Finger language recognition based on ensemble artificial neural network learning using armband EMG sensors

Seongjung Kim, Jongman Kim, Soonjae Ahn and Youngho Kim*
Department of Biomedical Engineering, Yonsei University, Wonju, Korea

Abstract.
BACKGROUND: Deaf people use sign or finger languages for communication, but these methods of communication are very specialized. For this reason, the deaf can suffer from social inequalities and financial losses due to their communication restrictions.
OBJECTIVE: In this study, we developed a finger language recognition algorithm based on an ensemble artificial neural network (E-ANN) using an armband system with 8-channel electromyography (EMG) sensors.
METHODS: The developed algorithm was composed of signal acquisition, filtering, segmentation, feature extraction and an E-ANN based classifier that was evaluated with the Korean finger language (14 consonants, 17 vowels and 7 numbers) in 17 subjects. E-ANN was categorized according to the number of classifiers (1 to 10) and size of training data (50 to 1500). The accuracy of the E-ANN-based classifier was obtained by 5-fold cross validation and compared with an artificial neural network (ANN)-based classifier.
RESULTS AND CONCLUSIONS: As the number of classifiers (1 to 8) and size of training data (50 to 300) increased, the average accuracy of the E-ANN-based classifier increased and the standard deviation decreased. The optimal E-ANN was composed with eight classifiers and 300 size of training data, and the accuracy of the E-ANN was significantly higher than that of the general ANN.

Keywords: Finger language recognition, armband sensor, surface electromyography (EMG)

1. Introduction

There are approximately 360 million deaf people in the world who use finger or sign languages to communicate [1]. Most of them suffer from financial losses and social inequalities in different fields of life, such as education, culture, arts, health, medical care and law, due to communication problems with normally hearing people [2]. In addition, it is even more challenging to communicate internationally, since each country uses its own finger or sign language.

Finger or sign language recognition systems translate hand gestures into voices or texts to facilitate communication between a deaf person and a hearing person [3], and they can play an important role in solving the language barrier caused by the use of different sign languages in each country. Many previous studies [4–13] have been performed on the development of finger or sign language recognition systems. Previous investigations of the systems can be categorized into two main groups: computer vision-based and wearable sensor-based techniques.

---

*Corresponding author: Youngho Kim, Department of Biomedical Engineering, Yonsei University, Wonju, Korea. Tel.: +82 33 760 2492; E-mail: younghokim@yonsei.ac.kr.
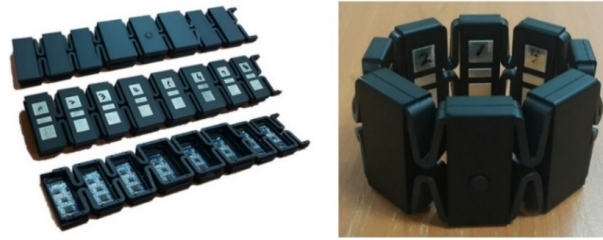
Fig. 1. An armband module with eight SEMG sensors.

First, a computer vision-based recognition system can track the position of bare hands and recognize gestures. Singha and Das [4] reported on a recognition system based on Euclidean classifier for 24 Indian finger languages using histogram matching of the skin-colored regions from the background. The accuracies for all of the finger languages were more than 96%. Dong et al. [5] classified 24 American finger languages by measuring the distance between a Kinect camera and the hand, and reported a classification accuracy of more than 90%. The computer vision-based recognition system can recognize gestures effectively without interferences like wearable gloves and bio-signal sensors. However, this would have the disadvantages that image recognition is very sensitive to background illumination and hand position [6]. In addition, the recognition method using video equipment would be not practical, because it is expensive and less portable. Kadous [7] reported on a system based on glove-type fiber optic sensors using a decision tree classifier to recognize 95 Australian Sign Language gestures with an average classification rate of 80%. Fang et al. [8] reported on a two-glove type system of 10 bending sensors and three electromagnetic sensors that used a fuzzy decision tree classifier to recognize 75 hand shapes with an average classification rate of 92%. Glove-type wearable systems are more practical than computer vision-based systems in terms of cost and can potentially be used without modification of the environment [7]. However, they would have the limitations of impeded hand movements and discomfort during daily life.

Differing from the approaches mentioned above, many studies on the use of non-invasive bio-signal sensors have been recently suggested. Most of these have used surface electromyography (SEMG) sensors for finger or sign language recognition [10–15]. They would be less sensitive to environments and more practical. In addition, electromyography (EMG), when measured with multiple sensors, could recognize various movements of the human upper limbs [9]. Zhang et al. [11] classified 40 Chinese finger languages using five channels of SEMG and hidden Markov model (HMM) classifier with an accuracy of 93%. Wu et al. [13] also classified 40 American finger languages using four channels of SEMG and support vector machine (SVM) classifier with an accuracy of 96%. Likewise, many studies involving a variety of classifier types and feature vectors of EMG have been performed. However, the sensors used in the previous studies did not maintain the performance and were easily detached, because they require expertise to use.

To distinguish the hand shapes in the finger languages effectively, EMG signals obtained in forearm muscles should be classified with the minimum error rate. There have been various kinds of pattern recognition techniques for classifying data such as decision tree [7], SVM [13] and HMM [14]. It has also been common to use artificial neural network (ANN) [17–20]. However, a single classifier model would be not appropriate to classify many different signals, because all classifier models always have an error rate. To minimize errors of the single classifier model, an ensemble learning algorithm to build multiple classifier models and to determine the class with the highest likelihood has been used [21]. However, most of the previous studies regarding finger or sign language recognition systems applied a single classifier.
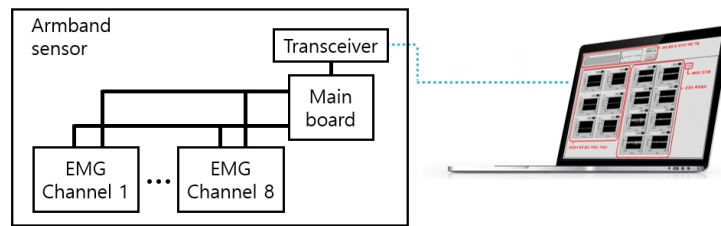
Fig. 2. The schematic of the armband system.

The purpose of the present study was to develop an armband-type wearable sensor consisting of multi-channel SEMG sensors and to evaluate a finger language recognition system using an ensemble learning algorithm based on an ANN classifier.

2. Hardware and software development

An armband module with eight SEMG sensors was developed in this study (Fig. 1). In each EMG sensor, there were two silver plate-shaped electrodes with 10 mm × 10 mm contact area and 10-mm electrode-to-electrode distance. Two 500-mAh lithium-ion batteries and a charging circuit were embedded on the main board. A power circuit was also embedded on the main board. Eight sensors were connected by a polyurethane strap so that it could be worn around the forearm. Low frequency noise was removed through a 15-Hz second-order high-pass filter circuit, and an instrument amplifier with 1000 × gain was installed in the system. The SEMG sensors were designed to record SEMGs synchronously. Because the main frequency band of EMG was from 15 Hz to 300 Hz, EMG signals from each channel were digitized at a sampling frequency of 600 Hz using an I2C-compatible 12-bit A/D converter (ADC121C021, Texas Instruments, USA). Digitized signals were transmitted to the PC using radio frequency wireless communications with a microcontroller (nRF52832, Nordic Semiconductor, Norway). A graphical user interface was developed using LabVIEW (Version 2011, National Instruments, USA) to measure and store multi-channel EMG signals and the battery level. Figure 2 shows the overall structure of the developed armband system.

3. Pattern recognition algorithm

3.1. Signal acquisition and filtering

EMG signals measured around the forearm muscles using the armband module provided information regarding wrist and finger movements. EMG signals were filtered with a 15–300 Hz 4th-order bandpass filter.

3.2. Segmentation

The onset and offset of muscle activations were determined to extract each subword from a sentence. Figure 3a shows filtered EMG signals that contain a lot of baseline noise. The segmentation of muscle activity was difficult to distinguish correctly because the signal-noise ratio was small. In this study, to
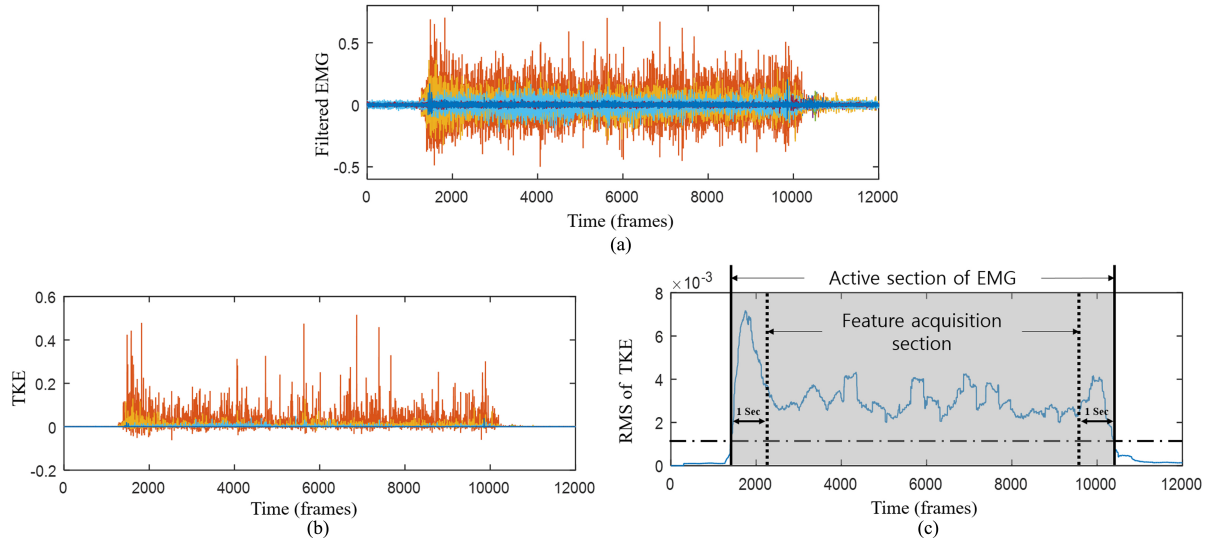
Fig. 3. Teager-Kaiser energy (TKE)-based muscle active segmentation process. (a) Filtered EMG signal; (b) TKE; (c) muscle active segmentation and feature vectors acquisition section.

minimize baseline noise for better segmentation, the Teager-Kaiser energy (TKE) operator [22] was applied, as shown in Fig. 3b. TKE was calculated as follows:

$$\text{TKE}_{\text{channel}}[n] = \text{EMG}_{\text{channel}}[n]^2 - \text{EMG}_{\text{channel}}[n+1]\text{EMG}_{\text{channel}}[n-1] \tag{1}$$

To smooth the TKE signals, the root mean square (RMS) at a 500-ms window length was calculated and the sum of each RMS was obtained as follows:

$$\text{RMS}(i) = \sum_{\text{channel}}^{8} \sqrt{\frac{\sum_{i-L}^{L-1} \text{TKE}_{\text{channel}}[n]}{L}}, \quad L = 500 \text{ ms} \tag{2}$$

The amplitude of RMS was compared with the threshold to detect the onset and offset of active EMG sections (Fig. 3c). Threshold value was experimentally determined because it differed per subject. The signals were removed for one second after onset and before offset, since the hand shape was not complete at those times. The extracted 8-channel EMGs in the feature acquisition section were used in the next step to calculate feature vectors.

3.3. Feature vector extraction

The classification accuracy of the pattern recognition system based on EMG depends on the feature vectors of measured signals [23]. The features of EMG can be obtained from time, frequency or wavelet domains. However, according to Englehart et al. [24], frequency and wavelet domain feature vectors were not suitable for real-time systems due to their computational complexity. For this reason, time domain feature vectors were most appropriate when processing a large amount of data such as EMG signals. Therefore, four features of EMG were selected in the time domain alone, and the window length was 500 ms. Table 1 shows selected EMG features and their formulas. L is defined as the number of samples in the window length ($L = 300$). A 32-dimension (8 channels $\times$ 4 features) feature vector of EMG was obtained per window.

Table 1
Feature vectors of EMG signals

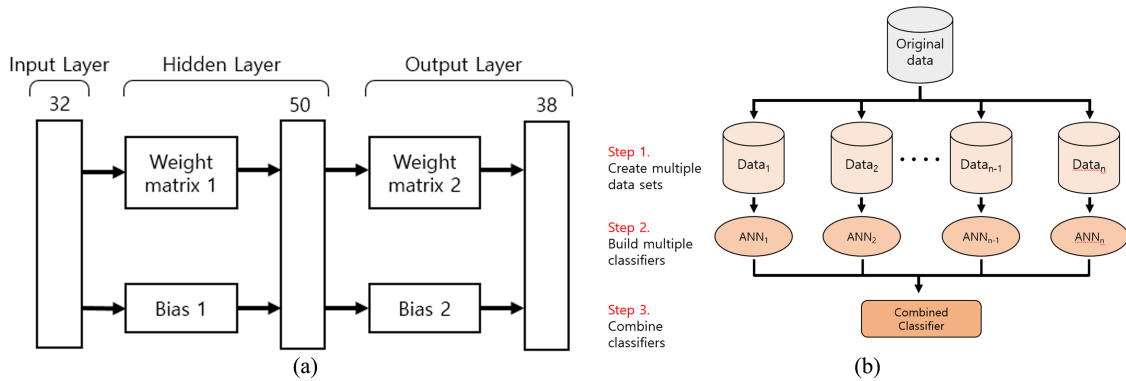| Feature vector	Formula
Mean of absolute value	$\mathrm{MAV} = \sum\limits_{i=1}^{L} \mathrm{EMG}_i$
Root mean square	$\mathrm{RMS} = \sqrt{\frac{\sum_{i=1}^{L} \mathrm{EMG}_i^2}{L}}$
Variance	$\mathrm{VAR} = \frac{1}{L-1} \sum\limits_{i=1}^{L} \mathrm{EMG}_i^2$
Waveform length	$\mathrm{WL} = \sum\limits_{i=2}^{L} |\mathrm{EMG}_i - \mathrm{EMG}_{i-1}|$

Fig. 4. Ensemble learning-based finger language recognition algorithm: (a) structure of artificial neural network; (b) ensemble artificial neural network classifier structure.

3.4. Classifier training

Figure 4a shows the basic structure of an ANN used as a single pattern classifier in this study. The ANN consists of three layers (input layer, hidden layer and output layer). The hidden and output layers contain a weight matrix (W) and bias vector (B) for the pattern classification of the feature vector. The number of nodes in the hidden layer was selected as 50. The W and B of the multi-layer ANN classifier were repeatedly trained using a scaled conjugate gradient backpropagation algorithm [25].

Ensemble learning is a method to analyze signal patterns by various results of two or more classifiers and obtain more accurate classification performance than that attained by a single classifier [26]. A random vector from the training data was used to create various ANN classifiers. An ensemble artificial neural network (E-ANN) structure was generated by combining ANN classifiers (Fig. 4b).

4. Experiment

The performance of the finger language recognition using the armband-type sensor was evaluated through classification accuracy verification. Right-handed subjects who had no musculoskeletal disorders were recruited. The subjects were twelve males and five females with an average age of 25.21 ± 1.14 years. Each subject wore an armband sensor on the right forearm. A reference sensor was placed at the belly of the flexor carpi radialis muscle, and the remaining sensors were positioned at regular intervals by the strap (Fig. 5). Korean finger language symbols (14 consonants, 17 vowels, 7 numbers)
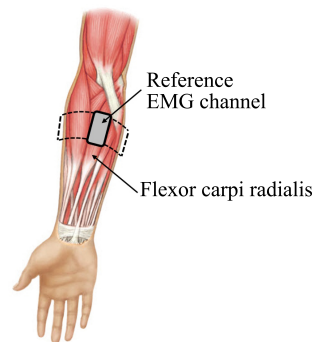
Fig. 5. Armband sensor mounting position using reference sensor.

were selected as experimental hand gestures. The numbers zero, two and six of the finger language were excluded because these were hand shapes similar to 'g, k (ㄱ)', 'n (ㄴ)', and 'a (ㅏ)'. EMG signals were measured for 15 seconds per selected finger language. To acquire a sufficient amount of feature vectors, the finger gestures were maintained for a long time. Because the muscle structure and EMG signals are measured differently in each subject, all sets of characteristic vectors are stored individually. The acquired feature vectors were divided such that 80% were used for training and 20% were used for classification evaluation. Signal pattern classification was performed using a five-fold cross validation. Each subject was informed about the purpose of the study and the experimental protocol, and provided written informed consent, which was approved by the Yonsei University Research Ethics Committee (1041849-201704-BM-018-01).

Algorithm evaluation was performed in two steps. First, we optimized an ensemble artificial neural network (E-ANN) classifier model. An optimal classifier structure should be selected, because the accuracy of the E-ANN-based algorithm depends on the number of multiple classifiers and the size of multiple data sets. Therefore, classification accuracies according to the number of ANN classifiers from 1 to 10 and multiple data set sizes from 50 to 1500 were compared. Second, we compared the performances between a general ANN classifier and an E-ANN classifier. All of the signal processing was performed using MATLAB (Version R2015a, Mathworks Inc., USA).

5. Statistical analysis

All values were expressed as means ± standard deviations. The statistical significance for each classification accuracy of E-ANN was analyzed using two-way (the number of multiple classifier and size of multiple data sets) repeated measures analysis of variance. The significance of differences between the general ANN and E-ANN groups was evaluated by paired t-tests. When a significant interaction between two main factors was observed, a Tukey's HSD post- hoc tests were carried out. All statistical analyses were performed using SPSS Statistics (Version 24, IBM, USA). A p-value less than 0.05 was considered statistically significant.

6. Results and discussion

6.1. E-ANN structure optimization

In the E-ANN-based finger language recognition (FLR) algorithm, the number of multiple classifiers is proportional to the amount of computation in the pattern classification process. An excessive amount
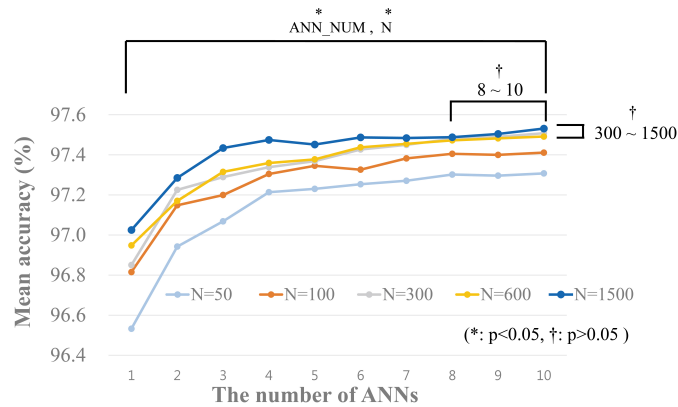
Fig. 6. Average recognition accuracy of finger language recognition classifiers according to the number of classifiers and size of training data.
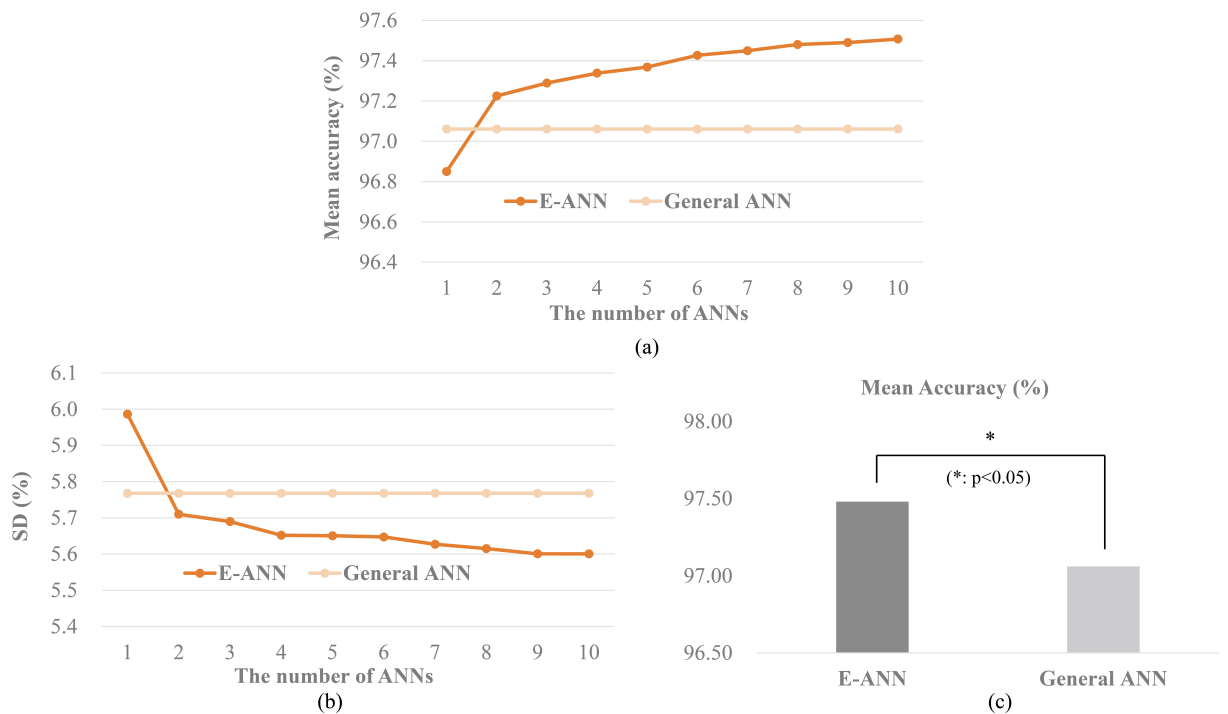

Fig. 7. Comparison between algorithms based on general ANN and E-ANN: (a) average classification accuracies for ANN and E-ANN ($N = 300$); (b) standard deviations of ANN and E-ANN ($N = 300$); (c) comparison of classification accuracies using optimal model E-ANN and general ANN.

of computation is a problem to be considered, because it is difficult to apply to a real-time system. Increasing the size of multiple data sets means acquiring additional data unnecessary for the classifier training, and it also requires unnecessary time to train the classifier. Therefore, the E-ANN structure needs to be applied to the algorithm after these two factors have been optimized.

Figure 6 shows a graph of the mean accuracies of the FLR according to the number of classifiers and the training data set size of the E-ANN-based algorithm. All of the trained classifiers showed accura-

Table 2
Comparison of studies on finger language recognition

Author	Sensor types	Recognition targets (the number of targets)	Classification accuracy
Singha and Das [4]	Video camera	American alphabet (24)	96.25%
Jiménez et al. [27]	Glove	Ecuadorian alphabet (30)	89.01%
Savur and Sahin [10]	EMG sensor	American alphabet (26)	91.73%
Our study	Armband-type EMG sensor	Korean finger language and numbers (38)	> 97.40%

cies of more than 96.53%, and the classification accuracies tended to increase as the number of ANN classifiers (ANN_NUM) and the size of the multiple data sets (N) increased. Statistical analysis showed that the accuracy was significantly affected by these two factors (ANN_NUM and N; $p < 0.05$). With ANN_NUM = 8 and $N = 300$, the accuracy of classification was more than 97.4%, and the accuracy did not increase further ($p > 0.05$). Therefore, the optimal structure of the E-ANN-based FLR algorithm was shown to be 8 ANN classifiers and 300 training data sets (500 ms).

6.2. Comparison of algorithms based on general ANN and E-ANN

Figure 7a and b show graphs comparing the mean classification accuracy and standard deviations of the general ANN and E-ANN algorithms ($N = 300$). The classification algorithm based on E-ANN showed that the average accuracy gradually increased and the standard deviation decreased as the number of classifiers increased. In contrast, an ANN-based algorithm, which is a single classifier, showed low accuracies and high standard deviations. Figure 7c is a graph comparing the classification accuracy of a general ANN and E-ANN with eight classifiers. The optimal structure of E-ANN showed significantly higher accuracy than the general ANN. The misclassification of the two algorithms may occur when the hand shapes are similar or mainly when external noises, such as power noise, ambient noise, motion artifact, etc., are measured in the process of calculating the feature vector. These two causes are inevitable problems that affect the error rate of the classifier. It was shown that this problem can be minimized by using a classification algorithm performed by an ensemble learning method.

6.3. Comparison with previous studies

Table 2 shows the results of previous studies on finger language or hand recognition technology, and the previous studies were compared with our present study. Although the number of finger languages in this study was larger than those in previous studies, the classification accuracy of the algorithm was higher. Since the classifier based on ensemble learning was applied, it was expected to display a more accurate and stable performance than other studies using a single classifier.

The armband type sensor that we developed is easy to wear and there are no spatial limitations, which are a big problem with camera-based systems. Furthermore, it is not necessary to put this device directly on a hand, and no interference in daily life is needed to recognize finger language. Furthermore, our system can be applied to portable devices such as cell phones and tablets.

7. Conclusion

Using an armband-type wearable multi-channel electromyography sensor, we developed a finger recognition system with a classification accuracy of more than 97.4%. A database containing EMG feature vectors (mean of absolute value, RMS, variance and waveform length) was obtained in the active

section, and a signal pattern classification algorithm based on an ensemble artificial neural network was developed. The classifier structure was optimized by comparing the effects of two factors (the number of classifiers and size of training data sets) on classification accuracy. We confirmed that the optimal performance was achieved with 8 ANN classifiers and size of 300 training data. In addition, we verified that the finger recognition algorithm of this study had a higher accuracy and more stable performance than a general ANN by comparing the average accuracies of the classifiers and their standard deviations. This armband-type wireless sensor system and finger language recognition algorithm have advantages in portability, convenience and marketability compared to previous research methods. Furthermore, the system developed in this study can be applied to various internet of things markets such as game interfaces, daily assistance systems, prostheses and machine control as well as finger language recognition technology.

Acknowledgments

Conflict of interest

None to report.

References

[1] World Health Organization. Multi-country assessment of national capacity to provide hearing care. Geneva: World Health Organization; 2013. Available from: http//www.who.int/pbd/publications/WHOReportHearingCare_Englishweb.pdf?ua=1&ua=1.

[2] Ong SC, Ranganath S. Automatic sign language analysis: A survey and the future beyond lexical meaning. IEEE Transactions on Pattern Analysis and Machine Intelligence 2005; 27(6): 873-891.

[3] Barberis D, Garazzino N, Prinetto, P, Tiotto, G, Savino, A, Shoaib, U, Ahmad, N. Language resources for computer assisted translation from italian to italian sign language of deaf people. Proceedings of Accessibility Reaching Everywhere AEGIS Workshop and International Conference; 2011; Nov 96-104; Brussels, Belgium.

[4] Singha J, Das K. Recognition of Indian sign language in live video. International Journal of Computer Applications 2013; 70(19): 17-22.

[5] Dong C, Leu MC, Yin Z. American sign language alphabet recognition using microsoft Kinect. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2015; Jun 45-52; Boston, MA.

[6] Weissmann J, Salomon R. Gesture recognition for virtual reality applications using data gloves and neural networks. Proceedings of International Joint Conference on Neural Networks (IJCNN); 1999; July 2043-2046; Washington, DC.

[7] Kadous MW. Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language. Proceedings of the Workshop on the Integration of Gesture in Language and Speech (WIGLS); 1996; Oct 165-174; Newark, DE.

[8] Fang G, Gao W, Zhao D. Large vocabulary sign language recognition based on fuzzy decision trees. Proceedings of IEEE Transactions on Systems, Man, and Cybernetics; 2004; Oct 305-314; Hague, Netherlands.

[9] Oskoei MA, Hu H. Myoelectric control systems – A survey. Biomedical Signal Processing and Control 2007; 2(4): 275-294.

[10] Savur C, Sahin F. Real-time American sign language recognition system by using surface EMG signal. Proceedings of IEEE 14th International Conference on Machine Learning and Applications; 2015 Dec 497-502; Miami, FL.

[11]   Zhang X, Chen X, Li Y, Lantz V, Wang K, Yang J. A framework for hand gesture recognition based on accelerometer and EMG sensors. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 2011; 41(6): 1064-1076.

[12]   Ahsan MR, Ibrahimy MI, Khalifa OO. Electromygraphy (EMG) signal based hand gesture recognition using artificial neural network (ANN). Proceedings of IEEE 4th International Conference on Mechatronics; 2011 May 1-6; Kuala Lumpur, Malaysia.

[13]   Wu J, Tian Z, Sun L, Estevez L, Jafari R. Real-time American sign language recognition using wrist-worn motion and surface EMG sensors. Proceedings of IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks; 2015 Jun 1-6; Cambridge, MA.

[14]   Li Y, Chen X, Zhang X, Wang K, Wang ZJ. A sign-component-based framework for Chinese sign language recognition using accelerometer and sEMG data. IEEE Transactions on Biomedical Engineering 2012; 59(10): 2695-2704.

[15]   Naik GR, Kumar DK. Identification of hand and finger movements using multi run ICA of surface electromyogram. Journal of Medical Systems 2012; 36(2): 841-851.

[16]   Chu JU, Moon I, Lee YJ, Kim SK, Mun MS. A supervised feature-projection-based real-time EMG pattern recognition for multifunction myoelectric hand control. IEEE/ASME Transactions on Mechatronics 2007; 12(3): 282-290.

[17]   Hiraiwa A, Shimohara K, Tokunaga Y. EMG pattern analysis and classification by neural network. Proceedings of IEEE International Conference on Systems, Man and Cybernetics; 1989; Nov 1113-1115; Cambridge, MA.

[18]   Naik GR, Kumar DK, Singh VP, Palaniswami M. Hand gestures for HCI using ICA of EMG. The HCSNet workshop on Use of vision in human-computer interaction; 2006; Nov 67-72; Canberra, Australia.

[19]   Englehart K, Hudgins B, Stevenson M, Parker PA. Dynamic feedforward neural network for subset classification of myoelectric signal patterns. Proceedings of IEEE 17th Annual Conference Engineering in Medicine and Biology Society; 1995 Sep 819-820; Montreal, Canada.

[20]   Kelly MF, Parker PA, Scott RN. The application of neural networks to myoelectric signal analysis: A preliminary study. IEEE Transactions on Biomedical Engineering 1990; 37(3): 221-230.

[21]   Yu L, Wang S, Lai KK. Credit risk assessment with a multistage neural network ensemble learning approach. Expert Systems with Applications 2008; 34(2): 1434-1444.

[22]   Solnik S, DeVita P, Rider P, Long B, Hortobágyi T. Teager-Kaiser operator improves the accuracy of EMG onset detection independent of signal-to-noise ratio. Acta Bioengineering and Biomechanics 2008; 10(2): 65-68.

[23]   Englehart K, Hudgins B. A robust, real time control scheme for multifunction myoelectric control. IEEE Transactions on Biomedical Engineering 2003; 50(7): 848-854.

[24]   Englehart K, Hudgins B, Parker PA, Stevenson M. Classification of the myoelectric signal using time-frequency based representations. Medical Engineering and Physics 1999; 21(6): 431-438.

[25]   Møller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 1993; 6(4): 525-533.

[26]   Dietterich TG. Ensemble methods in machine learning. Proceedings of International Workshop on Multiple Classifier Systems; 2000; June 1-15; Cagliari, Italy.

[27]   Jiménez LAE, Benalcázar ME, Sotomayor N. Gesture recognition and machine learning applied to sign language translation. Proceedings of VII Latin American Congress on Biomedical Engineering CLAIB 2016; 2016 Oct 233-236; Santander, Colombia.