Personal identification system based on multidimensional electroencephalographic signals

Eman A. Abdel-Gahffar, May A Salama
Department of Electrical Engineering, Faculty of Engineering Shoubra, Benha University, Cairo, Egypt

ABSTRACT

Personal authentication using electroencephalographic (EEG) signals, is one of the important applications in brain computer interface (BCI). In this work we investigate the use of EEG signals as a biometric trait. Multidimensional EEG signals were represented as symmetric positive-definite (SPD) matrices on a Riemannian manifold. Two experiments are performed in the first; we use minimum distance to Riemannian mean (MDRM) as a classifier. In the second; SPD matrices are vectorized, and the generated vectors are used to train various machine learning (ML) classifiers. MDRM classifier achieved a correct recognition rate (CRR) of 96.92%, while ML classifiers achieved CRR from 95.39% to 99.45%.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Eman A. Abdel-Gahffar
Department of Electrical Engineering, Faculty of Engineering Shoubra, Benha University
Cairo, Egypt
Email: eman.mohamed@feng.bu.edu.eg

1. INTRODUCTION

In the last twenty years, applications based on brain computer interface (BCI) have been attracting a lot of research work [1], [2]. Using brain waves as a biometric is one of the important applications of BCI. Electroencephalographic (EEG) signals offers high degree of stability, universality, uniqueness and are very difficult to be spoofed, which makes them a suitable biometric trait. Recently, the interest in investigating the impact of using EEG signals as a biometric trait has shown great rise. Most of the existing EEG based personal authentication systems can be categorized into three types: unsupervised, supervised and similarity based learning. Similarity based methods relay on measuring the distances among the features for decision making. Different similarity metrics have been investigated including; cross-correlation (CC), cosine similarity (CS), euclidean distance (ED), and mahalanobis distance (MD) [3], [4]. However, as the sample size increases, the computation efficiency and the biometric system accuracy decreases significantly.

In the supervised learning based methods, features are first extracted from the EEG signals, then supervised learning algorithms are deployed to train the model to perform prediction. Various features have been extracted including; power spectral density (PSD), sample entropy (SE), auto-regressive (AR) reflection coefficients, and others. For classification various machine learning techniques have been adopted including; support vector machines (SVM), random forest (RF), gaussian Naïve Bayes (GNB), hidden markov model (HMM), and frequency-weighted power (FWP) [5]-[9]. Although supervised learning based methods have shown excellent performance with respect to fair computation times, their performance is remarkably impacted by the selected features quality. As, EEG signals are non-stationary, complex, affected by human emotions and brain activities. If the feature does not accurately represent the data, the results will be unsatisfactory.

Journal homepage: http://ijeecs.iaescore.com
Recently, unsupervised learning-based methods are being used. The great success achieved by those models in the computer vision domain is now adopted in many BCI applications including brain biometric. Deep learning (DL) models are capable of automatically learning from raw EEG signals without the need for manually extracted features [10]-[12]. An extensive survey on the use of EEG signals for personal authentication can be found in [13]-[15]. In the last few decades, using Riemannian geometry (RG) in studying brain disorders and BCI applications attracted wide attention due to its robustness, simplicity and accuracy. Riemannian geometry have been used in different applications including; human emotion recognition [16]-[18], cryptographic key generation [19], and detecting brain disorders [20]-[22]. A thorough review on applying Riemannian geometry on BCI [23], [24]. The principal goal of the work at hand is to, build a personal identification system using multidimensional EEG signals as a biometric trait. In the proposed system, we avoided the effect of the quality of chosen features on the system performance, by using raw EEG signals without feature extraction. EEG signals recorded from N electrodes are represented as symmetric positive-definite (SPD) matrices on a Riemannian manifold. Two experiments are performed in the first; we use minimum distance to Riemannian mean (MDRM) as a classifier. In the second experiment; SPD matrices are vectorized, and the generated vectors are used to train various machine learning classifiers.

2. METHOD

The principal goal of the work at hand, is to build a personal identification system using multidimensional EEG signals as a biometric trait, based on Riemannian geometry. The basic concepts of Riemannian geometry are introduced in section 2.1. Section 2.2 gives an overview on DEAP dataset used in our study. Our personal identification experiments are introduced in section 2.3.

2.1. Riemannian geometry

Consider brain signals recorded from N channels. Let \( x_k(t), k = 1, ..., N \) be the time series obtained from each electrode. Each individual time series \( x_k(t) \) is broken up into \( m \) diminutive windows, a vector of \( n \) samples is enclosed in every window. Let \( W_{ik} \) refers to each window, where \( k = 1, ..., N \) and \( i = 1, ..., m \). \( m \) covariance matrices \( C_i, i = 1, ..., m \) are generated by convoluting each individual window with its corresponding windows from the \( N \) electrodes. The symmetric positive-definite matrices \( C_i, i = 1, ..., m \) forms a non-positive curvature smooth Riemannian manifold in the \( N(N+1)/2 \) dimensional Euclidean space.

Let \( X \) be the EEG signals obtained from \( N \) channels, each have \( n \) samples (\( X \in \mathbb{R}^{N \times n} \)). The covariance matrix \( C \in \mathbb{R}^{N \times N} \) is given by (1).

\[
C = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T
\]  

(1)

Where \( \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \).

2.1.1. Geodesic distance

The shortest distance between two points \( C_i, C_j \) on the SPD manifold, is measured using the geodesic distance [23], [25]:

\[
d(C_i, C_j) = \| \log \left( C_i^{\frac{1}{2}} C_j C_i^{\frac{1}{2}} \right) \|_F = \left( \sum_{i=1}^{n} \log^2 \lambda_i \right)^{1/2}
\]  

(2)

where \( \lambda_1, \ldots, \lambda_n \) are the eigenvalues of \( \left( C_i^{\frac{1}{2}} C_j C_i^{\frac{1}{2}} \right) \), and \( \| \cdot \|_F \) is the Frobenius norm.

2.1.2. The geometric mean

Geometric mean between \( m \) points \( (C_1, \ldots, C_m) \) on the SPD manifold is called Karcher mean and it is defined as [25], [26]:

\[
G(C_1, \ldots, C_m) = \arg\min_{C \in \mathcal{P}(n)} \sum_{i=1}^{m} d^2(Y, C_i)
\]  

(3)
where \( d \) is the geodesic distance calculated using (2), \( P(n) \) is the SPD matrices space. In (3) there is a unique point \( Y \) that represents the minimum, this point is the geometric mean that represents the solution for the (4):

\[
\sum_{i=1}^{m} \log \left( C^{-\frac{1}{2}}_i Y C^{-\frac{1}{2}}_i \right) = 0
\]  

(4)

for \( m > 2 \) iterative algorithms should be used as their is no closed-form solution for (4) [27], [28].

2.1.3. MDRM classifier

SPD matrices space is non-linear, so using most of the standard classifiers is not feasible. MDRM is a very simple and efficient classifier that can be used in SPD manifold. MDRM is based on nearest neighbor classifier. The number of classes and dimension of data do not have any effect on the way the MDRM operates.

Let \( z \) be the set of all labeled classes \( z_i \in \{ z_1, z_2, \ldots, z_k \} \), \( k \) indicates the number of classes. The class mean \( \hat{M}(z_i) \) is generated in the coarse of the training phase. During the test phase, we obtain \( M \) which is the new observation mean. calculating the distance between \( M \) and each class mean \( \hat{M}(z_i) \) is performed.

Based on the beneath classification rule (5), the class \( z \) to which the new observation belongs is determined.

\[
\hat{z} = \arg\min_{z \in z_1, z_2, \ldots, z_k} \left\{ d \left( M, \hat{M}(z) \right) \right\}
\]  

(5)

where \( \hat{z} \) is the predicted class label.

2.1.4. Vectorization

Each element \( c_{i,j} \) in the square covariance matrix \( C \in P(N) \) represents the covariance value between the signal recorded from the \( i \)th electrode and the \( j \)th electrode.

\[
C = \begin{bmatrix}
c_{1,1} & \cdots & c_{1,N} \\
\vdots & \ddots & \vdots \\
c_{N,1} & \cdots & c_{N,N}
\end{bmatrix}
\]  

(6)

As \( C \in R^{N \times N} \) is a symmetric matrix. \( C \) is flattened into an \( \frac{N(N+1)}{2} \times 1 \) vector [29]:

\[
V_C = \left[ c_{1,1}; \sqrt{2}c_{1,2}; \sqrt{2}c_{1,3}; \sqrt{2}c_{2,3}; c_{3,3}; \ldots; c_{N,N} \right]
\]  

(7)

where, \( \|C\|_F = \|V_C\|_2 \). To preserves the equality of norm non-diagonal elements has a coefficient of \( \sqrt{2} \).

2.2. Dataset

Our proposed system is examined using DEAP dataset [30] which is a publicly available dataset used to analyze the affective state of humans. EEG signals of 32 persons as well as their physiological signals were recorded when each individual was watching 40 videos. Recording was done using 32 electrodes at 512 Hz as a sampling rate. Placing the electrodes was done according the 10-20 international positioning system [31]. There exists pre-processed version of the dataset in which; artifacts were removed, down sampling to 128 Hz, and filtering (from 4-45 Hz) were performed. Every trial (observation) lasted 63 s. The first 3 s of the trial had been recorded before the participant began to engage in the trial experiment. In the work at hand the prepossessed version of DEAP dataset is used and the first 3 s are removed.

2.3. Personal identification experiment

In this work we propose a personal identification system using multidimensional EEG signals as a biometric trait. In the proposed system we use raw EEG signals without feature extraction. Brain signals recorded from \( N \) channels are represented as points on a Riemannian manifold. Two experiments are performed. In the first; we use MDRM as a classifier. In the second, SPD matrices are vectorized, and the generated vectors are used to train various machine learning classifiers. The proposed EEG based identification system consists of two stages; the training stage (enrollment stage) and the testing stage (identification stage).
During the enrollment stage Figure 1, each participant offers his ID, and M trials (each trial is recorded from N-electrodes). Each channel signal is broken up into 1s windows. Using (2.1.1), covariance matrix is calculated. Those covariance matrices are used to calculate the geometric mean for each trial (see section 2.1.2). The M geometric mean from the M trials offered during enrollment are used to calculate one common reference point (G) for each user using (3). G is placed in the user template in the system dataset. Then, G is vectorized (7) to form the participant common vector (V). Then, V_i (where i = 1, ..., P, P is the participants number) are used to train a ML classifier. Figure 2 is t-SNE figure, that shows the 32 participants vectors V_i (where i = 1, ..., 32).

Figure 1. Enrollment stage in which, the identification system is trained

Figure 2. Participants common vectors V_i (where i = 1, ..., P, P is the number of participants). For 32 users in DEAP dataset. Visualization produced through t-SNE method
During the identification stage presented in Figure 3, each user offers an observation (N-channels EEG signals). Signal obtained from every channel is broken up into windows 1 s each. Covariance matrix is generated by convoluting each individual window with its corresponding windows from the N electrodes. A common geometric mean ($\hat{G}$) is generated from the covariance matrices. ($G$) is vectorized using 2.1.4, to generate ($\hat{V}$).

In the first experiment, MDRM classifier is used, the distance between the new generated mean ($\hat{G}$) and participants means ($G_i, i = 1, ..., P, P$ is the participants number) stored in the system database will be computed. The new observation belongs to the user with the minimum distance. In the second experiment, the participant to which $\hat{V}$ belongs is decided by testing the ML classifiers trained using ($V_i, i = 1, ..., P$) during the enrollment stage.

![Identification Stage input](image)

Figure 3. Identification stage in which, each participant provides an observation (N-channels EEG signals)

3. RESULTS AND DISCUSSION

This study investigated the effect of using raw-multidimensional-EEG signals as a biometric trait. In the proposed system, we used raw EEG signals without feature extraction to avoid the impact of the selected features quality on the performance of the system. Brain signals recorded from N channels are represented as points on a Riemannian manifold.

3.1. Results

In the course of this work, two experiments were carried out. In one experiment, we use MDRM classifier to determine the participant to which $\hat{G}$ belongs. In the second experiment, we use different ML classifiers to determine the participant for which $\hat{V}$ belongs. We used five popular ML classifiers; K-nearest neighbours (KNN), SVM, RF, decision trees (DT), multi-layer perceptron (MLP). Data was splitted into training (70%) and testing (30%) using the train-test-split method in sklearn library, 10-fold cross validation was performed in both experiments. Multiple metrics can be used in evaluating biometric identification system performance; correct recognition rate (CRR) is the most commonly used metric [15]. The proposed system CRR is shown in Table 1. The MDRM achieved a CRR of 96.92%, while the best ML classifier is RF achieving CRR of 99.453%.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>MDRM</th>
<th>KNN</th>
<th>SVM</th>
<th>DT</th>
<th>MLP</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRR (± std)</td>
<td>96.92 ± 0.002</td>
<td>99.296 ± 0.004</td>
<td>99.299 ± 0.004</td>
<td>95.391 ± 0.014</td>
<td>97.578 ± 0.013</td>
<td>99.453 ± 0.004</td>
</tr>
</tbody>
</table>

3.2. Performance comparison

A comparison between our proposed system, which is based on representing multi-channel raw EEG signals as SPD matrices on a Riemannian manifold and other EEG based biometric identification systems is
shown in Table 2. Our study suggests that, the proposed system CRR is comparable to other existing EEG-based personal identification systems. When using raw EEG signals, our system CRR (with RF classifier) outperformed unsupervised supervised and similarity based learning systems.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Sbj Chs</th>
<th>EEG-protocols</th>
<th>Features</th>
<th>Classifier</th>
<th>Performance (CRR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[6]</td>
<td>54 62</td>
<td>Motor imagery (MI)</td>
<td>CSP, ERD/S, FFT, AR</td>
<td>SVM and GNB</td>
<td>SVM: 98.97% GNB: 97.47%</td>
</tr>
<tr>
<td>[7]</td>
<td>5 4</td>
<td>Photo stimuli</td>
<td>Hjorth descriptor</td>
<td>NNT</td>
<td>up to 100%</td>
</tr>
<tr>
<td>[8]</td>
<td>109 64</td>
<td>Eye open (EO) and Eye close(EC)</td>
<td>PSD, SE</td>
<td>Mahalanobis distance</td>
<td>EO: 99.7%</td>
</tr>
<tr>
<td>[11]</td>
<td>25 19</td>
<td>Visual evoked potentials</td>
<td>MFCCs, AR</td>
<td>Manhattan distance</td>
<td>MFCC: 95.87% - 96.0% AR: 91.47% - 94.53%</td>
</tr>
<tr>
<td>[13]</td>
<td>60 14</td>
<td>Audio stimuli</td>
<td>WPD</td>
<td>HMM and SVM</td>
<td>HMM: 97.5% SVM: 93.83%</td>
</tr>
<tr>
<td>[14]</td>
<td>50 1</td>
<td>Event-related potential</td>
<td>Raw signals</td>
<td>Cross-correlation</td>
<td>60% - 90%</td>
</tr>
<tr>
<td>[15]</td>
<td>25 14</td>
<td>Event-related potential</td>
<td>Raw signals</td>
<td>k-NN, SVM, LDA, DNN</td>
<td>72% - 96.7%</td>
</tr>
<tr>
<td>[16]</td>
<td>40 17</td>
<td>Visually evoked potential</td>
<td>Raw signals</td>
<td>CNN</td>
<td>80.649% - 98.81%</td>
</tr>
<tr>
<td>[17]</td>
<td>120 64</td>
<td>EEG in resting state</td>
<td>Raw signals</td>
<td>DNN</td>
<td>81.6% - 99.2%</td>
</tr>
<tr>
<td>Proposed system</td>
<td>32 32</td>
<td>Video stimuli</td>
<td>Raw signals</td>
<td>MDRM</td>
<td>MDRM: 96.92%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ML: 95.39% - 99.45%</td>
<td></td>
</tr>
</tbody>
</table>

3.3. Limitations and future research
Our study demonstrated that the proposed system offers a strong personal identification system based on multidimensional-raw-EEG signals. However, further studies are needed to explore the effect of human emotional state, and EEG signals recorded in temporally separated sessions on the EEG-based identification system performance.

4. CONCLUSION
In this work we offer a personal identification system based on multidimensional-raw-EEG signals. Brain signals recorded from N channels are represented as points on a Riemannian manifold. First we use MDRM as a classifier. Then to enhance the system performance, we used the vectorized geometric mean to train various machine learning classifiers. MDRM achieved a CRR of 96.92% , while ML classifiers achieved CRR from 95.39% to 99.45% the best ML classifier was RF. When using Raw EEG signals, our system CRR (with RF classifier) outperformed unsupervised, supervised and similarity based learning systems.

REFERENCES
Advances in Computational Mathematics


BIOGRAPHIES OF AUTHORS

Eman A. Abdel Ghaffar assistant professor in the Electrical Engineering Department, Faculty of Engineering, Benha University. Received the B.Sc., MSc., and Ph.D. degrees in electrical engineering from faculty of Engineering Shoubra, Benha University, Cairo, Egypt. Her research interests includes biometrics, brain computer interface, data compression, information security, signal, and image processing. She can be contacted at email: eman.mohamed@feng.bu.edu.eg.

May A Salama assistant professor in the Electrical Engineering Department, Faculty of Engineering, Benha University. She received her Ph.D. degree from Benha University, Egypt her MSc. and BSc. from Ain Shams University. Her research interests are in the areas of cognitive computing and information security and machine intelligence. She can be contacted at email: May.mohamed@feng.bu.edu.eg.