American Sign Language Recognition System Using CNN

[1]Jignesh Tailor, [2]Mithil Shah, [3]Siddhika Tibarewala, [4]Nilesh Rathod

[1,2,3]Student, [4]Assistant Professor, MCT's Rajiv Gandhi Institute Of Technology, Mumbai,

India, [1]jigzztailor@gmail.com, [2]mithilshah18@gmail.com, [3]siddhikat@gmail.com

[4]nilesh.rathod@mctrgit.ac.in

Abstract :America has a hard of hearing populace of an expected 10 lakhs individuals. The technique for correspondence among the hard of hearing network is communication via gestures. This paper portrays the method to catch the static signs (Which are alphabets and numbers) and afterward make an interpretation of that signs into texts. Different Image Processing Techniques are applied to these captured Images.Upon the completion of the various image processing techniques, the features are consigned by three different techniques.For training dataset convolutional neural network is used. There are 3 Image Processing Techniques that have been used in this paper. At last, the deciphered content yield for that sign in the English Language is shown.

I. INTRODUCTION

Sign Language is the skill of communique most of the deaf and mute community. Sign Language emerges and evolves naturally inside the listening to impaired community. Sign Language communique includes manual and non-manual signals where guide signals incorporate fingers, hands, palms and non-guide signals comprise face, head, eyes and frame. Sign Language is a well-based language with a phonology, morphology, syntax and grammar. Sign Language is a complete natural language that makes use of specific techniques of expression for conversation in ordinary life.The structure of the spoken language uses words sequentially whereas a sign language makes use of numerous body actions in parallel. Sign language interpretation system transfers the communication from human-human to human-laptop interaction. Sign language interpreters are used by manner of deaf and dumb people to talk with the hearing world. The purpose of the sign language interpretation system is to current an efficient and correct mechanism to transcribe text or speech, as a result the "dialog communication" between deaf and hearing person will be smooth. There is no standardized sign language for all deaf people at some stage in the world. However, signal languages are not universal, as with spoken languages, these fluctuate from location to region.

The significance of sign language is emphasised by the use of the developing public approval and budget for global projects. At this age of Technology the demand for a pc based device is relatively stressful for the dumb community. However, researchers had been attacking the hassle for pretty a while now and the outputs are showing a few promise. Interesting technology are being advanced for speech recognition but no actual industrial product for sign interpretation is in fact there in the current market.

The idea is to make computer structures to apprehend human language and broaden user pleasant human pc interfaces(HCI). Making a laptop apprehend speech, facial expressions and human gestures are some steps inside the path of it. Gestures are the non-verbally traded information. A consumer can carry out multitudinous indicators one after another. Since human gestures are seen through imaginative and prescient, it's far a topic of outstanding enthusiasm for pc imaginative and prescient researchers. The undertaking pursuits to determine human gestures by growing an HCI. Coding of those gestures into system language desires a complex programming algorithm. In our task we're that specialize in Image processing for finer output generation.

II. LITERATURE SURVEY

Many researchers have finished studies based on hand gesture popularity the usage of picture processing e.G. American Sign Language(ASL) interpretation system become introduced[1] in which the HSV coloration version is used to stumble on hand form the use of pores and skin color and part detection. In [2] HCI System for interpreting hand gestures and faces from a video camera are provided. In this technique head role and hand gestures are mixed to manipulate equipment. Head position is perceived the use of eye position, mouth and face center. Automatic gesture region segmentation and path standardization of the hand

gesture have an interpretation price 93.6%. In [3], a mixture of side detection set of rules and skin detection set of rules were delivered using MATLAB. For detecting points, the Canny edge detection algorithm is used at which image brightness adjusts sharply.Gesture recognition is accomplished the use of ANN set of rules for instant computation. Static hand gesture recognition includes three algorithms named K curvature, Convexity disorder and Part primarily based hand gesture popularity became supplied in [4]. In this microsoft's Kinect camera is used to capture pseudo-three-D photographs, effects section the input photo and tune the photo in 3D space. It limits with fee of Microsoft's kinect digital camera.

In [5], a static and dynamic hand gesture interpretation changed into evolved in depth statistics using dynamic time warping. A directional search algorithm allowed for complete hand contour, the K curvature algorithm become once employed to hit upon fingertips over that contour. Identification of Bengali Sign Language for 46 hand gestures become identified in [6]. Artificial Neural Network became trained through feature vectors of the fingertip finder algorithm. A dataset of 2300 photos of Bengali Sign Language was collected. The experiment confirmed an accuracy of 88.69%. In [7], Glove named "Velostat" made thru conductive material equipped with microcomputer and bluetooth module is developed for finger-gesture popularity. In [8], a wi-fi sensor glove is designed for American Sign Language finger spelling gesture recognition consisting of five flex sensors on arms and three-D accelerometer on the returned of the hand offers 80% accuracy.

In [9], proposed help vector Machine(SVM) for HSL reputation which gives 90% accuracy however the limitation occurs to separate W and O alphabet. In [10],the function extraction is carried out through manner of the use of Principle Component Analysis(PCA) and Fuzzy C-Means(FCM) for classification. In [11], the LAB view for deciphering the alphabet A-Z is used for feature extraction to calculate centroid of each stop of fingers. In [12]. a method to translate A-Z signal language is advanced.To compute the key point of every posture, SIFT set of rules has been used. Further three-D approach has been used for sign language translation. In [13], three-D shape and motion trajectory of hand from 2D video sequences of American Sign Language terms has been recovered. Myoung-Kyu Sohn et.Al [14] offered 3D hand movement trajectory method with depth camera normalized for translation invariant functions extraction and KNN classification. Hernandez-Rebollar et.Al [15] developed approach and equipment for translating hand gestures although which a signal language can be diagnosed and hand gestures may be translated to speech or written text. Various sensors are used in their system like hand, arm and shoulder that measures dynamic and static gestures. The information obtained thru these sensors is processed by means of

microprocessor to obtain insights and accurate results.A dataset of gestures is maintained which is utilized by microprocessor to create output signal. The output sign is then rehearsed to create synthetic sound.Accelerometers are installed on the hands and thumbs and at the returned hand gyroscopes are installed to come across any hand movement and forearm rotation. This information is transmitted to the microprocessor to scan for any orientation and form data of the arm related to the person In [16]. An ASL decoding glove using a flex sensor is presented. In this paper flex sensors are used to acquire sensory information for the popularity of ASL.Total range of alphabets of sign language is 26 and gloves are finished in such a way that it is able to stumble on all these extraordinary signs.
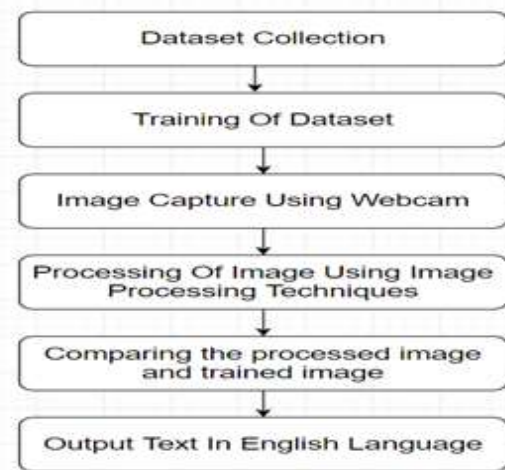
III.   PROPOSED METHODOLOGY

FIG.1 THE OVERVIEW OF PROPOSED SYSTEM

The proposed system comprises 6 stages. They are as per the following: Dataset Collection, Training of Dataset, Image capture using webcam, Processing of image using the image processing techniques, Comparing the processed image and trained image and display output text in english language. The proposed system is made in such a way that the user can write the entire sentence with the help of the system. The image is captured from the live video sequence which will detect the image in particular interval of time. After that image is processed and the output text is displayed on the screen. The processing is happening in the background meanwhile the user can make another gesture in front of the webcam or any image capturing device.

The system is made in such a way that the user can another gesture while the gesture which is previously made by the user is processed in the background. The system can run on any normal configuration machine. The only requirement of the system is that the user requires a image capturing device such as webcam, etc. to capture the image and continue the processing of the system. The dataset training is an important aspect of the system. The dataset training initially takes time to train such a huge dataset.

The image capturing device can affect the performance of the system as the image quality can affect the accuracy of the detection of gesture.

IV. IMPLEMENTED RESULTS

The dataset proposed system is implemented in such a way that all the stages of the system works in accurate manner. The live system is created so that the deaf community with the help of this system can write entire sentences.

A. Dataset Collection

The dataset collection is one of the important aspects of this system. The dataset collected of American Sign Language consists of 29 characters which includes 26 alphabets and 3 special characters which can help the deaf community in writing a sentence. There are 3000 images of every character of American Sign Language. All possible combinations of light, camera angles, camera distance(upto 2ft) is covered while collecting the dataset.The dataset has uniquely labelled every image of dataset for eg. The first image of letter A is A1.

B. Dataset Training

The model of Convolutional Neural Network model is created for the training of dataset. The accuracy achieved while training the dataset is 89%. Initially the time taken by the model to train the entire dataset is 2-3 days. After that it only takes 2-3 hours to train the dataset. The number of steps to train the dataset is 15 epochs. Bottlenecks of dataset are created of every image of dataset while training it with the CNN model.

C. Capturing The Image

The system is kept live so that the user can write the entire sentence with the help of gestures of American Sign Language. The system will not cease the operation until the deaf person manually closes the system. When the user starts the webcam there is one box in which the user has to keep his hand to make the gesture.The gesture will be captured by the system in a particular interval of time and process that gesture in the background.There are 2 screens when the user starts the system in which one is for the webcam or any other input device and the other screen is command prompt in which the user can see the output character of American Sign Language.The size of both the screens can be adjusted by the user in terms of width and height.

D. Image Processing Techniques

Diverse Image processing techniques are applied to the image captured using the webcam. The image processing techniques makes the system more accurate and effective.

1. Image Enhancement

The picture captured using the webcam is an RGB image.

The RGB snap shots are delicate to different mild conditions. So that photograph needs to transform into YCbCr picture where Y is the luma element which signifies luminance statistics of picture, Cb and Cr are the chromo components which offers shading statistics of picture red distinction and blue difference.

2. Color Filtering and Skin Segmentation

In real time, The photo captured through a webcam consists of a massive amount of frames. There is a want to transform RGB images into HSV photographs.Basically, the colour spaces differentiate into three components: Hue(H), Saturation(S) and Value(V). Segmentation is typically executed to find the hand object and limitations of pics, for this HSV aspect allows users to specify boundary of skin coloration in phrases of hue, saturation and fee.V cost offers brightness information so consequently it is straightforward to categorise skin color and non-skin shade records in pics[17]. In this approach adjusting the value of HSV within variety of zero to 255 to extract and get precise boundary of object.

3. Noise Removal: Gaussian Filtering Method

A Gaussian filter out is a linear filter. It's commonly used to blur the picture or to lessen noise and it may lessen the evaluation of image. The diploma of Smoothing is determined by the standard deviation of the Gaussian distribution.The distribution is postulated to have an average of 0.
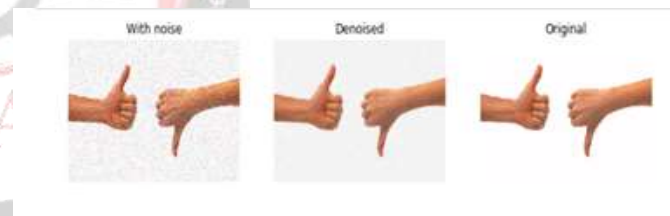

FIG.2 NOISE REMOVAL METHOD

4. Image Segmentation: Thresholding Method

It is the easiest technique of image segmentation. From a grayscale photo, thresholding technique may be habituated to engender binary photographs. In thresholding each pixel in photo supersede into black pixel, if photograph depth is extra preponderant than constant value. An essential belongings which pixels in photo can-apportion is its intensity. Hence in thresholding pics separates into areas relying on mild and dark areas.

5. Image Analysis: Blob Detection

In the field of computer based totally vision, blob detection refers to detection of regions in the captured picture which can be both brighter or darker than the encompassing regions. Fundamentally blob is described as a group of pixels prepared right into a structure. It is a detection of points in the captured photos which differs in functions like brightness and color. In this project, It is used to locate the

brighter regions inside the processed picture. The picture is processed with the aid of different picture processing techniques after that photo is analysed for the brighter regions.

6. Contour detection

In contour detection convexity hull set of rules makes use for drawing contour across the palm and finger elements detection. In convexity hull algorithm adaptive boosting algorithm use for hand detection. Initial step in the convexity hull algorithm is to segment pix in which hand is located. For this some features want to be assumed. The structure of the hand is assumed however that also can change according to the motion of the hand. Therefore skin coloration of hand is considered, because it is invariant to scale and movement of hand. The next step of a tracking gadget includes setting apart hand pixels from non-hand pixels.
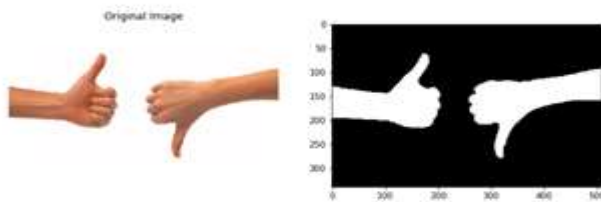

FIG.3 A) ORIGINAL IMAGE   B)CONTOUR EXTRACTION

In convexity hull algorithm inside the beginning computes maximum and minimum x & y coordinate factors and thru becoming a member of these factors form bounding rectangle in which incorporates hull. Like hulls there are different factors moreover gift i.E. Convex defects of hand, which can be current in among valley of two fingers[18].As proven in Fig. 4 then through taking not unusual of all such defects factors actually get a center of palm. So the radius of the palm is taken into consideration as a intensity of palm only. Then the ratio of palm radius and distance of hull factor from the middle aspect of palm should be extra or less to decide finger starting and ultimate position.
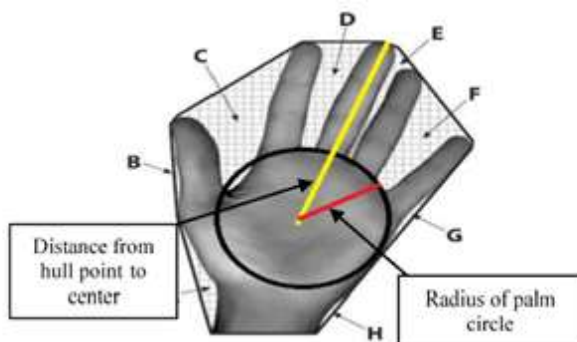

FIG.4 CONVEX HULL AND DEFECTS[19]

The factors which are a part of the boundary of hand are called convex hulls and letters A, B, C, D, E, F, G, H denotes gaps between fingers are known as convex defects. Hence for recognition of fingers position, distance (0) of convex hulls from the middle point have to be larger than

radius (R) of inner defects circle.Depending on value of D and R finger opening and closing can be determined. Therefore the convexity hull algorithm is a very handy and appropriate method for finger point detection and number recognition.

V.   PERFORMANCE ANALYSIS

To construct and achieve the results of the system we have collected a dataset of 29 characters of American Sign Language and then train that dataset with the help of Convolutional Neural Network Model.After training the dataset for the first time if we train the dataset for the second time it takes less time as compared to the first one. Initially dataset training takes time.

FIG.5 DATASET TRAINING

Fig. 5 shows the dataset training using CNN model. Initially the model will look for the images in the dataset directory. If images are not found in the dataset directory then the dataset will not be trained. The dataset directory contains folders for each character of American Sign Language. The model will look for the images in each folder.

VI.   CONCLUSION

The American Sign Language is the only communication method between auditorily impaired human beings of America with the mundane humans of america. The Sign Language plays a consequential role in auditorily impaired people lifestyles and they have to use sign language for their communication purposes. In this paper, we have constructed a live system, which can apprehend 29 characters of American Sign Language which consists of 26 alphabets and 3 special characters. The amalgamation of image processing techniques makes the system more structured. Using this system the deaf community can inscribe the entire sentence with the help of american sign language hand gestures.

REFERENCES

[1] Sharmila Konwar, Sagarika Borah and Dr. T. Tuithung, "An American sign language detection system using HSV color model and edge detection", International Conference on Communication and Signal Processing, IEEE, April 3-5, 2014, India.

[2] Yo-Jen Tu, Chung-Chieh Kao, Huei-Yung Lin,"Human Computer Interaction Using Face and Gesture Recognition", Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific, IEEE, Kaohsiung.

[3] Shweta. K. Yewale and Pankaj. K. bharne, "Hand gesture recognition system based on artificial neural network", Emerging Trends in Networks and Computer Communications (ETNCC), IEEE, 22-24 April, 2011.

[4] Marek Vanco, Ivan Minarik and Gregor Rozinaj, "Evaluation of static hand gesture recognition", International Conference on Signal and Image Processing (IWSSIP), IEEE, 12-15 May, 2014.

[5] 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)

[6] Angur M. Jarman, Samiul Arshad, Nashid Alam and Mohammed J.Islam, "An automated Bengali sign language recognition based on finger tip finder Algorithm", International journal of Electronics & Informatics, 2015.

[7] Eunseok Jeong, Jaehong Lee, and DaeEun Kim, "Fingergesture Recognition Glove using Velostat", 11th International Conference onControl, Automation and System, Oct. 26-29, 2011.

[8] Netchanok Tanyawiwat, and Surapa Thiemjarus, "Design of an Assistive Communication Glove using Combined Sensory Channels", Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference, 9-12 May 2012.

[9] M.S. Sinith, Soorej G Kamal, Nisha B.,Nayana S, Kiran Surendran, and Jith P S, "Sign Language Recognition Using Support Vector Machine", 2012 International Conference on Advances in Computing and Communications, 9-11 Aug.,2012.

[10] M. Ashraful Amin, and Hong Yan, "Sign Language Finger Alphabet Recognition From Gabor-PCA Representation of Hand Gestures", Proceeding of the 6th International Conference on Machine Learning and Cybernetics, 19-22 Aug. 2007.

[11] Andreas Domingo, Rini Akmeliawati, and Kuang Ye Chow, "Pattern Matching for Automatic Sign Language Translation System using LabVIEW", International Conference on Interlligent and Advanced Systems 2007, 25-28 Nov. 2007.

[12] Rudy Hartanto, Adhi Susanto, and P . Insap Santosa, "Preliminary Design of Static Indonesian Sign Language Recognition System", Information Technology and Electrical Engineering (ICITEE), 2013 International Conference, 7-8 Oct. 2013.

[13] Liya Ding and Aleix M. Martinez, "Three-Dimensional Shape and Motion Reconstruction for the Analysis of American Sign Language", Computer Vision and Pattern Recognition Workshop, 17-22 June 2006.

[14] Myoung-Kyu Sohn, Sang-Heon Lee, Dong-Ju Kim, Byungmin Kim, Hyunduk Kim, "3D Hand Gesture Recognition from One Example", 2013 IEEE International Conference on Consumer Electronis (ICCE), 11-14 Jan 2013.

[15] L. Hemandez-Rebollar, "Method and apparatus for translating hand gestures," US Patent 7.565.295, lui 2009.

[16] M. G. Kumar, M. K. Gurjar, and M. S. Singh, "American sign language translating glove using flex sensor," Imperial Journal of Interdisciplinary Research, vol. 2, no. 6, 2016.

[17] Aliaa A. A.Youssif, AmalElsayedAboutabl, HebaHamdy Ali,"Arabic Sign Language (ArSL) Recognition System Using HMM ", (lJACSA) International Journal of Advanced Computer Science and Applications, Vo1.2, Issue. 11,2011

[18] B. Toni, J. Darko, and Ipro, "A robust hand detection and tracking algorithm with application to natural user interface," 2012 pp. 1768-1774.

[19] S. Ganapathyraju and D. Ph, "Defects to Control an Industrial Robot," 2013 pp. 63-67.