Human Identification based on electroencephalography Signals using Sample Entropy and Horizontal Visibility Graphs

SHAYMAA ADNAN ABDULRAHMAN*, MOHAMED ROUSHDY**, ABDEL-BADEEH M.SALEM**

*Department of computer Engineering, Imam Ja’afar Al-Sadiq University, Baghdad,Iraq
PhD Student at Ain Shams University, Egypt
**Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

Shaymaa416@gmail.com, mroushy@cis.asu.edu.eg, absalem@cis.asu.edu.eg

Abstract:- Biometric development depends on electroencephalography (EEG) distinguishes people by utilizing individual qualities in human brainwaves. Two Essential features of EEG signals are Liveliness strength against adulteration. However, far reaching study on human authentication utilizing EEG signals is still remain. In this paper we propose a two-phase approach to distinguish EEG signals. The first phase, feature vectors are based on Sample Entropy (SaE) and Horizontal Visibility Graphs (HVG) to extract feature vector of EEG activities. The second phase performs a classification of these feature vectors using K-Nearest Neighbour (KNN) classifiers. We test the accuracy of the proposed approach on Machine Learning Repository (UCI) dataset. Experimental results on this dataset demonstrate significant improvement in the classification accuracy compared to other reported results. Our study applied two models, the first model using 13 channels to feature extraction. It was found that classifier with HVG had a much better performance giving the highest accuracy gave 94.8% compared to classifier with SaE gave 83.7% accuracy. The second model using all channels. The classification accuracy with HVG gave 97.4% and with SaE gave 92.6%.

Key-words: -- Brain Computer Interface, Horizontal Visibility Graphs, Sample Entropy, Graph Entropy, EEG Signals, K-Nearest Neighbor, Machine Learning, Biometrics

1. Introduction

Over the past few years numerous researchers have been analyzing the electroencephalographic (EEG), Electromyogram (EMG), Electrooculogram (EOG), Electrocorticography (ECG) signals. This signals using for different applications such as detection of human brain disorders like Alzheimer, Epilepsy, sleep disorder and brain stroke. Sometimes researchers using EEG equipments to provide services for the disable human like control toy car wheel chair or car driving[1]. Biometrics is the operation of uniquely determined individuals depends on one or more physical behavioral or cognitive characteristics [2]. There are widely utilized techniques for human identification like passwords, PIN, RF cards which are easily forget, stolen or lost. Existing technologies especially used fingerprints, speech, iris and signatures as a base for recognition. These traits however, are known to be vulnerable to forgery as it is possible to forge or steal. In this case Brain electroencephalogram (EEG) signals can be using as a viable biometric because of its robust against falsification. Through that brain signals can applied to remote healthcare service[3]. Biometrics is the operation of uniquely determined individuals depends on one or more physical, behavioral or cognitive characteristics [2]. There are widely utilized techniques for human identification like passwords, PINs, and RF cards, which are easily forgot, stolen or lost. Biometrics which indicate the technique app to identify individuals using individual biological features are more attractive Existing technologies especially used fingerprints, speech, facial
features, iris and signatures as a base for recognition. These traits however, are known to be vulnerable to forgery as it is possible to forge or steal. In this case Brain electroencephalogram (EEG) signals can be using as a viable biometric because of its robust against falsification. Biometrics which indicate the technique applied to identify individuals using individual biological features are more attractive alternatives. Existing technologies especially used fingerprints, speech, facial features, iris and signatures as a base for recognition. These traits however, are known to be vulnerable to forgery as it is possible to forge or steal. In this case Brain electroencephalogram (EEG) signals can be using as a viable biometric because of its robust against falsification. Through that brain signals can applied to remote healthcare services[3]. Physiological biometrics is refer to the physical characteristics of one or more body sections like (iris, face, or hand geometry). In behavioral biometrics using information extract from person’s behavior, such as typing rhythm, gait and signature. The Cognitive biometrics depend on the outputs of the central and peripheral nervous systems, such as the electroencephalogram and electrocardiogram[4]. Electrical action inside the brain can be influenced by different type of actions like movement of arms and legs, as well as visualization, problem solving or even just by fantasy.

The structure of this paper is as follows. Sections 2 and Section 3 represent Background and Related work, proposed work, respectively. Section 4 represent classification. Finally, the conclusion is presented in Section 5

2. Background and Related work

The human cerebrum comprises of billions of neurons and these are in charge of human brains electrical charge. EEG is an electrophysiological observing strategy that contains the data about the human cerebrum action. The EEG signal can be gotten by putting the sensors on the scalp or utilizing the intracranial electrodes. Measuring this electrical activity of the brain can be accomplished by using electrodes placed over the scalp (show figure 1).

![Electroencephalogram (EEG)](image)

Figure 1 commonly used signals recorded device

Usually signal is refer to a single-valued data of some task dependent on time, distance, position and carries some information of a phenomenon. In general term, we can defined the signal as the output of some sensing or measuring system or outcome of any mathematical operation. There are many kinds signals that we encounter in our day to day like instances, light, sound, electrical, physiological signals etc. Signals are consist of four types: deterministic (stationary), stochastic, fractal and chaotic [1].

(a) Deterministic signals are performed by some mathematical task, and its future value can be forecast exactly for any cycle or time if the past value of the signal is recognized.

(b) Stochastic signals are future value predict is impossible, as these signals show random nature, these signals are also called random signals or non-deterministic signals. These signals cannot be represent by a mathematical task, these signals are represented with the help of statistical properties or features. In the case of physiological signals non-stationary signals are:

(i) Electromyography (EMG): Which is generate by our nerve fiber in the muscles while performing any physical activity. It is used to access the health of the muscle in medical and in Human-Machine Interface used as a control signal for prosthetics.
(ii) Electroencephalography (EEG): is generate by the firing of neurons in human
brain due to ionic changes in our brain. Electrooculography (EOG): which measures the potential between the front and the back of the human eye.

(c) Fractal signals are contained in the biomedical signal processing. These signals possess interesting characteristics, that at every level of intensification they look very similar, this property is arbitrated as scale invariance.

(d) Chaotic signals are type of deterministic signals but its exact future prediction is not possible. The path of these signals echo the deterministic signals but these signals are so sensitive to their past value that determination of future value after a certain short period of time contains unignorably errors.

The brain signals can be obtained by using various approaches. It can be classified as invasive and noninvasive technique. The invasive method requires surgical intervention to putting electrodes under the scalp. Due to medical risks and researchers tend to avoid invasive approach. While when using noninvasive method the electrodes are placed on the scalp of human[5]. The EEG signals can be utilized for different purposes like emotion recognition[6]. The Negar Ahmadi, Mykola Pechenizkiy [15] proposed the horizontal visibility graph (HVG) method and used Phase-Lag-Index (PLI) as feature extraction. This method straightforward and fast to calculate synchronization between human brain signals. While Bashar and Khayrul [7] used investigated the fusion of ECG and EEG signals from low-cost devices with many classifiers (KNN, LDA, and ESAVM) used wavelet domain statistical feature. The feature vectors are computed from the transformed signal using statistical descriptors. The Rodriguez and Ricardo J [8] applied pass-band filter to delete noise and normalizing with 14 channels. Additionally, the usage of logarithmic band power processing combined with LDA as the machine learning approach provides higher accuracy when compared against common spatial patterns or windowed means processing in combination with GMM and SVM machine learning approach. Mahajan et al [9] presented a new unsupervised robust and computationally fast statistical algorithm that used Modified Multiscale Sample Entropy (mMSE) and Kurtosis to feature extraction. The ICA proposed in search.

Algorithm achieves an average sensitivity of 90%. The Ono et al [10] applied fractal dimension and sample entropy in search calculation and analyzed from 4 phases: (retrieval, resting, encoding, and calculation phase). The result fractal dimension analysis may be reflected the activity of the working memory better than sample entropy analysis. The compression result of existing works are shown in Table 1.

Table 1 Machine learning Techniques for pre-processing

<table>
<thead>
<tr>
<th>Authors</th>
<th>Preprocessing &amp; Feature Extraction</th>
<th>Machine Learning Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negar, Ahmad, Mykola Pechenizkiy (2016)</td>
<td>HVG, VVG, Phase-Lag-Index (PLI)</td>
<td>SVM</td>
</tr>
<tr>
<td>Md. Khayrul Bashar (2018)</td>
<td>wavelet domain statistical feature</td>
<td>LDA, and ESAVM</td>
</tr>
<tr>
<td>Ricardo J. Rodriguez, (2016)</td>
<td>discarding well-known artifacts</td>
<td>LDA</td>
</tr>
<tr>
<td>Ruhi Mahajan(2013)</td>
<td>Modified Multiscale Sample Entropy(mMSE)</td>
<td>ICA</td>
</tr>
<tr>
<td>Masaki Ono(2015)</td>
<td>Sample entropy, fractal dimension</td>
<td>statistical computations</td>
</tr>
</tbody>
</table>

Figure 2 shows our proposed methodology for biometric identification methods. Initial step contain the information acquisition from human body like (EEG, ECG, EMG signals). Raw data must be ready for its analysis. In this case using several technique such as re-sampling, smoothing or normalization. The more pertinent data of the EEG is represent by a collection of numerical parameters. This stage refer to feature extraction. Usually the dataset is divide into two part for training and testing respectively. The training set is refer to build the model while testing set are using to estimate the model. The classification accuracy may be high or low. In this case find various technique to reach higher accuracy.
3. Proposed work

3.1 Data acquisition process

Figure 3 display the proposed research methodology of this work. The human brain activity of the subject can be monitored effectively utilizing these signals. The main significance of utilizing these signals in BCI is that based on the human brain activity, the activity which a subject wishes to perform can be made possible.

The dataset utilized in this paper has been acquired from the UCI repository [11][12]. The dataset contains 12 input feature vectors and one target vector. The input feature vectors are obtained by applying the wavelet packet analysis on the original signal in the frequency band 7 to 13 Hz. The target vector represents relaxed or planning state. For experimental details and to know more about the EEG dataset one can refer[13]. Training data has 91 samples and the testing data also has 91 samples. Total samples are 182. Since the dataset used in [13] is 50% training data and 50% testing data a similar number of training and testing data are used in our study.

3.2 Feature Extraction process

3.2.1 Graph entropy (GE) methods

There are many types of GE calculation methods based on either vertex or edges[17]. This study know the (GE) with Shannon’s entropy formula (Clarke 1968). The Shannon’s entropy can be define in the equation 1. In this work we are using all channel as feature extraction. The Ahmad, et al [14] proposed Pz, P3, Pz, CP1, Pz, CP2, Pz, FC2, O1, P3, C4, P3, CP2, Fz, CP2 and Cz gave as better result. While Rodriguez and Ricardo J. [8] applied with 14 channel to feature extraction. The Zhu, et al [18] proposed to only eight channels. But the one channel used by Nakanishi, et al [19]. While Ono, et al [10] proposed Fp1, Fp2, F3, F4, Fz, T7, T8, C3, C4, Cz, P3, P4, Pz, P07, P08, and Oz. The [20] used Fp1, Fp2, F3, F4, C3, C4, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, Pz from 5 subjects used a headset with 19 electrodes. This paper we are selected the channels like(AF8, C1, C2, C3, C4, CP1, CP5, CP6, FC5, FC7, P8, PO8, PZ) as feature extraction. These channels are distributed according to electrode placement in the standard 10 - 20 system.

\[
h = - \sum_{i=1}^{n} p(k) \log (p(k)) \tag{1}
\]

where \( p(k) \) is the probability of \( i \)

3.2.2 Sample entropy (SaE) methods

SaE was proposed by Richman and Moorman in 2000 as the improved approximate entropy method which is a nonlinear dynamic parameter to measure sequence Complexity[21]. The Murphy, et al [22] utilized to evaluate intricacy present in the time arrangement information for short length. It is safe to the solid drifters obstructions e.g., spikes[14]. The algorithm of Sample entropy has three input parameters (1) \( m \): the embedded dimension, (2) \( r \): the similarity criterion, (3) \( n \): the length of a time series[18]. In our case to calculate sample entropy (SaE) two SE features Se1: \( m=2, \ r=0.15 \) Se2: \( m=2, \ r=0.2 \) of every epoch of EEG signals are extract. We applied SaE with all channels and with 13 channels as feature extraction. Sample entropy is the negative natural logarithm when given the conditional probability means any two sequences which are same for \( m \) points will remain same at the next point where \( r \) is defined as similar criterion and \( m \) is the length of data segment[23]. The sample entropy can be define in the equation 2.

\[
\text{Sample Entropy (m,r)} = \ln \left[ \frac{b_m(r)}{A^m(r)} \right] \ldots \ldots (2)
\]
Where

\[ A^m(r) = \frac{1}{N-m} \sum_{k=1}^{N-m} A^m_k(r) \quad \text{(3)} \]

\[ B^m(r) = \frac{1}{N-m} \sum_{k=1}^{N-m} B^m_k(r) \quad \text{(4)} \]

Where

\[ B^m(r) \text{ the probability of the two sequences which match for } m \text{ points} \]

\[ A^m(r) \text{ the probability for two sequences to match for } m+1 \text{ points} \quad \text{(27)} \]

\[ B_k^m(r) = \frac{1}{N-m-1} B_k \]

Where

Given N data points from a time series \( \{x(n)\} = x(1), x(2), \ldots, x(N) \).

Take \( m \) vectors \( X_m(1), \ldots, X_m(N-m+1) \) defined as \( X_m(i) = [x(k), x(k+1), \ldots, x(k+m-1)] \) for \( 1 \leq k \leq N-m+1 \). at the ith sample [27]

3.2.3 Horizontal visibility graphs (VG) Algorithm

The VG algorithm states that two signal samples have visibility following the vertical or horizontal criterion. The required steps to determine synchronization according to the HVG are presented in Figure 4. Horizontal Visibility Graph (HVG) is a type of complex networks. The visibility algorithm has been recently introduced as a mapping between time series and complex networks. This procedure allows us to applying procedure of complex network theory for characterizing time series[15].

![Figure 4 (a) Original Signal](image)

![Figure 4 (b) represent signal](image)

4. Classification Process

The experiments consist of two section(1) feature extraction EEGs depend on the HVG and SaE and(2) evaluating the classification accuracy by using HVG and SaE as feature extraction with KNN as classification.

4.1 Experimental data and Results

The compute feature are then nourished to the classifier for characterization between different states of the human/animal brain cerebrum. We have utilized KNN classifiers in our work. KNN classifier : K-Nearest Neighbour (K-NN) algorithm is chosen to conduct the binary classification. K-NN calculation is a traditional pattern recognition approach which is a statistical supervised classification. The idea is that given a
new test data T. The algorithm obtains the K nearest neighbours from the training set based on the distance between T and the training set. The most dominated class amongst these K neighbours is assigned as the class of T[12]. This section investigates the HVGE and SaE features of EEG signals for identifying human. In classification we are applied with all channels according to KNN classifier and applied with 13 channels like AF8,C1,C2,C3,C4,CP1,CP5,CP6,FC5,FT7,P8,PO8,PZ (shows figure 5). Comparison results between the existing works and proposed work are illustrated in table 2.

Table 2 comparison of various classification algorithms

<table>
<thead>
<tr>
<th>Authors</th>
<th>Data set</th>
<th>Method</th>
<th># channels Used</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mehmoord and Lee(2015)</td>
<td>IAPS Database</td>
<td>Statistical methods + KNN</td>
<td>5</td>
<td>61 %</td>
</tr>
<tr>
<td>Chinmays, etal,(2018)</td>
<td>EPOC Database</td>
<td>independent component analysis  + KNN</td>
<td>14</td>
<td>69.18 %</td>
</tr>
<tr>
<td>PARNIKAP. and PARANJAP.</td>
<td>BCI competition II dataset Ia</td>
<td>cross-correlation features, slow cortical potentials (SCP)+ KNN</td>
<td>2</td>
<td>94.5%</td>
</tr>
<tr>
<td>Paul and Goyal (2017)</td>
<td>UCI repository.</td>
<td>Statistical feature(SII, MAV, WL, RMS) + KNN</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>Proposed work</td>
<td>UCI Dataset</td>
<td>Sample entropy +KNN</td>
<td>13 all</td>
<td>83.7 92.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HVGE+ KNN</td>
<td>13 all</td>
<td>94.8 97.4</td>
</tr>
</tbody>
</table>

5. Conclusion and future work

The essential problem in the classification of EEG signals is the nature of the recorded signals which can be various during the experiment. These unwanted disturbances cannot be controlled since many activities are going on at the same time in the brain. The dataset used is taken from UCI repository[26]. In this study mainly focused on the feature extraction and classification techniques that could be used for EEG signal processing. We suggested SaE and HVGE methods was applied to feature extraction and then classifying with the help of KNN algorithm. The outcomes are encouraging. In our study two path applied, the first path used 13 channels like (AF8, C1, C2, C3, C4, CP1, CP5, CP6, FC5, FT7, P8, PO8, PZ) . The second path utilized all channels with SaE and HVGE methods as feature extraction. By using two path. The accuracy with SaE of all channels and 13 channels are 92.6 % and 83.7 % respectively. While accuracy with HVGE of all channels and 13 channels are 97.4% and 94.8% respectively. Our experimental results showed that the HVGE method is more reliable method than SaE in analyzing EEG signals.

Reference


Yogesh.” 4th IEEE International Conference on Signal Processing, Computing and Control (ISPCC 2k17), Sep21-23, 2017, Solan, India


[15] Negar Ahmadi, Mykola Pechenizkiy ," Application of Horizontal Visibility Graph as aRobust Measure of Neurophysiological Signals Synchrony "2016 IEEE 29th International Symposium on Computer-Based Medical Systems


[25] Helma, Christoph, Eva Gottmann, and Stefan Kramer. 2000. 9 Statistical Methods in Medical Research Knowledge Discovery and Data Mining in Toxicology.

[26] [URL: http://archive.ics.uci.edu/ml/]