Brainwaves feature classification by applying K-Means clustering using single-sensor EEG

Ahmad Azhari [a,1,*], Leonel Hernandez [b,2]

[a] Informatics Department, Universitas Ahmad Dahlan, Indonesia
[b] Institucion Universitaria – ITSA, Colombia
[1] ahmad.azhari@tif.uad.ac.id *; [2] lhernandezc@itsa.edu.co
* corresponding author

ARTICLE INFO

ABSTRACT

The use of brainwave signal is a step in the introduction of the individual identity using biometric technology based on characteristics of the body. Brainwave signal has unique characteristics and different on each individual because the brainwave cannot be read or copied by people so it is not possible to have a similarity of one person with another person. To be able to process the identification of individual characteristics, which obtained from the signal brainwave, required a pattern of brain activity that is prominent and constant. Cognitive activity testing using a single-sensor EEG (Electroencephalogram) divided into two categories, called the activity of cognitive involving the ability of the right brain (creativity, imagination, holistic thinking, intuition, arts, rhythms, nonverbal, feelings, visualization, tune of songs, daydreaming) and the left brain (logic, analysis, sequences, linear, mathematics, language, facts, think in words, word of songs, computation) give a different cluster based on two times the test on mathematical activities (no cluster slices of experiment 1 and experiment 2). The result showed that cognitive activity based on math activity can provide a signal characteristic that can be used as the basis for a brain-computer interface applications development by utilizing EEG single-sensor.

I. Introduction

The brainwaves signal has different typical and characteristic of the individual, brainwaves can not be imitated or read by people so it is not possible to have similarity. Identity recognition is necessary to distinguish the characteristics of an individual.

Introduction to identity aims to recognize the identity of the person appropriately. Development of the introduction of identity is done using biometrics technology that aims to do self-introduction by using parts of the body or human behavior. Identity recognition development is done using biometric technology to self-recognition by using parts of the body or human behavior. The use of brainwaves signal is one example of identity recognition using biometric technology based on body parts.

The use of EEG for identification evolves with the rapid development of science in the field of biomedicine [1]–[11], While the use of EEG for security authentication progressed due to the development of image processing in the field of biometrics [12], [13]. The use of brainwaves as a medium for authenticating users has several advantages over other biometric authentication systems such as fingerprints or iris scans, since brainwaves and minds can not be read by others [13].

Reference [12] mentioned that the latest biometrics technology is the emergence of cognitive (cognitive biometrics), i.e. the technology biometrics obtained based on the specific response of the human brain as a trigger, the response, in this case, arises from the cognitive aspect of brainwaves caused by granting some of tasks which stimulate the cognitive response of the brain. Characteristics of brainwaves become very strong when a person is exposed to specific stimuli. The signal measured by the frequency of brainwaves and differentiated based on recording time. To record and measure

the activity of brainwaves required EEG which is designed to acquire, process, and editor of brain signals in the waveform or another form [2]. The increasing availability of wireless devices with a varied number of EEG sensor deployment and possible use of EEG device repeatedly without the risks and the limitations of using non-invasive methods to be attractive to be developed as a medium for self-identity recognition. To recognize the patterns of brain activity is automatically required EEG signal pattern recognition. The selection method of EEG signal pattern recognition precisely specifies the information to be obtained and specific form of EEG signals [14].

II. Related Work

Research on identity recognition using EEG signal much done either in Indonesia and other countries. Klonovs et. al. conducted a study which aims to develop a mobile systems extraction characteristics and classification biometric authentication using EEG. The basis of this study is gaps between mobile web technology and EEG wireless equipment. Brainwaves measurement using the Emotiv Epoc EEG headset used 14 sensors. The research procedure used three tests component, memory based test, visualization test and stimuli based test. The signal is processed and separated from components contaminated by artifacts using blind source separation (BSS) method to get the main components that will be used for feature extraction. The feature extraction in this study using a zero-crossing rate (ZCR), coherence, power spectral density (PSD), cross correlation, wavelet transforms, and latencies. The results showed that there is a difference between different subjects based on the electric potential of biomedical obtained from statistical analysis that is mean and histogram. The most significant methods capable of indicating the typical feature of each individual is zero-crossing rate (ZCR) method, but the result is not so efficient therefor it takes more complex analysis techniques such as power spectral density (PSD) and wavelet analysis. PSD analysis retrieved the similarities in the histogram obtained from spectrogram image. The use of wavelet analysis can be used to measure the potential latency which was resurrected from the visual tests in the occipital lobe area [13].

Hong Bao using blind source separation (BSS) method for mental condition detection. This study aims to separate artifacts or noise from EEG data obtained from the EEG that uses many sensors (multi-channel). Other methods are also used to separate the signal from the artifacts such as event related potential (ERP), analysis of variance (ANOVA), and comparison method using Tukey-kramer. In this study feature extraction conducted by obtaining the average value of the amplitude, the average value of the power frequency, ratio frequency, and the correlation coefficient. As for the classification obtained using linear regression method, cross-validation, rank accuracy, significance, false discovery rate, class size imbalance, and resampling. By providing some tasks on individual the result shows that a classification with an accuracy of 31% for adults, 35% of adults and children together, and 24% for children [15].

III. Methodology

A. Data Collection

Data collection is conducted through data acquisition method that retrieves biometrics data from some people and process it into a signal for the following process. Data collection was performed using an EEG NeuroSky Mindset headset with a single sensor. Electrodes are placed in the FP1 position (frontal temporal) in of the skull or frontal lobe position. The process of EEG signals data retrieval process is done gradually and separately with 6 subjects with each subject performed two times data retrieval using a sampling frequency of 128 Hz per second and generate 2,560 data along for 20 seconds. EEG signals are generated from the stimulus in the form of cognitive tasks. There are nine types of cognitive tasks that used in EEG data collection process namely breath color, face, fingers, counting, objects, passthrough, singing, and sports.

In this research, nine types of cognitive task were used to be based on previous research [14], [16]–[18]. The whole task is based on psychological perceptions to get the imaginative and cognitive responses of the brain. The nine types of cognitive task are described below, each task in terms of its instructions for the subjects.

1. Breathing Task (Breath)

Subject closes the eyes and focuses on breathing for 20 seconds without the slightest movement of the limbs.

2. Object Counting Color Task (Color)

   Subjects are given the task of viewing the colors that appear on the screen and the subject is asked to remember the color and reapply the color column on the screen, this task lasts for 20 seconds silently.

3. Simulated Movement Finger (Finger)

   Subjects are asked to close their eyes and focus on simulating as if moving a finger without actually moving a finger for 20 seconds.

4. Simulated Facial Reconstruction (Face)

   The subject reconstructs a person's face in detail for 20 seconds, without making a sound and moving the limbs.

5. Simulated Object Reconstruction (Object)

   Subjects are given time to look at objects in detail then subjects are asked to close their eyes and focus to reconstruct the object in detail for 20 seconds without the slightest move the limbs.

6. Mathematical Counting Task (Math)

   The subject is given a matter of calculating the simple two-numbered multiplication, and solving the problem silently. For example: 15 x 16. The time given is 20 seconds. False and True is ignored in this task.

7. Simulated Password Recall Task (Pass-thought)

   The subject makes a sentence in the form of a combination of letters and numbers and subjects are asked to close their eyes and imagine the sentence for 10 seconds.

8. Song Recitation Task (Song)

   Subjects imagine for 20 seconds a song or sound with lyrics, without actually making a sound and moving the limbs.

9. Simulated Sport Task (Sport)

   The subject chooses a favorable sporting movement then the subject is asked to close his eyes and imagine in the mind of the exercise movement without actually moving the limbs. This task is done for 20 seconds.

B. Signal Processing

In terms of signal processing, biometric data is obtained from the result of the data collection using EEG will be grouped by subject, cognitive task, and time data collection. Furthermore, the data will be feature extraction which aims to get the characteristics differences that represent the main characteristics of the signal. The extraction performed by mean, standard deviation skewness, kurtosis, and entropy.

Mean measure of data distribution, standard deviation measures variations of data distribution, skewness measures the asymmetric level of data distribution, kurtosis measures how flat or high the distribution of data is to a normal distribution, and entropy is used to measure the randomness of the data distribution.

If mean ($\bar{x}$), standard deviation (σ), skewness (s), kurtosis (k), entropy (H) has the sum of data N and x_i is pixel data i then color moment can be calculated by equation (1-5).

$$mean = \bar{x} = \frac{1}{N}\sum_{i=0}^{N} x_i \tag{1}$$

$$standard\ deviation = \sigma = \sqrt{\frac{1}{(N-1)}\sum_{i=0}^{N}(x_i - \bar{x})^2} \tag{2}$$

$$skewness = s = \sum_{i=0}^{N}(x_i - \bar{x})^3/(N\sigma^3) \tag{3}$$

$$kurtosis = k = \sum_{i=0}^{N}(x_i - \bar{x})^4/(N\sigma^4) \tag{4}$$

$$entropy = H = E[-log_2 P(x)] = -\sum_{i=0}^{N} P(x)log_2 P(x) \ bits \tag{5}$$

C. Data Matching

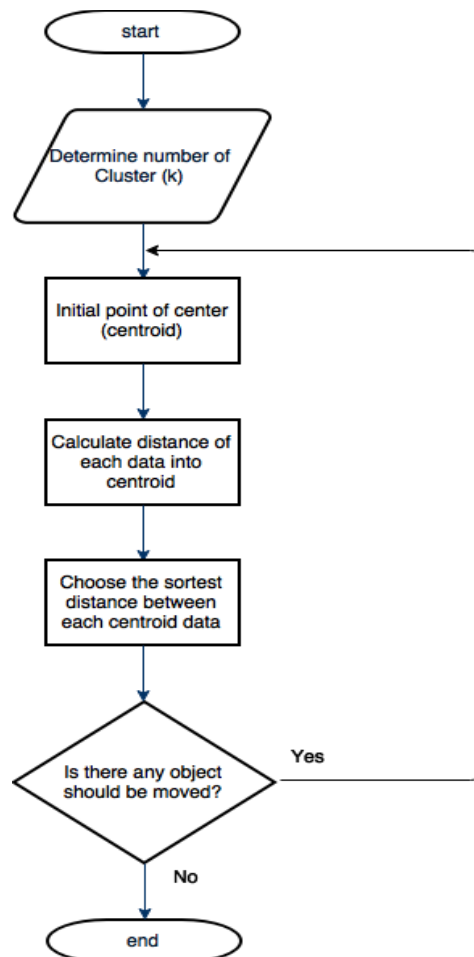Data matching is conducted through K-Means clustering. The results will be grouped into K clusters (Fig. 1).

Fig. 1. K-Means Clustering Flowchart

There are steps stages in performing data matching using the K-Means clustering:

a. The first step is to determine the number of clusters to be formed.

b. The second step is initialization point of the cluster (centroid) randomly from a set of data to be grouped.

c. The third step is to calculate the distance of any data to each centroid using Euclidean distance.

d. The fourth stage is to choose the shortest distance between each data with the centroid.

e. The fifth step is to determine the position of new centroid by calculating the average of data in the same centroid.

f. The sixth step is to make sure the algorithm converged with data in the new centroid same with the initial centroid. If the two centroids have the same data then the process end. If not, back to step 3.

IV. Results

Preliminary data obtained from the EEG will be extracted to produce characteristics. The results of the feature extraction are the value of the mean, the value of standard deviation, the value of skewness, the value of kurtosis, and the value of entropy.

The results grouped by subject, cognitive task, and time data retrieval. Extraction of characteristic results data will then be performed against the data normalization and divided into 4 clusters using K-Means Clustering. After the specified number of cluster, the next step is to determine the initial focal point centroid randomly. Initialization centroid can be seen in Table 1.

Table 1. Initialization Centroid

Cluster	Centroid Value
1	0.0395
2	0.8661
3	0.1242
4	0.2785

The next step is to determine the shortest distance between each data center point of each centroid. Determining the distance using the Euclidean method by selecting the closest distance or minimum value. The data is then grouped and distinguished by a cluster. Steps end when the data on the new centroid equal to data on the old centroid. Data clustering results can be shown in Table 2. Data in the same cluster group can be grouped and summed. The result of grouping data in a cluster shown in Table 3.

Table 2. The result of grouping data in the cluster

Cluster	*Centroid*	Sum
1	0.04	24
2	0.87	2
3	0.12	24
4	0.29	4

From the calculation results obtained four cluster groups. The first cluster has a cluster center at point 0.04 with the amount of same data 24 people. The second cluster, there are two people who have same data with a central cluster at the point of 0.87. The third cluster there are 24 people who have same data with a central cluster at the point of 0.12. The fourth cluster has a central cluster at the point of 0.29 with four people.

Tests of cognitive activity using a single-sensor EEG performed in this study is divided into two categories, namely the activity of cognitive involving the ability of the right brain (creativity, imagination, holistic thinking, intuition, arts, rhythms, nonverbal, feelings, visualization, tune of songs, daydreaming) and left brain (logic, analysis, sequencing, linear, mathematics, language, facts, think in words, word of songs, computation).

Based on research that has been conducted by three research subjects in two tests showed that mental activity involving the right brain's cognitive abilities (color, sing, sport). That can be shown in Table 3 and Fig. 2.

Table 3. Clustering Result

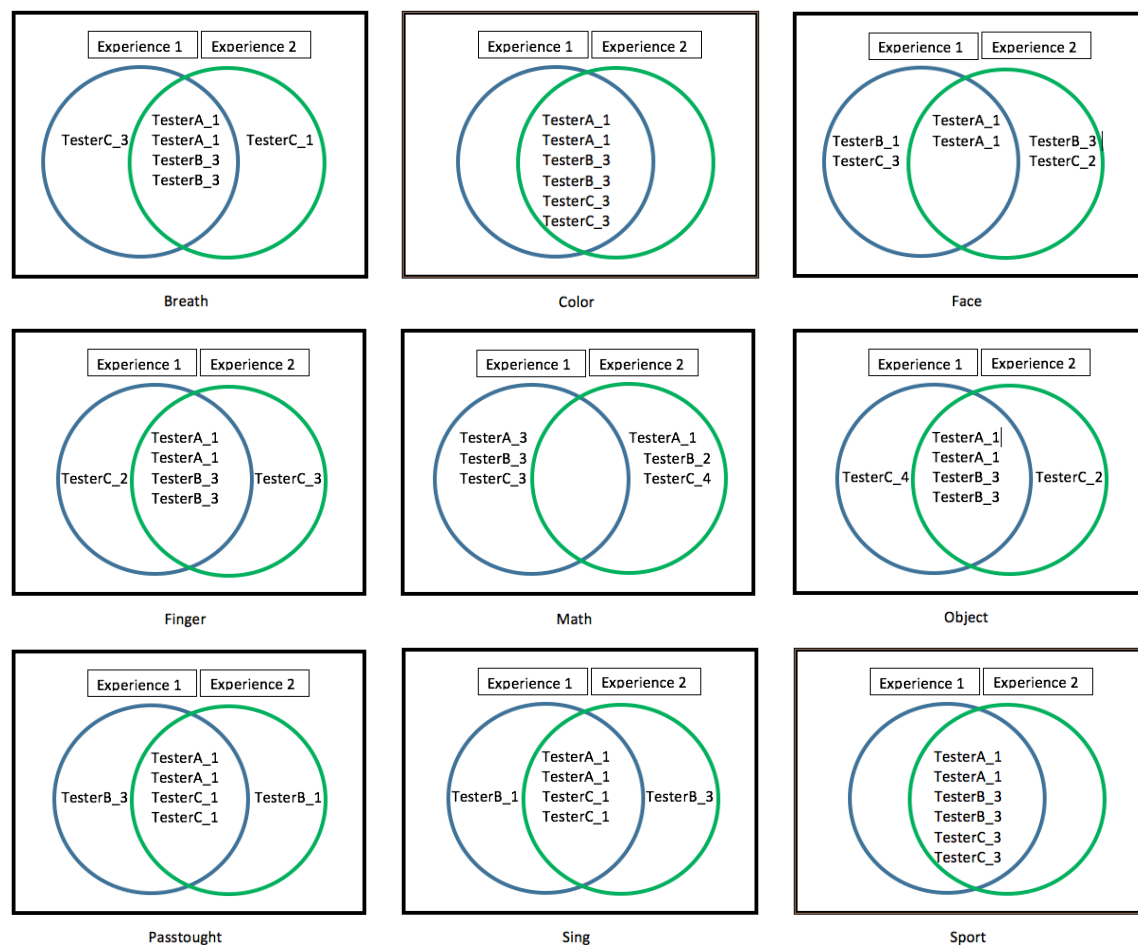| Subject	Experiment 1								
	'Breath'	*'Color'*	*'Face'*	*'Finger'*	*'Math'*	*'Object'*	*'PassThought'*	*'Sing'*	*'Sport'*
TesterA	1	1	1	1	3	1	1	1	1
TesterB	3	3	1	3	3	3	3	1	3
TesterC	3	3	3	2	3	4	1	1	3
Subject	Experiment 2								
	'Breath'	*'Color'*	*'Face'*	*'Finger'*	*'Math'*	*'Object'*	*'PassThought'*	*'Sing'*	*'Sport'*
TesterA	1	1	1	1	1	1	1	1	1
TesterB	3	3	3	3	2	3	1	3	3
TesterC	1	3	2	3	4	2	1	1	3

Fig. 2. Visualization of Clustering result

Cognitive activity "color" and the "sport" yield identical clusters of each subject in the two trials. Activities "breath", "finger", and "object" value consistent cluster in two trials, but only on the subject 1 and 2. While "passtought" activity, and "sing" give a consistent cluster in the two trials, only on the subject 1 and 3. Likewise with "finger" activity, which only gives consistent results on the subject 1.

Activities "math", is the only activity that provides different cluster results in two trials (no slices cluster of experiment 1 and experiment 2).

V. Conclusion

From the results of this study can be concluded that cognitive activity "Math" can provide signal characteristics that can be used as the basis of application development of brain computer interface by utilizing single sensors EEG using clustering method K-Means.

References

[1] I. Jayarathne, M. Cohen, and S. Amarakeerthi, "BrainID: Development of an EEG-based biometric authentication system," in *Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016 IEEE 7th Annual*, 2016, pp. 1–6.

[2] A. A. Ghodake and S. D. Shelke, "Brain controlled home automation system," in *Intelligent Systems and Control (ISCO), 2016 10th International Conference on*, 2016, pp. 1–4.

[3] Z. Pang, J. Li, H. Ji, and M. Li, "A new approach for EEG feature extraction for detecting error-related potentials," in *Progress in Electromagnetic Research Symposium (PIERS)*, 2016, pp. 3595–3597.

[4] J. Katona, I. Farkas, T. Ujbanyi, P. Dukan, and A. Kovari, "Evaluation of the NeuroSky MindFlex EEG headset brain waves data," in *Applied Machine Intelligence and Informatics (SAMI), 2014 IEEE 12th International Symposium on*, 2014, pp. 91–94.

[5] B. Trowbridge, C. Rodriguez, J. Prine, M. Holzemer, J. McCormack, and R. Integlia, "Gaming, fitness, and relaxation," in *Games Entertainment Media Conference (GEM), 2015 IEEE*, 2015, pp. 1–1.

[6] M. Varela, "Raw EEG signal processing for BCI control based on voluntary eye blinks," in *2015 IEEE Thirty Fifth Central American and Panama Convention (CONCAPAN XXXV)*, 2015, pp. 1–6.

[7] Y. Yoshimura *et al.*, "The Brain's Response to the Human Voice Depends on the Incidence of Autistic Traits in the General Population," *PLoS One*, vol. 8, no. 11, p. e80126, Nov. 2013.

[8] D. Bright, A. Nair, D. Salvekar, and S. Bhisikar, "EEG-based brain controlled prosthetic arm," in *Advances in Signal Processing (CASP), Conference on*, 2016, pp. 479–483.

[9] A. Patil, C. Deshmukh, and A. R. Panat, "Feature extraction of EEG for emotion recognition using Hjorth features and higher order crossings," in *Advances in Signal Processing (CASP), Conference on*, 2016, pp. 429–434.

[10] M. Diykh, Y. Li, and P. Wen, "EEG Sleep Stages Classification Based on Time Domain Features and Structural Graph Similarity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1159–1168, Nov. 2016.

[11] A. Azhari, A. Susanto, and I. Soesanti, "Studi Perbandingan: Cognitive Task Berdasarkan Hasil Ekstraksi Ciri Gelombang Otak.pdf," 2015, vol. 3.1, p. 7.

[12] D. Putra and K. Gede, "Sistem Verifikasi Biometrika Telapak Tangan dengan Metode Dimensi Fraktal dan Lacunarity," *Maj. Ilm. Teknol. Elektro*, vol. 8, no. 2, 2012.

[13] J. Klonovs, C. K. Petersen, H. Olesen, and J. S. Poulsen, "Development of a Mobile EEG-Based Feature Extraction and Classification System for Biometric Authentication," Aalborg University Copenhagen, 2012.

[14] A. Azhari, "Ekstraksi Ciri Gelombang Otak Menggunakan Alat Neurosky Mindset Berbasis Korelasi Silang.pdf," Universitas Gadjah Mada, Yogyakarta, 2015.

[15] B. H. Tan, "Using a low-cost eeg sensor to detect mental states," Carnegie Mellon University, 2012.

[16] B. Johnson, T. Maillart, and J. Chuang, "My thoughts are not your thoughts," 2014, pp. 1329–1338.

[17] A. Saha, A. Konar, P. Das, B. Sen Bhattacharya, and A. K. Nagar, "Data-point and feature selection of motor imagery EEG signals for neural classification of cognitive tasks in car-driving," in *2015 International Joint Conference on Neural Networks (IJCNN)*, 2015, pp. 1–8.

[18] R. Zafar, A. S. Malik, H. U. Amin, N. Kamel, S. Dass, and R. F. Ahmad, "EEG spectral analysis during complex cognitive task at occipital," in *Biomedical Engineering and Sciences (IECBES), 2014 IEEE Conference on*, 2014, pp. 907–910.