Real Time Sign Language Recognition and Translation to Text for Vocally and Hearing Impaired People

Khallikkunaisa
Associate Professor
Dept. Of CSE
HKBK College of Engineering
Bangalore, India

Arshiya Kulsoom A
Dept. Of CSE
HKBK College of Engineering,
Bangalore, India

Chandan Y P
Dept. Of CSE
HKBK College of Engineering
Bangalore, India

Fathima Farheen
Dept. Of CSE
HKBK College of Engineering,
Bangalore, India

Neha Halima
Dept. Of CSE
HKBK College of Engineering
Bangalore, India

***Abstract* - With a population of around 7.8 billion today communication is a strong means for understanding each other. Around 9,000 million individuals are vocally and hearing impaired. Because of the impediment there is a communication gap between the vocally handicapped people and the typical individuals. Gesture based communication is the fundamental method of communication for this section of our society. This language uses a set of representations which are finger sign, expression or mixture of both to precise their information among others. This system presents a completely unique approach of application based translation of sign-action analysis, recognition and generating a text description in English. Where it uses two important steps training and testing. In training set of fifty different domains of video samples are collected, each domain contains 5 samples and assigns a category of words to every video sample and it will be stored in the database. Wherein testing, test sample undergoes preprocessing using median filter, canny operator for edge detection, HOG (Histogram of Oriented Gradients) for feature extraction. SVM (Support Vector Machine) takes input as a HOG features and predicts the class label based on trained SVM model. Finally the text description will be generated in English language. The average computation time is minimum and with acceptable recognition rate and validates the performance efficiency over the normal model.**

Keywords- Sign language translation; vision based; preprocessing; feature extraction

I. INTRODUCTION

In regular day to day existence sign based communication is the main type of communication for the vocally impaired people. Communication through signs can't be grasped by every person since sign based communication has its own syntax and vocabulary. Sign based communication can be separated into two procedures ie vision based and sensor based system. This paper relies upon vision based technique.

Many researches have been directed on gesture based communication as it has gotten a part of acknowledgment. Prior communication through signing interpretation was generally subject to sensor based procedure. This method utilizes gloves with sensors which is interface with a recipient towards one side. But this strategy has its very own disadvantages. In this way numerous examples, profound learning, AI, convolutional neural system has improved the communication through signing interpretation.

The vast majority of the frameworks feeds the recorded video to the gadget, where the video is partitioned into numerous frames as shown in [1]. K. Bantupalli [1] utilized a CNN 3.

II. LITERATURE SURVEY

[1] built an informational index of a hundred kind of unique signs from the American Sign Language informational index. Each sign is played out on numerous occasions by a single endorser in changing lighting conditions and speed of marking. With the ultimate objective of consistency, the establishment in all of the accounts is the proportional. The recordings were recorded on a camera. The model stood up to issues in regards to the facial features and shade of the skin. While testing with various skin tones, the model dropped precision in the event that it hadn't been prepared on a specific skin color and was made to anticipate on it. The model likewise experienced loss of precision with the consideration of countenances, as appearances of endorsers shift, the model winds up preparing off base highlights from the recordings.

Morocho [2] considers the dataset as contribution from where the framework will remove the features and make sense of how to separate the different classes, later changing is applied to significant learning models that have been set up before on a substitute dataset. They proposed to empty the last plan of totally related layers of the present models, and displace them

with their new game plan of totally related layers Then, the framework is developed using a touch of learning rate all together for the new totally related layer to start taking in structures from the past convolutional layers in the fundamental period of the building. Applying this approach, they balanced a pre-arranged CNN to translate on new classes that were not set up on, along these lines showing up at a higher exactness.

M H Jaward [4] focused on making a mediator for communication by means of signs using a phone which is advantageous for utilization. In this paper the structure is executed reliant on the image handling technique to perceive the pictures of unique sign motions. It is prepared to decipher 16 various American communication by means of gestures signals with a precision of 97.13% In this paper, 16 static ASL letter sets are to be seen dynamically using Nokia Lumia 1520 cellphone with Windows Phone 8.1 working framework. Here the check is made over EmguCV library. All photographs caught are assessed into setup of 320 x 240 pixels in RGB(Red Green Blue) position. Canny edge identification and region developing system is utilized to determine the edges of hand motions. SURF highlights are gotten from edge recognized picture and are gathered into 16 classes of communication through signing utilizing K means clustering.

In [5] K Dixit shows a procedure which sees the Indian Sign Language (ISL) and converts into a conventional text. The perspective includes three phases, to be explicit a preparation arrange, a testing stage and a recognition state. A multi-class Support Vector Machine (MSVM) is used for planning and seeing signs of ISL. The ampleness of the proposed methodology is affirmed on a dataset having 720 pictures. Preliminary results show that the proposed system can viably see hand signal with 96% acknowledgment rate. The procedure involves preprocessing, which is applied to pictures before getting details from the hand motions. Segmentation of image is performed using Global Threshold Algorithm. By using Hu -Invariant and Structural Shape Descriptors feature extraction is done. It utilizes a Multi-Class Support Vector Machine (MSVM) to arrange the hand signals among various classes. Utilization of MSVM improves the acknowledgment rate in the framework.

K K Dutta[6] considered double handed Indian communication via gestures and it is seized as a progression of pictures and a while later these photos are dealt with using MATLAB, later converted into text and speech. Here the makers are actualizing a framework which gives a voice to vocally people. The system is altered with twofold hand gesture based communication by using least Eigen value calculation. Logitech web camera is used for picture catching and further handling is finished using MATLAB. The yield is seen in the wake of extricating Shi-Thomasi algorithm great highlights. This yield is the content which is additionally changed over into speech utilizing content to speech Synthesis.

In [7] charge-transfer touch sensors are embedded in a glove to translate the American Sign Language. In [7] this glove has a binary detection system which has a lot of digital touch sensors rather than the analog signs given by variable resistor, thus obtaining great recognition accuracy. This framework executes the recognition of 10 digits and every one of the 26 English letters in order utilized in American Sign Language. Here [7] is utilizing 8 autonomous capacitive touch sensors for gesture recognition, which gives yield as paired OFF/ON signals. The code for this framework is composed utilizing Python 3.0 under the Linux improvement condition and it has approx. 300 lines of code in it. Through 1080 trials, this system achieves an accuracy of 92 %. Here [7] is utilizing 8 autonomous capacitive contact sensors for motion acknowledgment, which gives yield as twofold OFF/ON signals. At the point when the capacitive sensors are brought inside 1.6 mm of human skin they get activated. To perceive the motion from the letters beginning to end and numbers 0 to 9 mapping of blend of 8 sensors are utilized. Furthermore, the principle motivation behind why the blend of 8 sensors are utilized is that it is unambiguous. PIC-116 capacitive sensor module is utilized to execute every one of these sensors. This PIC-116 module comprises of QT100 capacitive sensor coordinated circuit (Quantum Technologies) as its center. The motivation behind why QT100 is prepared is to lead remuneration, clamor crossing out and self-alignment which gives contact detecting high vigorous.

This [8] paper design and implements a smart glove for hearing and vocally impaired individuals. There have a few past researches to assist these individuals with communicating and convey to other people. The total structure comprises of a transmitter side, beneficiary side, microcontroller and TFT LCD. The transmitter area contains 5 flex sensors which distinguishes the bending and movements of each finger The product used to program the whole framework is Arduino Software. Here the framework comprises of both hardware and software. The whole implementation is glove based.

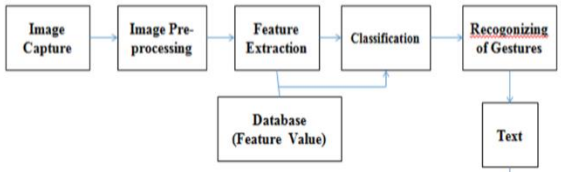
The [9] paper focuses on the pace of hearing impaired. This article subtleties the improvements and current structure of the Virtual Sign stage, a bidirectional communication through signing to content interpretation device that has been being created since 2015. The stage is divided into two principle parts, sign to content and content to sign and the two segments are depicted and clarified. Here the substance to discourse interpreter is utilized like customary online interpreters where the client enters message on a white box and immediately observe the outcome in other language. And the outcome is the 3D symbol that play out the motions relating to the composed content. For the sign to content interpretation it requires two unique parts that is a hand catch application to manufacture datasets for grouping, and the sign to content interpreter application itself. Here these segments utilize two bits of innovation to complete the interpretation that is a Microsoft Kinect sensor and a couple of information gloves. In [16] Automation system where users can use voice commands to control their electrical appliances, such as light, fan, television, heater etc.

III. METHODOLOGY

To represent the productivity of communication through signing, a lot of sign activities is considered in training to

create the content. This sign activity frames are then used to calculate the performance of sign recognition. Word handling is done as a recursive procedure of a sign activity image portrayal, where each frame information are prepared with HOG(Histogram of Oriented Gradients) highlights. In view of the frame reading rate the edges are separated and various edges are prepared in progressive arrangement to remove the area of intrigue. To perform the data handling, the fundamental methodology of the created framework is appeared in figure.

BLOCK DIAGRAM:



The principle motivation behind the given framework is to give a stage to vocally incapacitated people to share their perspectives among each one. The frames which are separated from the recordings are utilized for processing. There is a prerequisite to investigate the frame and that frame activity is distinguished. Preprocessing is utilized to expel the haziness from the picture. The general frameworks performs preprocessing, Feature extraction, Classification, Detecting the hand signal lastly creating text depiction.

This framework has two fundamental stages training and testing. Training procedure will be implemented for a created database. The training and testing follows underneath steps,

A. Video Acquisition

Video procurement is a procedure wherein the video delivered by the client is acquired which later experiences interpretation to change recordings into frames. At this point the whole arrangement of frames experiences preprocessing before feature extraction. In the meantime video is caught from the web camera and is split into various frames. Every individual picture is known as an frame. Video confining is the way toward removing outlines from the given video utilizing video traits like frame rate. For example the video length is the 2 min and 29 secs and casing rate will be 1000 FPS (Frame every second) at that point extricated edges will be 14900 frame.

B. *Pre-processing*

Preprocessing is performed on the video to get rid of blurriness and noise. Video contains enormous amount of frames which contain visual distortion like a video shot in low light area, voice distortion, light conditions etc. The preprocessing may be a common stage in every picture processing area. The principle motivation behind preprocessing is to reduce commotion on the edge and improve the image features for further handling. The median filter uses the nonlinear filtering techniques to remove blurriness from the input frame. Filtering techniques is used to upgrade image quality and output results. The main objective behind median filtering is replacing every entry with neighboring entry.

C. *Feature Extraction*

HOG is generally used to derive the feature from the input image. HOG is a picture processing algorithm. It will identify the object in the picture by utilizing feature vector. Whenever it derives a feature from the image it provides an output in the form of feature vectors. Feature contains the local shape information, that can be used for many tasks such as classification, detection of objects, tracking the object.

D. *Classification*

SVM(Support Vector Machine) characterization is a binary (two-class) order system, which must be modified to deal with the multiclass undertakings in certifiable circumstances. SVM order makes use of the features of the picture to the group. The characterization exploits prepared video and group testing video with specific depiction and provides the result.

E. *Text Generation*

Finally after the classification procedure is done the proportional content depiction will be delivered with the assistance of the class names that are allocated during the training stage.

IV. DESIGN

The framework is intended to visually perceive all the static indications of the Sign Language and all indications of letter sets utilizing uncovered hands. It is not necessary for the client/signer to wear the gloves or utilize any gadget to interface with the framework. Since various signers differ in their grasp shapes, body size, size, and activity propensities etc., which gets more challenging for recognition. In this way, it requires the need for endorser's autonomous communication via sign language recognition to improve the framework power and feasibility later on. The framework gives the examination of the three component extraction strategies utilized for Sign Language recognition and recommends a strategy dependent on acknowledgment rate. It depends on introducing the signal as a component vector that is translation, rotation and scale invariant. The blend of the feature extraction technique with phenomenal image preparing and neural systems capacities has prompted the effective improvement of Sign Language recognition framework. The framework has two stages: the component extraction stage and the classification stage.

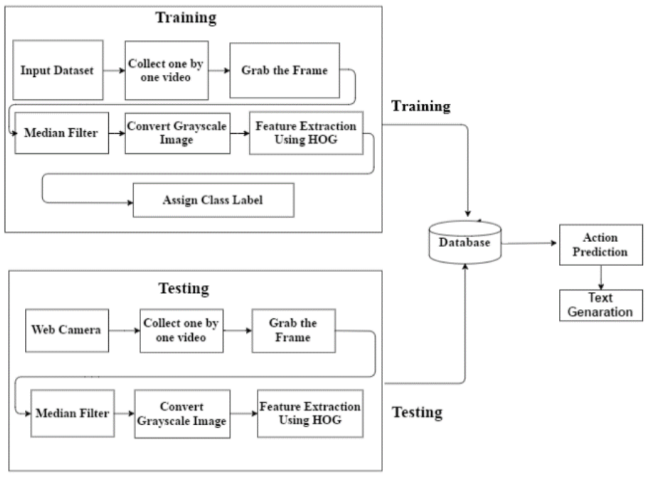
A. Extraction Phase

Images of signs were resized to 80 by 64, as a matter of course imresize utilizes closest neighbor interpolation to decide the estimations of pixels in the output images but other interpolation strategies can be determined. Here bicubic technique is utilized on the grounds that the predetermined

output size is smaller when compared to that than the of the size the input image, before reducing the aliasing using interpolation a low pass channel is applied by imresize. By which a channel size of 11 by 11 is achieved. To lighten the issue of various lighting states of signs taken and the HSV (Hue, Saturation, Brightness) non-linearity by taking out the HSV data while holding the luminance. The RGB shading space is transformed to a gray scale image and then to a binary image. A binary picture is one that comprises of pixels that can have one of precisely two hues, typically high contrast. The two qualities black and white are numerically represented as 0 and 1 respectively. Binary pictures are usually made by thresholding a dark scale or color picture from the background. This change achieved a sharp and clear details for the image. It is seen that the change of RGB shading space to HSV shading space then to a binary picture delivered pictures that needs numerous highlights of the sign. So edge detection is utilized to recognize the parameters of a bend that best fits a given arrangement of edge points. Objective of edge location is to make a line drawing of a scene from an image of that scene. Here canny edge detection strategy is utilized in light of the fact that it gives the ideal edge identification arrangement. Canny edge indicator prompts a far superior edge location contrasted with Sobel edge locator. The result of the edge locator characterizes what highlights are in the picture. Canny edge detection strategy is ideal, yet at times it gives additional details that are not required. To take care of this issue an edge of 0.25 is set after testing different threshold values and observing results on the general recognition system.

B. *Classification Phase*

The classification of neural system has 256 instances as its input vector, and 214 output neurons in the output layer arrangement stage incorporates network architecture making the network and preparing the system. Network of feed forward back propagation with supervised learning is utilized.



V. RESULT ANALYSIS

The presentation of the recognition framework is assessed by testing its capacity to classify signs for both preparing and testing set of information. The effect on the neural framework because of the quantity of input given is considered.

A. *User Input*

The client utilizes hand moments to make signals, which will act as the input for the framework of sign language recognition. During the training period of the system, the client makes a database of his/her hand sign gesture pictures. The training stage is finished when the framework has caught enough motions for which it is aware of the class, and the framework is then prepared to perceive gestures. The motions are caught by the webcam with a consistent division between the hand of the client and the camera. A dark shading board is placed behind to make the foundation steady and make the hand territory distinguishing procedure simpler.

B. *Data Set*

The dataset utilized for preparing and evaluating the recognition framework comprises of dark scale pictures for all the signs utilized in the trials. Additionally 8 samples of each sign are collected from 8 unique participants. Among all the 8 samples 5 signs will be utilized for preparing reason while the remaining signs were utilized for testing. The samples are collected by various separations using web camera along with various directions. In this manner a knowledge set is extracted with cases consisting of non-identical orientation and size and therefore can evaluate the proficiency of the feature extraction technique.

VI. CONCLUSION

A. *Conclusion*

The system processes the signs in real time to produce grammatically correct words. When several samples are considered for feature extraction, the class tag is allocated to extract features and finally produce text. The proposed approach results in higher recovery accuracy compared to the conventional processing system. This system provides results in a lower descriptive characteristic with less processing frames, it therefore obtains the objective of higher precision and lower processing overhead. The system can be further upgraded by reducing processing time and the high recognition rate, by applying a different technique.

People with hearing and vocal impairment depend on sign language interpreters to communicate. Due to the high costs and the difficulty in finding qualified interpreters, it is hard to be dependent on them. This system will help people with disabilities to improve their quality of life. The system is strong towards any changes in the gesture. Using the histogram technique we acquire poorly categorized results. Therefore the histogram technique is applicable simplest to small set of alphabets or gestures that are completely specific from each other. The main problem in this approach is how precisely differentiation is achieved this specifically depends on the image, but it additionally comes down to the algorithm. It can be upgraded using other image processing technique such as edge detection. In this project, Well-known edge detectors such as the Canny, Sobel and Prewitt operators are used to detect the sides with different thresholds. Good results are obtained with Canny with a threshold value of 0.25. A

good recognition rate is obtained using edge detection with the segmentation method. The system is made independent of background. As this is a sign to text translator likewise, in future a reverse text to sign translator can be developed.

B. Future Work

It can be unified with various search engines and text messaging applications such as Google, messenger, Skype, Google duo, etc. So that people with vocal impairment can use them to its maximum productivity.

This project is presently working on images, future development may include recognition of motion in a video sequence and allotting a meaningful sentence to it.

REFERENCES

[1] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018.

[2] K. Suri and R. Gupta, "Convolutional Neural Network Array for Sign Language Recognition Using Wearable Signal Processing and Integrated Networks *(SPIN* IMUs," 2019 6th International Conference on), Noida, India, 2019.

[3] C. M. Jin, Z. Omar and M. H. Jaward, "A mobile application of American sign language translation via image processing algorithms," 2016 IEEE Region 10 Symposium (TENSYMP), Bali, 2016.

[4] K. Dixit and A. S. Jalal, "Automatic Indian Sign Language recognition system," 2013 3rd IEEE International Advance Computing Conference (IACC), Ghaziabad, 2013

[5] Muhammad Rizwan Abid, Emil M. Petriu, Fellow, IEEE, and Ehsan Amjadian, "Dynamic Sign Language Recognition for Smart Home Interactive Application using Stochastic Linear Formal Grammer ", IEEE Transactions On Instrumentation And Measurement, Vol. 64, No. 3, March 2015.

[6] Houssem Lahiani, Mohamed Elleuch and Monji Kherallah, "Real Time Hand Gesture Recognition System for Android Devices", 2015, 15th International Conference on Intelligent Systems DeSign and Applications (ISDA).

[7] K. K. Dutta, Satheesh Kumar Raju K, Anil Kumar G S and Sunny Arokia Swamy B, "Double handed Indian Sign Language to speech and text," 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat.

[8] Rishabh Agrawal and Nikita Gupta, "Real Time Hand Gesture Recognition for Human Computer Interaction", 2016 IEEE 6th International Conference on Advanced Computing.

[9] K. S. Abhishek, L. C. K. Qubeley and D. Ho, "Glove-based hand gesture recognition sign language translator using capacitive touch sensor," 2016 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC), Hong Kong, 2016

[10] D. Abdulla, S. Abdulla, R. Manaf and A. H. Jarndal, "Design and implementation of a sign-to-speech/text system for deaf and dumb people," 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), Ras Al Khaimah, 2016

[11] Jian Wu, Student Member, IEEE, Lu Sun, and Roozbeh Jafari, Senior Member, IEEE, "A Wearable System for Recognizing American Sign Language in Real Time Using IMU and Surface EMG Sensors", IEEE Journal of Biomedical and Health Informatics, Vol. 20, No. 5, September 2016.

[12] Md. Mohiminul Islam, Sarah Siddiqua, and Jawata Afnan, "Real Time Hand Gesture Recognition using Different Algorithm Based on American Sign Language", ISBN.978-1- 5090-6004-7/17/ ©2017 IEEE.

[13] R. H. Goudar and S. S. Kulloli, "A effective communication solution for the hearing impaired persons: A novel approach using gesture and sentence formation," 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Bangalore, 2017.

[14] T. Oliveira, P. Escudeiro, N. Escudeiro, E. Rocha and F. M. Barbosa, "Automatic Sign Language Translation to Improve Communication," 2019 IEEE Global Engineering Education Conference (EDUCON), Dubai, United Arab Emirates, 2019.

[15] M. E. Morocho Cayamcela and W. Lim, "Fine-tuning a pre-trained Convolutional Neural Network Model to translate American Sign Language in Real-time," 2019 International Conference on Computing, Networking and Communications (ICNC), Honolulu, HI, USA, 2019.

[16] Parameshachari B D et. al "A Study on Smart Home Control System through Speech", International Journal of Computer Applications 69(19):30-39, May 2013. Published by Foundation of Computer Science, New York, USA