Smart Glove using Gesture Recognition Techniques for Speech Impaired and Deaf People

Dipon Sengupta
Department of Computer Science Engineering, SRMIST, Tamil Nadu, India.

M. Sanjanaa Sri
Department of Computer Science Engineering, SRMIST, Tamil Nadu, India.

S. Arun Kumar
Assistant Professor, CSE Department, SRMIST Ramapuram, Tamil Nadu, India.

Abstract – Speech impaired and deaf people in together have a population of 530 million across the globe. Hand gestures and facial expressions liaise between the dumb and the outside world. The main objective of the project is to help them discourse with other people. A smart glove once worn by a speech impaired will act as a translator which will convey his words in form of text and audio output.In this paper we focus mainly on the speech recognition techniques that can be inferred to improve the quality and accuracy of the smart glove. The system entirely comprises of the following modules: a smart glove with numerous sensors, data processing modules and a text to speech incorporated mobile application. The smart glove will be equipped with pressure sensors and flex-sensors in the finger joints. Kinect will be used for video spilling and recording the gestures other than the ones using the fingers. This project looks forward to interaction between computers and human which is a developing innovation.

Index Terms – Speech impaired, Gesture recognition, Sensors, Microcontroller, Wearable Computing, Kinect, Mobile Application

1. INTRODUCTION

Gesture recognition is a standout amongst the most imperative highlights in Human computer interaction automation system other than speech recognition. Its utilization significantly lies in the fact that speech impaired people cannot comprehend other than using gestures and their language symbolized using their fingers. Human Computer Interaction is evaluated as one of the forthcoming technological innovations in which a specific method is figured out in order to make the interaction user friendly as the machine has to achieve high accuracy rate so that the dialect used is not misapprehended. This project is to put forward assistance for the 7% of the world's population who are destined to be speech impaired or have the inability to hear.

Gesture based communications (otherwise called marked dialects) are dialects that utilization manual correspondence to pass on significance. This can incorporate all the while utilizing hand signals, development, introduction of the fingers, arms or body, and outward appearances to pass on a speaker's thoughts. Gesture based communications regularly share huge likenesses with their particular talked dialect, for example, American Sign Language (ASL) with American English). Punctuation and sentence structure, notwithstanding, may change to support effectiveness and ease in speaking. It is critical to take note of that in light of the fact that a talked dialect is coherent transnationally, for example, English in the United States and the United Kingdom, does not imply that the gesture based communications from those locales are also; ASL and British Sign Language (BSL) were shaped freely and are in this way unintelligible. Linguists consider both talked and marked correspondence to be kinds of common dialect, implying that both developed through a dynamic, extended maturing process and advanced after some time without fastidious arranging. Gesture based communication ought not be mistaken for "non-verbal communication", a sort of nonverbal communication. Wherever people group of hard of hearing individuals exist, communications via gestures have created, and are at the centers of nearby hard of hearing societies. In spite of the fact that marking is utilized fundamentally by the hard of hearing and in need of a hearing aide, it is likewise utilized by hearing people, for example, those unfit to physically talk, or the individuals who experience difficulty with talked dialect because of an incapacity or condition (augmentative and elective correspondence).

There has been a lot of research on this topic and different researchers have come up with different levels of accuracy either by providing hand fitting gadget or else by video streaming the gestures but the later has very poor accuracy rate. So in order to rectify the precision rate this paper initiates the research if both the smart glove and video trajectory can be implemented together using different gesture recognition algorithms. Thus it provides a righteous approach for handling this special kind of communication techniques that are pervasive across the globe. Usually people using sign

language and gestures need an interpreter who perfectly understands their language but this is to achieve a machine that will represent the sign language at a similar rate or rather more rapidly. In this project a noteworthy part is accomplished by sensors. The primary sensors used in this venture are namely flex sensors and pressure sensors.

Fig.1 The letters of the English alphabet in sign language

2. RELATED WORK

In the related work there are some smart working devices that have been presented earlier with but have some issues in its working procedure or accuracy.

Hand sign acknowledgment is the territory of research for some a long time. There are diverse techniques for hand sign acknowledgment utilizing distinctive techniques. Approaches that can perceive hand signals are of two kinds information glove based technique and vision based strategy. Sensor gadgets are used for digitizing hand developments into information in information glove technique .These estimations are exact and quick and are extremely costly. Vision based techniques requires a camera to accomplish regular cooperation without utilization of additional gadgets.

The vision based techniques can be of two classifications 3-D display based and appearance display based. 3D display construct strategy depends with respect to 3-D kinematic hand models. This decides the hand stances. In appearance based models extricate picture highlights are removed and the presence of hand pictures are displayed. This strategy is predominant over different systems for constant handling. Fragmenting out the hand areas from foundation is the initial step. Skin shaded areas from the pictures are recognized to be portion foundation. This strategy is well known, be that as it may, is receptive to lighting conditions and no other skin-like articles ought to be in the pictures. Other frequently utilized highlights are Haar-like highlights were proposed to perceive human countenances. Hand motion classifiers proposed were prepared with Adaboost what is more, Haar highlights to enhance the precision of grouping. The Adaboost learning calculation enhanced grouping precision by adaptively choosing the best highlights. Hand discovery with Adaboost that embraced changed Haar-like highlights and the recognition of scale-space highlights to discover palm and finger-like structures.

Numerous motion acknowledgment frameworks utilize the assistance of economical shading coded gloves for hand division, which improve the sifting through of different protests and foundations. Such methodologies have been tried and affirmed by numerous scientists. One of the frameworks utilized a glove with six hues. A specific method of shading was utilized to cover each finger and palm. Some different systems engraved examples in an ordinary fabric glove that disentangle the assessment of hand images. After division the following procedure to be performed is include extraction and characterization that should be possible utilizing distinctive mapping systems. One shape coordinating depends on the shape that make utilization of Fourier descriptors and wavelet descriptors. B-Spline is another shape coordinating strategy in light of form. Another shape coordinating strategy is nearby highlights based protest coordinating which utilized SURF and SIFT calculations. Two traditional methodologies for extricating the expository look and trimming the volume are Principal Segment examination (PCA) and Linear Projection Method (LPP) . Hand sign acknowledgment in view of PCA and LPP to remove highlight vectors after the hand shape tests are standardized in revolution and size. Another framework utilizes finger furthermore, palm courses of action for hand sign acknowledgment. It utilizes form focuses and arched frame or Jarvis calculation to consolidate the highlights for pose acknowledgment. Rectilinear and unbend signs can be perceived by this method.

The literature survey done along with the paper mainly focuses how we can translate a letter but nothing has been introduced so that a whole sentence can be translated at a time in both text and audio format. This thought of improvisation the related work brings us to the problem statement as described below.

3. PROBLEM STATEMENT

“A novel and smart wearable hand device with sign interpretation has been designed. After implementation and testing of the device it has two problems such that the device size is huge and uncomfortable for regular use and it only

reads and interprets letters of English language. Words, sentences and gestures are therefore not interpreted. Also instant Voice output components are missing." This is specifically the problem statement on which the smart glove is designed so that not only alphabets but also words and sentences will be understood and translated along with the gestures.

The following test is to choose how to catch these signs. It may appear glaringly evident that a camera can be utilized to catch pictures yet simply acquiring a picture won't give enough subtle elements as something in excess of an insignificant 2D portrayal of territory of intrigue might be required. So there is a need to decide on a gadget that additionally caters profundity alongside basic 2D shading pictures giving a significant to movement in 3D space.

4. SYSTEM OVERVIEW

The Proposes system will essentially comprise of the smart glove which is to be worn by the speech impaired or the deaf in order to help themselves to express and communicate with normal people. The sensors linked with a microcontroller.

Microcontroller will have the gestures and alphabets already fed into the system and needs to keep track of the arriving input. It will have an algorithm which will convert the input in the form of gestures as well as the sensor transmitted finger movements to a viable output. Text will be converted to audio outputs as well according to the need.

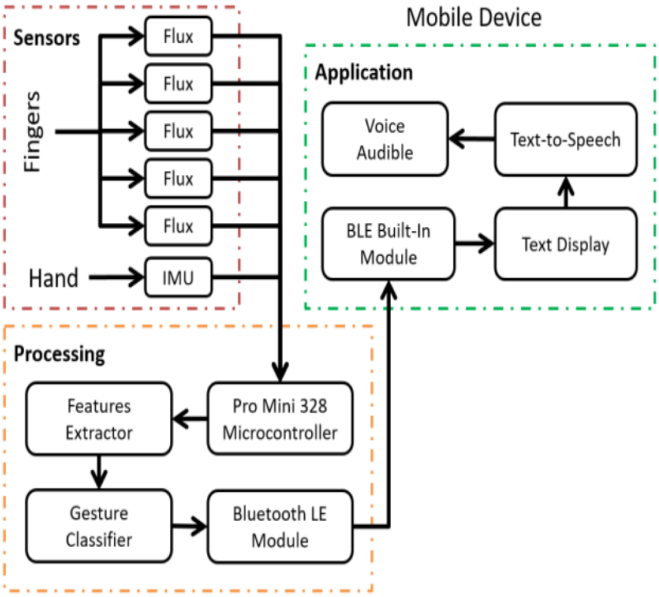

Fig 2: Three main modules of the smart glove.

To make a profitable sign language descriptor it is important to comprehend the structure and kinematics of the human hand. The thumb comprises of two connections, different fingers have three phalanges, all fingers are associated with the palm, and the palm to the wrist by joints. The thumb development has 5 DoF (Degree of Freedom), different fingers have 4 DoF and the wrist rotational and translational development has 6 DoF, this includes as a 27 DoF framework for the human hand.

Hagdil semantics separate 18 hand parameters. Ten parameters depict the opposite condition of the thumb with respect to the palm and the twisted conditions of finger phalanges. The following four signify the parallel position of the thumb relative to the palm and the crossed-shut spread condition of the fingers. Two parameters portray the outright introduction of the hand and another two give the situation of the wrist. Every parameter can have 16 unique esteems yet the convention leaves the chance to extend it to 256 for more modern application situations.

We likewise explored different avenues regarding an alternate approach for signal coding. Rather than taking the anatomic-kinematic model of the hand as a premise we examined the fingerspelling letter set and decided the base important arrangement of unmistakable states for each finger – we call this approach KML. We discovered different advantages and disadvantages when contrasting Hagdil with KML. The quantity of conceivable conditions of KML is significantly littler than that of Hagdil, making it less demanding to process. Notwithstanding, it scarcely enables clients to record custom motions to their own "finger speaking" lexicon. So Hagdil is a superior decision for an adaptable item, while KML may be a superior choice while making a control arrangement in light of predefined motions.

In this framework at the transmitter side we utilize a glove which must be worn by the client. This glove is mounted with 4 flex sensors each on the 4 fingers of the glove specifically thumb, list, centre and ring. The flex sensors give their yield as change in protection as indicated by the twist point. The yield from the flex sensors is given to the ADC channels of the microcontroller. The handled ADC esteems from the microcontroller are contrasted and the edge esteems for the acknowledgment of a specific signal. The specific signal is perceived and is given to the microcontroller which transmits them through the RF module in a serial way. For all esteem found at RF recipient, the microcontroller gives comparing charges to the LCD and the Voice Module. Along these lines we get the voice yield for each signal and show of each motion in type of content on the LCD show.

4.1. FLEX SENSORS:

In this gadget the hand motions are perceived utilizing flex sensor. These sensors are joined to the gloves. Flex sensors are like potentiometer, i.e. variable resistor. The protection of the sensor differs as indicated by the measure of its twisting, which assistant relies upon the development of finger. With a

specific end goal to unequivocally gauge the twisting flex sensor are utilized. The flex sensors have a normal level protection around 10k ohms. At the point when the sensor are bowed the protection offered by them increments.

4.2. ACCELEROMETER:

The ADXL335 is a little, thin and low power gadget fit for estimating complete 3-pivot increasing speed. The ADXL335 can gauge increasing speed with a base full scale scope of ±3g. It requires less power and gives yield motions regarding simple voltages that are relative to quickening. It can gauge the static increasing speed of gravity in tilt-detecting applications, and additionally unique quickening coming about because of movement, stun or vibration. The three tomahawks' sense bearings are exceptionally orthogonal and have minmal cross-hub affectability since it utilizes a solitary polysilicon surfacemicro-machined sensor structure for detecting X, Y and Z tomahawks. The client chooses the data transmission of the accelerometer utilizing the Cx, Cy, and Cz capacitors at the Xout, Yout, and Zout pins. Data transmissions can be chosen to suit the application, with a scope of 0.5 Hz to 1600 Hz for the X and Y tomahawks, and a scope of 0.5 Hz to 550 Hz for the Z pivot.

5. MODULE DESCRIPTION

The smart glove necessarily contains some modules which can be best explained by UML diagrams as given in Fig 3.1 or fig : 3.2. The modules are broken into three main parts mainly as shown in the system overview.

The main database to be matched with is the ASL or BSL which the user needs to choose accordingly to ease out the procedure for matching the sensors which are being already transmitting it to the algorithm which does the matching. Now in the smart glove is to be actuated with sensors gives the output as the following way as desired.

This algorithm assumes apriori that there are *'n'* Gaussian and then algorithm try to fits the data into the *'n'*Gaussian by expecting the classes of all data point and then maximizing the maximum likelihood of Gaussian centers.

The algorithm used in the matching module is namely Gaussian algorithm and is a type of clustering method which is set up This algorithm assumes apriori that there are *'n'* Gaussian and then algorithm try to fits the data into the *'n'*Gaussian by expecting the classes of all data point and then maximizing the maximum likelihood of Gaussian centers.

Let $X = \{x_1, x_2, x_3, ..., x_n\}$ be the set of data points

$V = \{\mu_1, \mu_2, \mu_3, ..., \mu_c\}$ be the set of means of Gaussian

$P = \{p_1, p_2, p_3, \ldots, p_c\}$ be the set of probability of occurrence of each Gaussian iteration initialize

$$\lambda_t = \{ \mu_1(t), \mu_2(t) \ldots \mu_c(t), \Sigma_1(t), \Sigma_2(t) \ldots \Sigma_c(t), p_1(t), p_2(t) \ldots p_c(t) \}$$

Compute the "expected" classes of all data points for each class using:

$$P(w_i|x_k, \lambda_t) = \frac{p(x_k|w_i, \lambda_t)P(w_i|\lambda_t)}{p(x_k|\lambda_t)} = \frac{p(x_k|w_i, \mu_i(t), \Sigma_i(t))p_i(t)}{\sum_{j=1}^{c} p(x_k|w_j, \mu_j(t), \Sigma_j(t))p_j(t)}$$

M-step.

Compute "maximum likelihood μ" given our data class membership distribution using:

$$\mu_i(t+1) = \frac{\sum_k P(w_i|x_k, \lambda_t)\, x_k}{\sum_k P(w_i|x_k, \lambda_t)}$$

$$p_i(t+1) = \frac{\sum_k P(w_i|x_k, \lambda_t)}{R}$$

where, *'R'* is the number of data points.

The advantage of using this type of clustering over others is it gives extremely useful result for the real world data set. It also superpowers the others with a validity of nearly 98.6% and a high accuracy rate too. The only disadvantage of this algorithm is that it is complex in nature.

The modules using the above mentioned algorithm can be implemented easily and the UML diagrams are attached as to how the data sets matching is implemented in real time.

The sequence diagram of the matching module as shown in fig 3.1 shows the interaction with the ASL (American Sign Language) database to match the data using this particular clustering method. It's a rapid procedure but it follows the sequence as mentioned in the diagram for a request to get accepted and after a data is matched it gets added in a sentence until then the next gesture is being read by the sensors and added to the sentence to make it achieve its perfection which is the main motive of the paper.

The activity diagram as shown in Fig 3.2 states the way in which the smart glove acts a working device to interpret the sign languages for the speech impaired using gesture recognition techniques with the help of the previously mentioned algorithm. The data flow occurs in a manner in which the gesture is directly converted to text and audio output. A speech impaired wearing the smart glove and

feeding the system with some particular gesture is instantly recorded and then by applying the clustering algorithm we check with which particular language the gesture is being mapped to and retrieve the data if matched. Then in the process it is being translated to the mapped sign language and thus gives the desired output.

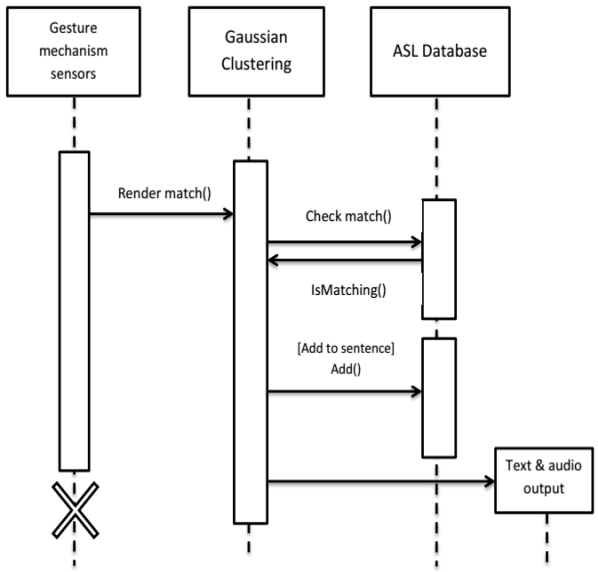

Fig 3.1: Sequence Diagram of the matching module.

The Kinect part which takes care of the joints to make the gesture sensing more accurate is investigating the beforehand proposed models to distinguish signals, we went over 2 noteworthy necessities, vigor furthermore, reaction time. A 3D direction is characterized by gathering a succession of static focuses assigned by an endorser. The focuses are procured utilizing skeleton class gave by Microsoft Kinect SDK. The joints of enthusiasm for our case are left/right elbow, left/right hand and left/right wrist. Skeleton class furnishes with purposes of joints in 3D space. The fundamental target here is to locate the best fit bend on the information purposes of each joint of intrigue and afterward find that bend's conic esteem or discriminant which will be characterized afterward.

Our component vector comprises of six conic estimations of six already specified joints of intrigue, to be specific Right and Left Hands, Right and Left Wrists and Right and Left Elbow joints. We have utilized bagging group technique for the arrangement of the element vector. Bagging is a "bootstrap" examining technique, it makes arbitrary redistribution of preparing set for each of its classifier. Each preparation set is created utilizing irregular drawing of tests from the populace with substitution. So a few cases are spoken to various circumstances while others are forgotten. Test-set mistake got by a solitary classifier is high be that as it may, is decreased when every one of the four classifiers are joined. Assorted variety in classifiers necessarily makes up for high blunder rate of any singular classifier.

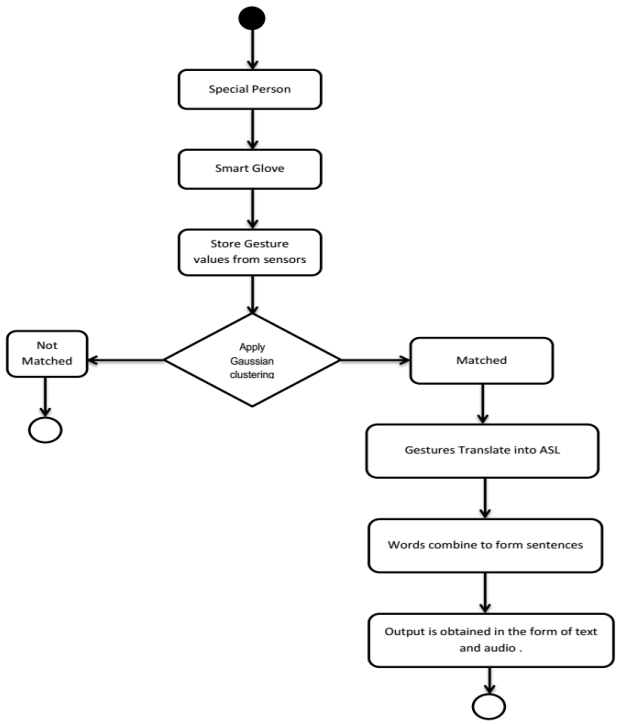

Fig 3.2: Activity Diagram of matching module.

The bagging algorithm:

1. Create a bootstrap test of highlight vector.

2. In the event that there are, N classifiers in the Bagging at that point rehash Step1

N times i.e., one example information collection is done for every classifier.

3. Total the bootstrap tests.

The bagging algorithm is nothing else than straightforward estimate whose outcomes can be enhanced by expanding the N (number of classifiers to be utilized). For additionally enhancing the exactness the motions were isolated into 3 classes to be specific "right hand motions" (which includes motions that include right hand just) "left hand gestures"(which included signals that include left hand just) and "both hand signals" (which includes motions which include the two hands). These classifications were made since if just a single hand was utilized as a part of the motion then the other hand information speaks to a subjective esteem that plays no noteworthy part in the acknowledgment of that signal.

6. CONCLUSION

The most normally utilized calculations utilized for movement/signal acknowledgment basically include layout coordinating procedures. A portable application was created to show the ease of use of the proposed shrewd wearable smart gloves with an accessible content to-discourse benefit. The taking an interest subjects gave a high rating to the proposed shrewd wearable sign elucidation framework as far as its solace, adaptability, and convey ability. The gadget holders were 3D-printed utilizing an adaptable fiber, and similar holders can fit distinctive hand and finger sizes, in this way taking out the need of hand crafted gadgets. Future chip away at the proposed shrewd wearable hand gadget will consider the plan of a littler estimated printed circuit board, the consideration of words and sentences at the gesture based communication level, and quickly capable of being heard voice yield parts.3D direction movement investigation calculation does not include layout coordinating procedures due to which its reaction time and exactness are vastly improved as contrasted with common movement acknowledgment calculations. ASLSI is gone for helping underwriters speak with society without the need of a human mediator. It can be conveyed at open places, for example, air terminals, malls and so forth.

7. FUTURE WORK

Thus the gesture recognition system designed using sensor fusion and gesture recognition techniques in this venture has a lot of future aspects that has to be taken into consideration in order to support the help for this differently abled people more. This smart glove readily banishes the required interpretation between a speech impaired and a normal person.

Future implementation can be made by enhancing the quality of the mobile application which can be used to produce a lot of technical quality research as in what is to be implemented to assist them more. It can be implemented in various fields like in airport and railway stations to assist the speech impaired. One more technical issue can be handled is to assist multi-gesture at a higher speed in which at times this device accuracy fails to reach the peak. Keeping in mind the end goal to enhance and encourage the more signal acknowledgment, movement handling unit can be introduced.

REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, “Automated Sign Language to Speech Interpreter”.
[2] B.G.Lee, “Smart Wearable Hand Device for Sign Language Interpretation System with Sensors Fusion”.
[3] Nikhita Praveen,Naveen Karnath,Megha m s “Sign Language Interpreter Using A Smart Glove” ,
[4] Geethu G Nath, “Embedded Sign Language Interpreter System For Deaf and Dumb People” unpublished.
[5] Smart glove implementation available online at https://www.slideshare.net/sapna_patil/smart-glove-63670479.
[6] Seungki Min et al., “Simple Glove-Based Korean Finger Spelling Recognition System”, Lecture Notes in Computer Science Volume 4705, 2007, 1063-1073.
[7] Craig L Taylor, Ph.D., Robert J. Schwarz, M.D, The Anatomy and Mechanics of the Human Hand.

Authors

Dipon Sengupta originating from West Bengal is currently pursuing his B.Tech Degree from SRM Institute of Science and Technology. He has been to National Taipei University of Technology for Semester Abroad Program. His main areas of interest are Information Security and Internet of Things.

M . Sanjanaa Sri hailing from Chennai, pursuing B.tech from SRM Institute of Science and Technology. She has been to National Tsing Hua University for Semester Abroad Program. She has been a part of a Major project in Taiwan. Her main area of interest is Internet of things (IoT).