IEEE JOURNAL OF
BIOMEDICAL AND
HEALTH INFORMATICS

A Wearable System for Recognizing American Sign Language in Real-time Using IMU and Surface EMG Sensors

Jian Wu, *Student member, IEEE*, Lu Sun, Roozbeh Jafari, *Senior Member, IEEE*

***Abstract*—A Sign Language Recognition (SLR) system translates signs performed by deaf individuals into text/speech in real time. Inertial measurement unit (IMU) and surface electromyography (sEMG) are both useful modalities to detect hand/arm gestures. They are able to capture signs and the fusion of these two complementary sensor modalities will enhance system performance. In this paper, a wearable system for recognizing American Sign Language (ASL) in real-time is proposed, fusing information from an inertial sensor and sEMG sensors. An information gain based feature selection scheme is used to select the best subset of features from a broad range of well-established features. Four popular classification algorithms are evaluated for 80 commonly used ASL signs on four subjects. The experimental results show 96.16% and 85.24% average accuracies for intra-subject and intra-subject cross session evaluation respectively, with the selected feature subset and a support vector machine classifier. The significance of adding sEMG for American Sign Language recognition is explored and the best channel of sEMG is highlighted.**

***Index Terms*—American Sign Language recognition; IMU sensor; surface EMG; feature selection; sensor fusion**

I. INTRODUCTION

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]. Fig. 1 shows a typical application of sign language recognition system. The system can be worn by deaf people who cannot talk and translates the signs performed to text or speech on the cell phone of the people who can hear and talk. The speech recognition system on deaf person's cell phone translates speech into sign language images/videos. The speech recognition part is not considered in this paper. The real-time translation enables them communicate in a more convenient and natural way.

There are different sign languages in different countries in different regions. Around 300 hundred sign languages are in use all over world today. Sign languages are natural languages and similar to spoken languages, they differ from each other. The system should be studied and designed for a specific sign language. In this paper, we focus on the recognition of ASL. There are thousands of signs in ASL dictionary but most of them are not commonly used. In our paper, 80 commonly used signs are chosen from 100 basic ASL signs [4, 5]. A sign consists of hand shape, hand location, hand orientation, hand and arm movement and facial expression. In our paper, facial expression is not considered when we design our system.

Vision-based and glove-based SLR systems are well-studied systems which capture signs using cameras and sensory glove devices, respectively [6, 7, 8, 9, 10].Vision-based techniques typically require cameras to be mounted in the environment which inherently suffer from a limited range of vision. Further, the required infrastructure may not be available at all of the desired locations or may be too expensive to implement. Issues associated with users' privacy also limit the utility of vision-based techniques. Due to high cost of glove devices, glove-based SLR systems are not ideal for use in daily life.

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.

Jian Wu is with department of Computer Science and Engineering, Texas A&M University, College Station, TX 77840 USA (e-mail:jian.wu@tamu.edu).

Lu Sun, was with University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: lxs@utdallas.edu).

Roozbeh Jafari is with Center of Remote Health Technologies and Systems, Departments of Biomedical Engineering, Computer Science and Engineering, and Electrical and Computer Engineering, Texas A&M University, College Station 77840 USA (e-mail: rjafari@tamu.edu).

Digital Object Identifier 10.1109/JBHI.2016.2598302

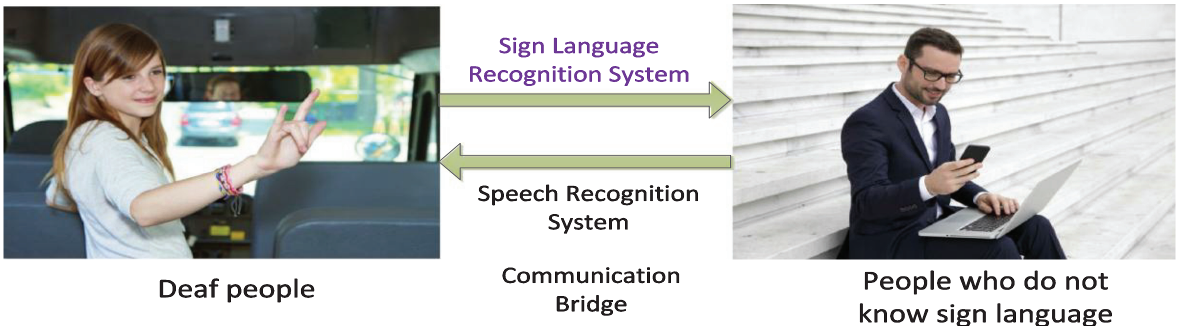


Fig. 1. Typical application of sign language recognition system.

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 measures acceleration and gravity with a 3-axis accelerometer and angular velocities with a 3-axis gyroscope. A surface electromyography (sEMG) sensor measures muscle electrical activity and it can be used to detect neuromuscular diseases and to analyze human kinetics. Different signs will generate different muscle electrical patterns and sEMG is able to capture this information to distinguish different gestures [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 each have their own advantages capturing different information about a sign. The fusion of these two complementary modalities will enhance the performance of an SLR system and thus enable the recognition of a large number of signs [15].

A wearable system for recognizing American Sign Language in real-time fusing information from inertial and sEMG sensors is proposed in this paper. 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 American Sign Language. In our work, an adaptive auto-segmentation technique using sEMG is proposed to define the beginning and ending of a sign. A broad range of well-studied features from both inertial and sEMG sensors are extracted from each segment and a best feature subset is selected using an information gain based feature selection approach. Four popular classification algorithms are evaluated for intra- and inter-subject testing and the significance of adding sEMG for SLR is explored.

The remainder of this paper is organized as follows. The related work is discussed in Section II. Our lab customized sEMG data acquisition and IMU hardware platforms are introduced in Section III. The details of our system are explained in Section IV, followed by the experimental setup in Section V. The experimental results are explained in Section VI and limitations are discussed in Section VII. At last, the paper is concluded in Section VIII.

II. RELATED WORK

SLR systems are well studied in the areas of computer vision and image processing. Two vision-based real-time ASL recognition systems are studied for sentence level continuous American Sign Language using Hidden Markov Model (HMM) [6]. In the first system, the camera is mounted on the desk while in the second system, the camera is mounted on a cap which is worn by the user. They are both tested for 40 signs and achieve 92% and 98% accuracy, respectively. A framework for recognizing the simultaneous aspects of ASL is proposed [7]. This framework targets at addressing the scalability issue associated with HMM. It breaks down the signs into their phonemes and modeling them with parallel HMM. In this way, the state space will decrease significantly as the number of signs increases. Another vision-based SLR system is studied for a medium vocabulary Chinese Sign Language [17]. Robust hand detection, background subtraction and pupil detection are implemented as the first module, followed by a tiered-mixture density HMM. With the aid of a colored glove, this system achieves 92.5% accuracy for 439 Chinese Sign Language words. A combination of three new vision based features are explored for ASL recognition [18]. Three features are mapped into four components of ASL: hand shape, place of articulation, hand orientation and movement. The proposed features achieve 10.90% error rate on an existing dataset.

Glove-based SLR systems implement multiple sensors on the glove and capture the physical features of the gestures. Unlike vision-based systems, they do not require cameras mounted around the user and the system can perform recognition at any place at any time with a wearable glove. A glove-based Australian SLR system is studied using two classifiers (*i.e.* Instance based classifier and decision tree classifier) with some simple features. 80% accuracy is achieved for 95 AUSLAN signs [19]. The performance of artificial neural networks is explored for an ASL recognition system using a sensory glove [9]. It achieves about 90% accuracy for 50 ASL signs.

The low cost wearable accelerometer and sEMG based SLR systems have the same advantages as glove-based systems compared to vision-based approach while they cost much less than glove based systems since they have fewer sensors deployed. Therefore, this kind of wearable SLR system is

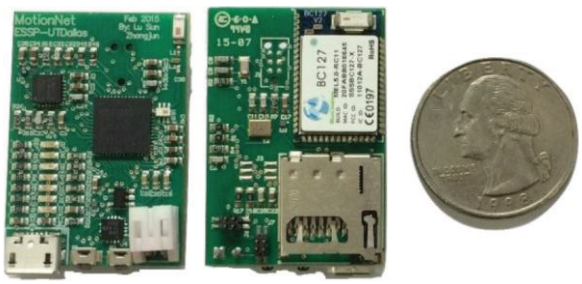


Fig. 2. Motion Sensor Board.

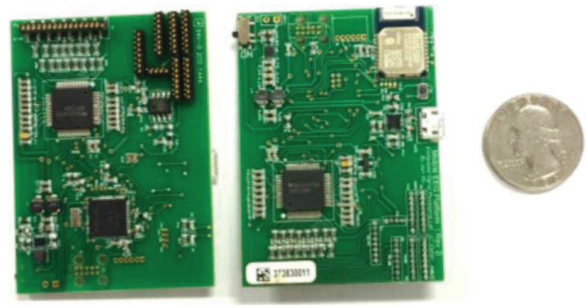


Fig. 3. 8-channel sEMG acquisition system.

gaining more and more popularity in recent years. SLR system has been explored in several studies fusing information from accelerometer and sEMG sensors. The comparison of accelerometer based and sEMG based gesture recognition systems is discussed [20]. It is suggested accelerometer and sEMG sensors are good at capturing different information of gestures and the performance enhancement combining these two modalities has been studied. The experiments show 5% - 10% accuracy improvement is obtained after fusing these two modalities [21]. An accuracy of 93% of recognizing 60 Greek Sign Language signs is achieved using only one effective sample entropy based feature set for both accelerometer and sEMG [22]. A Chinese SLR framework is proposed fusing data from an accelerometer and 4-channel sEMG sensors [16]. Auto segmentation is applied to extract sign words from continuous sentences according to sEMG signal intensity. Multiple classifiers are implemented at different stages and the decisions achieved by each individual classifier are fused. At the first stage, the linear discriminate analysis (LDA) classifier is applied for both sEMG and accelerometer data which are able to capture hand shape and hand orientation, respectively. All sEMG and accelerometer features are cascaded and fit into a multi-stream HMM to recognize signs. A Gaussian mixture model is applied to fuse decisions obtained in the first stage. Although this system obtains a 96.5% accuracy for 120 Chinese sign words with sensors deployed on two hands, multiple stages and multiple classifiers make it unfavorable for implementation on real-time wearable computers based applications which are constrained by limited computational resources. Another system is proposed to detect seven German sign words with 99.82% accuracy achieved using an accelerometer and one channel sEMG [23]. 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 paper is the first study for American Sign Language recognition fusing these two modalities.

III. Hardware description

A. IMU Sensor

Fig. 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 paper, the sampling rates for accelerometer and gyroscope are chosen to be 100 Hz which is sufficient for the sign language recognition system [24].

B. sEMG Acquisition System

sEMG measures the electrical activity generated by skeletal muscle. Fig. 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 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 [25].

Generally, sEMG signals are in the frequency range of 0Hz -500 Hz depending on the space between electrodes and muscle type [26]. To meet the Nyquist criterion, the sampling rate is chosen as 1K Hz, which is usually used in surface EMG based pattern recognition tasks [27].

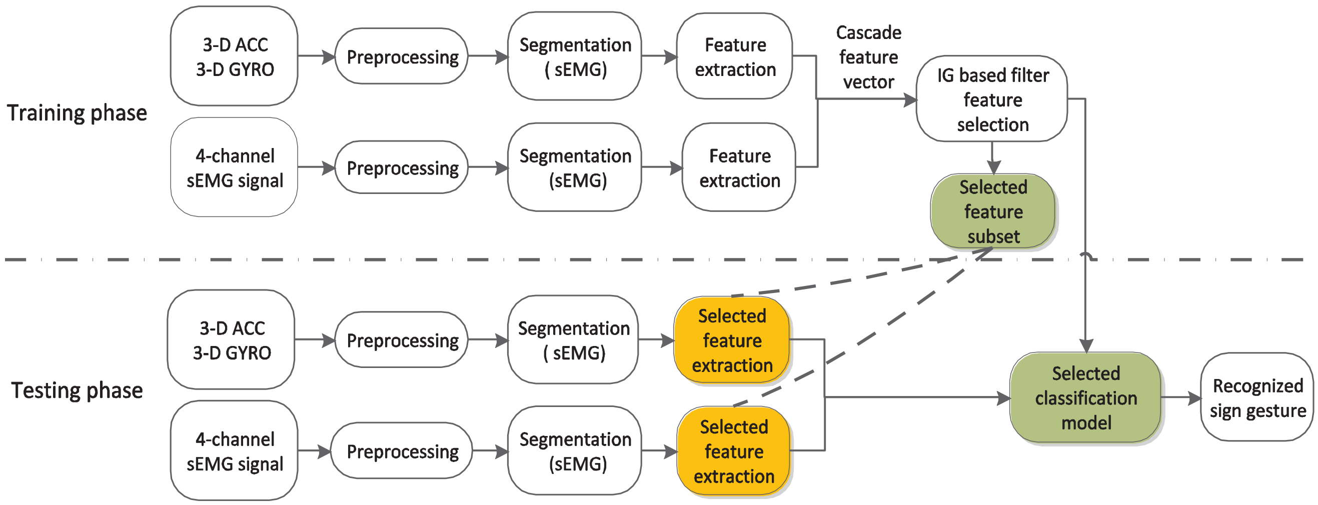


Fig. 4. Diagram of proposed system.

IV. 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.

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 (5-20ms) 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 5Hz IIR high pass filter is used since the frequency components of sEMG beyond the range of 5Hz – 450Hz are negligible [28]. 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 Equation (1). $S_c(i)$ denotes i^{th} sample of c^{th} 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 milliseconds, 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

TABLE I.
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 II.
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)	

have a 2-3 seconds quiet period at the beginning of system operation to have the system converge to a suitable threshold.

C. Feature Extraction

A large number of features have been proposed and studied for both sEMG and IMU sensors for detecting activities or gestures. We adopt some of these well-established features in our paper [29, 30, 31, 32, 33]. Table I and Table II show features from sEMG and IMU sensors, respectively. The dimension of each feature is also listed in the table. 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 gyroscope. It leads to a 192 dimension feature space. The features from sEMG and IMU sensors are combined into the final feature vector of dimension 268.

D. Feature Selection

Feature selection provides a way to select the most suitable feature subset for certain tasks from the well-established features. It reduces over fitting problems and information redundancy existing in the feature set. It can also suggest the best feature subset if a smaller feature set is required by applications with limited computational resources.

There are three different feature selection methods which are filter methods, wrapper methods, and embedded methods [34]. 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 Section V. Compared to wrapper methods, the features selected by filter methods will operate for any classifier instead of working only with a specific classifier.

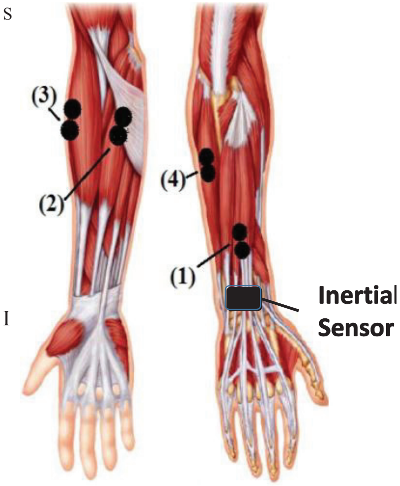


Fig. 5. Placement of sEMG electrodes.

E. Classification

Four popular classification algorithms are studied in this paper: decision tree (DT) [35], support vector machine (LibSVM) [36], nearest neighbor (NN) and NaiveBayes. Weka, a widely used open source machine learning tool, is applied for the implementations of these four algorithms [37]. The radial basis function (RBF) kernel is selected for the LibSVM and the best kernel parameters are tuned using a grid search algorithm. The default parameters are selected for the other three 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.

V. EXPERIMENTAL SETUP

A. Sensor Placement

The signs can involve one hand or two hands. In our paper, 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. Fig. 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 [38]. The electrode placements are also annotated in the figure.

TABLE III.
OPTIMAL DATA POINT OF FEATURE SELECTION

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

TABLE IV.
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

B. Data Collection

80 commonly used ASL signs in daily conversations are selected in our paper. Three male and one female volunteer are recruited for data collection. They are all first time learners and did not know ASL before. For each subject, the data is collected from three sessions on three different days and during each session, the subject repeats each sign 25 times. The dataset has 24000 instances in total.

C. Experiments

Four different experiments are conducted to test our system: intra-subject testing, all cross validation, inter-subject testing and intra-subject cross session testing. For intra-subject testing, the data collected from three sessions of same subject is put together and a 10-fold cross validation is done for the data collected from each subject separately. 10-fold cross validation means the data is split into 10 subsets randomly and the model is trained with 9 subsets and tested on the 10th subset. This process is repeated for 10 times and the average was taken over. For the all cross validation analyses, data from all four subjects are put together and a 10-fold cross validation is performed. For the inter-subject testing, the classifier is trained with data from three subjects and tested on the fourth subject. The performance is averaged for four tests. The feature selection for the first three experiments is carried out during all cross validation since it has data from all four subjects which makes it a good generalization for classification algorithms. For the intra-subject cross session testing, the feature selection is performed and the classifier is trained with two sessions from each subject and tested on the third session of the same subject. The process is repeated three times for each subject and the performance is averaged for each subject. This experiment would give an indication of how well the system will generalize to new data collected in future for the same subject.

VI. Experimental Results

A. Auto-segmentation

In our experiment, we do not have a gold standard (*e.g.* video record) and thus it is hard to measure the error of our automatic segmentation technique. However, we know the total number of signs each subject performed and the number of signs our algorithm recognized. An error rate (ER) is defined as:

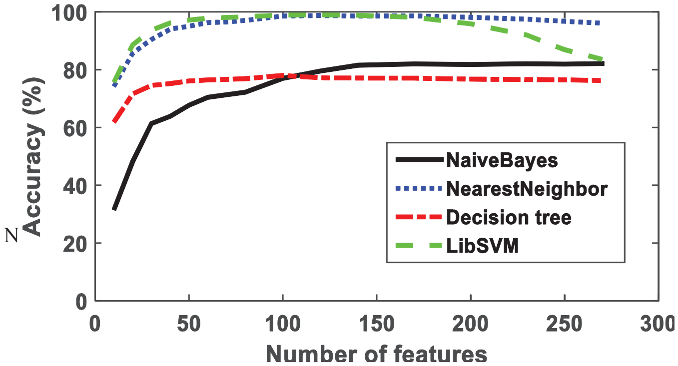


Fig. 6. Results of feature selection.

$$ER = \frac{|detected\ nums - performed\ nums|}{perfomed\ nums} \quad (2)$$

detected nums and *performed nums* are numbers of signs our algorithm detected and numbers of signs the user actually performed, respectively. The ER of our approach is 1.3% which indicates our segmentation technique achieves a good performance. The intra-subject classification results in section V.C also indicate suitable performance of the segmentation.

B. Feature Selection

All 268 features are ranked with a score obtained from information gain criterion. The highest ranked ones are selected to form the best 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.

Fig. 6 shows classification accuracies of four classifiers as the size of the best feature subset increases. It is seen from the figure that as the size of feature subset increases, the accuracies of all classifiers increase. However, when the feature number is bigger than 120 for LibSVM and nearest neighbor, their accuracies start to decrease as a result of over-fitting. This illustrates one of the reasons why feature selection is necessary. Table III lists four data points when classifiers achieve best performance.

Fig. 6 shows that when number of selected features becomes 40, LibSVM already achieves 96.16% accuracy. Due to the computational constraints associated with wearable systems, the feature size is thus selected to be 40. Among the 40 features, the numbers of features selected from different sensors are shown in Table IV. More than half of the features are selected from accelerometer which means accelerometer plays most important role in recognizing signs. Accelerometer measures both gravity and acceleration caused by movement. Gravity is usually the major part which is capable of capturing hand orientation information. It indicates hand orientation information is more significant than hand shape when distinguish different signs. Ten features from gyroscope are selected which means that the hand and arm rotation is also valuable information. Nine selected sEMG features make this modality necessary for our system.

TABLE V.

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

TABLE VI.

R

ESULTS 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%

To have a better understanding of the importance of different sensor features, forty selected features are listed in Table V along with their rankings. 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, we can see that most of the accelerometer features have very high rank which indicates accelerometer is the most important modality in our system.

The gyroscope features have higher ranks than sEMG features on average. Although the gyroscope is not as important as the accelerometer, it contributes more than sEMG. sEMG features are the least important among the three modalities which indicates it may not be significant in our system. 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 available, channel one would be selected and it can be integrated into a smart watch to capture the most important sEMG features.

C. *Classification results*

Table VI 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 does not scale well as the size of the training set increases since the testing instance needs to be compared to all instances in the training set. 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 applications. Therefore, LibSVM is the one we select for our system implementation. The results achieved for 80 signs are consistent with the results obtained for 40 signs in our prior investigation [39]. It indicates our technique scales well for intra-subject testing.

Table VII shows classification results of all cross validation. For all classifiers, the classification results with sEMG and without sEMG are given. The classification with sEMG means we use all 40 features while without sEMG means we only use 31 features from accelerometer and gyroscope. The performance improvement with adding sEMG is also listed in the table.

TABLE VII.
ESULTS 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%

TABLE VIII.
ESULTS OF INTRA-SUBJECT CROSS SESSION TESTING

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

Among four classifiers, LibSVM achieves the best performance in accuracy, precision, recall and F-score while NaïveBayes 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 [39]. This proves the scalability of approach for all cross validation test.

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.

Fig. 7 shows the average accuracy of inter-subject testing for both eighty sign words and forty sign words. It is seen from the figure, none of the classifiers offer good accuracy for recognizing 40 or 80 signs. LibSVM still offers the best performance among four classifiers. 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. For NaiveBayes, NN and LibSVM, the accuracy obtained from

R

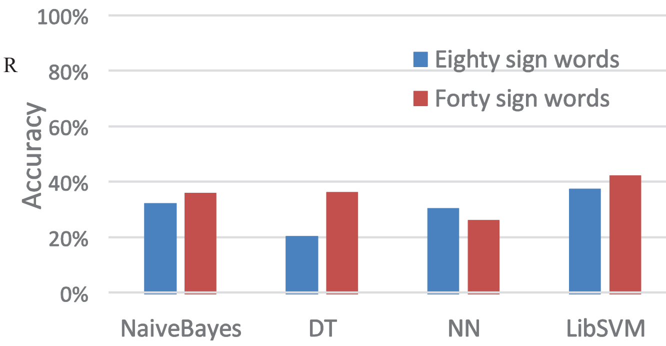


Fig. 7. Results of inter-subject testing.

R

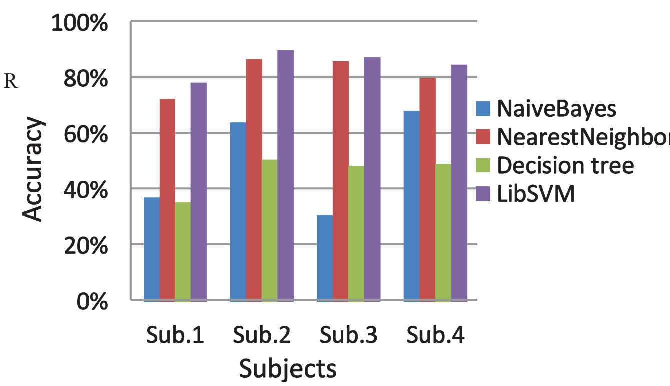
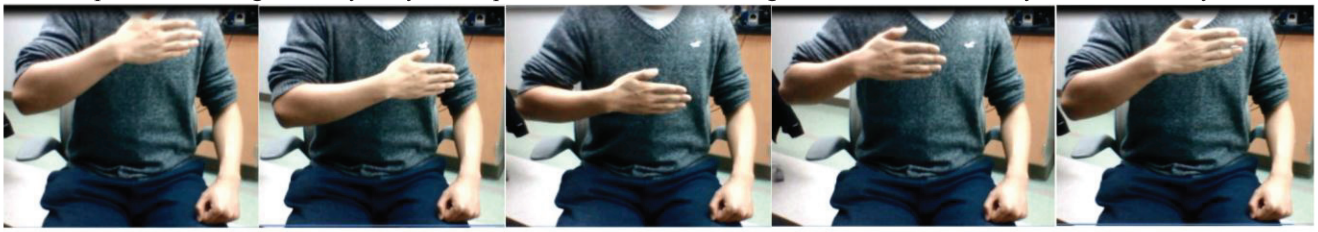


Fig. 8. Results of intra-subject cross session testing.

40 signs is higher than obtained from 80 signs. However, NN offers higher accuracy for 80 signs surprisingly. The results suggest our system is not suitable for inter-subject test. It is suggested that the system should be trained on each subject before using it to obtain a high accuracy.

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. Fig. 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 VIII 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 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.



(a)

(b)

Fig. 9. Sequence of postures when performing ‘Please’ and ‘Sorry’. (a). Sequence of postures when performing ‘Please’. (b). Sequence of postures when performing ‘Sorry’.

TABLE IX.
0 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%

D. *Significance of sEMG*

From the analysis of inter-subject testing in previous section, LibSVM achieves about 4% improvement for accuracy, precision, recall and F-score while the error rates for these metrics are reduced by about 50%. In this section, we further analyze the importance of sEMG. In American Sign Language, there are some signs which are very similar in arm movement and are different in hand shape and finger configurations (*e.g.* fist and palm). The sEMG is able to capture the difference of finger configuration and to distinguish these signs. If only inertial sensor is considered, the exactly same motion profile will make these signs confusing relative to each other. Fig. 9 shows an example of sequences of postures when the user is

have similar motion profile but different sEMG characteristics. Therefore, the sEMG is significant for our system.

VII. 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,

1

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.

Instead of average improvement, the improvement of true positive (TP) rate is analyzed to show how the sEMG impacts each individual sign. TP rate is rate of true positive and true positives are number of instances which are correctly classified as a given class. The improvement of TP rate of each sign with sEMG can tell how much sEMG will help for each individual signs. Fig. 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 IX, 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

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.

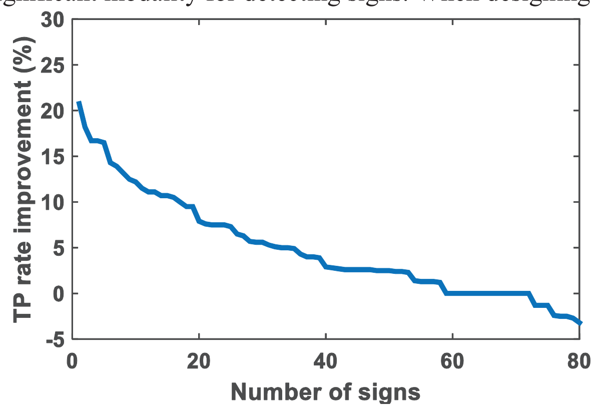


Fig. 10. TP rate improvement of all signs.

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.

VIII. Conclusion

A wearable real-time American Sign Language recognition system is proposed in our paper. This is a first study of American Sign Language recognition system fusing IMU sensor and sEMG signals which are complementary to each other. 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.

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] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] R. Merletti and P. Di Torino, “Standards for reporting emg data,” *J Electromyogr Kinesiol*, vol. 9, no. 1, pp. 3–4, 1999.

[29] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, “A novel feature extraction for robust emg pattern recognition,” *arXiv preprint arXiv:0912.3973*, 2009.

[30] 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.

[31] S. H. Khan and M. Sohail, “Activity monitoring of workers using single wearable inertial sensor,”

[32] 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.

[33] 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.

[34] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[35] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[36] 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.

[37] 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.

[38] M. Z. Jamal, “Signal acquisition using surface emg and circuit design considerations for robotic prosthesis,” 2012.

[39] 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.