Wearable Computers for Sign Language Recognition

Jian Wu and Roozbeh Jafari

Abstract A Sign Language Recognition (SLR) system translates signs performed by deaf individuals into text/speech in real time. Low cost sensor modalities, inertial measurement unit (IMU) and surface electromyography (sEMG), are both useful to detect hand/arm gestures. They are capable of capturing signs and are complementary to each other for recognizing signs. In this book chapter, we propose a wearable system for recognizing American Sign Language (ASL) in real-time, fusing information from an inertial sensor and sEMG sensors. The best subset of features from a wide range of well-studied features is selected using an information gain based feature selection approach. Four popular classification algorithms are evaluated for 80 commonly used ASL signs on four subjects. With the selected feature subset and a support vector machine classifier, our system achieves 96.16 and 85.24% average accuracies for intra-subject and intra-subject cross session evaluation respectively. The significance of adding sEMG for American Sign Language recognition is explored and the best channel of sEMG is highlighted.

Keywords American sign language recognition • IMU sensor • Surface EMG • Feature selection • Sensor fusion

J. Wu (✉)
Department of Computer Science and Engineering, Texas A&M University, College Station, USA
e-mail: jian.wu@tamu.edu

R. Jafari
Departments of Biomedical Engineering, Computer Science and Engineering, and Electrical and Computer Engineering, Center of Remote Health Technologies and Systems, Texas A&M University, College Station, USA
e-mail: rjafari@tamu.edu

S.U. Khan et al. (eds.), *Handbook of Large-Scale Distributed Computing in Smart Healthcare*, Scalable Computing and Communications,
DOI 10.1007/978-3-319-58280-1_14

1 Introduction

According to World Health Organization (WHO), over 5% of the world's population—360 million people—has disabling hearing loss (328 million adults and 32 million children) by March, 2015. Disabling hearing loss refers to hearing loss greater than 40 decibels (dB) in the better hearing ear in adults and a hearing loss greater than 30 dB in the better hearing ear in children. The majority of people with disabling hearing loss live in low- and middle-income countries. Hearing loss may result from genetic causes, complications at birth, certain infectious diseases, chronic ear infections, the use of particular drugs, exposure to excessive noise and ageing. 50% of hearing loss can be prevented by taking medicines, surgery and the use of hearing aids and other devices. However, there are still a large number of people who have profound hearing loss which is also defined as deafness. They often use sign language for communication.

A sign language is a language which uses manual communication to convey meaning, as opposed to acoustically conveyed sound patterns. It is a natural language widely used by deaf people to communicate with each other [1]. However, there are communication barriers between hearing people and deaf individuals either because signers may not be able to speak and hear or because hearing individuals may not be able to sign. This communication gap can cause a negative impact on lives and relationships of deaf people. Two traditional ways of communication between deaf persons and hearing individuals who do not know sign language exist: through interpreters or text writing. The interpreters are very expensive for daily conversations and their involvement will result in a loss of privacy and independence of deaf persons. The text writing is not an efficient way to communicate because writing is too slow compared to either spoken/sign language and the facial expressions during performing sign language or speaking will be lost. Thus, a low-cost, more efficient way of enabling communication between hearing people and deaf people is needed.

A sign language recognition (SLR) system is a useful tool to enable communication between deaf people and hearing people who do not know sign language by translating sign language into speech or text [2, 3]. Figure 1 shows a typical

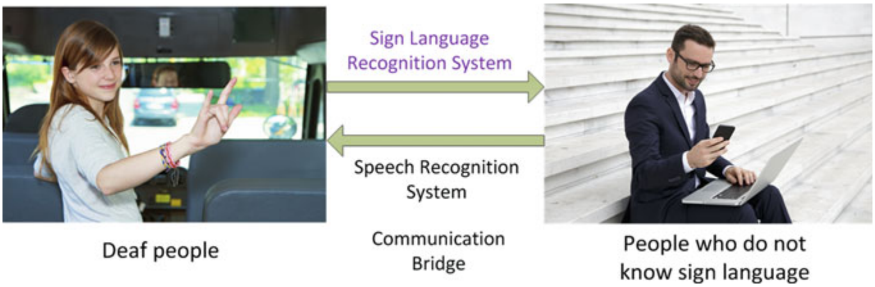


Fig. 1 Typical application of sign language recognition system

application of sign language recognition system. The SLR system worn by deaf people facilitates the translation of the signs to text or speech and transfer it to the smart phones of the people who can hear and speak. The spoken language of individuals who do not know sign language is translated into sign language images/videos by speech recognition systems. The speech recognition systems is not considered in this book chapter. The real-time translation of sign language enable deaf individual to communicate in a more convenient and natural way.

Similar to spoken languages, different countries have different sign languages. About 300 sign languages are currently being used all over the world. Due to the differences, the SLR should be trained and customized for every individual sign language. In our work, we have considered ASL. ASL dictionary includes thousands of signs, but most of them are not commonly used. In this chapter, 80 most commonly used signs are selected from 100 basic ASL signs [4, 5]. A sign is made up by five parts: hand shape, hand orientation, hand location, hand and arm movement and facial expression. Facial expression is more complicated and is not considered in this chapter.

Vision-based and glove-based SLR systems are well-studied systems which capture signs using cameras and sensory glove devices, respectively [6–10]. However, each of these two modalities has their own limitations. Vision-based systems suffer from occlusion due to light-of-sight factor. Moreover, cameras are mounted fixed in the environment and thus they can only be used in a limited range of vision. They are also considered to somewhat invasive to user's privacy. The glove-based systems are usually expensive which limits their usage in daily life.

Wearable inertial measurement unit (IMU) based gesture recognition systems attract much research attention due to their low cost, low power consumption and ubiquitous sensing ability [11, 12]. An IMU consists of a 3-axis accelerometer and a 3-axis gyroscope. The accelerometer measures 3-axis acceleration caused by motion and gravity while the gyroscope measures 3-axis angular velocity. A surface electromyography (sEMG) sensor is able to capture muscle electrical activities and can be used to distinguish different gestures since different gestures have different muscle activity patterns [13, 14]. For sign language recognition systems, the wrist worn IMU sensor is good at capturing hand orientations and hand and arm movements while sEMG does well in distinguishing different hand shapes and finger movements when the sensors are placed on the forearm. Thus, they are complementary to each other capturing different information of a sign and the fusion of them will improve the system performance [15]. Fortunately, the IoT platforms offer information from various sensor modalities and thus the performance of SLR would be enhanced by data fusion. However, additional sensor modalities will generate highly complex, multi-dimensional and larger volumes of data which introduce additional challenges. Challenges to address include increase in power consumption of wearable computers which will impact the battery life negatively and reducing the impact of modalities that appear to be too noisy and will degrade the performance of the classifiers.

In this book chapter, we propose a real-time wearable system for recognizing ASL by fusing inertial and sEMG sensors. Although such a system has been studied for Chinese Sign Language [16], to the best of the authors' knowledge this is the first time such a system is studied for the ASL. In this chapter, we first propose an adaptive auto-segmentation algorithm that determines the period during which the sign is performed. A wide range of well-established features are studied and the best subset of features are selected using an information gain based feature selection scheme. The feature selection determines the smallest feature subset which still provides good performance. It reduces the possibility of over-fitting and the smaller feature size is more suitable for wearable systems. Four commonly used classification algorithms are evaluated for intra- and inter-subject testing and the significance of adding sEMG for SLR is explored. When the best classifier is determined, the power consumption and the scalability of the classifiers are also considered.

The remainder of this book chapter is organized as follows. The related work is discussed in Sect. 2. Our lab customized sEMG data acquisition and IMU hardware platforms are introduced in Sect. 3. The details of our system are explained in Sect. 4, followed by the experimental setup in Sect. 5. The experimental results are explained in Sect. 6 and limitations are discussed in Sect. 7. At last, the chapter is concluded in Sect. 8.

2 Related Work

SLR systems are broadly studied in the field of computer vision with camera as a sensing modality. Two vision-based real-time ASL recognition systems are studied for sentence level continuous American Sign Language using Hidden Markov Model (HMM) [6]. The first system is evaluated for 40 signs and achieves 92% accuracy with camera mounted on the desk. The second system is also evaluated for 40 signs and achieves 98% accuracy with camera mounted on a cap worn by the user. A framework for recognizing the simultaneous aspects of ASL is proposed and it aims at solving the scalability issues of HMM [7]. The signs are broken down into phonemes and are modeled with parallel HMM. It reduces HMM state space dramatically as the number of signs increases. Another vision-based SLR system is studied for a medium vocabulary Chinese Sign Language [17]. It has two modules and the first module consists of three parts: robust hand detection, background subtraction and pupil detection. The second module is a tiered-mixture density HMM. With the aid of a colored glove, this system achieves 92.5% accuracy for 439 Chinese Sign Language words. In another work, three novel vision based features are learned for ASL recognition [18]. The relationship between these features and the four components of ASL is discussed. It yields 10.99% error rate on a published dataset. A Chinese Sign Language recognition system is proposed to address the issue of complex background in the environment [19]. The system is able to update the skin color model under various lighting conditions. A hierarchical classifier is used which integrates Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Principle Component Analysis (PCA).

Glove-based SLR systems recognize signs using multiple sensors on the glove. They are usually able to capture finger movements precisely. A glove-based Australian SLR system is proposed using some simple features and achieves 80% accuracy for 95 AUSLAN signs [20]. Another glove-based system is studied using artificial neural network classifier and it offers 90% accuracy for 50 ASL signs [9]. A flex sensor based glove is introduced recently that can be used to recognize 26 alphabets [21].

Similar to glove-based systems, the low cost wearable accelerometer and sEMG based SLR systems do not require cameras to be mounted at a certain location while they cost less than glove-based systems. Therefore, this kind of wearable SLR system is gaining more popularity. The importance of accelerometer and sEMG for recognizing gestures is studied [22]. The results show accelerometer and sEMG do well in capturing different information of a gesture and the fusion of them improve the system performance. In another work, 5–10% performance improvement is achieved after fusing these two modalities [23]. The sample entropy based feature set is proven to be effective for both accelerometer and sEMG and the system achieves 93% accuracy for 60 Greek Sign Language signs using this feature set [24]. A Chinese SLR framework is proposed fusing data from an accelerometer and 4-channel sEMG sensors [16]. It automatically determine the beginning and ending of a sign based on sEMG signal strength. Multi-stage classifications are applied to achieve an accuracy of 96.8% for 120 Chinese signs with sensors deployed on two hands. At the first stage, LDA is used for both accelerometer and sEMG to detect hand shape and hand orientation, respectively. In the meantime, a multi-stream HMM is applied for sEMG and accelerometer features. At the second stage, the decisions achieved from the first stage are fused with a Gaussian mixture model. Despite the good performance, multiple stages and multiple classifiers are not favorable for real-time wearable computers based applications. Recently, the same group proposes a component-based vocabulary-extensible sign language recognition system [25]. In this work, the sign is considered to be a combination of five common sign components, including hand shape, axis, orientation, rotation, and trajectory. There are two parts of this system. The first part is to obtain the component-based form of sign gestures and establish the code table of target sign gesture set using data from a reference subject. In the second part, which is designed for new users, component classifier are trained using a training set suggested by the reference subject and the classification of unknown gestures is performed with a code matching method. Another system is proposed to detect seven German sign words with 99.82% accuracy achieved using an accelerometer and one channel sEMG [26]. However, this work is not extensively evaluated for a large number of signs and does not include auto-segmentation which makes it difficult to operate in real time. The major differences between our work and the previous works are as follows: (1) An adaptive auto-segmentation is proposed to extract periods during which signs are performed using sEMG. (2) The best feature subset is selected from a broad range of features using information gain criterion and the selected features from different modalities (e.g. accelerometer, gyroscope and 4-channel sEMG) are discussed. (3) Gyroscope is incorporated and the significance of adding sEMG is analyzed. (4) Although such a system has been studied for Chinese Sign Language [16], our work is the first study for American Sign Language recognition fusing these two modalities.

3 Hardware Description

A. *IMU Sensor*

Figure 2 shows the 9-axis motion sensor customized in our lab. The InvenSense MPU9150, a combination of 3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer, severs as the IMU sensor. A Texas Instruments (TI) 32-bit microcontroller SoC, CC2538, is used to control the whole system. The board also includes a microSD storage unit and a dual mode Bluetooth module BC127 from BlueCreation. The system can be used for real-time data streaming or can store data for later analysis. It also has an 802.15.4 wireless module which can offer low power proximity measurement or ZigBee communication. In this book chapter, the sampling rates for accelerometer and gyroscope are chosen to be 100 Hz which is sufficient for the sign language recognition system [27].

B. *sEMG Acquisition System*

sEMG measures the electrical activity generated by skeletal muscle. Figure 3 shows a customized 16-channel Bluetooth-enabled physiological signal acquisition system. It can be used for ECG, sEMG and EEG data acquisition. The system is used as a four channel sEMG acquisition system in this study. A TI low power analog front end, the ADS1299, is used to capture four channel sEMG signals and a TI

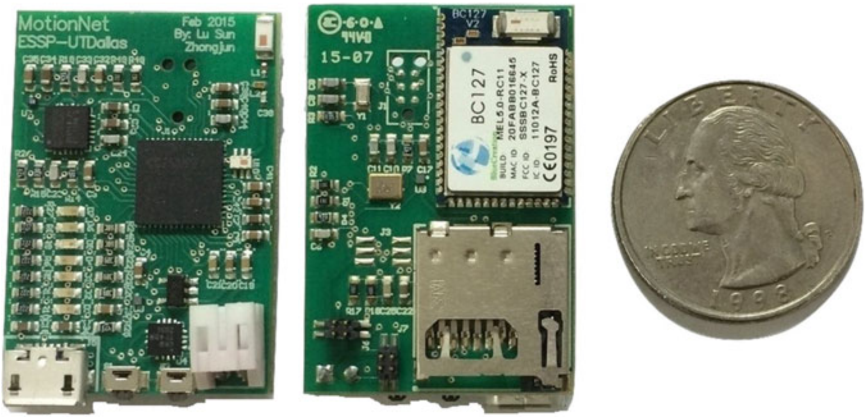


Fig. 2 Motion sensor board

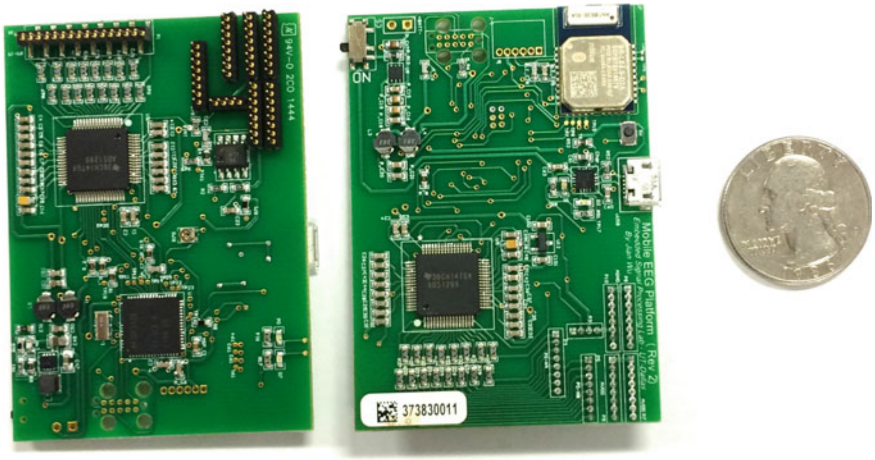


Fig. 3 8-channel sEMG acquisition system

MSP430 microcontroller is responsible for forwarding data to a PC via Bluetooth. A resolution of 0.4 μV is achieved setting a gain of 1 on the ADS1299. Covidien Kendall disposable surface EMG patches are attached to skin and the same electrodes are used as introduced in our previous work [28].

Generally, sEMG signals are in the frequency range of 0–500 Hz depending on the space between electrodes and muscle type [29]. To meet the Nyquist criterion, the sampling rate is chosen as 1 KHz, which is usually used in surface EMG based pattern recognition tasks [30].

4 Proposed SLR System

The block diagram of our proposed multi-modal ASL recognition system is shown in Fig. 4. Two phases are included: training phase and testing phase. In the training phase, the signals from 3-D accelerometer (ACC), 3-D gyroscope (GYRO) and four channel sEMG are preprocessed for noise rejection and synchronization purposes. The sEMG based auto-segmentation technique obtains the beginning and ending of a sign for both IMU and sEMG. As the segmentation is done, a broad set of well-established features are extracted for both IMU and sEMG signals. All extracted features are then put into one feature vector. The best feature subset is obtained using an information gain (IG) based feature selection scheme. Four different classifiers are evaluated (i.e. decision tree, support vector machine, NaïveBayes and nearest neighbor) on the selected feature subset and the best one is selected. In the testing phase, the same techniques are repeated for preprocessing and segmentation. The selected features are extracted and recognition of the sign is achieved by the chosen classifier.

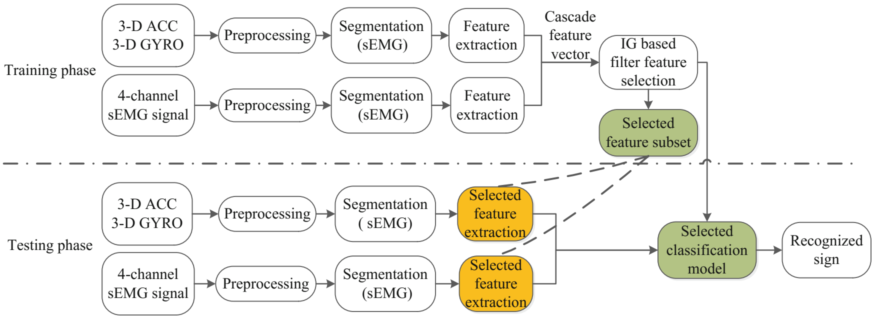


Fig. 4 Diagram of proposed system

A. *Preprocessing*

The synchronization between IMU and sEMG data is important for fusion. In our system, IMU data samples and sEMG data samples are sent to a PC via Bluetooth and time-stamped with the PC clock. The synchronization is done by aligning samples with the same PC clock. Bluetooth causes a transmission delay (5–20 ms) for both IMU and sEMG data and this small synchronization error is negligible for the purposes of our system. To remove low frequency noise in sEMG, a 5 Hz IIR high pass filter is used since the frequency components of sEMG beyond the range of 5–450 Hz are negligible [31]. The raw data is used for accelerometer and gyroscope.

B. *Segmentation*

Automatic segmentation is crucial for real-time applications. It extracts the period during which each sign word is performed such that the features can be extracted on the correct segment before classification is done. For certain parts of some signs, only finger movements are observed and no obvious motion signal can be detected from the wrist. Thus, sEMG signals are used for our automatic segmentation technique since sEMG signals can capture larger number of movements.

To explain our segmentation technique, we first define the average energy E of four sEMG channels in an n sample window in Eq. (1). $S_c(i)$ denotes ith sample of cth channel of sEMG. m is total number of channels which equals four in our case. A non-overlapping sliding window is used to calculate E in every window. The length of the window is set to 128 ms, which covers 128 samples with the 1000 Hz sampling frequency. If E in five continuous windows are all larger than a threshold T, the first sample of the first window will be taken as the beginning of a gesture. If E in four continuous windows are all smaller than the threshold, the last sample in the last window is considered to be the ending of this gesture.

$$E = \frac{1}{n} \sum_{i=1}^{n} \sum_{c=1}^{m} s_c^2(i) \tag{1}$$

Different people have different muscular strengths which will result in different E. A simple threshold may not be suitable for all subjects. An adaptive estimation technique is proposed to adjust the threshold according to different subjects and different noise levels on-line. The proposed approach is explained in two steps. In the first step, the average energy E is calculated for five continuous windows. If all five E is smaller than $a * T$, it is assumed no muscle activity is detected and the threshold is updated with $b * T$ in the second step. a is called the converge parameter and this reduces the threshold T when quiet periods are detected. b is the diverge parameter which enlarges the threshold T as the noise level increases. The values of a, b and T are set to be 0.5, 4 and 0.01 for the system empirically. 0.01 is much bigger than E for all subjects and the user is requested to have a 2–3 s quiet period at the beginning of system operation to have the system converge to a suitable threshold.

C. *Feature Extraction*

A broad range of features have been proposed and studied for both sEMG and IMU sensors for recognizing activities or gestures. In this chapter, these well-studied features are investigated [32–36]. Tables 1 and 2 list features and their dimensions from sEMG and IMU, respectively. The sEMG features are extracted for all four channel signals and the total dimension is 76. The IMU sensor features are extracted for 3-axis accelerometer, 3-axis gyroscope and the magnitude of accelerometer and

Table 1 sEMG features

Feature name (dimension)	Feature name (dimension)
Mean absolute value (1)	Variance (1)
Four order reflection coefficients (4)	Willison amplitude in 5 amplitude ranges (5)
Histogram (1)	Modified median frequency (1)
Root mean square (1)	Modified mean frequency (1)
Four order AR coefficients (4)	

Table 2 IMU sensor features

Feature name (dimension)	Feature name (dimension)
Mean (1)	Variance (1)
Standard deviation (1)	Integration (1)
Root mean square (1)	Zero cross rate (1)
Mean cross rate (1)	Skewness (1)
Kurtosis (1)	First three orders of 256-point FFT coefficients (3)
Entropy (1)	Signal magnitude area (1)
AR coefficients (10)	

gyroscope. The number of total IMU features is 192. The sEMG and IMU features are cascaded into a 268 dimension feature space.

D. *Feature Selection*

For classification, it is important to select the most useful features. There are usually two approaches to select the most useful features. The first approach is to define most useful and relevant features from a domain expert. For those experts who are familiar with their field, they usually know what the useful features are for certain tasks. The second approach is to select a certain subset features from an extensive number of features. Since even a domain expert may not be aware of all best features, thus the second approach is preferred. In this chapter, we use the second approach to select a subset of features from a wide range of features. It reduces over fitting problems and information redundancy in the feature set. It is also helpful if a small feature set is required by certain applications with limited computational power.

There are three different feature selection methods which are filter methods, wrapper methods, and embedded methods [37]. Wrapper methods generate scores for each feature subset based on a specific predictive model. Then, cross validation is done for each feature subset. Based on the prediction performance, each subset is assigned a score and the best subset is chosen. Filter methods use general measurement metrics of a dataset to score a feature subset instead of using the error rate of a predictive model. Some common measures are mutual information and inter/intra class distance. The embedded methods perform the feature subset selection in conjunction with the model construction. In our work, an information gain filter method is used in conjunction with a ranking algorithm to rank all the features. The best n features form the best feature subset which is evaluated with different classifiers. The choice of n is discussed in Sect. 5. Compared to wrapper methods, the features selected by filter methods will operate for any classifier instead of working only with a specific classifier.

E. *Classification*

Four commonly used classification algorithms are investigated in this chapter: decision tree (DT) [38], support vector machine (LibSVM) [39], nearest neighbor (NN) and NaiveBayes. The implementations of these classifiers are achieved by Weka, a popular open source machine learning tool [40]. LibSVM uses radial basis function (RBF) kernel and uses a grid search algorithm to determine the best kernel parameters. The default parameters are applied for other classifiers. In machine learning, it is usually hard to determine which classifier is more suitable for a specific application and thus it is worth testing several algorithms before we choose one.

5 Experimental Setup

A. *Sensor Placement*

The signs can involve one hand or two hands. In our work, we only look at the right hand movements for both one-hand or two-hand signs. If they system is deployed on two hands, it will increase the recognition accuracy. Figure 5 shows the sensor placement on right forearm of the user. Four major muscle groups are chosen to place four channel sEMG electrodes: (1) extensor digitorum, (2) flexor carpi radialis longus, (3) extensor carpi radialis longus and (4) extensor carpi ulnaris. The IMU sensor is worn on the wrist where a smart watch is usually placed. To improve signal-to-noise ratio of sEMG readings, a bi-polar configuration is applied for each channel and the space between two electrodes for each channel is set to 15 mm [41]. The electrode placements are also annotated in the figure.

B. *Data Collection*

In this chapter, we selected 80 most commonly used ASL signs in daily conversations. The data is collected from four subjects (three male subjects and one female subject). The subjects performed the signs for the first time and did not know the ASL prior to the experimentation. For each subject, the data collection includes three sessions which were performed on three different days. During each session, all signs were performed 25 times. The dataset has 24,000 instances in total.

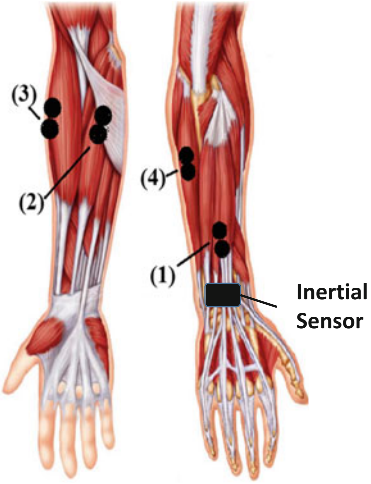


Fig. 5 Placement of sEMG electrodes

C. *Experiments*

To evaluate our system, four experiments are carried out: intra-subject testing, all cross validation, inter-subject testing and intra-subject cross session testing. In intra-subject testing, the data from the same subject from all sessions are combined and for each subject, a ten-fold cross validation is conducted. Ten-fold validation means that the data is split into 10 parts randomly and the model is trained with 9 parts and is tested on the 10th part. This process is carried out 10 times and the average performance outcome is considered the cross validation result. In all cross validation, all the data from different subjects from different days are combined. The ten-fold cross validation is performed similarly. In the inter-subject testing, the model is trained with data from three subjects and is tested on the fourth subject. This process is repeated four times. The feature selection for the first three experiments is performed during all cross validation since it has all the data and it will offer better generalization for classification models. The fourth experiment is called intra-subject cross session testing. The feature selection and model training are done with two sessions of data from the same subject and tested on the third session. This process is repeated three times for each subject and the average is taken over. The experiment indicates how well the model will perform with new data and a new subject.

6 Experimental Results

A. *Auto-segmentation*

In this chapter, no gold standard (e.g. video record) is included to determine the accuracy of our auto-segmentation technique. However, we approximately evaluate our auto-segmentation performance by looking at the difference in the number of signs each subject performed and the number of signs our system recognized. We define an error rate as in (2):

$$ER = \frac{|detected\ nums - performed\ nums|}{perfomed\ nums} \tag{2}$$

detected nums and *performed nums* are the numbers of signs our algorithm detected and numbers of signs the user actually performed, respectively. Our approach achieves 1.3% error rate which means our auto-segmentation algorithm performs well. The intra-subject classification results in Sect. 5. C also indicate suitable performance of the segmentation.

B. *Feature Selection*

All features are ranked with information gain criterion and the features with highest scores are chosen to form the best feature subset. To decide the size of best feature set, all cross validation is performed on four different classifiers as feature subset size increases from 10 to 268.

Figure 6 shows the accuracies of four classifier as the selected feature size increases. All classifier accuracies increase as the feature size increases. However, when the feature size is larger than 120 for the LibSVM and nearest neighbor, the accuracies decrease due to the over-fitting. This proves the feature selection is necessary. Table 3 shows the data points for four classifiers when they achieve the best accuracy.

It is shown in Fig. 6, when feature subset size becomes 40, LibSVM already offers 96.16% accuracy. The feature size is determined to be 40 in order to save computational cost for wearable systems. Among the 40 features, the numbers of features selected from different sensors are shown in Table 4. More than half of the features are selected from accelerometer which means accelerometer plays the principal role in recognizing signs. Accelerometers measure both gravity and acceleration due to the motion. Gravity is the major part of accelerometer measurements and captures the hand orientation information. It indicates hand orientation plays a more important role when recognizing different signs. 10 gyroscope

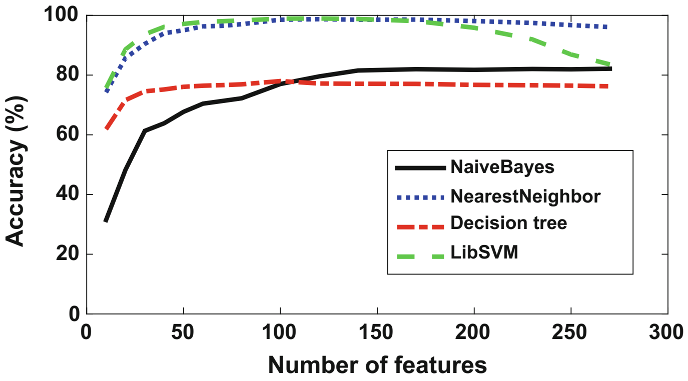


Fig. 6 Results of feature selection

Table 3 Optimal data point of feature selection

Classifier	Optimal point (feature number, accuracy) (%)
NaiveBayes	(270, 82.13)
Neareast neighbor	(120, 98.73)
Decision tree	(100, 78.00)
LibSVM	(120, 98.96)

Table 4 Number of features selected from different sensors

Sensor	Number of feature selected	Sensor	Number of feature selected
Accelerometer	21	sEMG2	2
Gyroscope	10	sEMG3	0
sEMG1	4	sEMG4	3

features are chosen which indicates the hand and arm rotation is also important. It is necessary to include sEMG sensors since nine features are selected from sEMG.

To have a better understanding of the importance of each individual feature, the rankings of 40 features are listed in Table 5. In the table, Acc_x, Acc_y and Acc_z represent accelerometer readings along *x-axis*, *y-axis* and *z-axis*, respectively. Similarly, Gyro_x, Gyro_y and Gyro_z are gyroscope readings along *x-axis*, *y-axis* and *z-axis*, respectively. From the table, the accelerometer contributes to the most highly ranked features which means the most significant modality of our system is the accelerometer. The gyroscope features are not as highly ranked as the accelerometer, but they have higher rankings than sEMG features. From the table, sEMG contribute least among all three. Among accelerometer and gyroscope features, the most important ones include mean, integration, standard deviation, RMS and variance. Mean absolute value, variance and RMS are valuable features for sEMG signal. One interesting observation of sEMG features is that four selected features from channel one have higher ranks than the others from channel two and channel four. Channel one is placed near the wrist where a smart watch is usually worn. In reality, if only one electrode is allowed, channel one could selected and it can be integrated into a smart watch without introducing a new device.

C. *Classification Results*

Table 6 shows the classification results of intra-subject testing on four subjects. In this experiment, each classifier is trained and tested with data from the same subject. We can see that nearest neighbor and LibSVM achieve high accuracies while decision tree classifier obtains the lowest accuracy. Nearest neighbor classifier is a lazy learning classifier and it does not require a trained model. In the testing phase, it compares the testing instance with all instances in the training set and assigns it a same class label as the most similar instance in the training set. It will require a large computation power as the number of training samples increase and thus is not suitable for our wearable SLR system. LibSVM trains a model based on training data. As the size of training set increases, it only increase the training time without affecting the time needs in testing phase. This is crucial for real time wearable computer based applications. Therefore, LibSVM is selected for our system. The results achieved for 80 signs are consistent with the results obtained for 40 signs in our prior investigation [42]. It indicates our technique scales well for intra-subject testing.

Table 7 shows classification results of all cross validation. For all classifiers, the classification results with sEMG and without sEMG are given. The performance with sEMG is when the performance achieved using all 40 selected features while

Table 5 Fourty selected features

Rank#	Feature name	Rank#	Feature name	Rank#	Feature name	Rank#	Feature name
1	Mean of Acc_y	11	RMS of Gyro_x	21	RMS of sEMG1	31	Signal magnitude area of Acc_x
2	Mean of Acc_z	12	RMS of amplitude of accelerometer	22	Zero cross rate of Acc_y	32	Variance of sEMG4
3	RMS of Acc_x	13	Mean of amplitude of accelerometer	23	Variance of Gyro_z	33	Entropy of Gyro_x
4	RMS of Acc_z	14	Mean of Acc_x	24	Standard deviation Of Gyro_z	34	RMS of sEMG4
5	RMS of Acc_y	15	Signal magnitude area of Acc_x	25	Variance of Acc_y	35	Signal magnitude area of Gyro_x
6	Integration of Acc_y	16	Standard deviation of Acc_z	26	Standard deviation of Acc_y	36	Zero cross rate of Acc_z
7	Integration of Acc_x	17	Variance of Acc_z	27	Modified mean frequency of sEMG1	37	Mean absolute value of sEMG4
8	Integration of Acc_z	18	Standard deviation of Gyro_z	28	Mean absolute value of sEMG1	38	Signal magnitude area of Gyro_z
9	Entropy of Acc_x	19	Variance of Gyro_x	29	First auto-regression coefficient of Acc_x	39	RMS of sEMG2
10	RMS of Gyro_z	20	Variance of sEMG1	30	Mean absolute value of sEMG2	40	Mean of amplitude of gyroscope

performance without sEMG is when the performance obtained using 31 features selected from accelerometer and gyroscope. The performance improvements by adding sEMG are also listed in the table. Among four classifiers, LibSVM achieves the best performance in accuracy, precision, recall and F-score while NaiveBayes gives the worst performance. The accuracy, precision, recall and F-score are very close to each other for all classifiers which indicates all classifiers achieve balanced performance on our dataset. With 40 features, LibSVM achieves 96.16% accuracy. It is consistent with the results (95.16%) we obtained for 40 sign words with 30 features in our prior study [42]. This proves the scalability of approach for all cross validation test.

Table 6 Results of intra-subject validation

	NaiveBayes (%)	DT (%)	NN (%)	LibSVM (%)
Subject 1	88.81	83.89	96.6	98.22
Subject 2	97.01	91.54	99.16	99.48
Subject 3	92.74	81.97	92.89	96.61
Subject 4	91.15	77.98	95.77	97.23
Average	93.68	83.85	96.11	97.89

Table 7 Results of all-cross validation

	NaiveBayes (%)	DT (%)	NN (%)	LibSVM (%)
Accuracy with sEMG	63.87	76.18	94.02	96.16
Accuracy without sEMG	48.75	68.93	87.62	92.29
Improvement	15.12	7.25	6.4	3.84
Precision with sEMG	66.9	76.3	94.0	96.7
Precision without sEMG	51.8	69.0	87.7	92.3
Improvement	15.1	7.3	6.3	4.4
Recall with sEMG	63.9	76.2	94.0	96.7
Recall without sEMG	48.8	68.9	87.7	92.3
Improvement	15.1	7.3	6.3	4.4
F-score with sEMG	63.6	76.2	94.0	96.7
F-score without sEMG	47.6	68.9	87.6	92.3
Improvement	16.0	7.3	6.4	4.4

The improvement after adding the sEMG modality is most significant for NaiveBayes classifier. It achieves about 15% improvement for all four classification performance metrics. However, for our chosen classifier LibSVM, the accuracy improvement is about 4% while the error rate is reduced by 50%. It indicates the sEMG is necessary and significant. The significance of sEMG is further analyzed in next section.

Figure 7 shows the average accuracy of inter-subject testing for both 80 sign words and 40 sign words. The figure shows none of four classifier achieves good performance. LibSVM is still the best classifier. There are three reasons for such low accuracies. First, different people perform the same signs in different ways. Second, all subjects in our experiment are first time ASL learners and never had experience with ASL before. Even though they follow the instructions, the gestures for the same signs are different from each other. Third, different subjects have very different muscular strength and thus leading to different sEMG features for same signs. From the comparison between accuracy of 40 signs and 80 signs, our technique offers low accuracy for all classifiers consistently. The low performance suggests our system is not suitable for inter-subject applications and it is recommended that our system should be trained on each individual to provide good performance.

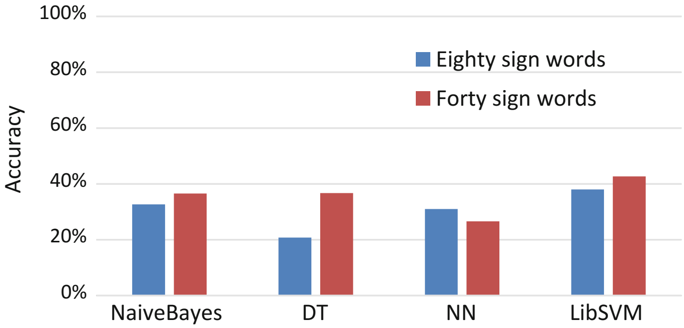
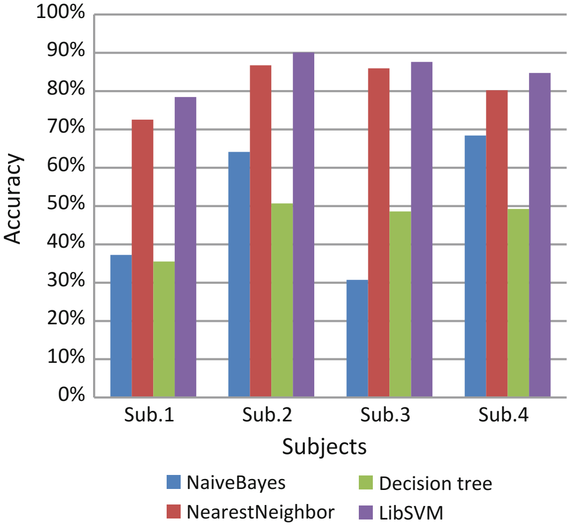


Fig. 7 Results of inter-subject testing

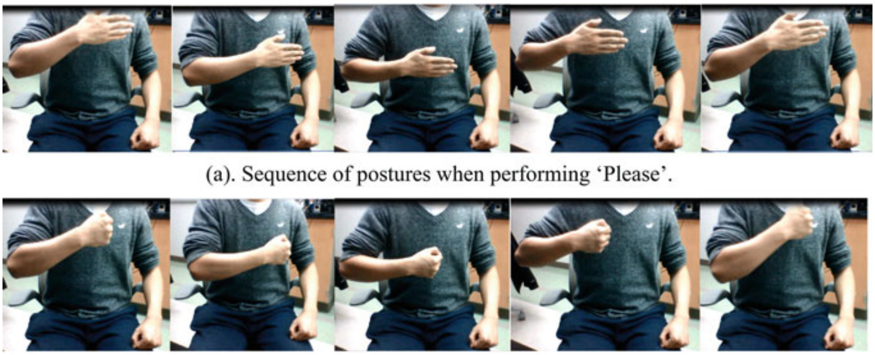
Fig. 8 Results of intra-subject cross session testing



The first three experiments show our system achieves suitable performance if the system is trained and tested for the same subject and the system obtains less ideal performance for inter-subject testing. We further investigate how well the system will generalize for new data collected in future for the same subject. Figure 8 shows the results of the intra-subject cross session testing in which the feature selection is performed and the classifier is trained with two days data from the same each subject and is tested on data of the third day for the same subject. This process is repeated three times for the same subject and the accuracy measures are averaged. We can see that both NaiveBayes and decision tree yield poor accuracies while LibSVM offers best accuracy. Table 8 shows the average accuracy of different classification algorithms between four subjects. LibSVM achieves 85.24% which is less suitable than the 96.16% of intra-subject testing. Two reasons may explain this performance decrease. The first reason is that the user may have placed the sensors

Table 8 Results of intra-subject cross session testing

Classifier	Accuracy (%)	Classifier	Accuracy (%)
NaiveBayes	50.11	NN	81.37
DT	46.01	LibSVM	85.24



(a). Sequence of postures when performing ‘Please’.

(b). Sequence of postures when performing ‘Sorry’.

Fig. 9 Sequence of postures when performing ‘Please’ and ‘Sorry’

at slightly different locations for the sEMG and IMU sensors, and with a slightly different orientation for the IMU sensor. The second reason is that all four subjects are first time learner who have not developed consistent patterns for signs. They may have performed the same signs somewhat differently on different days.

D. *Significance of sEMG*

From the analysis of inter-subject testing in previous section, the error rates for the accuracy, precision, recall and F-score are reduced by about 50%. In this section, we analyze the importance of sEMG in details. From the previous discussion, accelerometer and gyroscope are more important than sEMG. However, in ASL, there are some signs that have similar arm/hand movement and different hand shape and finger configurations (e.g. fist and palm). For these signs, they will have similar accelerometer and gyroscope readings and the IMU is not able to distinguish these signs. The sEMG is able to capture the difference of these signs since they will have different muscle activities. Figure 9 shows an example of sequences of postures when the user is performing two signs ‘Please’ and ‘Sorry’. We can see from the figures, the arm has the same movement which is drawing a circle in front of chest. The inertial sensor will offer same readings for these two different signs. However, the hand is closed (i.e. fist) when performing ‘Sorry’ while it is open (i.e. palm) when performing ‘Please’. This difference can be captured by sEMG and thus they will be distinguishable if sEMG is included.

In order to show how sEMG will enhance recognition performance of each individual sign, the improvement on the true positive (TP) rate of each individual sign is investigated. TP rate is rate of true positive and true positives are number of instances which are correctly classified as a given class. Figure 10 shows the TP rate improvement for 80 signs and the improvement is sorted in descend order. From the figure, we can see that for most of signs (last 29–80), the rate of improvement is within the range of [−5, 5]%. However, for the signs from 1 to 11, the improvement is bigger than 10% which is very helpful for recognizing these signs. In Table 9, 10 signs are listed with the highest TP rate improvement. We can see that ‘Sorry’ and ‘Please’ are both improved significantly since they are confused with each other. In reality, it is important to eliminate the confusion between signs which have similar motion profile but different sEMG characteristics. Therefore, the sEMG is significant for our system.

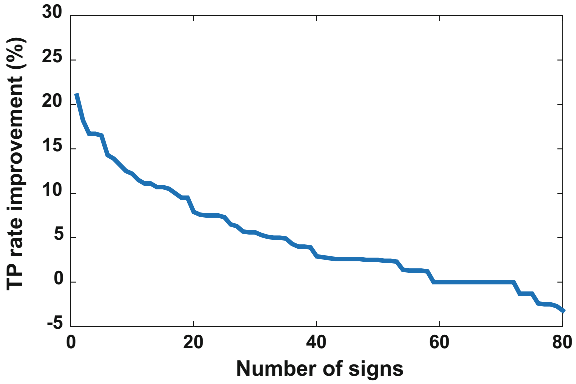


Fig. 10 TP rate improvement of all signs

Table 9 10 signs with most TP rate improvement

Sign ID	Sign	Improvement (%)
29	Thank	21
19	My	18.2
9	Have	16.7
24	Please	16.7
37	Work	16.5
57	Tall	14.3
67	Girl	13.9
26	Sorry	13.8
76	Doctor	12.5
66	Boy	12.5

7 Limitations and Discussion

The wearable inertial sensor and sEMG sensors based sign language recognition/gesture recognition systems have become more and more popular in recent years because of low-cost, privacy non-intrusive and ubiquitous sensing ability compared with vision-based approaches. They may not be as accurate as vision-based approaches. A vision-based approach achieves 92.5% accuracy for 439 frequently used Chinese Sign Language words [17]. Although we have not tested for such a large number of signs, it may be challenging with wearable inertial and sEMG systems to recognize such a big number of signs. Another disadvantage with wearable inertial sensor and sEMG based sign language recognition system is that the facial expression is not captured.

In our study, we observe that the accelerometer is the most significant modality for detecting signs. When designing such systems, if fusion of multiple modalities is not possible, the suggested choice order of these three are: accelerometer, gyroscope and sEMG. The significance of sEMG is to distinguish sets of signs which are similar in motion and this is crucial for sign language recognition. For some gesture recognition tasks, if gesture number is not big and there are no gestures which are very similar in motion, one inertial sensor may be sufficient for the task to reduce the system cost.

Our system offers high accuracy for both 40 signs and 80 signs for intra-subject testing and all cross validation. This shows our system is scalable for American Sign Language recognition if the system is trained and tested on the same subjects. However, very low accuracy is achieved for inter-subject testing which indicates our system is not very suitable for use on individuals if the system is not trained for them. We have talked to several experts of American Sign Language and they think it is reasonable to train for each individuals since even for expert, they will perform quite differently from each other for the same signs based on their preference and habits. This is the major limitation of sign language recognition systems. Our system is studied and designed to recognize individual signs assuming a pause exists between two sign words. However, in daily conversation, a whole sentence may be performed continuously without an obvious pause between each words. To recognize continuous sentence, a different segmentation technique or other possibility models should be considered.

Machine learning is a powerful tool for different applications and is gaining a lot of popularity in recent years in wearable computer based applications. However, it is important to use it in a correct way. For different applications, different features and different classifiers may have significantly different performance. It is suggested to try different approaches to determine the best one. The other point is that the classifier parameters should be carefully tuned. In our approach, if we do not choose the correct parameters for LibSVM, only 68% accuracy can be achieved.

As IoT emerges, the information from different sensing modalities could be explored. When designing applications, data from different sources should be considered and verified if they are complementary and the fusion of the modalities could potentially enhance the application performance.

8 Conclusion

A wearable real-time American Sign Language recognition system is proposed in this book chapter. This is a first study of American Sign Language recognition system fusing IMU sensor and sEMG signals which are complementary to each other. Our system design is an example of fusing different sensor modalities and addressing computation cost challenge of wearable computer based SLR due to the high-dimensional data. Feature selection is performed to select the best subset of features from a large number of well-established features and four popular classification algorithms are investigated for our system design. The system is evaluated with 80 commonly used ASL signs in daily conversation and an average accuracy of 96.16% is achieved with 40 selected features. The significance of sEMG to American Sign Language recognition task is explored.

Acknowledgements This work was supported in part by the National Science Foundation, under grants CNS-1150079 and ECCS-1509063. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations.

References

1. W. C. Stokoe, “Sign language structure: An outline of the visual communication systems of the American deaf,” *Journal of deaf studies and deaf education*, vol. 10, no. 1, pp. 3–37, 2005.
2. D. Barberis, N. Garazzino, P. Prinetto, G. Tiotto, A. Savino, U. Shoaib, and N. Ahmad, “Language resources for computer assisted translation from italian to italian sign language of deaf people,” in *Proceedings of Accessibility Reaching Everywhere AEGIS Workshop and International Conference, Brussels, Belgium (November 2011)*, 2011.
3. A. B. Grieve-Smith, “Signsynth: A sign language synthesis application using web3d and PERL,” in *Gesture and Sign Language in Human-Computer Interaction*, pp. 134–145, Springer, 2002.
4. B. Vicars, “Basic ASL: First 100 signs.”
5. E. Costello, *American sign language dictionary*. Random House Reference &, 2008.
6. T. Starner, J. Weaver, and A. Pentland, “Real-time american sign language recognition using desk and wearable computer based video,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 12, pp. 1371–1375, 1998.
7. C. Vogler and D. Metaxas, “A framework for recognizing the simultaneous aspects of American sign language,” *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 358–384, 2001.

8. T. E. Starner, “Visual recognition of American sign language using Hidden Markov models.,” tech. rep., DTIC Document, 1995.
9. C. Oz and M. C. Leu, “American sign language word recognition with a sensory glove using artificial neural networks,” *Engineering Applications of Artificial Intelligence*, vol. 24, no. 7, pp. 1204–1213, 2011.
10. E. Malaia, J. Borneman, and R. B. Wilbur, “Analysis of ASL motion capture data towards identification of verb type,” in *Proceedings of the 2008 Conference on Semantics in Text Processing*, pp. 155–164, Association for Computational Linguistics, 2008.
11. A. Y. Benbasat and J. A. Paradiso, “An inertial measurement framework for gesture recognition and applications,” in *Gesture and Sign Language in Human-Computer Interaction*, pp. 9–20, Springer, 2002.
12. O. Amft, H. Junker, and G. Troster, “Detection of eating and drinking arm gestures using inertial body-worn sensors,” in *Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on*, pp. 160–163, IEEE, 2005.
13. A. B. Ajiboye and R. F. Weir, “A heuristic fuzzy logic approach to EMG pattern recognition for multifunctional prosthesis control,” *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 13, no. 3, pp. 280–291, 2005.
14. J.-U. Chu, I. Moon, and M.-S. Mun, “A real-time EMG pattern recognition based on linear-nonlinear feature projection for multifunction myoelectric hand,” in *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*, pp. 295–298, IEEE, 2005.
15. Y. Li, X. Chen, X. Zhang, K. Wang, and J. Yang, “Interpreting sign components from accelerometer and sEMG data for automatic sign language recognition,” in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pp. 3358–3361, IEEE, 2011.
16. Y. Li, X. Chen, X. Zhang, K. Wang, and Z. J. Wang, “A sign-component-based framework for chinese sign language recognition using accelerometer and sEMG data,” *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 10, pp. 2695–2704, 2012.
17. L.-G. Zhang, Y. Chen, G. Fang, X. Chen, and W. Gao, “A vision-based sign language recognition system using tied-mixture density hmm,” in *Proceedings of the 6th international conference on Multimodal interfaces*, pp. 198–204, ACM, 2004.
18. M. M. Zaki and S. I. Shaheen, “Sign language recognition using a combination of new vision based features,” *Pattern Recognition Letters*, vol. 32, no. 4, pp. 572–577, 2011.
19. T.-Y. Pan, L.-Y. Lo, C.-W. Yeh, J.-W. Li, H.-T. Liu, and M.-C. Hu, “Real-time sign language recognition in complex background scene based on a hierarchical clustering classification method,” in *Multimedia Big Data (BigMM), 2016 IEEE Second International Conference on*, pp. 64–67, IEEE, 2016.
20. M. W. Kadous *et al.*, “Machine recognition of AUSLAN signs using powergloves: Towards large-Lexicon recognition of sign language,” in *Proceedings of the Workshop on the Integration of Gesture in Language and Speech*, pp. 165–174, Citeseer, 1996.
21. M. G. Kumar, M. K. Gurjar, and M. S. B. Singh, “American sign language translating glove using flex sensor,” *Imperial Journal of Interdisciplinary Research*, vol. 2, no. 6, 2016.
22. D. Sherrill, P. Bonato, and C. De Luca, “A neural network approach to monitor motor activities,” in *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, vol. 1, pp. 52–53, IEEE, 2002.
23. X. Chen, X. Zhang, Z.-Y. Zhao, J.-H. Yang, V. Lantz, and K.-Q. Wang, “Hand gesture recognition research based on surface EMG sensors and 2D-accelerometers,” in *Wearable Computers, 2007 11th IEEE International Symposium on*, pp. 11–14, IEEE, 2007.
24. V. E. Kosmidou and L. J. Hadjileontiadis, “Sign language recognition using intrinsic-mode sample entropy on sEMG and accelerometer data,” *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 12, pp. 2879–2890, 2009.
25. S. Wei, X. Chen, X. Yang, S. Cao, and X. Zhang, “A component-based vocabulary-extensible sign language gesture recognition framework,” *Sensors*, vol. 16, no. 4, p. 556, 2016.

26. J. Kim, J. Wagner, M. Rehm, and E. André, "Bi-channel sensor fusion for automatic sign language recognition," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pp. 1–6, IEEE, 2008.
27. J.-S. Wang and F.-C. Chuang, "An accelerometer-based digital pen with a trajectory recognition algorithm for handwritten digit and gesture recognition," *Industrial Electronics, IEEE Transactions on*, vol. 59, no. 7, pp. 2998–3007, 2012.
28. V. Nathan, J. Wu, C. Zong, Y. Zou, O. Dehzangi, M. Reagor, and R. Jafari, "A 16-channel bluetooth enabled wearable EEG platform with dry-contact electrodes for brain computer interface," in *Proceedings of the 4th Conference on Wireless Health*, p. 17, ACM, 2013.
29. C. J. De Luca, L. Donald Gilmore, M. Kuznetsov, and S. H. Roy, "Filtering the surface emg signal: Movement artifact and baseline noise contamination," *Journal of biomechanics*, vol. 43, no. 8, pp. 1573–1579, 2010.
30. I. Mesa, A. Rubio, I. Tubia, J. De No, and J. Diaz, "Channel and feature selection for a surface electromyographic pattern recognition task," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5190–5200, 2014.
31. R. Merletti and P. Di Torino, "Standards for reporting EMG data," *J Electromyogr Kinesiol*, vol. 9, no. 1, pp. 3–4, 1999.
32. A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A novel feature extraction for robust EMG pattern recognition," *arXiv preprint* arXiv:0912.3973, 2009.
33. M. Zhang and A. A. Sawchuk, "Human daily activity recognition with sparse representation using wearable sensors," *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, no. 3, pp. 553–560, 2013.
34. S. H. Khan and M. Sohail, "Activity monitoring of workers using single wearable inertial sensor."
35. O. Paiss and G. F. Inbar, "Autoregressive modeling of surface EMG and its spectrum with application to fatigue," *Biomedical Engineering, IEEE Transactions on*, no. 10, pp. 761–770, 1987.
36. A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, no. 5, pp. 1166–1172, 2010.
37. I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
38. J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.
39. C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
40. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
41. M. Z. Jamal, "Signal acquisition using surface EMG and circuit design considerations for robotic prosthesis," 2012.
42. J. Wu, Z. Tian, L. Sun, L. Estevez, and R. Jafari, "Real-time american sign language recognition using wrist-worn motion and surface EMG sensors," in *Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on*, pp. 1–6, IEEE, 2015.