International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

A Deep Learning Approach to Intelligent Gesture Recognition System for Deaf, Dumb and Blind Communication using KNN Algorithm on Tensor Flow Technique

Mr. N. Sowri Raja Pillai[1], Miss. V. Padmavathy[2], S. Nasrin[3]

[1]Head of the Department (Information Technology), Avvaiyar College of Engineering and Technology for Women.
[2,3]B.tech (Information Technology), Avvaiyar College of Engineering and Technology for Women.

ABSTRACT: *Communication is the fundamental channel to share thoughts. As of late the hard of hearing, stupid and visually impaired unfortunate casualties expanded. Since almost totally senseless can't speak with ordinary individual. New situation is the place the idiotic, hard of hearing speak with visually impaired individuals. Signal acknowledgment is the scientific explanation of a human movement by a registering gadget. Gesture based communication give best correspondence stage to the consultation hindered and moronic individual to speak with ordinary individual. The Deaf and dumb use hand motions to convey though dazzle individuals can hear just the voice and correspondence through voice. So change of hand signals to voice yield is the arrangement. So as to draw a stage nearer to these objective applications we use KNN calculation with Deep learning and Tensor flow method. The thoughts comprised of structuring and actualize a framework utilizing man-made brainpower, picture preparing and information mining ideas to accept contribution as hand signals and produce unmistakable yields as content and voice with 95% accuracy and above through proposed work.*

KEYWORDS: KNN, Deep learning, Tensor flow, Gesture,

I. INTRODUCTION

The present time is a zoom of innovation. Every single field has an effect of the innovative progressions onto it. One such quickly developing specialized progression is the expanding effect of cell phones on human life. The tremendous and consistently expanding Internet utilization alongside advanced cells has demonstrated a help to humankind. The PDA is a standout amongst the most critical electronic contraption in our life, since it is with the advanced mobile phone that we remain associated. In any case, in spite of these progressions, there is a sure piece of the general public which is denied of these advantages. The meeting crippled and quiet individuals can't blend with the social world on account of their physical incapacities. Inadvertently, they are treated in an uncommon way by whatever remains of the general public.

They can't be a piece of the get-togethers, state understudies can't consider in schools with ordinary understudies, elderly people can't work at work spots, and significantly more. Basic exercises like going and purchasing a ware from the basic supply shop is extremely confounded assignment for the almost totally senseless individual. The hole between typical people and almost totally senseless is wide and consistently expanding step by step. Today, the national check of hearing handicapped and quiet people all through India is roughly 17 lakh. Regardless of this extensive number, less research is done so as to connect the correspondence hindrance.

1.1 HANDGESTURE RECOGNITION SYSTEM

Gestures are a noteworthy type of human correspondence. Consequently motions can be observed to be an engaging method to interface with PCs, since they are as of now a characteristic piece of how individuals convey. An essential objective of motion acknowledgment is to make a framework which can distinguish explicit human signals and use them to pass on data for controlling gadget and by actualizing constant motion acknowledgment a client can control a PC by completing an explicit motion before a camcorder which is connected to the PC. An essential objective of this motion acknowledgment look into is to make a framework which can recognize explicit human motions for the control the traffic flags and mouse. This undertaking additionally covers different issues like what are motion, their grouping, their job in executing a motion acknowledgment framework for traffic and mouse control, framework engineering ideas for actualizing a motion acknowledgment framework, real issues associated with executing motion acknowledgment framework, and future extent of motion acknowledgment framework.

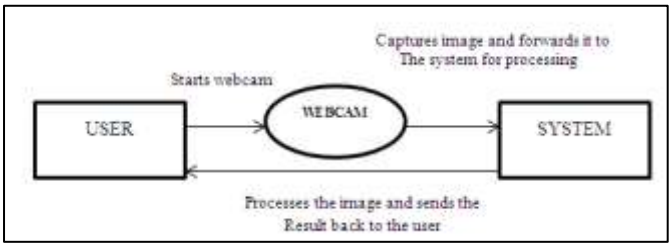

Fig 1.1Process of Hand Gesture Recognition System

1.2 SIGN LANGUAGE

The understanding of the complications of the handshapes utilized in marked languages is extremely simply starting. The human hand can make an immense range of conceivable shapes. All things being equal, gesture based communication will in general utilize just a predetermined number of handshapes to make the aggregate number of signs utilized in a given communication through signing. At the point when Deaf individuals comprise a moderately little extent of the all-inclusive community, Deaf people group regularly build up that are particular from the encompassing hearing network. These Deaf people group are extremely across the board on the planet, related particularly with gesture based communications utilized in urban zones and all through a country, and the way of life they have created are exceptionally rich.

1.2 INDIAN SIGN LANGUAGE

Indo-Pakistani Sign Language (IPSL) is the transcendent gesture based communication in South Asia, utilized by no less than a few hundred thousand hard of hearing endorsers (2003). Similarly as with many gesture based communications, it is hard to appraise numbers with any sureness, as the Census of India does not list communications through signing and most investigations have concentrated on the north and on urban regions.

(i)Number Signs: The numbers from zero to nine are framed in ISL by holding up a hand with the proper handshape for each number. From one to five the relating number of expanded fingers shapes the numeral sign, though for zero and the numbers from six to nine uncommon handshapes are utilized that get from composed numbers

(ii)Family Relationship: The signs for family relationship are gone before by the sign for 'male/man' and 'female/lady'

(iii)Sign families: Several signs have a place with same family on the off chance that they share at least one parameters including handshapes, place of enunciation and development.

II. LITERATURE REVEIW

Relatively few Researches have been done in this specific field, particularly in Binary Sign Language Recognition. Scarcely any looks into have been done on this issue however and some of them are as yet operational, yet no one had the ability to give an undeniable answer for the issue. Christopher Lee and Yangsheng Xu built up a glove-based motion acknowledgment framework that had the capacity to perceive 14 of the letters from the hand letters in order, learn new signals and ready to refresh the model of each motion in the framework in online mode, with an Intelligent Sign Language Recognition Using Image Processing rate of 10Hz.

2.1 ADAPTED CONVOLUTIONAL NEURAL NETWORK WITH DATA AUGMENTATION

Hand gestures give a characteristic method to people to collaborate with PCs to play out a wide range of uses. In any case, factors, for example, the intricacy of hand motion structures, contrasts close by size, hand act, and natural light can impact the execution of hand signal acknowledgment calculations. Ongoing advances in Deep Learning have essentially propelled the execution of picture acknowledgment frameworks. Specifically, the Deep Convolutional Neural Network has shown predominant execution in picture portrayal and arrangement, contrasted with ordinary machine learning approaches. This paper proposes an Adapted Deep Convolutional Neural Network (ADCNN) reasonable for hand signal acknowledgment undertakings. Information expansion is at first connected which shifts pictures both evenly and vertically to a degree of 20% of the first measurements haphazardly, so as to numerically build the span of the dataset and to include the strength required for a profound learning approach.

Methods	No. Epoch	Precision	Recall	F1 Score	Accuracy (%)
Baseline CNN	10	0.96	0.96	0.96	95.73
Proposed ADCNN	10	1	1	1	99.73

Table 2.1 Classification results of ADCNN

2.2 MACHINE LEARNING USING BAG OF WORDS

Recent works are endeavoring to tackle the issue of hand signals acknowledgment utilizing machine learning strategies. Some of them are putting on a show to accomplish elite.

Notwithstanding, few of them are considering compulsory necessities to apply the work process of a learning model, basically information unbalance, display determination and speculation execution metric decision.

In this work, a machine learning technique for continuous acknowledgment of 16 gestures of client hands utilizing the Kinect sensor that regards such necessities. The acknowledgment is activated just when there is a moving hand signal. The technique depends on the preparation of a Support Vector Machine show available profundity information from which pack of expressions of SIFT and SURF descriptors are removed.

2.3 DEEP LEARNING ON SKELETAL DATA

In this paper, we present another 3D hand motion acknowledgment approach dependent on a profound learning model. We propose another Convolutional Neural Network (CNN) where arrangements of hand-skeletal joints' positions are prepared by parallel convolutions; we at that point research the execution of this model available motion succession characterization errands. Our model just uses hand-skeletal information and no profundity picture. On the DHG dataset, our model accomplishes a 91.28% classification precision for the 14 motion class's case and an 84.35% classification exactness for the 28 signal class's case

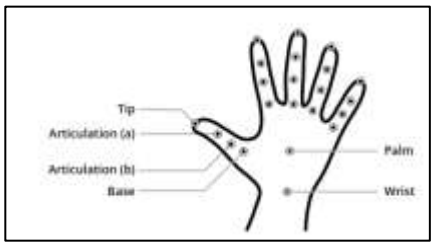

Fig 2.3 Hand skeleton returned by Intel Real Sense camera[6].

2.4 CONVOLUTIONAL NEURAL NETWORK

Hand gesture plays an important role in nonverbal communication and natural human-computer interaction. However, the complex hand gesture structure and various environment factors lead to low recognition rate. For instance, hand gesture depends on individuals, and different individuals' hands are with different sizes and postures, in addition, unconstrained environmental illumination also influences hand gesture recognition performance. Therefore, hand gesture recognition is still a challenging issue. This paper proposes a robust method for hand gesture recognition based on convolutional neural network, which is utilized to automatically extract the spatial and semantic feature of hand gesture. Our method consists of a modified Convolutional Neural Network structure and data preprocessing, which corporately increase hand gesture recognition performance. In the convolution layer, the feature map of the previous layer is

convoluted with a trained convolution kernel, and then the result is output through an activation function to get the feature map of this layer.

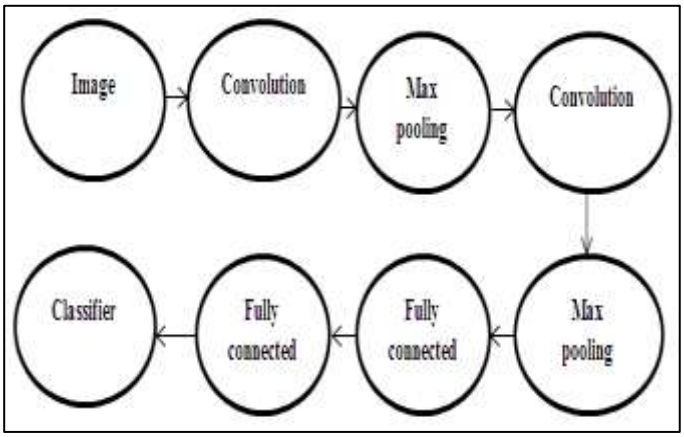

Fig 2.4 Architecture of Convolutional Neural Network

2.5 SPECTRAL-SPATIAL CONVOLUTIONAL NEURAL NETWORK

Hyperspectral picture (HSI) grouping has been a functioning theme as of late. In the course of recent decades, countless have been proposed to manage this issue. Anyway among these techniques, profound learning based strategies are uncommon. Enlivened by the magnificent execution of profound convolutional neural system (DCNN) in visual picture characterization, in this paper, we bring DCNN into HSI arrangement. Rather than utilizing two-measurement parts as DCNN is utilized in two-measurement picture grouping, one-measurement pieces is embraced in our DCNN to fit the HSI setting. The proposed technique is contrasted and the best in class profound learning based HSI order strategies, assessed on two prevalent datasets, and creates better arrangement results. Completely obliged unfixing issue by augmenting the l2 standard of the meager vector utilizing the ADMM strategy, and give the

reasons of utilizing the l2 regularization to supplant the l1 regularization and in addition demonstrate the sparsity of l2 regularization under ASC requirement.

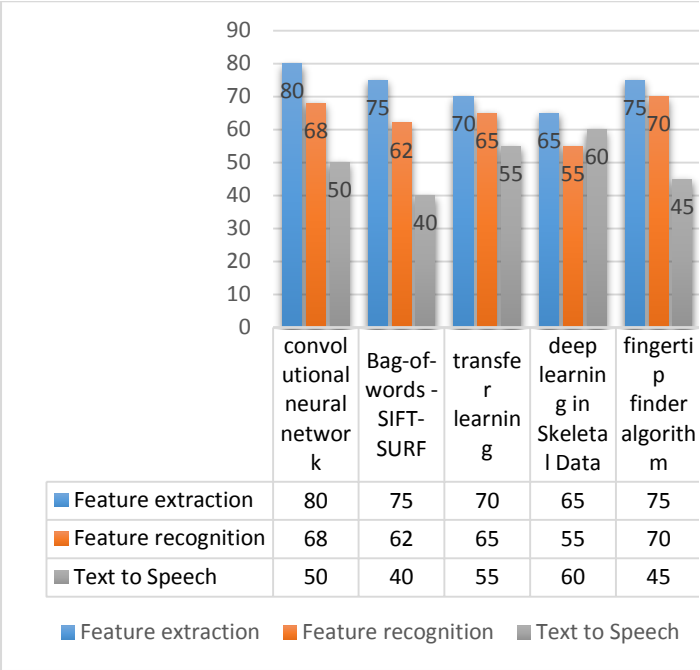

Table 2.5 Analysis of the various algorithms for Text to speech conversion

III. EXISTING METHOD

3.1 FINGER-TIP FINDER ALGORITHM

Combination of K shape and curved body: Convex body is making a polygon around the divided hand area. It encases all purposes of the district. The form of the polygon is identified by Sobel edge discovery calculation. The shape of the first hand motion is additionally distinguished. At that point the AND task is performed between these two pictures to locate the normal outskirts. This decreases the preparing time definitely in light of the fact that there is no compelling reason to utilize the K bend calculation to all form pixels of the hand shape. It likewise diminishes the likelihood of distinguishing false fingertip.

(i)PIXEL SEGMENTATION: In this calculation, a picture is divided into 16 hinders as appeared. Each square contains 256 pixels. At that point the quantity of white pixels in each square is determined. Along these lines, 16 squares give 16 highlight vectors containing the quantity of white pixels. In this framework, this calculation is utilized as a component for HGR.

(ii)ECCENTRICITY: The real hub of a picture is characterized as the line portion associating the two extraordinary focuses which makes the greatest Euclidean separation inside a shape. The minor hub of the shape is characterized as the line opposite to the significant hub and of such length that if a case goes through the external four of the two tomahawks, it totally encases the shape. The proportion of the real pivot to the minor hub is known as the flightiness of the shape.

(iii)ELONGATEDNESS: Elongatedness of an item is characterized as the proportion of the littler side to the bigger side of the base bouncing square shape along the article demonstrates the elongatedness of the hand shape

(iv)ROTATION: In ASL, there are a few signs which are fundamentally the same as. The main contrast between them is the turn point. In such cases, the revolution point is the main component that can separate those signs. In our examination, some mind boggling letters in order indications of ASL are supplanted by other pivoted ASL signs for better acknowledgment rate.

3.2 METHODOLOGY

The framework is intended to outwardly perceive every single static motion of American Sign Language (ASL) with uncovered hand. Distinctive clients have diverse hand shapes and skin hues, making it increasingly troublesome for the framework to perceive a motion. The framework consolidates five component extraction calculations for client autonomous and powerful hand motion acknowledgment. The entire framework works in four stages for signal acknowledgment, for example, picture obtaining, preprocessing, highlight extraction and highlight acknowledgment.

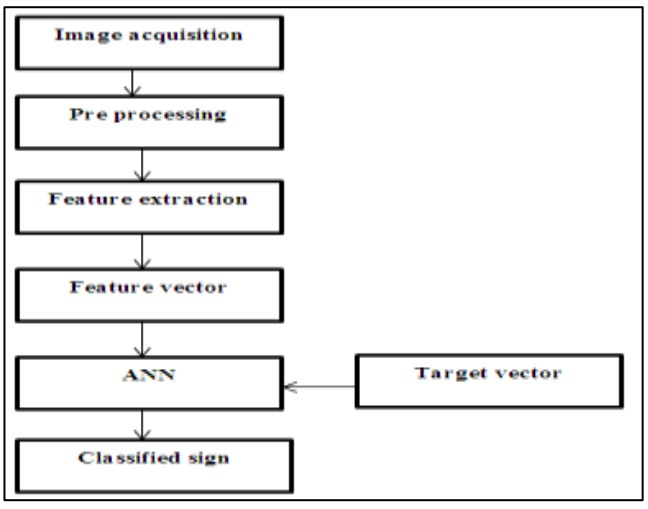

Fig 3.2 Existing block Diagram for Detection System

(i)IMAGE ACQUISITION: A sum of 50 picture tests of each indication of ASL is gathered from various individuals. A database of 1850 pictures of 37 signs is made to separate element vectors. The signs for all letter sets and quantities of ASL. A neural system is prepared utilizing this database.

(ii)PREPROCESSING: Pre-preparing of a picture is essential as the picture gotten from the camera may contain diverse clamors. At first, the pictures are resized to 260×260 pixels. At that point, they are changed over from RGB to parallel by Global histogram edge utilizing Otsu's strategy. At that point middle sifting is improved the situation expelling commotion and saving the edges. Morphological framework and outline are utilized for filling gaps and smoothing the edges. For sign identification, the part from wrist to fingers of a hand is required. So whatever remains of the part is wiped out from the picture by trimming it. At that point the fingers of the pictures are should have been adjusted vertically upward. So the pictures are turned from 0 to 360 degrees regarding hand wrist position. Something like 15 back to back white pixels are looked at the base of a picture to find the wrist. On the off chance that white pixels are found at the base of the picture, no pivot is required. On the off chance that not, the picture is pivoted by 90 degrees clockwise and checked once more. Along these lines, the circle proceeds until the point when the wrist is found.

(iii)FEATURE EXTRACTION: A picture can be recognized and arranged by a few of intrigue or set of qualities called the highlights.

Existing paper, five unmistakable highlights, for example, fingertip discoverer, unconventionality, elongatedness, pixel division and pivot are utilized for highlight extraction. Highlight Recognition Multilayer feed forward neural system with back proliferation preparing calculation is utilized to distinguish American numbers and letter sets. ANN emulates organic neural system framework. As a matter of fact neural systems are prepared with the goal that a specific info prompts an explicit target yield by modifying the estimations of the associations between components. In the preparation period of the neural system, input vector is equivalent to the element vectors and target vector is equivalent to the quantity of signs utilized for acknowledgment. Each sign is spoken to by a vector containing 30 highlights. ANN has three layers, for example, input, covered up and yield. Concealed layer has a variable number of neurons with outspread premise work. Here 20 concealed layers are utilized with experimentation reason for best execution. Fifty example pictures of each sign from various individual are taken to prepare, test and approve the ANN.

IV. PROBLEM DEFINITION

Task of hand gesture acknowledgment is one the imperative and basic issue. With ongoing advances human association frameworks are to be manufacture. So the current framework can deliver the content yield from the hand signal contribution through webcam. At that point there is a database associated in the current one where we can just utilize predetermined number of hand signals .Since we have set number of sign that can be performed through hands. Shape that we take upon the productivity issue. Hand gesture has 3 process.

(i)FEATURE EXTRACTION: In the existing one the background errors are to be removed much more efficiently

(ii)FEATURE RECOGNITION: In the existing one training and implementation phase can be done by a single person only different color of the human hand are not accepted.

(iii)TEXT TO SPEECH CONVERSION: The main drawback is the blind people can only hear but cannot see. So in the existing system does not have a facility to enable the voice output if provided also the text to speech voice is to be recorded. To record the numerous number of voices for each and every word or sentence are difficult.

V. PROPOSED SYSTEM

The whole system works in four steps for gesture recognition such as image acquisition, preprocessing, feature extraction and gesture recognition. The flow of the proposed system is the aim of this project is to develop a system that can convert the hand gestures into voice. The focus of this project is to place the pictures in the temporary database and with temporary database matching the image is converted into speech. The detection involves observation of hand movement.

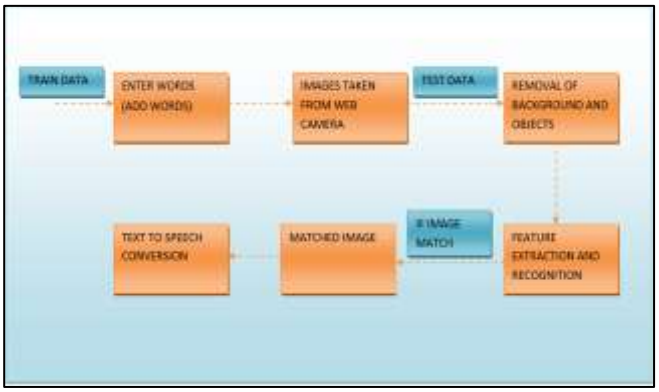

Fig 4.1 Architecture of the proposed system

(i)TRAIN DATA: Preparing and testing are two regular ideas in machine learning. Preparing and testing are all the more effortlessly clarified in the system of administered realizing, where you have a preparation dataset for which you know both info information and in addition extra credits that you need to foresee. Preparing comprises in taking in a connection among

information and traits from a small amount of the preparation dataset, and testing comprises in testing forecasts of this connection on another piece of the dataset.

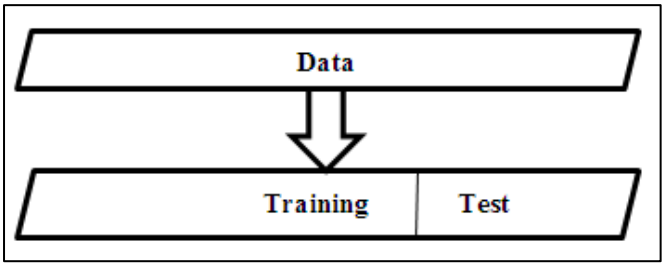

Fig 4.2 Data sent for training phase

(ii)ENTER WORDS (ADD WORDS): In that preparation information we have to add words to the relating signs. Entered words are coordinated to the signs. It will be perceived when (testing information) the sign is coordinated then the comparing content is changed over to voice.

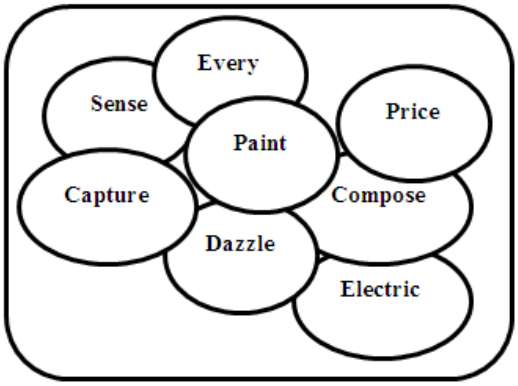

Fig 4.3 Input for the application

(iii)PICTURES TAKEN FROM WEB CAMERA: Webcam as opposed to going to costly sessions with an educator. In contrast to the mirror, the framework can likewise give moment input on the accuracy of the hand motion.

Our decision of utilizing webcam pictures as our info arrange likewise makes some trouble in effectively grouping hand shapes. Contingent upon the nature of the webcam, caught picture information normally contains commotion from pressure. Since we can't compel where the client's hand will be in connection to the camera, our acknowledgment framework should be invariant to scale and position.

(iv)EVACUATION OF BACKGROUND AND OBJECTS: Expelling a foundation from a picture is a typical errand that each originator needs to do on a relatively normal premise. There are such huge numbers of ways that you can remove a subject from its experience to wind up with a straightforward foundation. With all the distinctive courses, there is truly is no correct technique that you should take without fail. Contingent upon the foundation or multifaceted nature is the means by which you ought to figure out which technique you should take.

In the event that you have the advantage of being the creator and also the picture taker, at that point you have finish control of the complexity between the subject and the foundation. Endeavor to make the foundation as plain as conceivable with profoundly characterized edges. This will make any of the strategies beneath a simple procedure in evacuating the foundation. More often than not you will be provided a picture or need to discover one on a stock site. Shockingly, if the foundation is quite unpredictable, you will have all the more a tedious street in front of you. You ought to survey your picture and endeavor to make sense of which strategy will be the most effortless and cleanest route workable for your plan. Turning into an ace at extraction is extremely valuable in all plan employments. Experience the diverse choices beneath and discover the strategy that you like the best. More than likely you will end up utilizing a blend of various methods to accomplish an ideal extraction.

(v)FEATURE EXTRACTION AND RECOGNITION: Highlight extraction begins from an underlying arrangement of estimated information and fabricates inferred values (highlights) planned to be educational and non-repetitive, encouraging the consequent learning and speculation steps, and now and again prompting better human translations. Highlight extraction is identified with decrease. At the point when the information to a calculation is too huge to be in any way prepared and it is suspected to be repetitive (e.g. a similar estimation in the two feet and meters, or the redundancy of pictures introduced as pixels), at that point it tends to be changed into a diminished arrangement of highlights (additionally named an element vector). Deciding a subset of the underlying highlights is called include determination. Highlight extraction includes diminishing the measure of assets required to depict an extensive arrangement of information. When performing investigation of complex information one of the significant issues originates from the quantity of factors included. Examination with

countless for the most part requires a lot of memory and calculation control.

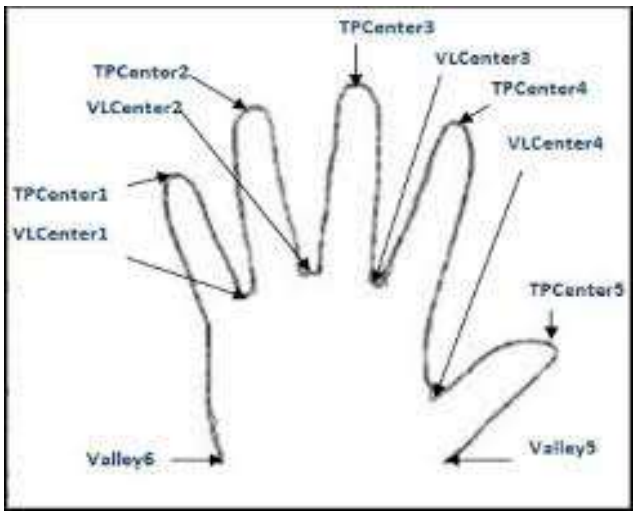

Fig 4.3 Feature extraction from the image

(vi)MATCHED IMAGE: Picture and performs solid coordinating between two pictures having distinctive scales. We have demonstrated that the highlights are invariant to scale, enlightenment, jumbles and give hearty coordinating to acknowledgment. The procured highlights are exceptionally particular to such an extent that a solitary component can be accurately coordinated with high likelihood.

(vii)TEXT TO SPEECH CONVERSION: This framework can be utilized for changing over communication via gestures to voice and furthermore voice to gesture based communication. A movement catch framework is utilized for communication via gestures change and a voice acknowledgment framework for voice transformation. It catches the signs and manages on the screen as composing. It additionally catches the voice and shows the communication through signing significance on the screen as motioned picture or video.

VI. ALGORITHM USED

A. Training phase

In the training phase the extraction algorithms are used to extract the values from the images.

(i). KNN algorithm

Utilizing the crude information removed from the sensor, the highlight first venture of making the element vector was to make it with a length of 10, one for each finger utilizing two hands. By figuring the length from each fingertip to the palm of the particular hand, a relative portrayal of each fingertip position was made as found in fig 4.3.

Tip point:

d (s,t) = Square root [(s1 - t1)2+ (s1 - t1)2]

To acquire these separations, the Euclidean separation in condition was determined, bringing about a variety of 10 esteems. As a technique for reflection, to contemplate the connection between the fingers, paying little mind to the span of the hand, the information was standardized.

Valley point:

d (s,v) = Square root [(s1 - v1)2+ (s1 - v1)2]

In any case, since an imperative perspective was to ration the co-connection between the fingers on each hand, every one of the 10 separations were standardized in regard to their hand utilizing the standardization recipe condition 2.The subsequent standardization had values in the interim.

(ii). Deep learning

The picture will be changed over to greyscale (scope of dim shades from white to dark) the PC will allocate every pixel an esteem dependent on how dim it is. Every one of the numbers are put into a cluster and the PC does calculations on that array. Colors could be spoken to as RGB values (a mix of red, green and blue running from 0 to 255). PCs could then concentrate the RGB estimation of every pixel and put the outcome in a cluster for elucidation.

At the point when the PC translates another picture, it will change over the picture to an exhibit by utilizing a similar system, which at that point looks at the examples of numbers against the definitely known items. The PC at that point allocates certainty scores for each class. The class with the most elevated certainty score is normally the anticipated one.

B. Implementation phase

In this phase the image given at the real time is been detected and the according text and voice output is seen.

(i)Deep learning with Tensor flow

We report a classification precision of 95% and computational efficiency of 110ms per sequence. Given a video succession containing hand motion motion, with T outlines, the hand signal acknowledgment calculation doles out a motion mark g from a lot of pre-defined motion names to the video. In this paper, the LRCN

profound learning system [1] is embraced to evaluate the hand signal name. In the first LRCN calculation, the creators test D < N outlines from the first video arrangement, and acquire the examined video grouping. The inspected arrangement is given as the contribution to the LRCN. In our proposed calculation, we enhance the precision and classification exactness of the LRCN structure, by removing T < S agent outlines from the motion based video arrangement, acquiring the delegate video sequence. These delegate outlines are then given as a contribution to the LRCN to evaluate the hand signal name. The issue of order of different sorts of hand signal stances can't be fathomed by just bolstering the system with the pictures of different kinds of postures. To add to the burdens, even the different sorts of information growths won't have the capacity to help you in powerfully unravelling this problem. A great method for tackling this issue will be by perceiving different milestones present in a human hand. This incorporates not just distinguishing the five fingers present in a human hand yet in addition recognizing the joints of fingers present. When you get the tourist spots and areas of the fingers, attempt to set up the geometry it is making and afterward order the posture of the hand.

This specific arrangement will likewise be exceptionally powerful, if at all the sorts of hand postures to order changes and our model will require less changes to adjust to new classes.

(ii)Text to speech engine

Here we use the Alex Net api for the conversion of the text to speech. There are actually 3 levels involved, which will refer as

- Text to words

So the underlying stage in discourse combination, which is for the most part called pre-handling or standardization, is tied in with diminishing uncertainty: it's tied in with narrowing down the a wide range of ways you could peruse a bit of content into the one that is the most suitable.

- Words to phonetic

For each word, we would require a rundown of the phonemes that make up its sound.

- Phonetic to sound

Another is for the PC to produce the phonemes itself by creating essential sound frequencies (somewhat like a music synthesizer), and a third methodology is to imitate the instrument of the human voice.

VII. PERFORMANCE ANALYSIS

To do a real time hand gesture recognition performance analysis, each sign is taken from five different people. Two images are collected from each people while a black background and proper illumination are maintained. Features are extracted from the images and tested in the real-time database-trained neural network. The number of correct responses out of 10 times of testing of each sign is shown in Table4.4

VIII. CONCLUSION AND FUTURE ENHANCEMENTS

The proposed framework is a communicator among Deaf and Dumb-Blind individuals and in addition Normal People goes for crossing over the correspondence hole between two conditions of the general public. Significant measure of work has been done already in this area, however this venture includes emit correspondence in a productive way in light of the fact that the framework will be executed in people in general host so that is can be utilized in versatile entry too. In this way, it genuinely serves its needs in all perspectives .methodologies demonstrate productive regarding time and exactness. It gives a productivity of about 95%. Further improvements, can be made as far as executing the communicator with the other way around idea (i.e.) changing over the voice contribution to image output.

Inputs	Correct training Response(10)	Extraction Rate(training)	Correct testing Response(10)	Recognition Rate(testing)	Average Voice Output rate
1-9	9.2	92%	9.7	97%	95%
A-Z	9.8	98%	10	100%	99%
a-z	9.5	95%	9.7	97%	96%
Special characters	9.7	97%	9.8	98%	97.5%
Two letter word	9.3	93%	9.8	98%	96%
Three letter word	9.4	94%	9.7	97%	95%
Four letter word	9.2	92%	9.6	96%	94%
One sentence	9.0	90%	9.5	95%	92%

Table 4.4 Average Voice Output Analysis Rate for the Proposed system = 95.56%

REFERENCES

[1] Md. Mohiminul Islam, Sarah Siddiqua, and Jawata Afnan " Real Time Hand Gesture Recognition Using Different Algorithms Based on American Sign Language", in 2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (ICIVPR).

[2] Guillaume Plouffe and Ana-Maria Cretu, "Static and dynamic hand gesture recognition system in depth data using dynamic time warping" IEEE Transactions on Instrumentation and Measurement (Volume: 65, Issue: 2, Feb. 2016).

[3] A. A. Alani, G. Cosma, A. Taherkhani and T. M. McGinnity, "Hand gesture recognition using an adapted convolutional neural network with data augmentation," *2018 4th International Conference on Information Management (ICIM)*, Oxford, 2018, pp. 5-12.

[4] M. F. Wahid, R. Tafreshi, M. Al-Sowaidi and R. Langari, "An efficient approach to recognize hand gestures using machine-learning algorithms," *2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME)*, Tunis, 2018, pp. 171-176.

[5] M. Benmoussa and A. Mahmoudi, "Machine learning for hand gesture recognition using bag-of-words," *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, Fez, 2018, pp. 1-7.

[6] G. Devineau, F. Moutarde, W. Xi and J. Yang, "Deep Learning for Hand Gesture Recognition on Skeletal Data," *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, 2018, pp. 106-113.

[7] S. Hussain, R. Saxena, X. Han, J. A. Khan and H. Shin, "Hand gesture recognition using deep learning," *2017 International SoC Design Conference (ISOCC)*, Seoul, 2017, pp. 48-49.

[8] Li, Gongfa & Tang, Heng & Sun, Ying & Kong, Jianyi & Jiang, Guozhang & Jiang, Du & Tao, Bo & Xu, Shuang & Liu, Honghai. (2017). Hand gesture recognition based on convolution neural network. Cluster Computing. 10.1007/s10586-017-1435-x.

[9] A. Liu, N. Xu, W. Nie, Y. Su, Y. Wong and M. Kankanhalli, "Benchmarking a Multimodal and Multiview and Interactive Dataset for Human Action Recognition," in *IEEE Transactions on Cybernetics*, vol. 47, no. 7, pp. 1781-1794, July 2017.P.

[10] S. Bambach, S. Lee, D. J. Crandall, and C. Yu, "Lending a hand: Detecting hands and recognizing activities in complex egocentric interactions," in The IEEE International Conference on Computer Vision (ICCV), December 2015.

[11] Y.Huang,X.Liu,X.Zhang,andL.Jin,"Apointinggesture based egocentric interaction system: Dataset, approach and application," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 16–23.

[12] G. Rogez, J. S. Supancic, and D. Ramanan, "Understanding everyday hands in action from rgb-d images," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3889–3897.

[13] L. Baraldi, F. Paci, G. Serra, L. Benini, and R. Cucchiara, "Gesture recognition in ego-centric videos using dense trajectories and hand segmentation."

[14] S. Ruffieux, D. Lalanne, and E. Mugellini, "Chairgest: a challenge for multimodal mid-air gesture recognition for close hci," in Proceedings of the 15th ACM on International conference on multimodal interaction. ACM, 2013, pp. 483– 488.

[15] S. Escalera, X. Bar´o, J. Gonz`alez, M. ´A. Bautista, M. Madadi, M. Reyes, V. Ponce-L´opez, H. J. Escalante, J. Shotton, and I. Guyon, "Chalearn looking at people challenge 2014: Dataset and results," in Computer Vision - ECCV 2014 Workshops, 2014, pp. 459–473.

[16] E. Ohn Bar and M. M. Trivedi, "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 6, pp. 2368–2377, 2014.

[17] P. Molchanov, X. Yang, S. D. Mello, K. Kim, S. Tyree, and J. Kautz, "Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural networks," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016.

[18] T. Starner, J. Weaver, and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based video," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371–1375, 1998.

[19] S. Karaman, J. Benois-Pineau, V. Dovgalecs, R. M´egret, J. Pinquier, R. Andr´e-Obrecht, Y. Ga¨estel, and J.-F. Dartigues, "Hierarchical hidden markov model in detecting activities of daily living in wearable videos for studies of dementia," Multimedia tools and applications, vol. 69, no. 3, pp. 743–771, 2014.