This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSII.2016.2635674, IEEE Transactions on Circuits and Systems II: Express Briefs

The Virtual Trackpad: an Electromyography-based, Wireless, Real-time, Low-Power, Embedded Hand Gesture Recognition System using an Event-driven Artificial Neural Network

Xilin Liu, Jacob Sacks, Milin Zhang, Andrew G. Richardson, Timothy H. Lucas, and Jan Van der Spiegel

***Abstract*—This work presents a wireless, low-power embedded system that recognizes hand gestures by decoding surface electromyography (EMG) signals. Ten hand gestures used on commercial trackpads, including pinch, stretch, swipe left, swipe right, scroll up, scroll down, single click, double click, pat, and ok, can be recognized in real time. Features from four differential EMG channels are extracted in multiple time windows. Unlike traditional data segmentation methods, an event-driven method is proposed, with the gesture event detected in the hardware. Feature extraction is triggered only when an event is detected, minimizing computation, memory, and system power. A time-delayed artificial neural network (ANN) is used to predict the gesture from the transient EMG features instead of traditional steady-state features. The ANN is implemented in the microcontroller with a processing time less than 0.2ms. The detection results are sent wirelessly to a computer. The device weights 15.2g. A 4.6g battery supports up to 40 hours continuous operation. To our knowledge, this work shows the first real-time, embedded hand gesture recognition system using only transient EMG signals. Experiments with four subjects show that the device can achieve a recognition of 10 gestures with an average accuracy of 94%.**

***Index Terms*—EMG, gesture recognition, ANN, event-driven, low power, real-time**

I. Introduction

Hand movements and gestures are used as natural means of human communication. The invention of trackpads (or touchpads) expands this communication to machines [1]. Large and multi-touchpoint trackpads have become a typical feature of laptops, as well as some desktops. However, users are still constrained by the size and location of the trackpad, and the gestures are limited by the number of touch points. In this work, a virtual trackpad is proposed by decoding surface electromyography (EMG) signal. The advantages of the virtual trackpad are three-fold. First, there are no constraints on where the gestures are performed. For example, users can use gestures to control the play of a video, tune the music volume, or scroll a webpage without being confined to a desk. Second, there are no limits on the number of touch points. The virtual trackpad allows users to use true 3D gestures. Third, the response is nearly instantaneous. Since EMG signals precede actual movement, the recognition system can be even trained to predict the hand movement, resulting in a fluent user experience.

Research on EMG-based pattern recognition and control have been reported in the literature. The EMG-based gesture recognition algorithm can be summarised as classifier-based algorithms [2–5], artificial neural network (ANN)-based algorithms [6, 7], fuzzy logic-based algorithms [8], and probabilistic model-based algorithms [9]. Portable, wireless EMG acquisition sensors have also been reported recently [2] and are available on the market [10]. However, most existing gesture recognition algorithms were running in computers. The data streaming of multiple channel EMG signal places a burden in the wireless link at the price of power consumption, and also introduces delay in the processing. Dedicated algorithm programs running on computer increases the requirements in the computer, thus limiting the pervasive use of these devices. A low-power hand gesture and finger recognition device with the ability to process the detection in local hardware is highly desirable. Neural network is one of the well studied algorithms that is suitable for an implementation in portable devices [11, 12]. It has been used for various applications in literature, i.e. wireless communication [13], motion sensing [14].

This paper presents a neural network based highly integrated design with all recognition tasks included in a battery-powered device. This work is the first wireless virtual trackpad based on EMG, to the best of our knowledge. The proposed design significantly reduces the latency, wireless data rate, and the software dependency on the receiver end. Multiple transient-state EMGs are used for decoding instead of steady-state signals in order to achieve minimum latency. Compared with steady state signal, the difficulty of classifying transient state signals lies in its nature of short duration and non-repeatability. Steady state EMG proves to be easier for classification, especially when majority voting is used [15]. This work is also the first real-time gesture recognition device based on purely multiple state instantaneous EMG signals, to the best of our knowledge.

The work was done as part the National Science Foundation grants 1404041.

Xilin Liu, Jan Van der Spiegel are with the Department of Electrical and Systems Engineering (ESE), University of Pennsylvania, Philadelphia, PA, 19104 USA.

Milin Zhang is with the Department of Electronic Engineering, Tsinghua University, Beijing, China, 100084. Corresponding author e-mail: zhangmilin@seas.upenn.edu.

Jacob Sacks is with the Department of Biomedical Engineering, University of Texas at Austin, Austin, TX, USA.

Andrew G. Richardson and Timothy H. Lucas are with the Department of Neurosurgery, University of Pennsylvania, Philadelphia, PA, 19104 USA.

Digital Object Identifier

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSII.2016.2635674, IEEE Transactions on Circuits and Systems II: Express Briefs

The paper is organized as follows. Section II presents the system design. Section III presents the data acquisition and the algorithm. The measurement results and discussion are given in section IV. Section V concludes the paper.

II. The Virtual Trackpad System Design

The system architecture is illustrated in Fig. 1. The system consists of analog front-end circuits, a microcontroller, a wireless module, and power management circuits. Conductive solid gel surface electrodes (Rochester Electro-Medical, Inc.) are used during the experiments. The four differential electrode pairs are placed on selected muscles: extensor digitorum, flexor digitorum superficialis, flexor carpi ulnaris, and extensor carpi ulnaris, respectively. The ground electrode is placed on the elbow.

The analog front-end circuits consist of an instrumentation amplifier (AD623, Analog Devices, Inc.), and several low-power amplifiers for the implementation of the filters. The instrumentation amplifier is capacitor coupled (CCIA) to remove the DC offset from the surface electrode. Large biasing resistors (200MΩ) are used to keep the input impedance high. A DC servo-loop is used to further remove the low frequency components in the signal under 10Hz, and set the output in the common mode voltage [16]. A second-order low pass filter is used to condition the signal with a cut-off frequency of 500Hz. The common mode voltage of the amplifier is sent back to regulate the human body ground after a phase shift.

A 32-bit microcontroller with ARM Cortex-M4 core (TM4C123GH6PM, Texas Instruments, Inc.) is used as the central processor [16]. The 4-channel amplified neural signal is sampled by the integrated 12-bit ADC at a rate of 1.6 kHz. The gesture detection algorithm is performed in the microcontroller including event detection, feature extraction, and the gesture recognition algorithm. The detection result is sent wirelessly to a computer host through a low-power 2.4GHz wireless module pair (nRF24L01+, Nordic Semiconductor). The system is powered by a 3.7V lithium-ion battery.

III. The Gesture Recognition Algorithm and the Hardware Implementation

A. Feature Extraction

This work focuses on ten hand gestures used commonly on a trackpad: pinch, stretch, swipe left, swipe right, scroll up, scroll down, single click, double click, pat, and ok. Fig. 2 shows the gestures with corresponding signals in the four EMG channels.

The feature extraction process is illustrated in Fig. 3, with the active time for the major hardware modules highlighted. At the start of the recording, the mean signal value $\overline{V}[s]$ (common mode voltage) and standard deviations $\sigma[s]$ of each channel s are calculated within a 0.1s period. A threshold $V_{th}[s] = 5\sigma[s]$ is set in each channel for gesture event detection. Once the thresholds are set, digital comparators are used in each channel for event detection, and the microcontroller's CPU is put in the deep sleep mode to save power. Once an event is detected by the comparator in one channel, an interrupt is generated to wake up the CPU, and starts the feature extraction procedure. Among all the time-, frequency-, and time-frequency domain features, mean absolute value (MAV) is chosen in this work since it provides good performance and low computational cost [6]. The MAV feature $f_x[s]$ of each channel is extracted and normalized across the four channels in a pre-defined window. The procedure can be expressed as:

$$M[s,k] = |V[s,k] - \overline{V[s]}| \tag{1}$$

$$X[s] = \sum_{k=n-T+1}^{n} \frac{M[s,k]}{\overline{M[k]}} \tag{2}$$

$$f_x[s] = \frac{X[s] - (X_{max} - X_{min})/2}{(X_{max} - X_{min})/2} \tag{3}$$

where k is the time point, T is the window size, s is the channel number, N_s is the total number of channels, V is the signal amplitude, and $\overline{V[s]}$ is the mean value for each channel. If the resulting feature $f_x[s]$ in the event channel is lower than a pre-defined threshold, the event is taken as a false alarm, and the data will be abandoned. If the result feature is higher than the threshold, a second feature for each channel will be calculated. The microcontroller's most peripheral module clocks are turned off. The wake up time for the CPU is much shorter than one sample period.

The extracted features for the ten gestures are scatter plotted in Fig. 4. Experimental results show that these two features give reliable information for the classification algorithms.

B. Event Driven Artificial Neural Network

Using the extracted features, several popular recognition algorithms were evaluated: linear discriminant analysis (LDA), native Bayes classifier, k-nearest neighbours (kNN), support vector machine (SVM), and multi-layer, feedforward artificial neural network (ANN). There were convergence difficulties while using LDA with these data. SVM's computational cost increased significantly as the class number increased. The cross-validation performance of the remaining classifiers are shown in Fig. 5(a). The time delayed ANN showed superior performance in this task.

The non-linear transfer function and number of neurons of the ANN hidden layers were further evaluated. The results are shown in Fig. 5(b). Different non-linear transfer functions showed similar performance in this task. The positive saturating linear (satlin) function was chosen given its superior computational efficiency. Twelve neurons were chosen as a tradeoff between accuracy and computation cost.

The block diagram of the overall algorithm is shown in Fig. 6(a). An example of the network's computational implementation is illustrated in Fig. 6(b). The algorithm consists of four blocks: an event detection unit, a time-delayed feature extraction unit, a three-layer feedforward ANN, and an output judgement unit. The output target vector is set as

$$Y_i = \begin{cases} 1 & \text{if } i = k \\ 0 & elsewise \end{cases} \tag{4}$$

where i is the ith element of the output vector, and k indicates the target gesture. A winner takes all method, which picks the highest probability gesture, is used to decide the final output.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSII.2016.2635674, IEEE Transactions on Circuits and Systems II: Express Briefs

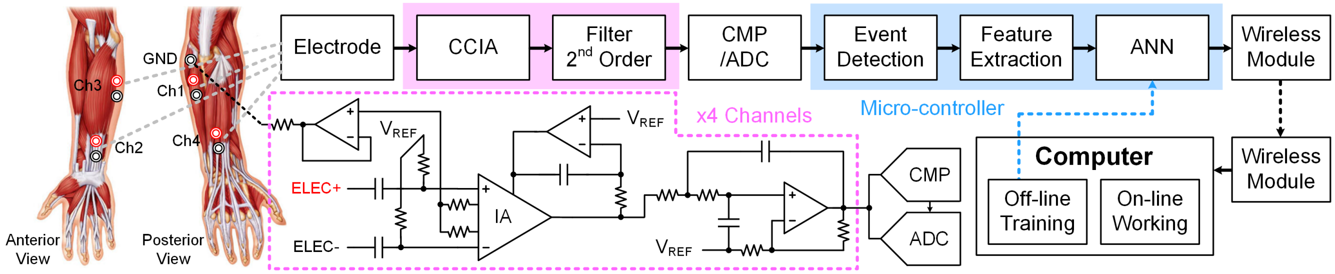


Fig. 1. Overview of the system. The placement of the four channel surface EMG electrodes is shown in the left. The block diagram of the system is in the middle, where the analog front-end circuits boxed. The training of the algorithm is performed in the computer, while during the working mode, only gesture indicators are sent to computer via a wireless module.

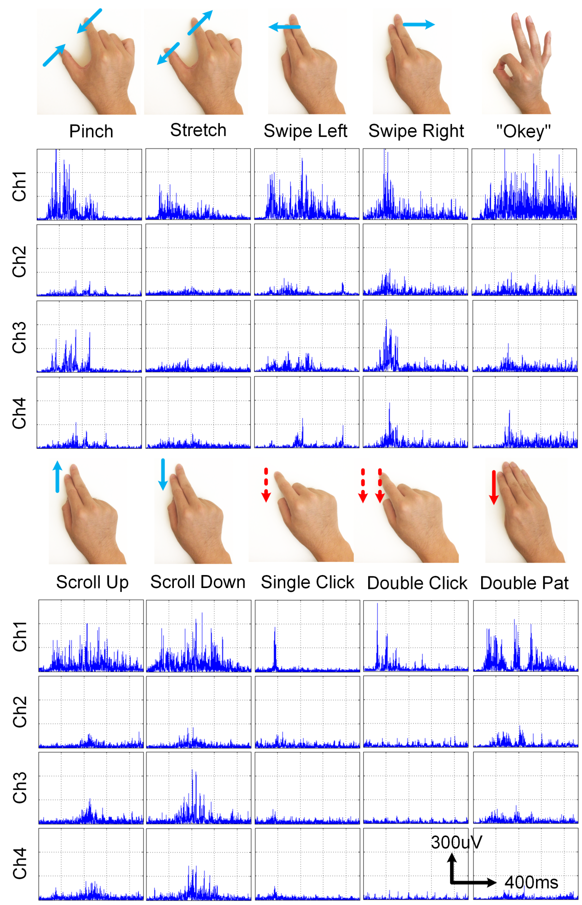


Fig. 2. EMG signal patterns in different gestures. The rectified EMG signal in the four channels are shown.

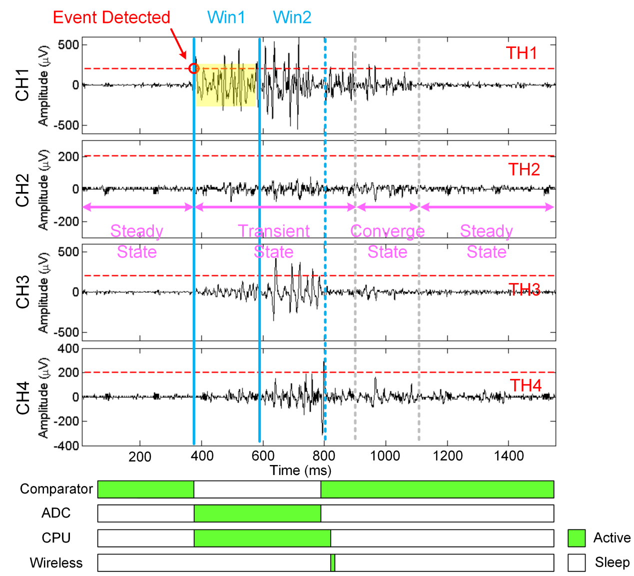


Fig. 3. Illustration of the event-driven feature extraction. The green bar shows the active time of the different modules in the system.

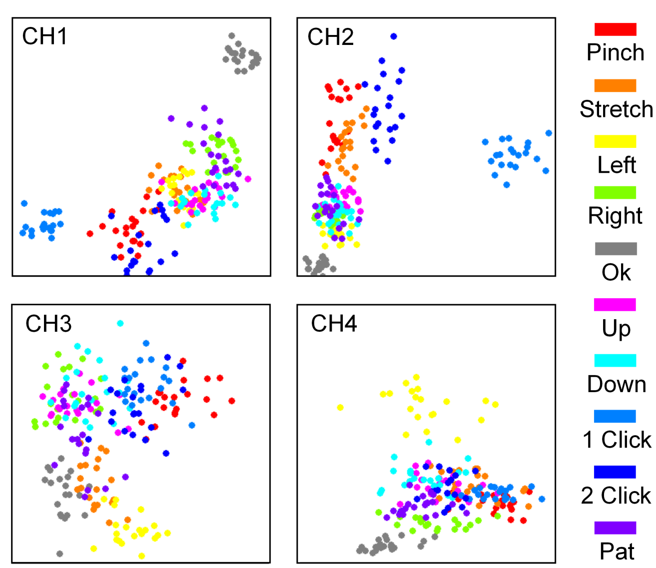


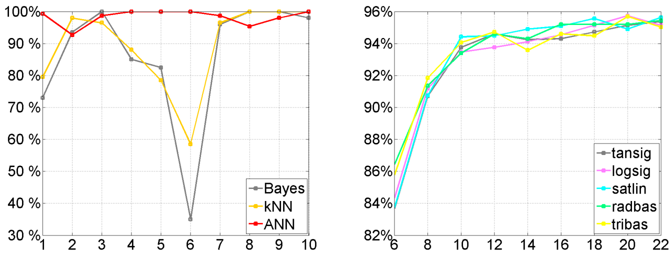
Fig. 4. Scatter plot of the normalized features in each channels. X-, Y-axis presents the features in window 1 and 2, respectively.

The ANN was trained off-line through supervised learning by using the hardware extracted features. The subjects were requested to perform each gesture thirty times as the training data set. There was no synchronization signal in the collecting phase, and the hardware automatically detected the gesture event. After the features for each gesture were collected, the data was randomly divided into 5 folds for cross-validation. The optimal weighting matric and biasing vectors were solved using the back-propagation algorithm [6]. The validation error was less than 5%. The parameters were then programmed in the microcontroller to perform real-time recognition. Single precision floating point format was used, according to IEEE Standard 754.

IV. Experiment and Discussion

The fabricated device is shown in Fig. 7(a). The dimension of the device is 34mm×25mm. A wristband was designed to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSII.2016.2635674, IEEE Transactions on Circuits and Systems II: Express Briefs



(a) (b)

Fig. 5. (a) Comparison of different algorithms. X-, Y-axis presents the gesture tags and the overall recognition accuracy in a cross validation, respectively. (b) Comparison of different non-linear transfer functions and number of neurons in the hidden layer. X-, Y-axis presents the number of neurons, and the overall recognition accuracy, respectively.

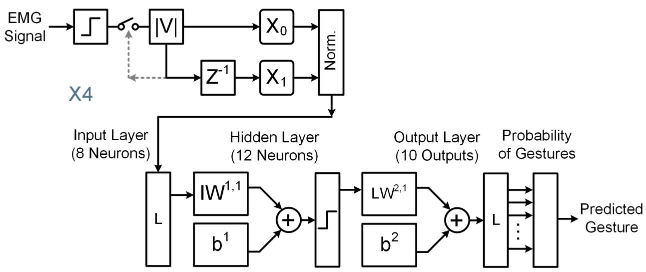


Fig. 6. Block diagram of the event-driven gesture recognition algorithm. The threshold for event detection is trained online. The weighting matrices are trained off-line, and programmed into the microcontroller for real-time gesture recognition.

mount the device easily to the subject, and electrodes were connected to the board using flexible wires.

A few bench-top tests were first performed to verify the function of the system. The input-referred noise of the analog front-end was 3.64μVrms, and the thermal noise with electrodes was 4.45μVrms. The measured bandwidth of the front-end is 9.57Hz to 511Hz, with a mid-band gain of 66dB. The wireless module shows a BER lower than 10^{-6} in a distance of 5m using custom protocol and 5 times retry. The microcontroller runs at the system clock of 80MHz, and built-in floating point unit (FPU) is used to perform the network computation. The system processes the feedforward neural network in less than 0.2ms per event. The system is put in sleep mode when no event is detected to save power. The measured power consumption in different modes are shown in Table I.

(a) (b)

Fig. 7. (a) Photograph of the device, with a dimension of 34mm×25mm. (b) Photograph of the experiment setup.

TABLE I
MEASURED POWER CONSUMPTION OF THE SYSTEM

	IA	Filter	MCU	RF	Total
Event Detection	6mW	0.6mW	6mW	0.2mW	12.8mW
Processing			69mW		75.8mW
Wireless			60mW	34mW	100.6mW

The system have been tested with four subjects (two males, two females). The ANN was trained for each individual subject. During the experiment, the subjects were asked to perform the gestures with the arm resting comfortably. A computer program displayed one of the ten pre-defined gestures on the screen in a random sequence, and the subject performed the gesture accordingly. Notice that the detection algorithm was running real-time in the microcontroller, and there was no synchronization between the computer and the microcontroller. The detected gesture was sent to the computer wirelessly. The feature window size was chosen between 200ms$\sim$250ms, optimized for each subject, mainly because each subject performs the gestures in slightly different speed.

The recognition accuracy metric was defined as:

$$ACC = \frac{T_P}{T_P + F_N + F_P} \tag{5}$$

where T_P was the number of correct recognitions, F_P was the number of wrong recognitions, and F_N was the number of missed gestures. The test results are shown in Fig. 8. Robust recognition was achieved. The overall recognition accuracy of the ten gestures for each subject was 93.3%, 93.2%, 98.3%, and 92.0%. The recognition for subject one was less accurate in discriminating scrolling up or down, that was because when the subject wanted to scroll up, he moved his fingers down first, and the event-driven system catched the down gesture. Similar situation happens for subject two in discriminating swiping left or right. This could be improved in the future work by including more time-delayed features, so the system can capture the complete movement as up-down-up, or left-right-left.

Majority vote has been widely used in steady state-based EMG classification algorithms, which significantly improves the detection accuracy if the subjects are required to hold their gestures for a long duration [7, 17]. However, natural hand gestures never pause for more than 0.5 second, and most gestures are sequences of muscle movements. In this work, we explore the use of EMG features in a time sequence for decoding transient gesture. For most gestures used, the detection results come out before the EMG signal is converged, which gives the virtual trackpad an advantage over the traditional physical trackpads.

A comparison of the proposed work with recent publications is given in Table II. Notice that as the gesture signal duration reduces, the detection challenge increases significantly. Computers have much stronger computation ability and memory compared to embedded system, however, the transmitting of the raw signal causes significant latency.

Moreoever, the ten gestures chosen in this work are limited to hand and finger movements in a resting arm to fit the requirement of the trackpad application. The subtle differences

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSII.2016.2635674, IEEE Transactions on Circuits and Systems II: Express Briefs

TABLE II
COMPARISON WITH sEMG BASED GESTURE RECOGNITION WORK

	Kondo *et al*	Chen *et al*	Hasan *et al*	Rossi *et al*	Savur *et al*	Naik *et al*	Benatti *et al*	This work
Year & Publication	2010 [6]	2013 [2]	2014 [7]	2015 [4]	2015 [5]	2015 [17]	2015 [3]	2015
# of EMG Ch	5	4	1	4	8	11	8	4
# of Gestures	5	10	4	6	26	11	7	10
Gesture set	Typing	Chinese Num	Hand gesture	Hand gesture	Hand Sign	Finger	Finger	Trackpad
Gesture state	transient+ converge	steady (hold 0.8s)	steady (hold 30s)	steady	steady (hold 2s)	steady (hold 3-4s)	steady (hold 3s)	transient
Recognition algo.	ANN	SVM	ANN	HMM/SVM	SVM	LDA	SVM	ANN
Wireless	-	Yes (sensor)	-	-	-	-	-	Yes
On-line	Yes	Yes	-	Yes	Yes	-	Yes	Yes
Algo. implement	Computer	Computer	Computer	Embedded	Computer	Computer	Embedded	Embedded
Detect accuracy	85%	>90%	74%	92%	82%	>95%	90%	94%

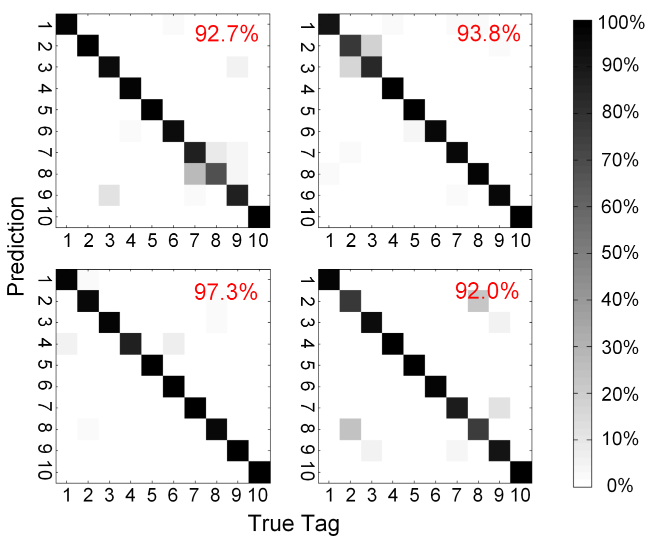


Fig. 8. Confusion matrices of the testing results of the four subjects, the numbers in cells are in percentage. The tags stands for: 1-pinch, 2-swipe left, 3-swipe right, 4-stretch, 5-single click, 6-double click, 7-scroll up, 8-scroll down, 9-pat, 10-ok. The average overall recognition accuracy is ∼94%.

in these gestures also pose a challenge for the detection task compared with other works.

The experiment demonstrated that the proposed system is capable of diverse gesture recognition in low cost, low power hardware with fast response. The testing results suggest that more gestures can be included in future work. Another extension in the future work will be to include more time delayed features, also in different window size. Features with shorter window size will improve the time resolution, and will capture more complex gesture. However, too small window size might increase the variation which leads to convergency problem in training. A careful trade-off needs to be resolved, and might need to be optimized for individual user.

V. CONCLUSION

In this paper, we presented the design of a low-power, wireless, real-time, embedded virtual trackpad system based on four channel sEMG. An energy-efficient, event-driven, time-delayed ANN was implemented in the embedded system, and achieved robust, low-latency, and high-accuracy gesture recognition. 10 gestures commonly used on a commercial trackpad were decoded from the transient signal instead of steady state. The device has been tested with four subjects, achieving an average recognition accuracy of 94%. The system is compact and lightweight, and supports a battery life up to 40 hours in continuous operation. The design achieves the state-of-the-art recognition accuracy for sEMG decoding, with more challenging transient gestures. The proposed method shows promising applications in rehabilitation and real-time human-computer interface systems.

REFERENCES

[1] Thryft, Ann R. “More Than a Mouse,” Computer Product Development, EBN Extra, November 14, 1994, pp. E16-E20
[2] X. Chen, *et al*, “Pattern recognition of number gestures based on a wireless surface EMG system,” *Biomed. Signal Process. Control*, vol. 8, no. 2, pp. 184-192, Mar. 2013.
[3] S. Benatti, *et al*, “A Versatile Embedded Platform for EMG Acquisition and Gesture Recognition,” *TBioCAS*, vol. 9, no. 5, pp. 620-630, 2015.
[4] M. Rossi, *et al*, “Hybrid EMG classifier based on HMM and SVM for hand gesture recognition in prosthetics,” *Int. Conf. Ind. Technol.*, 2015.
[5] C. Savur, *et al*, “American Sign Language Reconition System by Using Surface EMG Signal,” *Int. Mach. Learn. Appl. Conf.* , 2015.
[6] Kondo, G., *et al*, “Classification of Individual Finger Motions Hybridizing Electromyogram in Transient and Converged States.” *In IEEE Intl. Conf. on Robotics and Automation*, pp. 2909-2915, 2010.
[7] M. Hasan, *et al*, “Human hand gesture detection based on EMG signal using ANN,” *Intl. Conf. on Informatics, Electronics, and Vision*, 2014.
[8] A. B. Ajiboye, *et al*, “A heuristic fuzzy logic approach to EMG pattern recognition for multifunctional prosthesis control.,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 3, pp. 280-91, Sep. 2005.
[9] Y. Huang, *et al*, “A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses,” *IEEE Trans. Biomed. Eng.*, vol. 52, no. 11, pp. 1801-1811, 2005.
[10] http://www.delsys.com/products/wireless-emg/
[11] S. H. Bang, *et al*, “A hardware annealing method for optimal solutions on cellular neural networks”, *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 43, no. 6, pp. 409-421, 1996.
[12] J. Yuan, *et al*, “Cort-X II: The Low-Power Element Design for a Dynamic Neural Network” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 12, pp. 1130-1134, 2007.
[13] N. Naskas; Y. Papananos, “Neural-network-based adaptive baseband predistortion method for RF power amplifiers” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 51, no. 11, pp. 619-623, 2004.
[14] B. E. Shi, *et al*, “Design of linear cellular neural networks for motion sensitive filtering” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 40, no. 5, pp. 320-331, 1993.
[15] M. Asghari Oskoei, *et al*, “Myoelectric control systems - A survey,” *Biomed. Signal Process. Control*, vol. 2, no. 4, pp. 275-294, Oct. 2007.
[16] X. Liu, *et al*, “The PennBMBI : A General Purpose Wireless Brain-Machine-Brain Interface System for Unrestrained Animals,” *IEEE International Symposium on Circuits and Systems*, 2014.
[17] G. Naik, *et al*, “Transradial Amputee Gesture Classification Using an Optimal Number of sEMG Sensors: An Approach Using ICA Clustering.,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2015.