International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

A Survey on Hand Gesture Recognition Using Machine Learning

N. Gopinath[1], J. ANUJA[2], S. ANUSHA[3], V. MONISHA[4]

[1]Assistant professor, Department of Computer Science and Engineering, Sri Manakula Vinayagar Engineering College, Puducherry, India
[2-3]Department of Computer Science and Engineering, Sri Manakula Vinayagar Engineering College, Puducherry, India

---

Abstract - *Gesture recognition is one of the language technologies with the objective of converting gestures through mathematical algorithms that are obtained from various body movements. Gestures are usually derived from any motion which emerges from face or hand. Several approaches have been proposed with the help of hand gloves and cameras to interpret signs. The proposed literature deals with a wearable sensor-based gloves. To obtain gesture's information Sensor-based recognition systems are used. In existing system, it is very complex to perform reverse process and conversion of text to sign. It needs further usage of gloves which is connected with sensors. It is also a complex and costly process. This system will provide a two-way communication in which dumb or deaf and individual can communicate with each other without the aid of translator. This prototype is an application of hand gesture technology. The hand movements (gestures) are uploaded into developed application and by using various algorithms the models are trained from which sign languages are converted into text which can be understood by normal people.*

***Key Words*: *Gesture recognition* 1, *image processing* 2, *calculation and text output* 3, *fusion sensor* 4, CNN and DCGAN 5**

1. INTRODUCTION

Artificial Intelligence (AI) has become a major part in today's growing world. This is because the computing resources to people. The developers can able to create new machines and train them accordingly for better results. Machine Learning is the process in which system can improve performance based on experiences without being programmed explicitly. The learning process begins with gathering experiences or observations. It is difficult to take decisions based on inputs. To solve this problem, several algorithms were introduced. Machine Learning algorithms are categorized into Supervised Learning, Unsupervised learning and Reinforcement learning [16]. The most commonly used algorithms are Supervised Learning and Unsupervised learning. Supervised Learning consists of both input and desired outputs. They have the labelled data which are already trained and processed. Supervised Learning is further classified into Classification and Regression. Classification is a technique in which model is constructed based on trained dataset. Regression is a technique in which the output is limited within a range. Unsupervised Learning is a technique in which it contains only input. In this learning, it is not necessary to define the output and consists of unlabeled data. It is further classified into Association and Clustering. Clustering is the process to identify the structure or any pattern in unlabeled data. Association is the process of finding relationship among data. Reinforcement Learning is the technique in which both inputs and outputs are not given. It is mostly based on the interaction with environment. It is mostly concerned with software agents. Machine Learning has been used in search engines like google, Bing etc. so that it has learnt how to rank pages. It is also used in Facebook for tagging friends with the use of face recognition algorithm [17]. By using machine learning it is easy to handle in general deaf and dumb people make use of gesture language for communication, but they find it complicated in communicating with others who don't understand the gestures. Due to which the communication between deaf-mute people and normal people have always been a challenging task. Hand gesture represents several ideas and actions using different hand shapes, orientations or finger patterns being resolved by gesture recognition system and converts them into equivalent response. The hand gesture images are collected to form a data set which is used to train the software. Then those data can be classified by using the CNN and NLP algorithms [18]. Then the Deaf and Dumb people can interact the normal people without any hesitations by using a web application. Here we use Machine Learning to recognize hand gestures made by deaf and dumb people and convert them into text so that the normal people can able to easily understand the gestures made.

1.1 Literature Survey

1.1.1 Su Min Lee, Boon Giin Lee implemented Sign Language Interpretation System with Sensors in 2017[2]. In this paper they designed the system for Deaf and dumb people use Gestures for interaction with others. The present System is a tool for them to blend into the society without any obstacle. This system uses wearable hand device for communication. This System translates the sign into American Sign Language. This System consists of 3 modules. Wearable device with sensor has following steps Processing, Module, and Output. The Wearable device is built using a 3D printer to hold the hardware components. The holder for wearable device is printed using flexible filaments [10]. This flexible filament is used to hold the flex sensor. It is placed on the first joint of each finger. These holders can accommodate different finger sizes of users. The sensor used with the device is flex sensor.

Flex sensor is connected to the top of each finger. It is used to measure the amount of deflection or bending. It consists of unidirectional and bidirectional type. This System use bidirectional flex sensor. This type of flex sensor is used to change the resistance when it bends both upward and downward direction. Processing module has following steps. The IMU Data which is collected by flex sensor is transferred to Arduino Pro mini 328. It acts as microprocessor (small computer on a single chip) which is used to get the data from the sensors. After receiving the data, the feature is extracted from the information and this information is passed as input to SVM classifier. In this model, 26 gestures pattern are used to represent 26 alphabets in ASL language [7]. The Gesture is classified and converted into text Application. After the detecting the sign, the text is transferred to developed android application through Bluetooth. This application converts the text into audio. The resistance of the flex sensor varies based on different hand size that is smaller hand size has lesser variation whereas larger hand size has larger variation to solve this drawback the values which is extracted from the sensor is normalized based on deviation values of each flex sensor. Since everyone hand size is different, it generates different flex sensor values. So, the deviation values of flex sensor are calculated for each target based on sensors readings. The hand movements are inherited from the IMU in three orientations that is Roll, Pitch, and Yaw. The IMU Data and flex sensor are gathered at a frequency of 100 Hz. To simplify the coding implementation, flexion degree in tabular format was used. The degree of flex sensor is categorized into three parts. The first part is "no bend" which is related with a normalized value within the range [0.0, 0.3). The second part is "partial bend" which is connected with a normalized value within the range [0.3, 0.7). And the last part is "complete bend" within the range of [0.7, 1.0). The signs are classified into 28 classes using SVM (support vector machine). It is a binary supervised learning, that is, the class labels can have the values of +1 and -1. The training process use quadratic optimization algorithm to separate the dataset. For training and testing part in SVM "leave one subject out" (LOO) method is used. In LOO method, it uses same data in training and testing so that it makes the model to know more about the data. The accuracy of each module is calculated using, AC= (TP+TN)/ (TP+TN+FP+FN'). The resistance of the flex sensor varies based on different hand size that is smaller hand size has lesser variation whereas larger hand size has larger variation. It is not portable that is it is difficult to carry the wearable device. In this system, it was designed and implemented in smart wearable hand device. Mobile application is developed in order to achieve the usability of the proposed smart wearable device with a goal of converting the text to speech.

1.1.2 Rajesh M Autee, Shweta S. Shinde implemented a model for Deaf and dumb to communicate with normal people using image processing in 2016. In this paper they designed a model for Deaf and dumb people face many barriers to communicate with normal people. They mostly use gesture (physical movements) to communicate, but it is difficult to understand by normal people. Hence the implemented system is based on hand gesture recognition system which makes the impaired people to feel comfortable while communicating. Identification of hand gestures is done then background segmentation, feature extraction, peak and angle of gestures is calculated. Finally, the Hand Gestures are recognized and converted it into voice and vice versa. All the process is done in MATLAB software [11]. MATLAB is one of high-level language. It contains many image processing operations so it makes easy process the image. The first step is to capture the image. Hence the image is captured using 20MP web camera using pre defined MATLAB command [1]. Image processing is done to process the image. So, this step is important in this process. The captured image is converted into gray image. Background Segmentation: It is used to separate the Hand Gestures from the Background. The Final Step in Image Preprocessing is noise elimination. The Feature is extracted using peak and angle calculation. It calculated the Number of raising and folding fingers. For each hand gestures 12bit hand gestures is assigned. The last stage is converting the recognized hand gesture is converted into Speech. In REVERSE PROCESS for feature extraction Mel Frequency Cepstral Coefficient [24] is used. This Technique is used for both testing and training phase. Dynamic wrapping is one Algorithm which is used to distinguish the feature vector of the signals. The disadvantage in this system is each user voice is unique, so the system cannot have template for each user. The communication time between normal people and impaired people is high. This system makes the impaired people to communicate with the normal people. It is achieved by using converting the speech into gestures and vice versa.

1.1.3 Communication system for silent speaker using hand gesture recognition was proposed by M. Nithyanandham, N. Sriram in the year 2013. It was implemented using hand gloves sensors. Technique used: simple template matching technique considering the maximum and minimum voltage values according to the accelerometer orientation towards gravity. Glove design using three axis-accelerometer. Microcontroller (ATMEGA 2560). Bluetooth chip used for wireless data transmission to the android mobile. The ASL signs, the alphabets A to Z are generally used by the silent speakers. They are analyzed by Flex sensors and accelerometer sensors [12]. Flex sensor are usually used in detecting the bending of fingers and accelerometer sensors are used in sensing the orientations. There are several sensors used for distinguishing and hence gloves are used. The Data gloves senses the finger movements "N" using sensor technology. The sensors are placed on glove at the back of the hand. The design is made by fixing accelerometers on each finger. The microcontroller with the Bluetooth is placed in rear end of the hand. The design is compact and simple. The requirement of this method is the recognition of ASL signs which involves stretching, bending and orientation of fingers and palm. Various kind of sensors are used to differentiate the movements and hence gloves

are used. Finger movements are measured using glove which are embedded with sensors [16]. For some alphabets like A, B, E, H and Z the readings may be varied with respect to the gravity in each of the plane and can produce maximum values. The entire alphabets are grouped into five categories. They can be recognized individually. The values can be maximum or minimum in any one of the planes. The values obtained are unique so that they can be differentiated easily. Hence with orientation analysis with the sensors placed at each finger, the sequence table is made for coding with the help of analysis. The obtained values from different orientations are gathered and tabulated. The ADC values may vary for same sign with difference in height. So, it may be difficult to code with the obtained values. Fixing of range is also a tedious task and it can result in errors.

Decoding of Gestures

When the controller gets the input sequence, it computes and compares them with coded values. When all the conditions are satisfied, the characters are sent to Bluetooth else the process is repeated until a valid sequence is obtained. The gestures that are recognized are sent to the android mobile through Bluetooth. "Blank space" is also assigned a code [20]. The developed android application which is named as speaking gestures, stores the alphabets in an array format. Then they are displayed as a word [6]. These words are given artificial voice which is to be understood by the dumb people. Certain signs (S, M and N) and (U, V) may overlap. So, the auto-correction option is provided to resolve it. Thus, a more accurate system with less use of sensors has been developed. A portable communication system using the accelerometer sensors, microcontroller and an android mobile has been designed. The computation process consists of simple back conversion of ADC to voltage for coding and decoding. The major drawback of existing system which has limited communication capacity has been overcome. The accuracy and the vocabulary are also good in comparison with other speech synthesizers. Thus, the dumb people can able to effectively communicate with others with the help of this artificial tongue (Android phone) [3].

1.1.4 Real-time gesture recognition system was proposed by Carlos R. del-Blanco, Tomas Mantec´on, Fernando Jaureguizar, Narciso Garcı´a in the year 2019 by using the near-infrared imagery. Leap motion (using infra-red) is used to track the hand skeleton information which is used to recognize the hand gesture. This method consists of sensors which uses the skeleton information processing because it simpler than the image processing method [19]. Thus, using the skeleton information method works faster for deploying the Human and Machine Interaction applications. Then the classifier is used to recognize the respective hand gesture. Two different methods are proposed for classifying the different gestures: Random Forest and Support Vector Machine (SVM). The purpose of device characteristics are background objects filtering like arms, wrists and chests etc., which are generally connected to the hands [4]. The feature extraction consists of four stages they are Hand segmentation normalization Block division of the hand region Computing the feature descriptors for each block Computation of the final hand descriptors. Initially, the hand segmentation normalization is used to normalize every segmented hand region to form a probable size. Next, the dividing the blocks for hand region separates the splited region which is normalized for avoiding the overlap on other blocks. Then the third stage is computation of feature descriptors per block, it is used to characterize every block's structural content. The final stage is computation of the final feature vectors, it makes a stack memory to store the hand gestures which is finally extracted. The dynamic hand gesture recognition uses Support Vector Machine (SVM) algorithm for classifying the different hand gestures, in both cases that it may be static or dynamic. The algorithms used in this implementation are**:** For classifying the different hand gestures, they used two types of algorithms they are:s multiclass SVM Nearest Neighbor (KNN). Artificial Neural Network (ANN) is used to differentiate among the various hand gesture. Convolutional Neural Network (CNN) is used for extracting the feature based on depth imagery. The drawbacks are they use leap motion sensors to track the human skeleton information which costs more. Using the infrared sensors, we can capture only the images which are closer to the sensor. It only recognizes and converts the gesture into text but not converting it from text to gestures. Finally, the conclusion says that the leap motion sensors with infra-red imagery are used for recognizing the gestures in real time. This process increases the performance on hand gesture recognition system [13]. And also, it eliminates the issues related to the classifying the hand poses. It uses the feature descriptors like HOG and LBP for feature extractions. The final result shows that the gesture recognition is achieved for the individual hand poses then the multiclass SVM [18] performs the recognition based on the histograms of previous hand gesture.

1.1.5 Gesture Recognition for Calculation and Text Output was proposed and implemented by Wei Fang, Yewen Ding, Feihong Zhang and Jack Sheng in the year 2019 based on CNN and DCGAN. The images are captured through web camera for collecting the data directly. But the sample data might be affected by the illuminations and noises. So initially the light detection process is done [5]. This process is carried out to get the high-quality data. By adjusting the camera angle and light intensity the illuminations can be eliminated. The next stage is image Pre-processing – Pre-processing is done with adding some filters to the images which helps to eliminate the noise and the background objects from the images to make the machine recognize the gesture with more accuracy. Then, the training/testing- The data set is separated into two parts for training and testing process. From the whole data set, the 80 percent of data is used for training and the remaining part of data is used to test the accuracy of the machine on gesture recognition [17]. The gesture Recognition was used for recognizing the gestures. Here, the recognition technique consists of two types as non-

contact and contact. The contact type interaction method mostly gets three-dimensional gesture information with the help of equipment like gloves, but the usage of hardware peripherals may reduce the comfort for interaction between the human and the computer and also the operator may feel it inconvenient. The non-contact interaction is typically known as a visual-based method, it helps the operators to establish the human-computer interaction more comfortable and natural without wearing any peripherals. They used DCGAN and CNN. Then the model is evaluated with various real datasets to examine that their developed model that can obtain good results. It can be tested by applying the recognition model that whether they can effectively recognize the real definition of the gesture or not. So that the new model can gain fully automated interaction and it can acquire high amount of accuracy [14]. The Calculation and Text Conversion was made based on the good results, the conclusion can be made that the same recognition and identification model can be applied even for complex tasks. The obtained data and developed identification model can be used for the calculation of gesture and the gesture is converted into text as output. Algorithms here are Convolution Neural Network (CNN) classification of the gestures during training/testing process. Deep Convolution Generative Adversarial Networks (DCGAN) consists of two models namely Generator - It is a neural network which creates the images that resolves the over fitting problem. Discriminator – It is also a neural network that examines the images to determine whether they are real or not. The main drawbacks**:** They trying to recognize both the expression and the gestures this action may leads to some complex problems. Because, the machine may get confused between the data of expression and gestures. The model can only perform recognizing the gesture and convert it to text. It has no ability to do the reverse process. It is used to communicate between the computer and human and not for the interaction between deaf and dumb person to normal people. The conclusion tells that the hand gesture recognition process in real time is completely based upon the algorithms DCGAN and CNN [23]. Initially, the particular gesture is recognized by using the user developed recognition model and it gives the specific meaning for the respective gestures [8]. The experiment made from the model proves that it can provide a favorable result. And also, it gives a way to achieve the automation with high accuracy. The DCGAN algorithm [21] is used to eliminate the over fitting problem by separating the training data from the available similar images. The good result with more accuracy will be achieved only after performing the pre-processing stages which helps to exclude the illumination from the images.

Table -1:

YEAR	AUTHOR	TITLE	IMPLEMENTATION
2016	Shweta S. Shinde, Rajesh M. Autee, Vitthal K. Bhosale	Two-way communication approaches for impaired person based on image processing in Real Time	It is implemented using image processing techniques using MATLAB.
2017	Boon Giin Lee, Su Min Lee	Interpretation of Sign Language using Wearable Hand Device	It is implemented by using the wearable gloves which is fixed with Sensor
2013	N. Sriram, M. Nithiyanandham	Communication System for Silent Speakers using Hand Gesture Recognition	It is implemented with the help of sensor gloves based on analysis of ASL signs.
2019	Carlos R. del-Blanco, Fernando Jaureguizar, Narciso García, Tom´as Mantec´on	Gesture Recognition System using Near Infrared Imagery	The Gesture Recognition is implemented using Leap Motion (near infra-red imagery)
2019	Yewen Ding, Feihong Zhang and Jack Sheng, Wei Fang	Gesture Recognition using CNN and DCGAN algorithms	Deep CNN and DCGAN algorithms used to implement the gesture recognition.

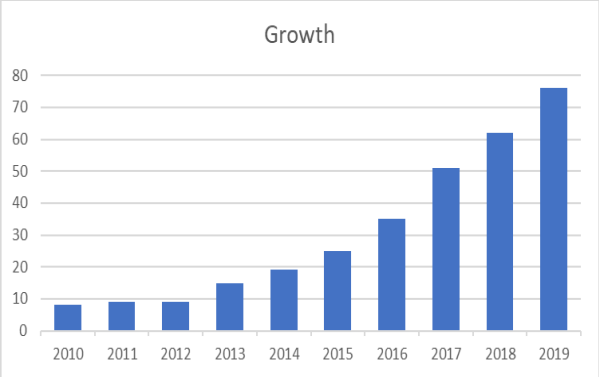

Chart -1: Growth of Machine Learning

3. CONCLUSION

With this technology Hand Gesture can be recognized with the CNN algorithm will provide us the best result other than using the sensors. The Hand Gesture Recognition will provide a two-way communication which helps to interact between the impaired people to normal people without any difficulties. Here, the NLP is used to feed the data and CNN is used to classify the different hand gestures.

REFERENCES

[1] Shweta S. Shinde, Rajesh M. Autee, Vitthal K. Bhosale, "Real time two-way communication approaches for hearing impaired and dumb person based on image processing" ,2016.

[2] Boon Giin Lee, Su Min Lee, "Smart Wearable Hand Device for Sign Language Interpretation System with Sensors Fusion" ,2017.

[3] N. Sriram, M. Nithiyanandham, "A Hand Gesture Recognition Based Communication System for Silent Speakers" ,2013.

[4] Tom´as Mantec´on, Carlos R. del-Blanco, Fernando Jaureguizar, Narciso Garcı´a, "A real-time gesture recognition system using near-infrared imagery",2019.

[5] Wei Fang, Yewen Ding, Feihong Zhang and Jack Sheng, "Gesture Recognition Based on CNN and DCGAN for Calculation and Text Output", 2019

[6] P. Sai Prasanth, A swathy Gopalakrishnan, Oviya Sivakumar, A. Aruna, "Enhancing User Experience Using Hand - Gesture Control", 2019.

[7] Jay Prakash, Uma Kant Gautam, "Hand Gesture Recognition", 2019.

[8] Wang Jingqiu , Zhang Ting, "An ARM-based embedded gesture recognition system using a data glove" ,2014.

[9] C.M. Travieso, J.B. Alonso, M.A. Ferrer, "Sign language to text by SVM" ,2003.

[10] V Bhavana, G Mohana Surya Mouli, G Venkata Lakshmi Lokesh, "Hand Gesture Recognition Using Otsu's Method", 2017.

[11] Guillaume Plouffe, Ana-Maria Cretu ,"Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Wrapping", 2015.

[12] Mark Billinghurst, Tham Piumsomboon, Huidong Bai, "Hands in Space: Gesture Interaction with Augmented-Reality Interfaces", 2014.

[13] Darya Frolova, Helman Stern, Sigal Berman, "Most Probable Longest Common Subsequence for Recognition of Gesture Character Input" ,2013.

[14] "Contour Model Based Hand Gesture Recognition Using the Kinetic Sensor" 2014.

[15] Cao Dong, Ming C.Leu,"American Sign Language Alphabet Recognition using Microsoft Kinect" ,2015.

[16] "Albanion Sign Language(AlbSL) Number Recognition from both hand's Gestures"

[17] Eriglen Gani , Alda Kika, , "Albanian Sign Language (AlbSL) Number Recognition from Both Hand's Gestures Acquired by Kinect Sensors" Vol. 7, No. 7, 2016.

[18] Miada A. Almasre, Hana Al-Nuaim "A Real-Time Letter Recognition Model for Arabic Sign Language Using Kinect and Leap Motion Controller v2", Vol-2, Issue-5,ISSN: 2454-1311,May 2016. [19] Rashmi. B. Hiremath, Ramesh. M. Kagalkar, "Methodology for Sign Language Video Interpretation in Hindi Text Language", Vol. 4, Issue 5, May 2016.

[20] Paulo Trigueiros, Ferando Ribeiro, Luis Paulo Reis, " Vision Based Portuguese Sign Language Recognition System", New Perspectives in Information System and Technologies, Volume 1,Advances in Intelligent System and Computing, Springer International Publishing Switzerland,pp-605-608,2014.

[21] Vajjarapu Lavanya, M.S. Akulapravin, Madhan Mohan, "Hand Gesture Recognition and Voice Conversion System Using Sign Language Transcription System", IJECT Vol. 5, Issue 4, Oct - Dec 2014.

[22] K. Sangeetha, L. Barathi Krishna,"Gesture Detection For Deaf and Dumb People ", International Journal of Development Research Vol. 4, Issue, 3, pp. 749-752, March, 2014.

[23] M.M.Gharasuie, H.Seyedarabi , "Real-time Dynamic Hand Gesture Recognition using Hidden Markov Models", 8th Iranian Conference on Machine Vision and Image Processing (MVIP), 978-1-4673-6184- 2/13/$31.00,2013.

[24] Aditi Kalsh, N.S. Garewal , "Sign Language Recognition for Deaf & Dumb", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 9, September 2013.

[25] Bhupinder Singh, Neha Kapur, Puneet Kaur , "Speach Recognition with Hidden Markov Model: A Review" International Journal of Advanced Research in Computer and Software Engineering, Vol. 2, Issue 3, March 2012.