SMART WEARABLE HAND DEVICE FOR SIGN LANGUAGE INTERPRETATION SYSTEM WITH SENSORS FUSION

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

SREELAKSHMI HARI

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering Kothamangalam

SMART WEARABLE HAND DEVICE FOR SIGN LANGUAGE INTERPRETATION SYSTEM WITH SENSORS FUSION

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

SREELAKSHMI HARI

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering Kothamangalam

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MAR ATHANASIUS COLLEGE OF ENGINEERING
KOTHAMANGALAM

CERTIFICATE

This is to certify that the report entitled **Smart Wearable Hand Device for Sign Language Interpretation System Using Sensors Fusion** *submitted by* **Ms.SREELAKSHMI HARI**, *Reg. No.* **MAC15CS056** *towards partial fulfillment of the requirement for the award of Degree of Bachelor of Technology in Computer science and Engineering from APJ Abdul Kalam Technological University for December 2018 is a bonafide record of the seminar carried out by him under our supervision and guidance.*

....	....	....
Prof. Joby George	**Prof. Neethu Subash**	**Dr. Surekha Mariam Varghese**
Faculty Guide	Faculty Guide	***Head of the Department***

Date: Dept. Seal

ACKNOWLEDGEMENT

First and foremost, I sincerely thank the God Almighty for his grace for the successful and timely completion of the seminar.

I express my sincere gratitude and thanks to Dr. Solly George, Principal and Dr. Surekha Mariam Varghese, Head Of the Department for providing the necessary facilities and their encouragement and support.

I owe special thanks to the staff-in-charge Prof. Joby George, Prof. Neethu Subash and Prof Joby Anu Mathew for their corrections, suggestions and sincere efforts to co-ordinate the seminar under a tight schedule.

I express my sincere thanks to staff members in the Department of Computer Science and Engineering who have taken sincere efforts in helping me to conduct this seminar.

Finally, I would like to acknowledge the heartfelt efforts, comments, criticisms, co-operation and tremendous support given to me by my dear friends during the preparation of the seminar and also during the presentation without whose support this work would have been all the more difficult to accomplish.

ABSTRACT

Sign language plays a vital role for deaf and mute people to communicate with normal people in a non verbal manner. This serve as an auxiliary tool for deaf and mute people to blend into society. To meet this purpose, the smart wearable device is introduced. It utilizes variety of sensors such as five flex sensors, two pressure sensors and three axis inertial motion sensors to distinguish the characters in American Sign Language Alphabet. These datas are analyzed using a support vector machine classifier. An Android based mobile application was developed with a text-speech function that converts the text into audible voice output. It demonstrates the usability of proposed smart wearable sign interpretation system in terms of comfort, flexibility and portability. The proposed system outperforms the existing method, for instance, although background lights, and other factors are crucial to a vision based processing method, they are not for proposed system.

Contents

List of Figures

List of abbreviation

IMU	Inertial Motion Unit
SVM	Support Vector Machine
ASL	American Sign Language
MMSE	Minimum Mean Square Error
PCA	Principal Component Analysis
SIFT	Scale invariance Fourier transform
LOO	Leave One Out
JEM	Jet Engine Modulation

Introduction

Sign language plays a vital role for deaf and mute people to communicate among themselves or with normal people in a non-verbal manner. Gestures are the primary method to convey messages, which are usually conducted in a three-dimensional space, known as a signing space[1], through an integration of manual and non-manual signals. Manual signals commonly correspond to hand motions and hand posturing, whereas non-manual signals correspond to an external appearance such as mouth movements, facial expressions, and body orientation. Nevertheless, sign language has not been standardized globally. Each nation has developed its own sign language,such as the American Sign Language (ASL) and Germany Sign Language (GSL). However, each sign language varies slightly within different regions of the same country. Hence, it can be a challenge to develop a standardized sign language interpretation system for use worldwide.Sign languages have been recognized using two major techniques, i.e., vision and non-vision approaches.

A direct approach detects the hand gestures based on the RGB color spaces of the skin color. For instance,identified Indian Sign Language (ISL) using the scale invariance Fourier transform (SIFT) algorithm by searching the matched key points between the input image and images stored in a database. A similar method was also applied.Using the SIFT algorithm, which further reduces the dimensions of the feature vector using a principal component analysis (PCA) algorithm for speeding up the processing time. To detect the dynamic hand gestures used in Japanese Sign Language (JSL) proposed the use of recurrent neural networks capable of recognizing the JSL finger alphabet, which has 42 symbols. In contrast, used a simple scanning method to compute the orientation and movement of fingers in binary converted images captured from a web camera proposed a more sophisticated gesture recognition system using digital images, including image filtering (pre-processing), image segmentation, color segmentation, skin detection (finger and hand detection using binary images), and template matching.Indirect approach identifies the fingers and hand gestures based on the RGB color spaces segmented based on different colors for each finger using a data glove. A possible segmentation method using RGB color spaces along with a hand glove for gesture detection was proposed. This method utilizes the RGB color space values extracted from a captured hand gesture image of a data glove, and compares the values with those stored in a database.

The exploitation of a vision-based method is greatly affected by the processing of the images, such as image filtering, background cancellation, color segmentation, and boundary detection. For instance, diverse and uncontrolled background images can influence the skin color segmentation or movement detection. Indeed, many researchers have failed to address these complications, and no solid solutions have yet been proposed. Consequently, a non-vision based method is an alternative approach. This method typically utilizes flex and motion sensors to measure the flexion of fingers and the orientation of the hand, respectively,attached five flex sensors on a glove with respect to each finger,to identify hand gestures by matching the motions with those in a stored motion database,similar technique of using flex sensors, but mapped the sensor data to a character set, which was implemented using a minimum mean square error (MMSE) algorithm for gesture recognition. The results are displayed as text on an LCD screen. This technique was improved by the bending of each sensor is further divided into three flexions,namely, a complete bend (finger close), partial bend, and straightening (finger open). Each ASL alphabet is then mapped according to the bend flexions to be used for template matching.

1.1 Minimum Mean Square Error

In statistics and signal processing, a minimum mean square error (MMSE) estimator is an estimation method which minimizes the mean square error (MSE), which is a common measure of estimator quality, of the fitted values of a dependent variable. In the Bayesian setting, the term MMSE more specifically refers to estimation with quadratic loss function. In such case, the MMSE estimator is given by the posterior mean of the parameter to be estimated. Since the posterior mean is cumbersome to calculate, the form of the MMSE estimator is usually constrained to be within a certain class of functions. Linear MMSE estimators are a popular choice since they are easy to use, easy to calculate, and very versatile. It has given rise to many popular estimators.

The term MMSE more specifically refers to estimation in a Bayesian setting with quadratic cost function. The basic idea behind the Bayesian approach to estimation stems from practical situations where we often have some prior information about the parameter to be estimated. For instance, we may have prior information about the range that the parameter can assume; or we may have an old estimate of the parameter that we want to modify when a new observation is made available; or the statistics of an actual random signal such as speech. This is in contrast to the non-Bayesian approach like minimum-variance unbiased estimator (MVUE) where absolutely nothing is assumed to be known about the parameter in advance and which does not account for such situations. In the Bayesian approach, such prior information is captured by the prior probability density function of the parameters; and based directly on Bayes theorem, it allows us to make better posterior estimates as more observations become available. Thus unlike non-Bayesian approach where parameters of interest are assumed to be deterministic, but unknown constants, the Bayesian estimator seeks to estimate a parameter that is itself a random variable. Furthermore, Bayesian estimation can also deal with situations where the sequence of observations are not necessarily independent. Thus Bayesian estimation provides yet another alternative to the MVUE. This is useful when the MVUE does not exist or cannot be found.

The study observed that only a single channel of sEMG located on the wrist is sufficient for the ASL recognition. Likewise, Wu et al. [30] fused the information from an inertial sensor and EMG sensor which are placed on a wearable system to recognize 80 commonly ASL signs with selected feature subset and processed by a support vector machine classifier.This study aims at the development of a sign language interpretation system by analyzing hand and finger gestures from a smart wearable device. The finger gestures are observed through the flexion of the flex sensors, whereas the hand gestures are examined based on the hand motion through the orientation derived from an inertial motion sensor. The gestures are recognized using a support vector machine (SVM) modelimplemented in the wearable device. The gestures are then received using our developed mobile application through a wireless Bluetooth transmission, and text is displayed on the mobile device screen. Moreover, a text-to-speech service is also available in the mobile application, which instantly converts the received texts into audible outputs.

Related Works

Traditionally, the technology of gesture recognition was divided into two categories, vision -based and glove-based methods and also the colored marker approaches. In vision-based methods, computer camera is the input device for observing the information supplied by various gestures of hands fingers. In glove based systems data gloves are used which can archive the accurate positions of hand gestures as its positions are directly measured, reviewed various methods and techniques which are provided by different authors for recognition of hand gestures. Hand gesture recognition is carried out by the methods like pixel by pixel comparison,edge method, using orientation histogram thinning method. Present technologies for recognizing gestures are divided into vision based,data glove based, and colored marker approaches. A novel gesture spotting algorithm which is accurate and efficient, is purely vision-based. This system robustly recognizes gestures, even when the user gestures without any aiding devices in front of a complex background.The Prototype version, in which system the user forms a sign and holds it for two seconds to ensure recognition.Real-time hidden Markov model based systems for recognizing sentence-level continuous American Sign Language (ASL) using a single camera to track the users unadorned hands.

An instrumented data glove approaches system in which sensor devices used for capturing hand position, and motion. In this approach, detection of hand is eliminated by the sensors on the hand and it can easily provide exact coordinates of palm and fingers location and orientation, and hand configurations. However using data gloves become a better approach than camera as the user has the flexibility of moving the hand around freely, unlike the camera where the user has to stay in position before the camera. Light, electric or magnetic fields or any other disturbance does not affect the performance of the glove and the proposed a method where a glove generates commands based on position measurements. When the angles of the fingers change the output of the sensors will change. The combined sensor outputs form a pattern that corresponds to different finger flexions. Different finger flexions generates different Commands. The glove is simple and it generates sufficient signals for a fuzzy control system.

2.1 Description

Generally dumb people use sign language for communication but they find difficulty in communicating with others who dont understand sign language. The aim of proposed system is to reduce the barrier in communication. Now days its a need of developing an electronic support

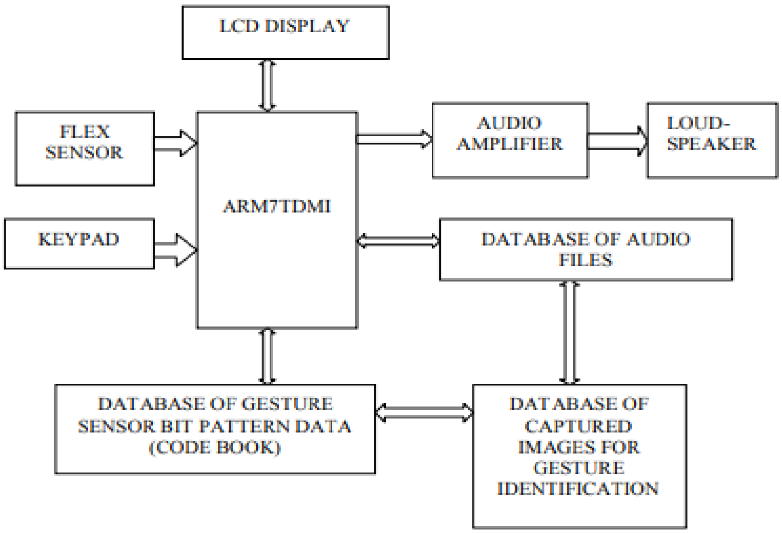


Fig. 2.1: Model for Electronic Support System

system that can translate sign language into text and speech in order to make the communication between the deaf dumb communities with the general public. A data gloves is used which is normal cloth driving gloves fitted with flex sensors along the length of each finger and the thumb. Deaf dumb people will use the gloves to perform hand gesture and it will be converted into text and speech so that normal people can understand their expression.

In this electronic support system data glove is implemented to capture the hand gestures of a user. s the major role, Flex sensors are sensors that change in resistance depending on the amount of bend on the sensor, The data glove is fitted with flex sensors along the length of each finger and the thumb. The flex sensors output is a stream of data that varies with degree of bend. The analog outputs from the sensors then fed to the ARM7TDMI[9]. It processes the signals and perform analog to digital signal conversion. The resulting digital signal is fed to the gesture recognition. In this section the gesture is recognized and the corresponding text information identified. Text to speech conversion takes place in the voice section and play out through the speaker.

Flex sensors are fitted on hand gloves. As per the hand gesture movement it will bend the flex sensors of all fingers. The value of bending is in resistance.All fingers give different resistance value depending on bending. The output of flex sensor is given to the ADC of ARM7TDMI LPC-2148 which used to convert analog signal into digital signal. In the proposed support system we have recognized alphabets (A-Z), Numbers (0-9) and some words by using hand gloves with flex sensor. The proposed support system is useful for communication between deaf and dumb people with normal person.

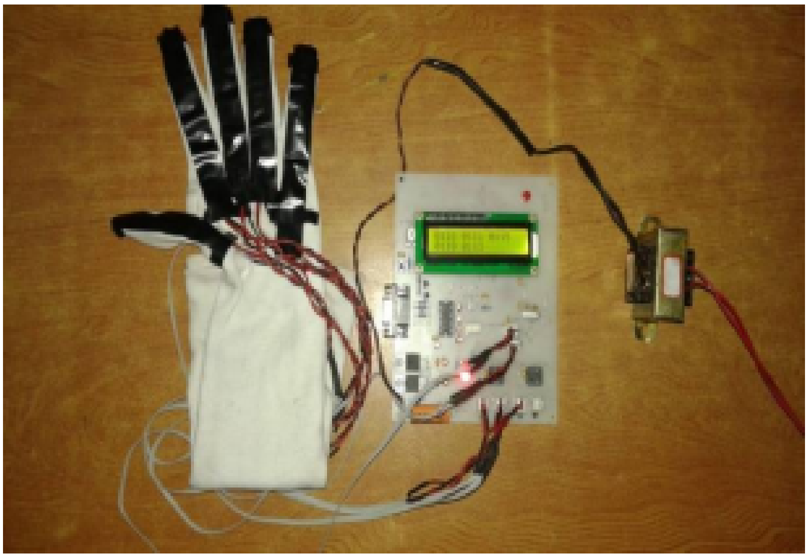


Fig. 2.2: Hardware components

The frequency of terahertz radar ranges from 0.1 THz to 10 THz, which is situated between microwaves and infrared waves. High-resolution range profiles and Doppler signatures can be achieved easily due to the high frequency of terahertz waves. As a result, a terahertz wave has advantages with regards to target detection and recognition. With the development of terahertz radar systems, terahertz imaging, and detection [4], there has been much interest in the study of terahertz radar. However, the research on practical applications is rare and gesture recognition using terahertz radar is still an unexplored field.A great deal of the research into hand-gesture recognition is based on computer vision and contact-based gesture classifications[5]. The performance of vision-based approaches depends strongly on lighting conditions. Contact-based gesture recognition demands individuals to be accustomed with the usage of the interface device, which is not adaptable to new users. As a result, the application of vision-based and contact-based gesture recognition has many limits. In contrast, terahertz radar can not only provide full-time observation of targets, but can also work without wearable devices. In addition, terahertz radar can be used for speed and distance detection. It is applicable to the recognition of hand gestures by detecting changes in distance and speed. In recent years, centimeter-wave radar (frequency in the 330 GHz range) has been used in gesture recognition systems. However, large-scale motion has usually been studied, which is confined to low resolution. Very recently, a gesture recognition system Soli was designed for some specific usage scenarios based on millimeter-wave radar with a short range (frequency in 60 GHz). Compared with the lower frequency radar system, the terahertz radar has a higher carrier frequency. Its easy for terahertz radar to achieve wider bandwidth and provide better range resolution, which can precisely capture a change of hand gesture. Information for tracking, Jet Engine Modulation (JEM), polarization, Doppler shifts, HRRP, and radar images are usually utilized to perform target recognition.We focus on gesture recognition for terahertz radar using multi-modal signals. Multi-modal signals in terahertz systems include HRRP and Doppler signatures. A range profile of terahertz radar is actually a one-dimensional terahertz radar image. Since terahertz radar has sufficient bandwidth, the shape of the returned wave from a target can easily describe the geometric shape and structure of a target. As a result, a change in targets will definitely lead to a change in range profiles. In addition to the change in the target itself, aspect changes are shown in dynamic gestures. Furthermore, a range profile of a single aspect is sensitive to aspect changes . Therefore, range profiles have been widely used in the target recognition. However, most of the previous studies focused on the target itself. HRRP sequences in continuous time reflect movement characteristics, but gesture recognition is rarely discussed. On the other hand, terahertz radar systems used the theory of Doppler speed detection to measure the offset of a frequency. Doppler signatures obtained from terahertz radar are the velocity information of the target motion, which can be used in the hand-gesture recognition field. Multi-modal signals, including HRRP and Doppler signatures in terahertz radar systems, provide target information for both image and velocity. This property allows terahertz radar signals to carry much more information than single sensors, such as camera, infrared sensors, data gloves, and so on. Since gesture recognition has its advantages, thanks to the characteristics of terahertz waves, gesture recognition represents a promising future development for terahertz radar systems. To our knowledge, there have been no reports regarding gesture recognition in the terahertz region.

Communication via gestures as a sort of signal is a standout amongst the most characteristic methods for exchanging information for most hard of hearing individuals. The objective of gesture based communication, sign language recognition (SLR) is to give a productive and exact mechanism to translate gesture based communication into the content or discourse so that correspondence amongst hard of hearing and listening to society can be more helpful. SLR, as unitary of the important research areas of human-computer interaction (HCI), Being as mind boggling as any talked language, SL has numerous signs, framed by outward appearances and important signals, including physical developments of the fingers, hands, wrists, arms, and head, each contrasting from another by minor changes close by movement, shape, position, and outward appearance. Appropriately, a communication via gestures can be considered as a gathering of important and easy to use hand motions, developments, and stances. Hand motion acknowledgment is the most ordinarily utilized methodology among other correspondence modalities in humancomputer communication.

Dynamic hand motion correspondence is a more common and humanoid mode of correspondence with PCs, as contrasted and static hand motion. Dynamic sign language recognition for smart home interactive application using stochastic linear formal grammar. A phone based video calling framework for American sign language requires real time catch, encoding, and disentangling of advanced video on a cell gadget for transmission over the U.S. cell system. Cell gadgets have constrained handling force and battery life, forcing limitations on the multifaceted nature of the encoding and translating calculations. A computational intelligibility model for assessment and compression of american sign language video. . As per a standard Korean sign language (KSL) word reference, the 45-year-old Korean gesture based communication contains around 6,000 vocabulary words. Be that as it may, they are framed by consolidating a generally little number of essential motions. In addition, two sorts of motions of hands and fingers are utilized: one is static and the other is dynamic gestures. The previous comprises 31 particular stances communicating the dactylology while the last is made up with evolving designs, constituting the primary body of the KSL and communicating distinctive implications of vocabulary words.

The most regularly utilized measurements are hand shape/introduction, changes fit as a fiddle/introduction, hand area, developments of hand areas, hand-hand touching, hand-body touching (for the most part particular areas on the face), lip developments, outward appearance, and middle/shoulder posture and developments. Moreover, as a rule, connection is key to extraordinarily characterize the importance of a sign. In any case, that does not as a matter of course imply that the dynamical parts of gesture based communication have the same conduct and play the same semantic part as flow in talked languages. No less than three critical qualifications must be considered. Most importantly, the one-dimensionality of discourse makes it consecutive in nature.

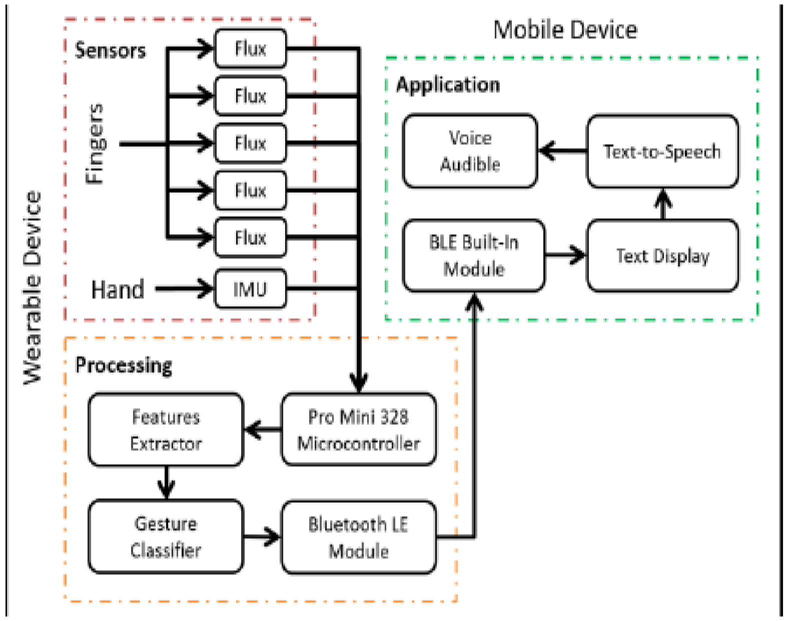
The Proposed Method

The wearable device holder is printed using flexible filaments with good elasticity. These filaments enable functional hinges, joints, and shaped parts, allowing the device to fit different hand sizes. Five finger holders were also designed using a flexible filament placed on the first joint of each finger to hold the flex sensors. Similarly, these flexible holders can also accommodate different finger sizes of different users. Finger gestures are exploited through the flexion of flex sensors placed on the top of the finger. The flex sensors used in this study are either 4.5 inches or 2.2 inches in size. The shorter length flex sensor is suitable for the pinky finger, whereas the longer length flex sensor is used for the other four fingers. The flex sensor for the thumb has a longer connection distance to the microcontroller board as compared to the other fingers, and thus a longer length flex sensor is used instead of a shorter length one. In fact, the flex sensor consists of both omnidirectional and bidirectional types. The omnidirectional flex sensor changes its resistance when it bends in one direction only, whereas the bidirectional type changes its resistance when it bends in both the upward and downward directions. This study utilizes a bidirectional type, and each flex sensor is connected with a 10 k resistor.

The proposed sign interpretation system is divided into three distinct modules: a sensor module, processing module, and application module. The sensor and processing modules are implemented in the smart wearable hand device, whereas the application module operates in an Android-based mobile device. The flex sensor and IMU data are collected using an Arduino Pro Mini 328, which is operated at 5 V with an ATmega328 processor running at 16 MHz with an external resonator (0.5 percent tolerance). The features are extracted from the sensor data and serve as inputs to the built-in SVM classifier to determine the sign language alphabet letters. In this study, there are a total of 28 gesture patterns, which refer to the 26 alphabet letters of ASL, a neutral state, and an invalid sign. The detected sign is translated into text and transmitted to a mobile device using a Bluetooth 4.0 module, i.e., Bluetooth low energy. The received text is displayed on the mobile screen.Concurrently, the automated text-to-speech service translates the text into an audible voice to be played by the built-in speaker of the mobile device.

3.1 Flex Sensors

Bend sensors have been developed based on conductive ink, optical fiber, and electronic textiles. Each type has advantages and disadvantages in terms of performance, ease of use, and cost. This study proposes a new and low-cost bend sensor that can measure a wide range of accumulated bend angles with large curvatures. This bend sensor utilizes a Bowden-cable,



which consists of a coil sheath and an inner wire. Displacement changes of the Bowden-cables inner wire, when the shape of the sheath changes, have been considered to be a position error in previous studies. However, this study takes advantage of this position error to detect the bend angle of the sheath. The bend angle of the sensor can be calculated from the displacement measurement of the sensing wire using a Hall-effect sensor or a potentiometer. Simulations and experiments have shown that the accumulated bend angle of the sensor is linearly related to the sensor signal, with an R-square value up to 0.9969 and a root mean square error of 2percent of the full sensing range. The proposed sensor is not affected by a bend curvature unlike previous bend sensors. The proposed sensor is expected to be useful for various applications, including motion capture devices, wearable robots, surgical devices, or generally any device that requires an affordable and low-cost bend sensor.

3.2 Implementation

Twelve subjects were recruited from a university campus and participated in the experiments voluntarily. The participating subjects do not possess any muscular diseases or neuromuscular disorders that would affect their sign gestures. The subjects were requested to perform all sign gestures for 20 times with each approximately 10 s each. In addition, a neutral state was also recorded in which all of their fingers were wide open regardless of the hand orientation. All data were recorded during the experiments using a desktop computer application.

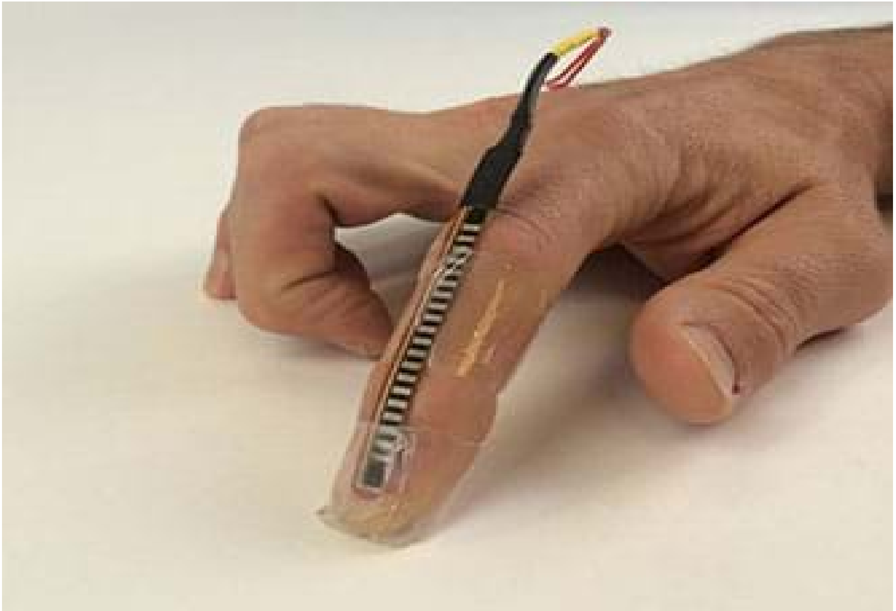


Fig. 3.1: Flex Sensors

The data were saved in a text file format for easy accessibility. The computer application is not described in detail herein because it only served as a recording device. Twelve subjects were recruited from a university campus and participated in the experiments voluntarily. The participating subjects do not possess any muscular diseases or neuromuscular disorders that would affect their sign gestures. The subjects were requested to perform all sign gestures for 20 times with each approximately 10 s each.In addition, a neutral state was also recorded in which all of their fingers were wide open regardless of the hand orientation. All data were recorded during the experiments using a desktop computer application. The data were saved in a text file format for easy accessibility. The computer application is not described in detail herein because it only served as a recording device.

3.3 Preprocessing

Throughout the experiments, it was observed that the flexion values varied among the different subjects owing to the different hand sizes. A smaller hand size has a lower discrepancy in terms of the flex sensor values with respect to particular signs. In other words, the resistance of the flex sensors for a smaller hand has less variation as compared to the resistance with a larger hand. Thus, the differences among the signs are not discernable from the raw flex sensor values. To solve this issue, the sensor values were normalized based on the computed mean and standard deviation of each flex sensor for each subject into a range of [0, 1].

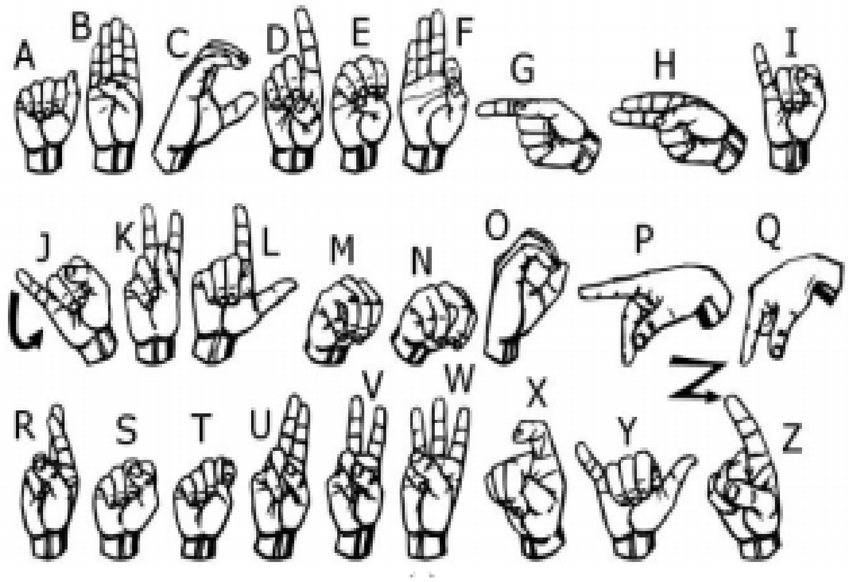


Fig. 3.2: Hand Gestures

The grasping of fingers into fist shape generate different maximum flex sensors values. Thus, the mean and SDs of flex sensors are computed separately for each subject based on the subjects sensors readings. This method eliminated the necessity for the new user to perform all the sign gestures before the new user started to use the proposed system. Meanwhile, each hand movement was derived from the IMU in three orientations: pitch, roll, and yaw. The calculation of each orientation is similar depending on the axis used. The angle agli of each axis is computed using a complementary filter method at the i-th time,

$$agli = 0.98(agli-1+gyri/gyrHZ)+aglci0.02, aglci = arctan(Au, Av)180/\pi \quad \text{(equ 3.1)}$$

where gyri, gyrHZ , and aglci are the raw gyroscope sensor reading, gyroscope sensor sampling rate, and the angular acceleration speed at the i-th time In addition, Au, and Av are the coordinates of the exclusive raw linear accelerometer readings that do not correspond to the angles being computed, e.g., to compute the pitch angle, Au and Av are denoted as the y-axis and z-axis of the raw linear accelerometer readings, respe ctively. The multiplier constant in converts the angle from radians into degrees () for ease of analysis. Further details of this computation can be found in our previous related study. The flex sensors and IMU data are gathered at a sampling rate of 100 Hz.

3.4 Feature Extraction

To simplify and optimize the coding implementation, a vector of the flexion degree in tabular format was considered. The flexion degree is split into three regions in each vector. The first region is denoted as no bend or slight bend, which is associated with a normalized flexion value within the range of [0.0, 0.3). The second region is considered as a partial bend with the associated normalized flexion value within the range of [0.3, 0.7), and the last region is a complete bend with associated normalized flexion value within the range of [0.7, 1.0]. These regions are abbreviated in order as OR (open region), PR (partially open or closed region), and CR (closed region). Table 1 shows a mapping of these regions for all 26 alphabet letters in ASL, as well as a neutral gesture.

It can be seen that some of the signs exhibit the same regions, for instance, the letters U and V, K and P, and I and J, as well as G, L, Q, and Z. To further distinguish these signs, sensor level fusion is adopted. In this study, recognition with only flex sensors and inertial sensor are denoted as 1st version and addition of pressure sensors for sensor level fusion are designated as 2nd version. To support the sensor fusion, two Flexiforce pressure sensors were considered. The resistance of a Flexiforce pressure sensor is reduced when the surface is pressed, but the resistance does not change while being flexed. In fact, the resistance changes only when the pressure is applied to the round area at the end of the sensor, and ranges from 0 to 25 lbs of pressure. These sensors are placed below and on the left side (right side for a left-handed user) of the first joint of the middle finger. The pressure sensors are connected to the digital inputs in processing module which produced reading of 0 if pressure sensor surface is not pressed, and produced reading of 1 if the pressure is sensed on the surface of the sensor. The inclusion of the pressure sensors on the middle finger managed to solve the issues for the letters U and V, whereas the second pressure sensor resistance value is high for letter U and low for the letter V. Moreover, the letters G, L, and Q can also be distinguing shed based on the resistance of the respective pressure sensors.

However, the letters L and Z still exhibit a similar pattern and can be differentiated through a hand motion. The letter L has static motion, whereas letter Z has dynamic motion over time. The same behavior is applied to discriminate the letters I and J. Finally, the letters K and P can only be differentiated based on the hand orientation. To determine the hand motion, standard deviation (SD) of the angular reading from the IMU sensors is computed for each axis. The SD is a measure that is used to quantify the amount of variation or dispersion of the motion readings. For instance, the preliminary results had indicated that the mean SDs of letter I are approximately 0.0369 (pitch), 0.0151 (roll), and 0.0375 (yaw) while the mean SDs of letter J are 1.3968 (pitch), 0.5603 (roll), and 1.1548 (yaw) respectively. Thus, in this study, the computed SDs are sufficient to observe the occurrence of hand motions.

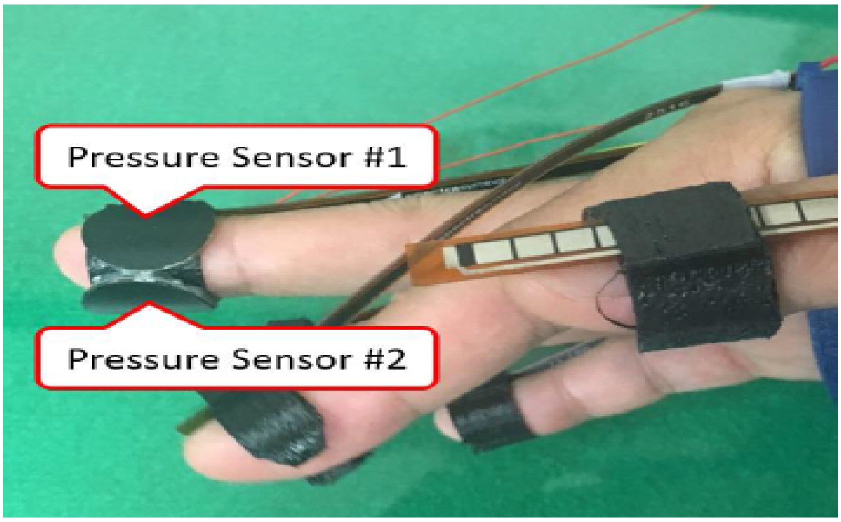


Fig. 3.3: Pressure Sensors

3.5 Accelerometer Sensor

Accelerometers are electromechanical devices that sense either static or dynamic forces of acceleration. Static forces include gravity, while dynamic forces can include vibrations and movement.Accelerometers can measure acceleration on one, two, or three axes. 3-axis units are becoming more common as the cost of development for them decreases.With an accelerometer you can either get a really ”noisy” info output that is responsive, or you can get a ”clean” output that’s sluggish. But when you combine the 3-axis accelerometer with a 3-axis gyro, you get an output that is both clean and responsive in the same time.”Generally, accelerometers contain capacitive plates internally. Some of these are fixed, while others are attached to minuscule springs that move internally as acceleration forces act upon the sensor. As these plates move in relation to each other, the capacitance between them changes. From these changes in capacitance, the acceleration can be determined.Other accelerometers can be centered around piezoelectric materials. These tiny crystal structures output electrical charge when placed under mechanical stress (e.g. acceleration).

3.6 Pressure Sensors

A pressure sensor is a device for pressure measurement of gases or liquids. Pressure is an expression of the force required to stop a fluid from expanding, and is usually stated in terms of force per unit area. A pressure sensor usually acts as a transducer; it generates a signal as a function of the pressure imposed. For the purposes of this article, such a signal is electrical. Pressure sensors are used for control and monitoring in thousands of everyday applications. Pressure sensors can also be used to indirectly measure other variables such as fluid/gas flow, speed, water level, and altitude. Pressure sensors can alternatively be called pressure transducers, pressure transmitters, pressure senders, pressure indicators, piezometers and manometers, among other names.Pressure sensors can vary drastically in technology, design, performance, application suitability and cost. A conservative estimate would be that there may be over 50 technologies and at least 300 companies making pressure sensors worldwide.There is also a category of pressure sensors that are designed to measure in a dynamic mode for capturing very high speed changes in pressure. Example applications for this type of sensor would be in the measuring of combustion pressure in an engine cylinder or in a gas turbine. These sensors are commonly manufactured out of piezoelectric materials such as quartz.

This sensor measures the difference between two pressures, one connected to each side of the sensor. Differential pressure sensors are used to measure many properties, such as pressure drops across oil filters or air filters, fluid levels (by comparing the pressure above and below the liquid) or flow rates (by measuring the change in pressure across a restriction). Technically speaking, most pressure sensors are really differential pressure sensors; for example a gauge pressure sensor is merely a differential pressure sensor in which one side is open to the ambient atmosphere.

3.7 Sign Classifier

The signs are classified into 28 classes using a support vector machine (SVM). An SVM is a binary supervised learning classifier, that is, the class labels can only take the values of +1 and -1. The training procedure used a quadratic optimization algorithm to derive structural axes to separate the training dataset into n numbers of a hyperplane. Assume the i-th training sample using

(xi, yi), yi 1,+1, i = 1,2,3, ... , n,

where xiis the feature vector and yi is the training label in accordance to the feature vectors of the i-th training datasets. The decision boundary is defined as

$$f(x) = w \cdot x - b \qquad \text{(equ 3.2)}$$

where the i-th feature is classified as positive (+1) if f(x) ¿ 0, and negative (-1) if f(x) ¡ 0. The separating hyperplane line is structured at f(x) = 0 . The points positioned around the separating hyperplane line are known as support vectors (SVs) and their distance to the hyperplane line is known as the margin.

However, in this study, there are more than two classes that are being classified. To obtain an M-class of the SVM classifiers, a set of binary classifiers need to be constructed, where each binary classifier is trained to distinguish one class from the rest. The results are then integrated to form a multi-class classification according to the maximal output of each binary classifier xj, which is also known as the confidence value and j is referred to each alphabet binary classifier. Thus, x belongs to the class with the largest confidence value. In fact, there will be gestures that belong to none of the sign classes or the aforementioned neutral class in a real-world application, and thus any gesture with a confidence value of less than 0.5 (50 percent) is considered as an invalid class. In this study, a feature vector x is built using a total of ten features, which are the normalized flex sensor data for five fingers, two pressure sensor data, and finally, three computed SDs of angular readings from the IMU sensor. Sliding windows of 3 s are adopted to construct the feature vector for every 10 s sensor data in order to accommodate the changes in hand movement within a 3 s time period, particularly for the letters J and Z. The values of the ten features in 3 s time period are averaged for the feature vector construction. There are total of 6,480,000 datasets (12 subjects 20 times 10 s 100 Hz 27 signs) collected in this experiments.

The SVM is trained and tested using the leave one subject out (LOO) method. Here, the SVM is trained using n 1 subject datasets, where n is the total number of subjects. Subsequently, the trained model is tested using the leave-out subject dataset that did not participate in the training process. This process was repeated n times, where each subject was treated as a leave-out subject once when the testing dataset. The accuracy of each trained model

$$AC = TP + TN/TP + TN + FP + FN \qquad \text{(equ 3.3)}$$

where TP, TN, FP, and FN refer to a true positive (the number of signs correctly identified), true negative (the number of signs correctly rejected), false positive (the number of signs incorrectly identified), and false negative (the number of signs incorrectly rejected).

The overall accuracy of the SVM is eventually averaged. From the cross-validation (CV) model selection point of view, concluded that LOO method is best to utilize for model training with high computational complexity which is proportional to the number of data splits. Thus, it is suitable method to estimate the risk when learning a model.

3.8 Support Vector Machine

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

SVMs can be used to solve various real world problems: SVMs are helpful in text and hypertext categorization as their application can significantly reduce the need for labeled training instances in both the standard inductive and transductive settings. Classification of images can also be performed using SVMs. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes after just three to four rounds of relevance feedback. This is also true of image segmentation systems, including those using a modified version SVM that uses the privileged approach as suggested by Vapnik.[1] Handwritten characters can be recognized using SVM[8][citation needed]. The SVM algorithm has been widely applied in the biological and other sciences. They have been used to classify proteins with up to 90 percent of the compounds classified correctly. Permutation tests based on SVM weights have been suggested as a mechanism for interpretation of SVM models. Support vector machine weights have also been used to interpret SVM models in the past.Posthoc interpretation of support vector machine models in order to identify features used by the model to make predictions is a relatively new area of research with special significance in the biological sciences.

In machine learning and statistics, classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. Examples are assigning a given email to the ”spam” or ”non-spam” class, and assigning a diagnosis to a given patient based on observed characteristics of the patient (sex, blood pressure, presence or absence of certain symptoms, etc.). Classification is an example of pattern recognition.

In the terminology of machine learning, classification is considered an instance of supervised learning, i.e., learning where a training set of correctly identified observations is available. The corresponding unsupervised procedure is known as clustering, and involves grouping data into categories based on some measure of inherent similarity or distance.

Often, the individual observations are analyzed into a set of quantifiable properties, known variously as explanatory variables or features. These properties may variously be categorical (e.g. ”A”, ”B”, ”AB” or ”O”, for blood type), ordinal (e.g. ”large”, ”medium” or ”small”), integer-valued (e.g. the number of occurrences of a particular word in an email) or real-valued (e.g. a measurement of blood pressure). Other classifiers work by comparing observations to previous observations by means of a similarity or distance function. An algorithm that implements classification, especially in a concrete implementation, is known as a classifier. The term ”classifier” sometimes also refers to the mathematical function, implemented by a classification algorithm, that maps input data to a category.

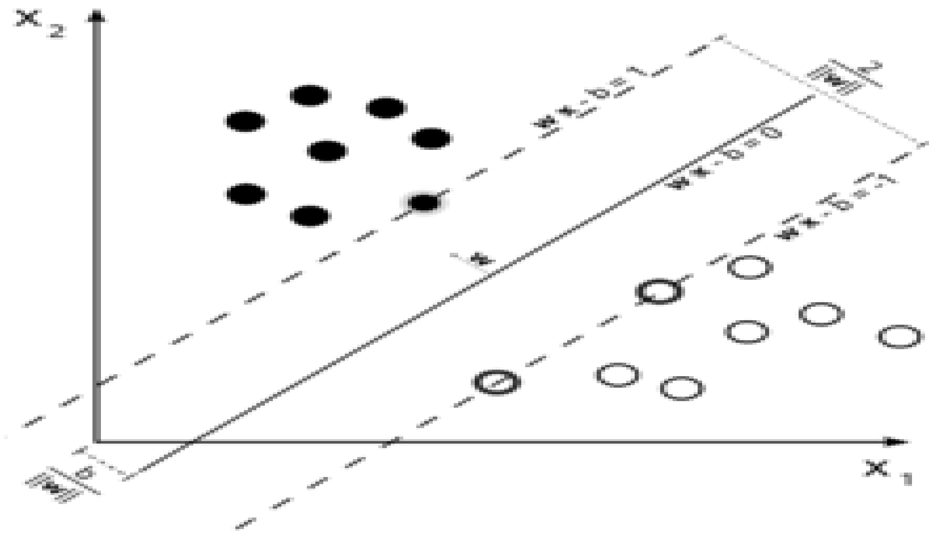
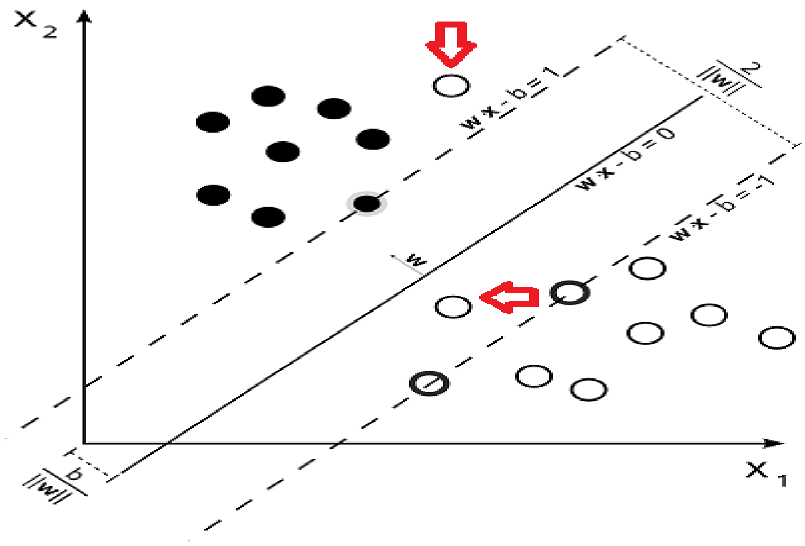


Fig. 3.4: Hyperplanes

In order to be able to use this decision rule for the classification of company j, the SVM has to learn the values of the score parameters w and b on a training sample. Assume this consists of a set of n companies i = 1, 2, ,n. From a geometric point of view, calculating the value of the parameters w and b means looking for a hyperplane that best separates solvent from insolvent companies according to some criterion. The criterion used by SVMs is based on margin maximization between the two data classes of solvent and insolvent companies. The



margin is the distance between the hyperplanes bounding each class, where in the hypothetical perfectly separable case no observation may lie. By maximising the margin, we search for the classification function that can most safely separate the classes of solvent and insolvent companies. The graph below represents a binary space with two input variables. Here crosses represent the solvent companies of the training sample and circles the insolvent ones. The threshold separating solvent and insolvent companies is the line in the middle between the two margin boundaries.In a non-perfectly separable case the margin is soft. This means that in-sample classification errors occur and also have to be minimized. Let i be a non-negative slack variable for in-sample misclassifications. In most cases i =0, that means companies are being correctly classified. In the case of a positive i the company i of the training sample is being misclassified. A further criterion used by SVMs for calculating w and b is that all misclassifications of the training sample have to be minimized.

3.9 Result

The sensor value of index finger for the alphabetA (red), B (green),and C (blue) respectively. It was clearly observed that the flexion region for alphabet A of index finger is at CR region which is verified. The similar results also applied to the alphabet B (in OR region) and alphabet C(PR). The signs of the classification results for versions of the proposed system. The results clearly signified a large difference of 32.5 percent between the first and second versions of the system. The accuracy in the first version was low owing to similar patterns occurring among several of the signs, which caused negative classifications. The accuracy improved significantly after the pressure sensors were added to the system. However, there are still minor misclassifications in the second version of the system. An analysis indicated that the incorrect pattern recognitions for all subjects occurred more commonly between the letters E and S. This is due to the significantly lower differences in the flexion values for subjects with a smaller hand size. Thus, the system misinterpreted the thumb region as a PR instead of a CR, or vice versa. A similar issue appeared for letters M and N as well, with an incorrectly identified region for the ring finger. It was noticed that even though the flexion sensor of the sensor value of ring finger for alphabet N had higher value than the value for alphabet M, they are both fall into same PR region which produce the false classification between both alphabets.

The alphabet R, U, and V. The accuracy rate of signs recognition for alphabet U increased significantly from mean AC of 57.25to 97.52 percent when two pressure sensors data are included for the classification. Likewise, the mean AC for alphabet R and V increased dramatically from 57.78 to 97.28 percent and 57.22 to 97.49 percent respectively.Even though the AC increased, but still, there was a slight misclassification between the letters R and U, where subjects with thicker fingers tended to touch the surface of the second pressure sensor for the letter R, e.g., subjects 4, 5, and 12. On the contradictory, subjects with thinner fingers did not tend to touch the surface of the second pressure sensor.Nevertheless, the other signs did not incur this same issue. The inclusion of the first pressure sensor surface showed significant differences for the signs between the letters U and V.The differences of alphabet I and J in term of hand motion. The result indicated that there was no motion in pitch, roll, and yaw angles when performing gesture for the alphabet I throughout the time. Meanwhile, the action of performing the sign gesture for alphabet J showed dynamic changes of angular which are distinguishable from the alphabet I.

On the other hand, proposed system with previous existing methods. In fact, there have been a number of studies over the past decades focusing on gesture recognition using image processing techniques.Many sophisticated algorithms have been proposed to distinguish gestures patterns, e.g., neural networks and decision trees. As semiconductor and electronic component technologies advance, hardware components of a smaller size, higher performance, and lower power consumption have created alternate methods for gesture recognition. Consequently, a data glove with small-scale sensors has been introduced. Recent studies have shifted focus from using image techniques to using a flex sensor, motion sensor, tilt sensor, or optical sensor for gesture recognition. Moreover, current studies are literally focusing solely on alphabets, numbers, and simple gestures. The average accuracy of sign or gesture recognition observed from these studies is over 90 percent for letters and numbers at the language level, but is slightly lower for more complicated gesture patterns. Conclusively, although the wearable device proposed in this study is currently only targeted at the alphabet level, its accuracy is significantly higher

TABLE I
VECTOR FOR FLEXION VALUES SEPARATED BY REGIONS

Alphabets	TB	IN	MD	RG	PK	PS1	PS2	HM
A	OR	CR	CR	CR	CR	O	O	X
B	PR	OR	OR	OR	OR	X	O	X
C	OR	PR	PR	PR	PR	X	O	X
D	PR	OR	CR	CR	CR	O	O	X
E	CR	CR	CR	CR	CR	O	O	X
F	PR	PR	OR	OR	OR	X	O	X
G	OR	OR	CR	CR	CR	O	O	X
H	OR	OR	OR	CR	CR	X	O	X
I	PR	CR	CR	CR	OR	O	O	X
J	PR	CR	CR	CR	OR	O	O	O
K	OR	OR	OR	CR	CR	X	X	X
L	OR	OR	CR	CR	CR	X	X	X
M	OR	PR	PR	PR	CR	O	O	X
N	OR	PR	PR	CR	CR	O	O	X
O	PR	PR	PR	PR	PR	X	O	X
P	OR	OR	OR	CR	CR	X	X	X
Q	OR	OR	CR	CR	CR	X	O	X
R	OR	OR	OR	CR	CR	O	X	X
S	PR	CR	CR	CR	CR	O	O	X
T	OR	PR	CR	CR	CR	O	O	X
U	PR	OR	OR	CR	CR	X	O	X
V	PR	OR	OR	CR	CR	X	X	X
W	PR	OR	OR	OR	CR	X	X	X
X	PR	PR	CR	CR	CR	X	X	X
Y	OR	CR	CR	CR	OR	O	O	X
Z	OR	OR	CR	CR	CR	X	X	O
NT	OR	OR	OR	OR	OR	X	X	X

TB: Thumb finger
IN: Index finger
MD: Middle finger
RG: Ring finger
PK: Pinky finger
PS1: Pressure Sensor 1
PS2: Pressure Sensor 2
HM: Hand in motion
OR: Open region
PR: Partially open/close region
CR: Close region
O: Yes
X: No

Fig. 3.5: FLEXION VALUES OF THE REGIONS

3.10 Sign Interpretatio Application

The classified sign gestures from the proposed smart wearable device is transmitted to the sign interpretation system. The sign interpretation system was built on an Android-based mobile device, a Google Nexus 6p [10]. First, the application searches for available Bluetooth devices and displays the search results through a card list. Next, the selected Bluetooth device (HMSoft) initiates a wireless connection with the proposed wearable device. Once the connection is successfully established, the application will start to receive data in a text format, ranging from A to Z, or in a neutral state, i.e., NT, or as an invalid sign, i.e., INV,. The received text is displayed at the center of the screen, whereas the past history of received texts is displayed on the right side of the screen with the newest text shown at the bottom of the list. Moreover, a text-to-speech service is also implemented in the application. The service converts the received text into an output, which is played back concurrently by the mobile device speaker.

In this study, the Android-based sign interpretation application is merely utilized for receiving classified sign gestures from the smart wearable device and further displaying the results on the screen.

TABLE III
COMPARISON OF CLASSIFICATION RESULTS FOR SIGN RECOGNITION FOR ALPHABET 'R', 'U' AND 'V'

Alphabet \ Subject	R	U	V
1	65.1 / 97.2	66.2 / 97.5	65.8 / 98.5
2	51.2 / 96.4	52.1 / 97.1	50.5 / 98.2
3	53.4 / 97.4	52.8 / 98.6	53.5 / 97.2
4	56.1 / 95.5	55.2 / 96.4	55.3 / 97.5
5	61.5 / 95.8	60.8 / 97.5	61.1 / 96.7
6	62.8 / 97.9	61.8 / 97.8	60.5 / 96.9
7	60.8 / 96.4	61.1 / 98.4	61.0 / 97.5
8	69.5 / 98.7	68.4 / 96.7	67.5 / 98.5
9	51.6 / 97.5	50.8 / 97.2	51.1 / 96.9
10	52.7 / 98.6	51.8 / 97.8	50.8 / 97.6
11	53.8 / 99.5	52.8 / 99.4	53.1 / 98.2
12	54.8 / 96.4	53.2 / 95.8	54.2 / 96.1

/ – Accuracy rate for *1st version vs. (2nd) version in %

Fig. 3.6: COMPARISON OF CLASSIFICATION RESULTS FOR SIGN RECOGNITION FOR ALPHABET R, U AND V

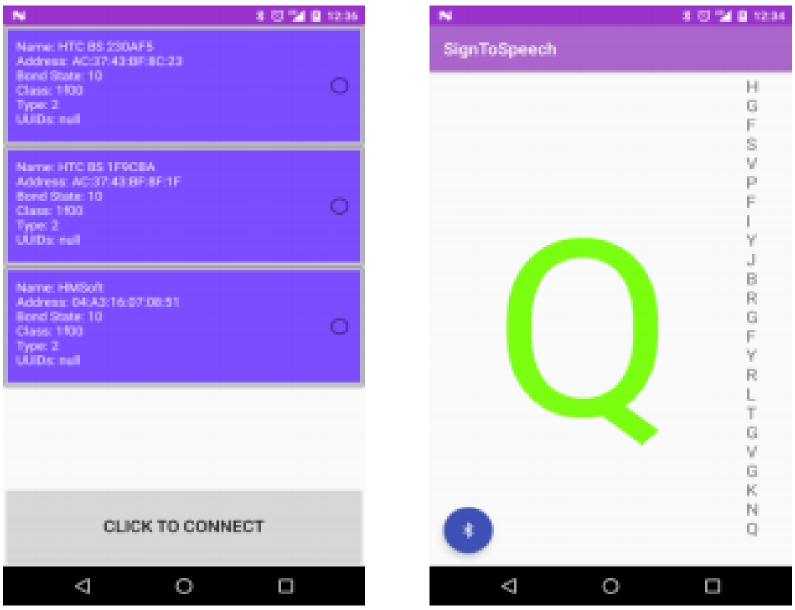


Fig. 3.7: Text to speech service

Sign interpretation system showing (a) Bluetooth search screen and (b) text received from the proposed wearable device through a Bluetooth connection, as well as a text-to-speech service converting the text into audible output played back simultaneously using the built-in mobile speaker.

Conclusion

smart wearable hand device as a sign interpretation system using a built-in SVM classifier. An Android-based mobile application was developed to demonstrate the usability of the proposed smart wearable device with an available text-to-speech service. The participating subjects gave a high rating to the proposed smart wearable sign interpretation system in terms of its comfort, flexibility, and portability. The device holders were 3D-printed using a flexible filament, and the same holders are able to fit different hand and finger sizes, thus eliminating the necessity of custom-made devices. Future work on the proposed smart wearable hand device will consider the design of a smaller sized printed circuit board, the inclusion of words and sentences at the sign language level, and instantly audible voice output components.

REFERENCES

[1] J. Wang and T. Zhang, An ARM-based embedded gesture recognition system using a data glove, presented at the 26th Chinese Control and Decision Conf., Changsa, China, May 31 - June2, 2014.

[2] A. Z. Shukor, M. F. Miskon, M. H. Jamaluddin, F. A. Ibrahim, M. F. Asyraf and M. B. Bahar, A new data glove approach for Malaysian sign language detection, Procedia Comput. Sci., vol. 76, no. 1, pp. 60-67, Dec. 2015.

[3] N. Sriram and M. Nithiyanandham, A hand gesture recognition based communication system for silent speakers, presented at the Int. Conf. Human Comput. Interact., Chennai, India, Aug. 23-24, 2013.

[4] J. Galka, M. Masior, M. Zaborski, K. Barczewska, Inertial motion sensing glove for sign language gesture acquisition and recognition, IEEE Sensors J., vol. 16, no. 16, pp. 6310-6316, Aug. 2016.

[5] Y. L. Hsu, C. L. Chu, Y. J. Tsai and J. S. Wang, An inertial pen with dynamic time warping recognizer for handwriting and gesture recognition, IEEE Sensors J., vol. 15, no. 1, pp. 154-163, Jan. 2015.

[6] N. Caporusso, L. Biasi, G. Cinquepalmi, G. F. Trotta and A. Brunetti, A wearable device supporting multiple touch- and gesture-based languages for the deaf-blind, presented at Int. Conf. Appl. Human Factors and Ergonomics, LA, CA, USA, Jul. 17-21 2017.