

Hand Gesture Interpretation Using Sensing Glove Integrated with Machine Learning Algorithms

Aqsa Ali, Aleem Mushtaq, Attaullah Memon, Monna

Abstract—In this paper, we present a low cost design for a smart glove that can perform sign language recognition to assist the speech impaired people. Specifically, we have designed and developed an Assistive Hand Gesture Interpreter that recognizes hand movements relevant to the American Sign Language (ASL) and translates them into text for display on a Thin-Film-Transistor Liquid Crystal Display (TFT LCD) screen as well as synthetic speech. Linear Bayes Classifiers and Multilayer Neural Networks have been used to classify 11 feature vectors obtained from the sensors on the glove into one of the 27 ASL alphabets and a predefined gesture for space. Three types of features are used; bending using six bend sensors, orientation in three dimensions using accelerometers and contacts at vital points using contact sensors. To gauge the performance of the presented design, the training database was prepared using five volunteers. The accuracy of the current version on the prepared dataset was found to be up to 99.3% for target user. The solution combines electronics, e-textile technology, sensor technology, embedded system and machine learning techniques to build a low cost wearable glove that is scrupulous, elegant and portable.

Keywords—American sign language, assistive hand gesture interpreter, human-machine interface, machine learning, sensing glove.

I. INTRODUCTION

HAND gesture recognition is an emerging and field having a wide variety of applications, such as human computer interaction, automation control, computer games, 3D animations, sign language recognition, traffic signal control using gestures, 3D mouse, virtual keyboard, etc. The specific application that we are addressing in this project is sign language recognition to assist the speech impaired people. The fact sheet [1] provided by the WHO (World Health Organization) tells that over 5% of the world population is impaired from speech and hearing, and they use gestural language to communicate. Our device provides the speech impaired people with a medium that can convert their gestures into audible speech.

The three popular methods for hand gesture recognition are vision based, EMG (Electromyography) based and glove based. In Vision based approach, a camera is used to capture the gestures of hand. The captured image is processed by using different algorithms such as, 3D model-based algorithms, Skeletal-based algorithms and Appearance-based models [2]-[9]. This approach has serious limitations due to the fact that images are captured using a 2D camera while the movements

are in 3D. Moreover, it has much environmental trepidation like the place, position and focus of camera, lightening, background condition and other interferences.

In EMG based approach, EMG sensors are used to sense the muscle activity [10]. These sensors extract multiple EMG signals, called electromyograms, from hand and wrist prostheses. The signals are measured by attaching conductive electrodes to the skin surface [11]-[13]. This is a nascent technique and suffers from the drawback that the signals are very weak, and vary significantly under different conditions. Consequently, these signals are very difficult to classify.

The third approach is glove-based approach. This approach comprises of different sensors placed on hand glove to sense the bending, orientation and position of hand, and contact and spacing between fingers [14]. Different types of sensor are used on the glove that includes flexible tubes with light, capacitive electrode, flex sensors and magnetic sensors to detect the curl of finger. Accelerometer, Gyroscope and other tilt sensors are used for the orientation and position of hand, and proximity and touch sensors for the detection of contacts. The technique is independent of surrounding which makes it robust.

Parvini [15] used 22 flex sensors that could detect the bending of fingers. The paper proposed the utilization of Multi-Layer Perceptron (MLP) and Gesture Recognition by Utilizing Bio-Mechanical Characteristics (GRUBC) algorithm for recognition of ASL. The maximum accuracy was 82% using GRUBC for 20 alphabets. Rajamohan [16] used five flex sensors but added one accelerometer and few tactile sensors in the glove. Using this glove, they were able to classify 13 gestures. Khan [17] used gloves of 5DT Company, containing seven flex sensors. Five flex sensors were mounted on each of the finger, one sensor were used to measure the tilt of the hand and one for the rotation of the hand. Neural Networks were used for classification, and 88% accuracy was obtained. Raut [18] also used a glove comprising of flex sensors and accelerometer. In this solution, an audio file corresponding to the recognized gesture was fetched from a pre-stored database for playing. Only 10 gestures were recognized in this system.

A. Ali, Dr. A. Mushtaq, Dr. A. Memon, and Monna are with the Department of Electrical and Power Engineering, National University of Sciences and Technology, Karachi, Pakistan (e-mails: aqsasyed966@gmail.com, aleem@pnec.nust.edu.pk, attazz@gmail.com, mn.khatr@hotmail.com).

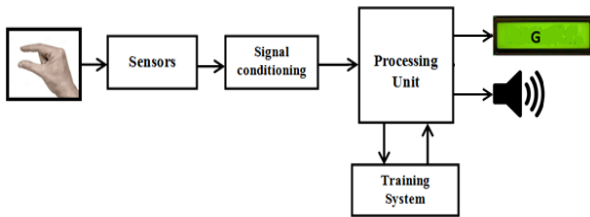


Fig. 1 Block Diagram

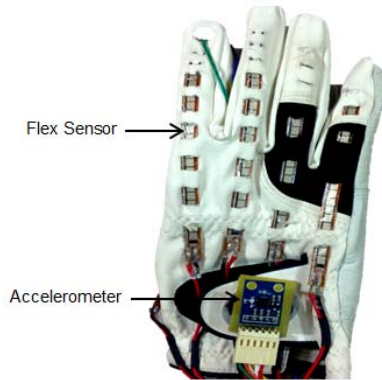


Fig. 2 Sensors Mounted Glove

Our design, described in this report, combines flex sensors, accelerometers and contact sensors to capture all aspects of the hand movements during gesture formation. The information from the sensors is fused together to form a feature vector whose dimension is 11. Linear Bayes classification algorithms are then applied on the feature vector to make a robust decision on the gesture made. The number of classes has been increased to 27, and still an efficiency of up to 99.3% is achieved when training data from the individual being tested is available. Another attractive feature of our glove is that it is user friendly and easy to wear.

This report is organized into seven sections. Section I reviews various approaches to tackle the gesture recognition problem, and argues why the glove based approach is the best. It also gives a survey of the glove based solutions available in literature. In Section II, the complete design of our smart glove is presented. Section III describes the classification algorithms used for classification of gestures. The experimental setup to validate the performance of the glove and results obtained are presented in Sections IV and V respectively. The cost of the glove is given in Section VI before concluding with future plans for the project in Section VII.

II. GLOVE DESIGN

In this project, flex sensors are used for sensing the extent of finger bending while contact detectors are used for the detection of contacts between fingers. Movements of dynamic nature and the orientation of the hand for various gestures are measured with accelerometers. The signals obtained from the sensors are refined by signal conditioning unit to make them compatible with the processing unit. Conditioned signals are then given to processing unit, where they are classified using machine

learning/pattern recognition algorithms. Finally, the recognized letter is displayed on a display screen carried by the user. Fig. 1 shows the overview of the project.

The glove recognizes many gestures by detecting the extent of curl of fingers by using five 4.5" flex sensors on each finger and a 2.2" flex sensor for the inner side of index finger for precision. These sensors vary their resistance in unidirectional bending with respect to bend angle. The signals from the flex sensors are conditioned to match with the full scale of ADCs. As the flex sensors provide resistance as output signal, so the signal conditioning unit comprises of a voltage divider circuit to convert resistive form of signal into voltage signal and LM358 operational amplifier which works as impedance buffer. For those gestures, which have similar flex of fingers, but different orientation and tilt of hand, such as 'G' and 'Q', 'H' and 'U', and 'K' and 'P', an accelerometer is used. ADXL335 is an analog accelerometer that gives 3D orientation of object i.e. roll, pitch and yaw. Gestures for U, V and R are having the same orientation of hand and bending of fingers but distinct in fingers contacts. Therefore, contact detectors are used to distinguish them.

The Atmel Atmega2560 micro-controller is working as a processing unit of the system. It takes the analog data from flex sensors and accelerometer and converts it into digital signals for the further processing. The data from contact detectors captured through GPIO pins. The processor classifies the captured signals into alphabets. These alphabets are then translated into words and output to the person being spoken by using a TFT screen display and a speech synthesizer. The Emic 2 Text-to-Speech Module is used for speech synthesis.

A DC battery of 11.1V/2200mA·h (7,920A·sec) is installed for portability of the device.

III. THE CLASSIFICATION ALGORITHMS

The setup consists of multiple measurements from sensors. These measurements are comprised of nine attributes for each instance, and therefore nine dimension feature vectors. In this work, we have used Bayesian Classifier for classification of the sensor measurement into one of the 27 classes. Since the processing needs to be done on a resource limited microcontroller, some assumptions are made about the class distributions. If each class is considered to be Gaussian same covariance matrix, then we can use computationally efficient Mahalanobis Distance Classifier given in (1),

$$D_M(x) = (x - \bar{x})^T \Sigma^{-1} (x - \bar{x}) \quad (1)$$

where Σ is the Covariance matrix. If further constraint is put on the distribution, and it is assumed that the variance of each feature is same and cross covariance between features is zero, then Euclidean Distance Classifier can be used, which is computationally even more efficient than the Mahalanobis Distance Classifier. The Euclidean distance is given in (2),

$$\text{Euclidean Distance} = \sqrt{\sum_{i=1}^9 (x_i - \bar{x}_i)^2} \quad (2)$$

Means and variances for each class is obtained using (3), and Covariance matrix is obtained using (4), of each attribute for respective classes.

$$\text{Mean: } \bar{x} = \frac{\sum_{i=1}^N x_i}{N} \quad (3)$$

$$\text{Covariance Matrix: } \sigma_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{N} \quad (4)$$

where x_i and y_i is the i^{th} sample data of two classes (x, y), N is total number of samples for each class, \bar{x} is the mean value of respective class and σ_{ij} is the standard deviation of 'x' row and 'y' column of Covariance Matrix.

These classifiers measure the distance between the test instance and the computed mean values of all attributes of each class. The resulting classified class is the one which is minimum distant from the test sample. Mahalanobis distance classifier takes into account the Covariance of the classes in addition to the mean, and is therefore more powerful given there is enough data to estimate covariance.

Neural network is one of the high computational classifiers been using in the area of recognition. It has capability of storing knowledge in training phase using synaptic weights. The number of neurons in input layer is equal to the number of features in each sample vector. The network processes the test data with stored data and computes the output. We used 9 nodes in input layer, 50 nodes in first hidden layer and 27 in other while 27 nodes comprised the output layer. Like Mahalanobis classifier, neural network also works on large datasets. It gives more precision with hidden layers added at the trade of high computational power.

IV. EXPERIMENTAL SETUP

A. Database Collection

To build a training corpus for testing the complete solution, gestures from multiple individuals were recorded. The glove is connected to PC with the help of serial cable at a baud rate of 9600bits/sec. The data is collected in the form of '.txt' files from the glove into a PC using MATLAB through serial port provided on arduino board. This serial port is also used for loading of classifier parameters in the EEPROM of Atmega2560 on arduino and debugging of the embedded system on the glove.

In this project, we have total eleven features; six from flex sensors, three from accelerometer and two from contact sensors. Moreover, there are 27 classes, out of which 26 are ASL alphabets and one class is self-made gesture for space. The training corpus was collected from students at NUST. For each user, twelve sample sets are collected, where each set has twenty seven classes (C). The data samples (S) are taken from five users (N), so that the total number of samples can be calculated by (5),

$$NSC = 1620 \text{ samples} \quad (5)$$

B. Mathematical Model

Let us define variables for the mathematical modeling of data:

- N = number of users [5]
- n = index for user [1..5]
- C = number of classes [27]
- c = index for class [1..27]
- S = number of samples [12]
- s = index for sample [1..12]
- F = features [9]
- f = index for feature [1..9]
- d = feature data

Data vector (D) comprises of all features, can be represented by (6):

$$D = \{d_1, d_2, d_3, \dots, d_F\} \quad (6)$$

Training data (\bar{U}) of all users is empirically formulated as given in (7),

$$\bar{U} = \{D_{c,s}^n : n \in \{1,2,\dots,N\}, c \in \{1,2,\dots,C\}, s \in \{1,2,\dots,S\}\} \quad (7)$$

while the test data is an unknown data vector (8),

$$\Gamma = \{D_u = \{d_1, d_2, d_3, \dots, d_{11}\} : u \text{ denotes unknown class}\} \quad (8)$$

Three of the five individuals (users 1, 2 and 5) were well aware with ASL while the remaining two (users 3 and 4) were new to ASL. Moreover, the same glove was used by all individuals, and therefore, it did not necessarily fit them perfectly.

A PC is used to determine the parameters of the classifiers, which include the Euclidean classifier and the Mahalanobis classifier. These classifiers were first tested on MATLAB, and then implemented on the processing unit. The processing unit, i.e., the glove, becomes a standalone unit once it has been loaded with the classifier parameters.

V. RESULTS

Table I shows the results for the case when testing is done using user specific data, i.e., when testing user 1, only training data from user 1 himself is used. In this case, Multilayer Neural Network clearly performs the best, whereas the Euclidean distance classifier also provides high accuracy results. The Mahalanobis classifier does not perform well because there is insufficient data to estimate all the parameters involved while the thresholding based classifier works efficiently only for small data corpus.

In the second configuration, we used data from multiple users to train the models, and therefore, the training corpus is more generic. The results for this scheme are given in Table II. Since the size of training set is larger, and there is sufficient data to estimate the parameters of Mahalanobis classifier, the performance of the Mahalanobis classifier improves significantly (63.52% to 89.26%). While the Neural Networks performed better than the other three classifiers. The improvement in the Mahalanobis and Neural Networks

classifiers as the number of training samples is increased is further emphasized by the results of Fig. 3, where the overall performance of all classifiers is depicted against the number of samples used for training.

TABLE I
PERFORMANCE FOR USER SPECIFIC DATA

Dataset	Thresholding	Euclidean Distance	Mahalanobis Distance	Multilayer Neural Networks
User-1	77.47%	92.90%	61.42%	93.9%
User-2	75.25%	96.30%	63.88%	99.3%
User-3	36.53%	84.25%	58.95%	95.1%
User-4	60.67%	85.80%	69.75%	98.6%
User-5	62.03%	96.60%	63.58%	95.7%
Average	62.39%	91.17%	63.52%	96.52%

TABLE II
PERFORMANCE FOR GENERAL SEEN DATA

Dataset	Thresholding	Euclidean Distance	Mahalanobis Distance	Multilayer Neural Networks
User-1	24.92%	81.48%	91.67%	95.4%
User-2	24.86%	92.90%	91.67%	97.8%
User-3	21.22%	80.24%	83.33%	92.4%
User-4	22.66%	80.86%	86.72%	95.2%
User-5	22.77%	91.67%	92.90%	90.7%
Average	23.29%	85.43%	89.26%	94.04%

TABLE III
PERFORMANCE FOR GENERAL UNSEEN

Dataset	Thresholding	Euclidean Distance	Mahalanobis Distance	Multilayer Neural Networks
User-1	25.63%	76.85%	71.29%	82.4%
User-2	27.06%	91.04%	83.64%	94.8%
User-3	32.39%	77.77%	71.91%	81.2%
User-4	25.93%	80.25%	64.50%	86.6%
User-5	25.06%	90.12%	88.27%	90.2%
Average	27.21%	83.21%	75.92%	87.04%

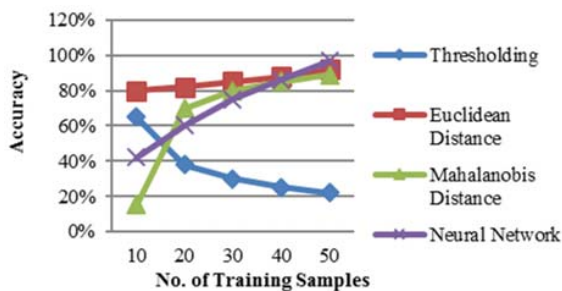


Fig. 3 Performance of each classifier vs. number of training samples

In the third case, complete data set is used for training, but the data from the user being tested is omitted. This is an attractive configuration from the user friendliness prospective because the user would not have to provide his own samples before he starts using the glove. The results are shown in Table III. The Euclidean classifier performs the best for this configuration. An interesting observation for this configuration is that the performance of Mahalanobis classifier deteriorates more than Euclidean when compared with the second configuration given above. The Neural Networks classifier

clearly performs better when the training corpus is without user specific data. The authors conjecture that Mahalanobis enforces more constraints of the variances of features compared to the Euclidean classifier, and these variations constraints are not consistent with the training data in this scheme. The bending of fingers and orientation of hands may differ for different users, and therefore, data from one set of users may not conform well to another set of users. The observation is reinforced by the fact that performance of Mahalanobis classifier for user 4, who is not well trained on ASL, is the poorest because his gestures are different from other users.

TABLE IV
COST CHAR

S. No.	Components	Qty.	Cost per unit (PKR)	Total (PKR)
1	Flex sensor 4.5''	5	1,900	9,500
2	Flex sensor 2.2''	1	1,100	1,100
3	Accelerometer ADXL335	1	750	750
4	Arduino Mega 2560 R3	1	1,300	1,300
5	TFT screen 2.8''	1	2,300	2,300
6	Emic2 Speech synthesizer	1	10,000	10,000
7	Miscellaneous	-	-	2000
Total				26,950

Overall, we observe that the best performance is achieved with Multilayer Neural Network classifier when the size of training data is large enough and user's own data is part of the training corpus.

VI. COSTING

The device should be economic for the common people to access it easily. So comparing with currently available marketed Sensing Gloves, such as 5DT Glove 5 Ultra from Virtual Reality, costs US\$995 (104,948 PKR approx.) for single hand, Immersion Corp., offers CyberGlove II which is available with 18 and 22 sensors costs US\$12,295 (12,86,979 PKR approx.) and US\$17,795 (18,62,691.62 PKR approx.), respectively. ShapHand data glove from Measurand Inc., is also a commercial product for capturing motion and costs US\$9,900 (10,36,282 PKR approx.). Hence, our Smart Glove presents an affordable and cost effective solution as stated in Table IV.

VII. CONCLUSION

In this paper, we have presented a low cost sensing glove that will be used for hand gesture interpretation of ASL. The glove was built using off the shelf sensors and hardware. Four different machine learning algorithms were implemented to recognize the signals received from the 11 sensors installed at various locations on the glove. A database of gestures was collected from five different individuals to test the competence of the glove. The results show that the accuracy of the system is up to 99.3% for well-trained users using the Neural Networks algorithm for 27 hand gestures.

The design of the glove is being improved to incorporate dynamic gesture recognition so that signal for capture of gesture signals is not required. Moreover, more powerful machine

learning algorithms are being applied to improve the accuracy of the smart glove.

Other than sign language recognition, the applications of sensing glove include telesurgery which facilitates the doctor to remotely control the robot surgical system at operation location by wearing smart glove without being present at the same location. Another application is wheel chair controlling using self defined gestures for physically challenged people. Virtual reality and gaming are also widely evolving areas in gesture recognition.

ACKNOWLEDGMENT

The project was supported by ICT-R&D Grass Root Initiative funding.

REFERENCES

- [1] WHO, "Deafness and hearing loss," in World Health Organization, World Health Organization, 2016. (Online). Available: <http://www.who.int/mediacentre/factsheets/fs300/en/#>.
- [2] Ms. Nutan D Sonwane, Prof. S. A. Chhabria, Dr. R. V. Dharaskar, "Self Organizing Markov Map for Speech and Gesture Recognition", International Journal of Advanced Research in Computer Science and Electronics Engineering Volume 1, Issue 2, April 2012.
- [3] Daniel Stein, Philippe Dreuw, Hermann Ney, Sara Morrissey, Andy Way Hand in Hand: Automatic Sign Language to English Translation.
- [4] Mohamed Alsheakhali, Ahmed Skaik, Mohammed Aldahdouh, Mahmoud Alhelou, "Hand Gesture Recognition System", Computer Engineering Department, The Islamic University of Gaza, Gaza Strip, Palestine, 2011.
- [5] Sanjay Meena, "A Study on Hand Gesture Recognition Technique", Department of Electronics and Communication Engineering, National Institute of Technology, Rourkela Orissa 769 008, India 2011.
- [6] Yu Shi, Ronnie Taib and Serge Lichman, "GestureCam: A Smart Camera for Gesture Recognition and Gesture-Controlled Web Navigation".
- [7] Joyeeta Singhal, Karen Das2, "Hand Gesture Recognition Based on Karhunen-Loeve Transform".
- [8] Kumo, Yoshiri, Tomoyo Ishiyama, Kang-Hyun Jo, Nobutaka Shimada and Yoshiaki Shirai, Vision-Based Human Interface System: "Selectively recognizing intentiona Hand Gesture". In Proceedings of the IASTED International Conference on computer Graphics and Imaging, 219-223, 1998.
- [9] Utsumi, Akira, Jun Kurumisawa, Takahiro Otsuka, and Jun Ohya, "Direct Manipulation Scene Creation in 3D", SIGGRAPH'97 Electronic Garden, 1997.
- [10] Muhammad Zahak Jamal, "Signal Acquisition Using Surface EMG and Circuit Design Considerations for Robotic Prosthesis".
- [11] T. Scott Saponas1, Desney S. Tan2, Dan Morris2, Ravin Balakrishnan3, "Demonstrating the Feasibility of Using Forearm Electromyography for Muscle-Computer Interfaces".
- [12] Eldin Henry Shroffe, D P. Manimegalai "Hand Gesture Recognition Based On Emg Signals Using Ann", International Journal of Computer Application, Issue 3, Volume 2 (April 2013).
- [13] Xu Zhang, Xiang Chen, Associate Member, IEEE, Yun Li, Vuokko Lantz, Kongqiao Wang, and Jihai Yang, "A Framework for Hand Gesture Recognition Based on Accelerometer and EMG Sensors".
- [14] Laura Dipietro, Angelo M. Sabatini, Senior Member, IEEE, and Paolo Dario, Fellow, IEEE, "A Survey of Glove-Based Systems and Their Applications", IEEE transactions on systems, man, and cybernetics—vol. 38, no. 4, July 2008.
- [15] Farid Parvini, Dennis McLeod, Cyrus Shahabi, Bahareh Navai, Baharak Zali, Shahram Ghandeharizadeh, "An Approach to Glove-Based Gesture Recognition", 13th International Conference, Human Computer Interface International 2009, San Diego, CA, USA, July 19-24, 2009, Proceedings, Part II.
- [16] Anbarasi Rajamohan, Hemavathy R., Dhanalakshmi M., "Deaf-Mute Communication Interpreter", International Journal of Scientific Engineering and Technology (ISSN: 2277-1581), Volume 2 Issue 5, pp: 336-341 1 May 2013.
- [17] Yasir Niaz Khan, Syed Atif Mehdi, "Sign Language Recognition using Sensor Gloves", FAST-National University of Computer and Emerging Sciences, Lahore.
- [18] Ajinkya Raut, Vineeta Singh, Vikrant Rajput, Ruchika Mahale, "Hand Sign Interpreter", The International Journal of Engineering And Science (IJES), 2012.