Electromyogram-based algorithm using bagged trees for biometric person authentication and motion recognition

Siho Shin, Jaehyo Jung, Mingu Kang, Kyeung Ho Kang and Youn Tae Kim+
IT Fusion Technology Research Center, Dept. of IT Fusion Technology, Chosun University, Gwangju61452, Korea, petruskim@chosun.ac.kr

Abstract—With the recent rising interest in health and convenience, bio-signal measurement, and sensor technology are developing rapidly. Using bio-signals, one can calculate motion and other variables, can authenticate individuals using fingerprints, and can control objects using gestures. These bio-signals are convenient because they can be measured easily, anytime, anywhere. However, there are cases of misuse of bio-signal based technology. To solve this problem, we develop an electromyogram (EMG)-based algorithm that can be used for biometric authentication. The proposed algorithm uses two channels to acquire EMG data, and hand motion recognition and authentication are carried out through signal processing. As a result of dividing 50 data sets into artificial neural networks and applying the ensemble technique, the authentication success rate achieved by the proposed algorithm was 83.8%.

I. INTRODUCTION

With the development of technology, interest in research that can improve the quality of life is increasing. The wearables market is growing rapidly owing to the popularization of smartphones since the 2000s; devices such as health bands and smartwatches, which can determine user health, have emerged. According to the Internet market research firm Internet Data Centers, the global wearables market is expected to grow from 113 million units in 2017 to 222 million units in 2021.[1]

Biometrics technology is rapidly developing, based on advances in sensor technology, to the extent that, by using various sensors, it is possible to check user identity with various biometrics, such as an electrocardiogram, brainwave or vein.[2] An electromyogram (EMG) can be measured easily, anywhere, anytime; thus, it is applied not only to user authentication, but also to other applications, such as rehabilitation, object control, etc.[3][4]

However, along with these technological advances, there have been a growing number of cases of hacking and misuse of personal information. [5] For example, hackers often profit inappropriately by copying fingerprints, or a person’s iris, or by hacking passwords. There are reported studies on certifying individuals using their gait. However, this method requires that the subjects wear large devices on their thighs. This causes users to feel uncomfortable and is an inefficient system because of its reduced portability.

An EMG signal generates different signals, depending on the individual’s physical condition, so that, although people perform the same actions, their EMG signal has different characteristics. Thus, authentication based on an EMG signal should be able to resolve the aforementioned problems.

Based on this conjecture, we proposed an EMG-based algorithm for use in personal authentication.

EMG data were obtained using a two channel-based EMG measurement module. Subsequently, signal processing and data classification were carried out using Matlab, and then, based on filtered EMG data, operations that required individual authentication—such as validating an individual’s identity, opening a door, or turning on / off a bulb, for example—could proceed.

A. EMG (Electromyogram)

When muscles move, muscle fibers in the muscle repeat contraction and relaxation actions, [6] and an EMG is a record of the electrical current changes that occur during this process. [5]

B. Ensemble technique

Ensemble methods combine several decision tree classifiers to produce predictive performance better than that available from a single decision tree classifier. The main principle behind the ensemble model is that a group of weak learners come together to form a strong learner, thus increasing model accuracy. [11]

When we try to predict the target variable using any machine learning technique, the main causes of differences between actual and predicted values are noise, variance, and bias. The ensemble method helps to reduce these factors (except noise,
II. METHOD

A. Paper Identification Line

The algorithm is designed for hand motion classification, and personal authentication. EMG signals were acquired from one arm of the user. We extracted five feature points for personal authentication and pattern recognition, so that personal authentication and pattern recognition could be performed based on these extracted feature points. To classify the hand movements, an artificial neural network (ANN), which has simple and excellent performance, was applied, while personal authentication was performed by applying bagged trees that combined Bootstrap with decision trees.

B. EMG signal acquisition

When people move, they simultaneously use multiple muscles, which requires that multiple channels must be configured to obtain the EMG data from any action. [12] In our study, EMG signals to be used for individual certification were acquired from 41 men and 9 women, using six Ag/AgCl wet electrodes per person, attached to areas with relatively large muscle distribution.

C. Feature extraction

There are several parameters that can be extracted from EMG signals, including peak to peak value, signal length, signal average, median frequency, and several other parameters. In this study, five parameters were extracted from the filtered EMG signal for pattern recognition—the Integrated EMG (IEMG), the Simple Square Integral (SSI), the Waveform Length (WL), the VAR (Variance), and the RMS (Root Mean Square). [13][14]

\[
RMS = \sqrt{\frac{1}{N} \sum_{n=1}^{N} x_n^2} \tag{1}
\]

The RMS feature point is a parameter obtained by performing a square * average * square root operation on the EMG signal. This is a variable that appears primarily to be highly correlated with muscular strength, and which can be simply calculated from EMG signals, as shown in Eqn (1).

\[
WL = \sum_{n=1}^{N-1} |x_{n+1} - x_n| \tag{2}
\]

WL (Eqn 2) is a parameter that expresses surface EMG signal complexity. It can be defined as the accumulated length of the surface EMG signal waveform over a period of time. There is also correlation with the amplitude, frequency, and time of the waveform.

\[
IEMG = \sum_{n=1}^{N} |x_n| \tag{3}
\]

IEMG (Eqn 3) is a widely used parameter for estimating muscle use and strength. Following a rectification process on the raw EMG signal, the area of the extracted signal is calculated.

\[
SSI = \sum_{n=1}^{N} |x_n|^2 \tag{4}
\]

SSI (Eqn 4) is an indicator of the surface potential energy and refers to the cumulative value of the squared EMG signal amplitude.

\[
VAR = \frac{1}{N-1} \sum_{n=1}^{N} x_n^2 \tag{5}
\]

VAR (Eqn 5) represent the dispersion values of the EMG signals.

D. Pattern recognition

The experiment used data from six hand motions—1. Making a fist, 2. Hands up, 3. Hands down, 4. Double tap, 5. Fingers spread, 6. Twist wrist. We extracted the parameters for each motion using Matlab, which resulted in extraction of 600 parameters—100 for each action. The 100 parameters consisted of 20 each of IEMG, SSI, WL, VAR, and RMS, and each action was repeated for five seconds.

E. Bagged trees

Bagging techniques are simple and powerful ensemble techniques, and became known in 1996 after publication by Breiman. [15] In general, less training data reduces algorithm accuracy, and can lead to results converging on one side. To compensate for this, bagging, which combines the Bootstrap (random sampling) technique with a decision tree, was used. [16]
III. RESULT

The top panel in Fig. 3 shows the EMG signals created from various hand motions, and the bottom panel shows the confusion matrix created using the ANN.

Green blocks in the RHS show the number and percent class samples in the data set, while red blocks show misclassifications. Grey blocks show total classification percentages for each pattern, and the blue block shows the overall average. [17]

Fig. 4 Classification results: Bagged trees

The graph on the left of Fig. 4 is a confusion matrix, showing the EMG data classification result. The row of the confusion matrix indicates true class, and column indicates predicted class. Red indicates a misclassified result, while green means that the true class matches the predicted class. A total of 50 EMG data were classified, using the Bagged trees. A total of 1,500 parameters were applied to the Bagged trees and 83.8% (1,257) probability separated individuals.

Fig. 5 Receiver Operating Characteristic curve
The Receiver Operating Characteristic (ROC) curve is a graph illustrating classifier performance; the Area Under the Curve (AUC) is the area under the ROC curve. The closer the AUC is to 1, the higher the accuracy of the classifier, while if the AUC is less than 0.5, it means that the performance of the classifier is bad. The Bagged trees applied in this study have exhibited good performance, as illustrated by their AUC value of 0.94.

IV. CONCLUSION

Research was carried out on EMG-based pattern recognition and individual authentication. To improve the accuracy of EMG-based data classification results, we developed an EMG module that could effectively minimize noise. A total of 1,500 parameters were extracted using Matlab, and based on these data, were grouped, and the authentication test results applying two simple movements were 92.7% and 83.8%, respectively. In the field of personal certification, the authentication success rate of 83.8% is considered very low accuracy. However, there are no existing research related to personal authentication studies using EMG signals measured at the wrist. Further, these studies are meaningful considering that they are in the early stages of research. As further research is being carried out to modify the structure of the artificial neural network used in this study, the accuracy is expected to improve in the future.

ACKNOWLEDGMENT

This research was supported by the Mid-career Researcher Program (No. 2016R1A2B3009423) through NRF grant funded by the MSIT (Ministry of Science and ICT) and the Basic Science Research Program through the NRF funded by the Ministry of Education of Korea (No. 2018R1A6A1A03015496).

REFERENCE

[5] About fingerprint duplication [Internet]. Available from https://www.youtube.com/watch?v=w3Yexpw6k3w