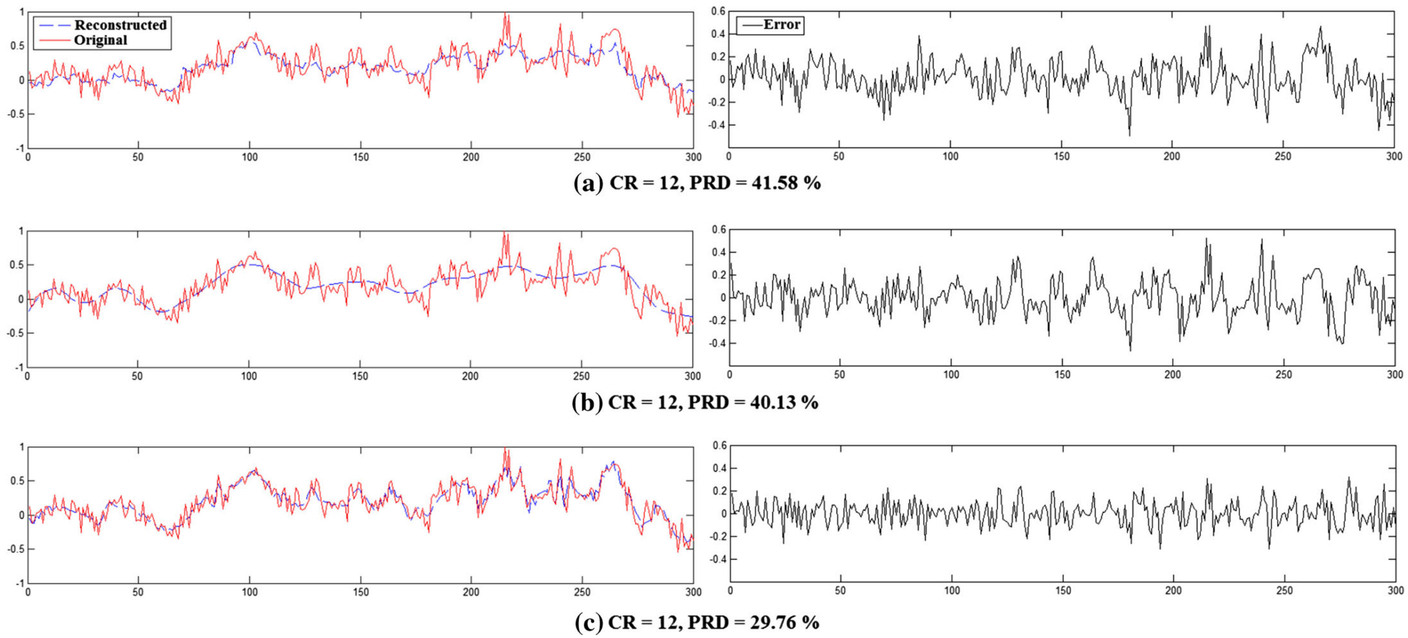

Fig. 10 Reconstructed and error signals for an example of EEG signal using different compression methods (CR = 12). **a** Fractal coding, **b** modified wavelet coding, and **c** the proposed hybrid methods

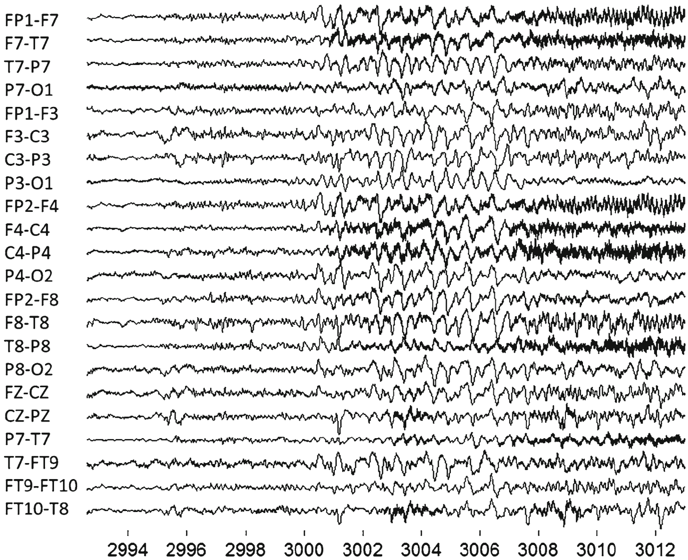

Fig. 11 Example of a seizure with the scalp EEG of a patient in the dataset [20], the seizure begins at 3,000 s

show the performance on the perseveration of these phenomena in the reconstructed EEG signals. For this experiment, the EEG dataset described in [20] is applied (see Fig. 11). This database, collected at the Children's Hospital Boston, consists of EEG recordings from pediatric subjects with intractable seizures. Subjects were monitored for up to several days following withdrawal of anti-seizure medication in order to characterize their seizures and assess their candidacy for surgical intervention. Spectrogram is an important feature, which is used for detecting the seizure in the EEG signal [20]. Here, we have visually compared the spectrograms of reconstructed EEG signals obtained from different algorithms. Compression results are shown in Fig. 12. It should be mentioned that in the original EEG sample, a seizure begins at 5 s. It can be observed from Fig. 12 that the seizure can be visually detected in the spectrogram of the reconstructed EEG signal from the proposed method. However, for other algorithms including fractal coding and wavelet-based compression, it is hard to determine when the seizure is happened.

It is also interesting to evaluate the various compression techniques from a practical point of view that is the computation time. With the proposed hybrid method, the total process (including coding and decoding) lasts approximately 2.67 s for compressing $64 \times 1{,}024$ EEG samples. With this computation time, the proposed hybrid algorithm is real time. This is because the EEG was recorded at a sampling rate of 160 Hz, and therefore, 1,024 EEG samples are equal to the 6.4 s. However, it is possible to reduce the computation time using graphic processing unit (GPU). GPU provides increased processing capability for applications with a high degree of data parallelism. In the past few years, GPUs have become readily available in the commercial market, and off-the-shelf programming tools (e.g., CUDA from the NVIDIA Corporation and Jacket from Accelereyes, LLC) have made them more accessible to the technical community.

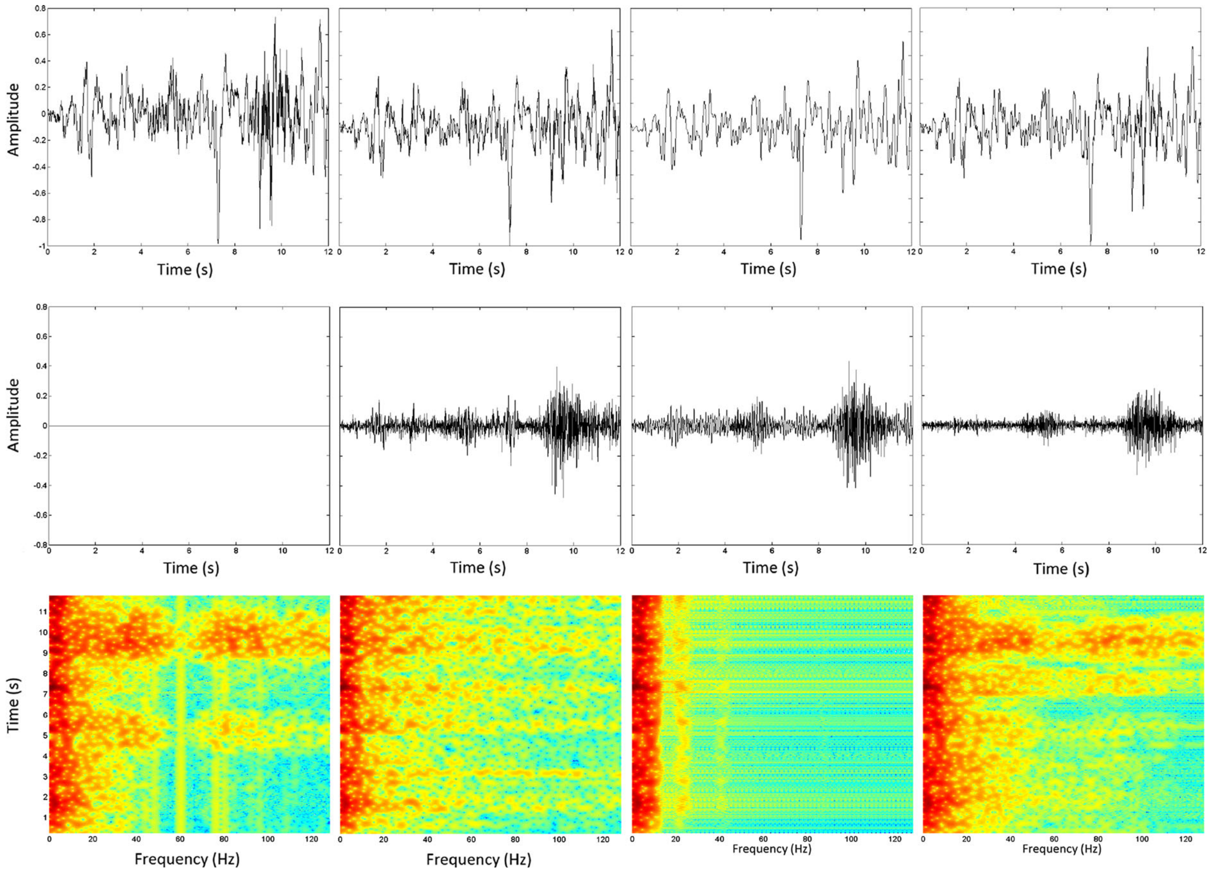

Fig. 12 Compression of EEG signal with a seizure using different compression methods (CR $=$ 12). *Top row, left to right*: Original EEG signal, fractal coding ($PRD =$ 32.14 %), wavelet coding (PRD = 33.24 %), and the proposed hybrid method (PRD = 22.31 %)

Table 3 Relative computation time of various compression techniques for coding and decoding of $64 \times 1{,}024$ EEG samples

Method	Unite of time [a] coding and decoding
Fractal coding	12.78
Wavelet-based coding	0.45
Proposed hybrid coding	2.67

[a] The computation times have been averaged over ten runs

Table 3 summarizes the relative computation time of the various methods considered in this paper. All computations have been performed on a Pentium IV personal computer, using a 2.00 GHz processor, running windows Seven. Note that our method is fully implemented in Matlab software. As it can be seen in Table 3, computation time of our hybrid approach is far better than the fractal method.

6 Conclusions

In this paper, a new hybrid compression method is presented based on fractal and wavelet coding for multi-channel EEG signal. Decorrelation of multi-channel EEG signal using PCA, and using a hybrid method based on fractal and WPT to efficiently compress EEG signal are the main novelties of this paper. The experimental results demonstrated that the proposed method outperforms the fractal and wavelet-based coding.

Acknowledgments The valuable comments of the anonymous reviewers obviously contributed to the improvement of this paper.

References

1. S.D. Abhijeet, *A Wavelet Packet Based Sifting Process and its Application for Structural Health Monitoring*. Master Thesis, Faculty of Worcester Polytechnic Institute (2004)
2. G. Antoniol, P. Tonella, EEG data compression techniques. IEEE Trans. Biomed. Eng. **44**(1), 105–114 (1997)
3. M.F. Barnsley, *Fractals Everywhere* (Academic, New York, 1988)
4. J.J. Benedetto, M.W. Frazier, *WAVELETS: Mathematics and Applications* (CRC Press, Boca Raton, 1994)
5. G.A. Edger, *Measure, Topology, and Fractal Geometry* (Springer, New York, 1990)
6. X. Fan, M.Z. Zuo, Gearbox fault detection using Hilbert and wavelet packet transform. Mech. Syst. Signal Process. **20**, 966–982 (2006)
7. E.B. Fgee, W.J. Phillips, W. Robertson, Comparing audio compression using wavelets with other audio compression schemes. in *IEEE Canadian Conference on Electrical and Computer Engineering* (Edmonton, Canada, 1999), pp. 698–701
8. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd edn. (Academic Press, New York, 1990)
9. E. Goldberg, *Genetic Algorithms in Search, Optimization & Machine Learning* (Addison-Wesley, Reading, MA, 1989)

10. A.L. Goldberger, L. Amaral, L. Glass, J.M. Hausdorff, R.G. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation **101**(23), e215–e220 (2000)
11. G. Higgins, S. Faul, R.P. McEvoy, B. McGinley, EEG Compression Using JPEG2000: How Much Loss Is Too Much? in *32nd Annual International Conference of the IEEE EMBS Buenos Aires*, Argentina, pp. 614–617 (2010)
12. I.T. Jolliffe, *Principal component analysis* (Springer, New York, 2002)
13. R. Kastantin, D. Stefanoiu, G. Feng, N. Martin, M. Mrayati, Optimal Wavelets for High Quality Speech Coding, in *Proceedings of EUSIPCO-94, European Association for Signal Processing*, Edinburgh, Scotland, pp. 399–402 (1994)
14. W. Kinsner, A. Langi, Speech and Image Signal Compression with Wavelets, in *IEEE Wescanex Conference Proceedings*, New York, pp. 368–375 (1993)
15. A. Konak, D.W. Coit, A.E. Smith, Multi-objective optimization using genetic algorithms: a tutorial. Reliab. Eng. Syst. Saf. **9**, 992–1007 (2006)
16. S.K. Mitra, S.N. Sarbadhikari, Iterative function system and genetic algorithm-based EEG compression. Med. Eng. Phys. **19**(3), 605–617 (1997)
17. R.T. Ogden, *Essential Wavelets for Statistical Applications and Data Analysis* (Birkhauser, Boston, 1997)
18. S. Sanei, J. Chambers, *EEG Signal Processing* (Wiley Interscience, New York, 2007)
19. G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, BCI2000: A general-purpose brain-computer interface (BCI) system. IEEE Trans. Biomed. Eng. **51**(2), 1034–1043 (2004)
20. A. Shoeb, *Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment*. PhD Thesis, Massachusetts Institute of Technology (2009)
21. K. Srinivasan, M.R. Reddy, Efficient preprocessing technique for lossless real-time EEG compression. Electron. Lett. **46**(1), 26–27 (2010)
22. N. Sriraam, C. Eswaran, An adaptive error modeling scheme for the lossless compression of EEG signals. IEEE Trans. Inf. Technol. Biomed. **12**(1), 587–594 (2008)
23. G. Strang, T. Nguyen, *Wavelets and Filter Banks* (Wellesley-Cambridge Press, Wellesley, 1996)
24. L. Thomas, F. Deravi, Region-based fractal image compression using heuristic search. IEEE Trans. Image Process. **4**(2), 832–838 (1995)
25. S. Welstead, *Fractal and Wavelet Image Compression Techniques'* (SPIE, Bellingham, 1999)