Towards EMG Based Gesture Recognition for Indian Sign Language Interpretation Using Artificial Neural Networks

Abhiroop Kaginalkar[(✉)] and Anita Agrawal

Birla Institute of Technology and Science – Pilani, K.K Birla Goa Campus, Goa, India
akaginalkar@gmail.com, aagrawal@goa.bits-pilani.ac.in

Abstract. There are several techniques of data measurement for gesture recognition, with applications ranging from prosthetic or autonomous control to human-computer interfacing. Most of the typical techniques depend on image processing, and might face portability hurdles. This paper discusses a method to classify gestures based on the surface EMG (sEMG) readings, thereby allowing user portability. These sEMG readings acquired from the upper forearm provide a direction towards gesture recognition for Indian Sign Language (ISL) interpretation. An Artificial Neural Network (ANN) based on the Scaled Conjugate Gradient (SCG) assisted learning is used to process the data and classify gestures with an accuracy of 97.5 %. The training involved 120 samples corresponding to four distinct wrist gestures. Additionally, the foundations for user-independent adaptability have been laid in this paper.

Keywords: Human-computer interaction · Biomedical electronics · Artificial neural networks · Sign language interpretation · EMG

1 Introduction

Indian Sign Language (ISL) comprises of a common ground for a variety of different dialects specific to various regions over India. It comprises of multiple hand gestures, coupled with simple or complex motions. A single word might not necessarily be gesticulated by one distinct motion. In the light of such a complexity, it is highly imperative to develop an interpretation system that will efficiently process all the nuances in the gestures, and extract the meaning with minimal computation and inconvenience.

Most of the existing approaches to this issue are based on image processing [1] or wearable flex-sensing technologies. Rajam et al. [2] make use of an edge detection algorithm to convert the wrist images into a binary classification of finger positions by relying on Euclidian distances with respect to a fixed base-point. Adithya et al. [3] also depend on Euclidian distances and the Fourier descriptors of their projection vectors, making the system robust with respect to noise. Agrawal et al. [4] implement a feature extraction approach that uses a Support Vector Machine for classification based on the extracted data.

The major issue associated with an image based gesture recognition system is the portability limitation. It is very tedious and inconvenient for the user to have an image capture system aptly set in place. The relative distances between these image capturing devices and the hand performing the gestures affect the clarity and introduce inconsistencies. Most of such measurements require a good contrasting background to detect the hand and its shape, which is something that might not always hold. Approaches that depend on edge detection might face trouble when gestures consist of overlapping hand or figure scenarios. In a lab-setting, many of the mentioned approaches, quite effectively, fulfill the gesture recognition task, but fail to take into account the practicality of the usage. This paper proposes an approach keeping the usability and convenience as the primary motive. Some novel EMG based technologies have been proposed [5, 6] that attempt to bypass the above mentioned issue with existing gesture recognition solutions. Some of the approaches make use of k-NN and Bayes classifiers [7] and Statistical Feature Extraction from EMG waveforms [8]. This paper proposes a Scaled Conjugate Gradient (SCG) based approach to the EMG based gesture recognition problem specific to the ISL. The assisted learning ANN is used to distinguish among 4 distinct wrist gestures from the underlying noise. The intention of this research was to recognize an ANN based gesture recognition approach that can be applied to Indian Sign Language interpretation and in other hand-motion based control scenarios. This methodology of data acquisition and classification, coupled with sensor-based hand motion/orientation detecting algorithms, will pave the way for a practical solution to the sign language interpretation problem.

2   Methodology

2.1   Surface EMG Interfacing

The electrodes used for EMG signal acquisition were standard Ag/AgCl button electrodes connected with a multi-stranded shielded cable. A single channel was used to measure the myo-electrical activity on the surface of the upper-arm. After a couple of initial trials an electrode placement directly over the Flexor Carpi Radialis was chosen (owing to best relative voltage readings and minimal noise encountered). The distance between the two measurement electrodes was maintained at 5.5 cms. The reference electrode was placed over the elbow bone so as to provide minimum interference.

2.2   Digital Signal Conditioning

Data received (at 1 kHz) via the Arduino was processed sequentially through a scalar Kalman Filter (sKF) with Q and R values empirically tuned to 0.0001 and 0.01 respectively. An algorithm was constructed to select a temporal region of activity. This reduced the computational and storage requirements by restricting the analysis to an activity window. A particular activity threshold in the recorded voltage levels was recognized during an initial training/setup period. This threshold was used as a trigger

to identify the window of activity, and subsequently, isolate it for further analysis. The window captured any voltage fluctuations associated with the motion within a time-span of 3.5 s.

3   Gesture Recognition

3.1   Artificial Neural Network

AANs mimic complex arithmetic equations, which, when fed with inputs and desired outputs (in the case of assisted learning) adjust the free parameters (weights) so as to reduce the net error. For this study an ANN based on a SCG assisted learning approach has been implemented. This type of learning is more robust and efficient when it comes to pattern recognition [9], as opposed to the standard steepest gradient and conjugate gradient methods. This particular learning technique relies on the steepest gradient along consecutively conjugate vectors, eliminating the directional redundancies and decreasing the number of iterations required to converge. The SCG approach implies an optimization problem that utilizes mutually orthogonal gradients and conjugate directions for minimizing the cost function, which is the error function. The SCG optimization algorithm works with the second order approximations and avoids the line search per learning iteration by using the Levenberg-Marquardt approach to scale the step size [9]. The ANN implemented had 350 inputs that were fed with the EMG waveform transformed to the frequency domain. The output of this ANN had 4 outputs each associated with a wrist motion – fist clench, wrist flick, double wrist clench and no operation. The output layer of the ANN was implemented with trans-sigmoid activation functions. An algorithm then selected the output with the highest confidence score (ranging from 0 to 1).

3.2   Training

In this particular experiment, we used data recorded from the right forearm measured from multiple users. Users were trained to perform specific wrist gestures in a time-dependent fashion based on the visual cues provided by the graphical user interface. The training used a semi-batch approach, wherein, the data collected during each activity window was processed and fed into the ANN for training. The training was planned in two phases:

3.2.1   Phase 1
The first where, a background noise reading was measured (each time a new user wore the electrodes) and smoothened using a cubic spline function approximation to obtain an upper threshold (Highest_Noise_Voltage) of noise. This was then used to detect the activity window in contrast to the underlying noise. The detection algorithm was heuristically programmed to detect any voltage fluctuation above 1.2 times the Highest_Noise_Voltage. The user was provided with a visual cue to maintain the forearm in a relaxed position to allow the accurate recording of the inherent noise in the

measurement setup. The noise readings were recorded over a period of 5 s before the beginning of each new trial.

3.2.2 Phase 2

The second phase involved training the ANN based on SCG supervised learning. A sample set of 120 distinct wrist motions (among the ones mentioned above) were fed into the ANN with a template of the expected outputs. After a number of trials the best performance (for 4 distinct wrist gestures) in terms of the fastest convergence was obtained for an ANN with 10 hidden layers. The training was a batch process based on pre-measured data.
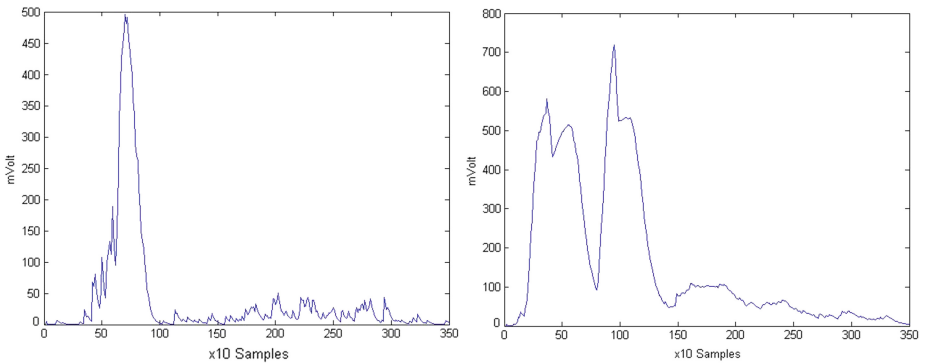
Fig. 1. The waveform on the left represents a single wrist clench and the waveform on the right represents a double wrist clench.

4   Results

4.1   sEMG Waveforms

The surface EMG waveforms obtained after being processed through the scalar Kalman filter are shown in Fig. 1. Each activity window was successfully identified and the measurement noise was successfully filtered so as to obtain visually distinct waveforms associated with each wrist motion. On the left is the resultant waveform when the user quickly clenched and released the fist of the right hand. On the right is the resultant waveform when the user repeated the clench-release cycle twice in quick succession. Visually the waveforms are distinguishable, and the ANN was programmed to recognize this difference. We have included the detection of a no-operation waveform in the ANN. The primary reason for this inclusion is to allow the ANN to effectively discern between the underlying noise readings and a no-operation period (either between two signs or when the arm is relaxed). By doing so, we have managed to reduce the effect of motion artifacts, which some-times introduce spikes within an order of magnitude of the actual EMG voltages, by a considerable extent.

4.2   Performance Parameters

The performance was quantified based on the Confusion Matrices for each trial and test pair. The training phase was stabilized with the lowest gradient measure of 0.005316 in 61 epochs of the 120 input set (Fig. 2).
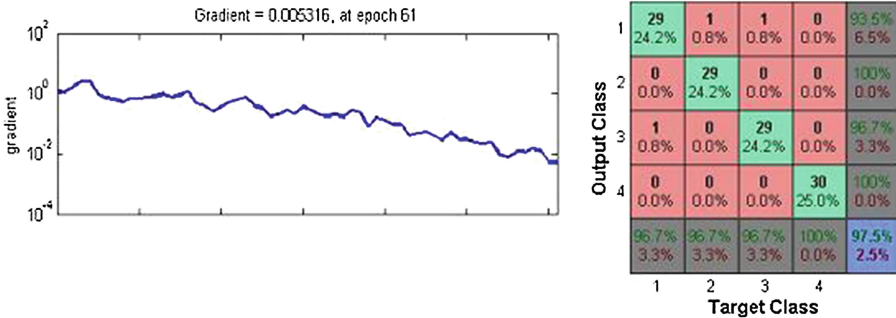
Fig. 2. Left: Gradient minimization during the training phase. Right: Confusion matrix for 4 distinct gestures.

The confusion matrix provides a graphical 'scoreboard' that depicts the input class, the expected output class and the variance in the mapping. Once the gradient reading stabilized to the minimum acceptable value, the training phase was stopped and the test phase ensued. The test phase involved presenting 120 input samples from the collected dataset in a random manner. The results were then plotted and analyzed based on the confusion matrix. The Fig. 2 depicts the performance of one of the trials conducted:

The following are the observations based on this matrix:

- The first, second and third classes correspond to the single wrist clench, double wrist clench and wrist flick actions respectively. Out of the 30 samples for each class, one of the samples led to an erroneous prediction.
- The fourth class corresponding to the no-operation input resulted in favorable results. This was expected as the no-operation waveform was significantly different from the others (more passive), leading to a better classification.
- The overall performance for 120 randomly ordered samples led to an accuracy of 97.5 %.
- The predictions can be made more accurate by increasing the complexity of the ANN, but compromising on the computation and trial time. Hence, an expected tradeoff exists between the complexity, computation time, storage and accuracy. Each of these can be adjusted depending on the application and acceptable ranges.

5   Conclusion

Through this research we have developed a fundamental base for using SCG based ANNs in gesture recognition for the interpretation of sign language. We have managed to distinguish between 4 gesture types, thereby, establishing a proof-of-concept methodology. The next step will be to establish an extensive database of ISL signs and construct a corresponding ANN for the same. This will involve scaling the existing network by increasing the layer count and input/output parameters. The data collected can be used with other motion sensors to design an integrated sign language recognition system. The ANN proposed in this paper can be replicated using a microprocessor and, thereby, be used in a portable sign language interpretation solution.

References

1. Ong, S.C.W., Ranganath, S.: Automatic sign language analysis: a survey and the future beyond lexical meaning. IEEE Trans. Pattern Anal. Mach. Intell. **27**(6), 873–891 (2005)
2. Rajam, P.S., Balakrishnan, G.: Real time indian sign language recognition system to aid deaf-dumb people. In: 2011 IEEE 13th International Conference on Communication Technology (ICCT), pp. 737–742 (2011)
3. Adithya, V., Vinod, P.R., Gopalakrishnan, U.: Artificial neural network based method for Indian sign language recognition. In: 2013 IEEE Conference on Information and Communication Technologies (ICT), vol., no., pp. 1080–1085 (2013)
4. Agrawal, S.C., Jalal, A.S., Bhatnagar, C.: Recognition of Indian Sign Language using feature fusion. In: 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI), pp. 1−5 (2012)
5. Mitra, S., Acharya, T.: Gesture recognition: a survey. IEEE Trans. Syst. Man Cybern. Part C-Appl. Rev. **37**(3), 311–324 (2007)
6. Yun, L., Chen, X., et. al.: Automatic recognition of sign language subwords based on portable accelerometer and EMG sensors. In: Proceedings of the International Conf. on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, 17 (2010)
7. Kim, J., Mastnik, S., Andre, E.: EMG-based hand gesture recognition for realtime biosignal interfacing. In: Proceedings of the 13th International Conference on Intelligent User Interfaces, pp. 30–39 (2008)
8. Shroffe, E.H., Manimegalai, P.: Hand gesture recognition based on EMG signals using ANN. Int. J. Comput. Appl. **2**(3), 31–39 (2013)
9. Moller, M.F.: A scaled conjugate gradient algorithm for fast supervised learning. J. Neural Netw. **6**(4), 525–533 (1993)