Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 7, Issue 07, July -2020

Real Time Indian Sign Language Recognition System using CNN Classifier

Sign to speech translator for Deaf and Dumb People

[1]Ms.Savita More, [2]Prof.Ashwini Kokate

[1]*Electronics &Telecommunication Engineering, Siddhant College of Engineering,Pune*
[2]*Electronics & Telecommunication Engineering, Siddhant College of Engineering,Pune*

ABSTRACT: *Several works has been conducted in the field of human computer interaction (HCI). The HCI solution is given by one of the system sign language recognition. The system we design called the system for the recognition of sign language. This gives the solution to build the HCI where the computer is used as interpreters. Using Python, these systems are used to recognize the convention of static & dynamic signs in real time. Signs are collected using a web camera and CNN technique is used to retrieve information. Due to the use of simple logic condition applied to recognize the sign, this proposed system can be used for real time application.*

Keywords- *database; Covariance; CNN; Web Camera*

I. INTRODUCTION

Communication in day-to-day life is the main part of exchanging information. Normal people can clearly communicate with each other, but Sign language Understanding or translator is important for the contact between people who sign and regular people. Some 60 million people in India are mute & deaf. During communication with these people most people use signs. The knowledge between them is very difficult to grasp & share. So there is a need for an translator to communicate with them through this distance. This translator can be used to translate the gesture into text or voice A sign language is a language that uses gestures or movement to communicate instead of words, but there is a difficulty when they communicate as this sign language is useful to all cultures for communicating purposes in the absence of a real translator. This will include incentives for employment in the industry for this population, employment in the IT sector, and jobs for the government. This Indian sign language is more appropriate than a traditional system of data flow. The typical system can consist mainly of four modules acquiring, segmenting and tracking management, extraction and description features, classification and recognition. Sign language is usually understandable for the signer and the person who knows sign language, but it is so difficult for those who do not know sign language or any gesture's meaning[4].

A system for the recognition of sign language is developed primarily into two steps first is the acquisition of the database and second is the classification. We used Indian sign language images as database in our project and the system displays the English alphabet that the mute, deaf people want to say. To obtain this sign a standard 5MP camera is used. We need to sign database with 26 English alphabets sign & 9 Numeric sign with proper pictures to develop this project. Each number or alphabet is assigned to a given image. These images are in a .jpg format. In feature extraction information which consists of large quantities of data can be collected from the image automatically and it is also the method which is useful for capturing feature and classification of images. The original image is transformed into a gray scale in the extraction method of app. These excreted images are necessary as classification inputs. A variety of classification techniques are available and are used to classify the image. Classifier is the identification of training data with input data set. We use CNN classifier to classify images in our work. CNN is a vector supporting computer used for the supervised learning model, with a related algorithm that analyzes data used for classification and processing. CNN classifier is the process by which classification tasks are performed. For such systems a total of 30 seconds timer is used. First 5 seconds are used to capture background and next 25 seconds will be used to capture the image in real time. These captured images are then compared to the trained database and on- screen displayof results.

II. LITERATURE REVIEW

There are various theories used by different authors to present Indian Sign Convention. The Sign Convention literature under survey is as follows:

Rekha **J**, [2] proposed the static and dynamic alphabet signing scheme. This system was used to

collect 23 static ISL alphabet signs from 40 different signers as training samples, and use 25 videos as test samples. The images are extracted using the Principle Curvature Based Region Detector method. Multi class CNN, DTW & non-linear KNN are used as classifiers for signs. The results of the experiment were static 94.4% and 86.4% for dynamics.

Geetha, M, and U C Manjusha[4] suggested a vision-based identification of Indian Sign Language characters and numerals using approximations of B-Spine. This data set consists of 50 samples of each alphabet and numbers The boundary extracted from the region of interest from the image signs to a B-Spline curve by taking the Maximum Curvature Points (MCPs) as the Control Points are used as the extraction technique of features. Support Vector Machine classifier is used to classify input sign and the recognition result is approximately 90.00 percent.
Around 5 percent of the world 's community is using sign language as a communication medium in all parts of the world[5]. Any downside is discovered during review of research by different authors. The first drawback is a dataset that is not as standard. Due to non-standard researcher-led data set experiment, it is not clear. The system that we are planning to use in public places presents different types of noises or background in the acquisition of sign. The second drawback we found is that some laboratory developed all the databases. The laboratory therefore needed more processing power, or higher costs.

III. METHODOLOGY

The main objective of the project is "To analyze the recognition of the indigenous sign language by ordinary people." The objectives of the proposed system are the following:

1) Image Capturing

2) Image pre-processing

3) Segmentation

4) Input database

5) Feature extraction

6) Classification

7) Result and discussion

Figure 1 shows the proposed system. If no standard data set is available to automatically check ISL gesture recognition, then two ISL character sign data sets are generated. In the first set the signs are alphanumumic, and in the second set the Indian Sign Language is divided into ten digits. In the data set development portion, the details about acquiring the data set are given.

Pictures are captured on each white background number by using a digital camera. These registered images are stored in 680X480 RGB pixel format as an input file. After starting the webcam, we have adjusted 30 second time to get the picture. These photos are taken for each character by a single mute male. The webcam is used to capture a video as we take the command and gives the freedom to move the hand gesture, a very important component of the machine communication. We use those images in the JPG format, since extracting the image in various hardware and software environments is very simple. This dataset requires almost 40 KB of memory. These pictures are flat in shape, rectangular. The "back color space" property is used to make the database web camera. The color space we want in the image box is defined in this property. If you start your webcam, use a capture frame with one trigger.fig to display the captured image with the numbers 1 , 2 and 5.

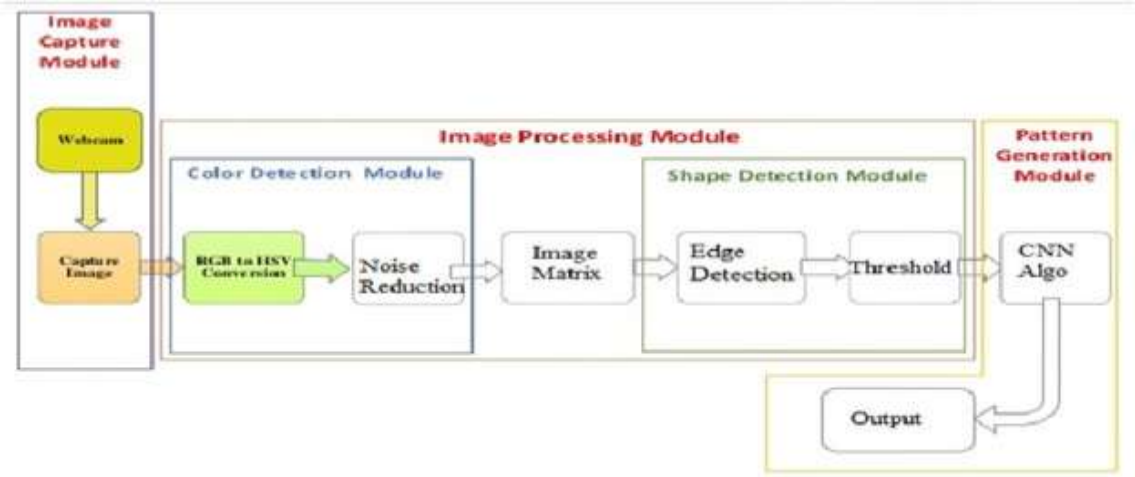

Figure 1 shows the proposed system

Image capturing–

Two types of dataset are creating for this experiment. In this dataset 10 images of single & double handed images are captured using a digital camera for each number with white background. These captured images are saved as input image in the form of 680X480 RGB pixel size. We adjusted 30 second time period for acquiring the image after start the web camera. These images are collected from single mute male person for each character. The webcam is used to capture the video as we give the command and it has given the degree of freedom to get the movement of hand gesture which is very essential part of communication with the machine. We use this images in the JPG format because it is very easy to extract the image in different hardware and software environment. The memory required for this dataset is nearly 40 KB. These images are flat rectangular shape in structure. For making of database web camera is activated with the "returned color space" property. This property is specified the color space as we want in the image box. When webcam will start it uses capture one frame with one trigger. Fig show the captured image for the number of one, two and five.

Fig 2.Real time Captured image of One, Two and Five

Preprocessing–

We need to preprocess this image after we collect the database from the user. We use the pre-processing approach to eliminate low-frequency background noise by using PYTHON. First, we transform RGB images into gray scale (rgb to gray converter) images. It transforms RGB pictures to Gray scale images of high intensity. In this stage, it is possible to remove noise and to segment. The primary aim of preprocessing is to eliminate undesired distortion and to boost input data (sign language images). The technique for image preprocessing uses considerable image redundancy. The adjacent pixel in the real picture has some or similar luminosity value changed. Median filters are used to minimize preprocessing Images of the sound of "salt and pepper." The sharpening and edge enhancement process is carried out using median filter smoothing. It can either run via signal entry or substitute each entrance with the median of neighboring entries, which is the primary use of the median filter. Each pixel output contains the median value around the appropriate pixel in the input image in the area of three by three.

Segmentation:-

With the dithering, the picture taken turns the gray image into a binary image (Blue and White). The image output replaces all the pixels in the input image to a point above 1 (white) and substitutes for all the pixels to 0. The image is resized by the bilinear method by 256X256 row numbers and columns.

These methods can convert the mathematical mapping variable analog to the digital filter conversion. It is the standard way of mapping the analog plane into a Z digital plane in digital filtering.

Fig3.Segmentation result of Captured image of One, Two and Five

Input Database:-

The database of input is part of image training. For each number used in a particular number with a scaling rotation, all numbers consist of 10 pre-processed pictures. These images are stored in a folder called a trained database. For the training process, each picture is saved in white background to obtain more accurate results with an extension of .bmp (eg 1 1.bmp,1 2.bmp, 1 *.bmp etc.) We adjust the time of 30 seconds after starting the webcam video to get the picture as input data. The background is sensed for the first 5 seconds.

Fig4.Database for input image one

Feature Extraction:

Each image consists of large quantities of data, which are called as functional extraction to decrease these data automatically extracted from the images. The key component analysis tool used here is to extract image features. The processed input data shall be transformed into a reduced range of functions. This is known as the technique of function extraction. It is essential to extract certain characteristics because they are unique to every gesture or sign[8]. The Feature extraction stage is necessary.

Classification:

Several forms of technology are available to conduct a classification function, but we use multi-class CNN classification, that is. Machine vector support. Multi CNN is a training set with an appropriate group vector, and classifies a certain test set using the one vs all relationship CNN classifier. First, we consider the unique value of the group train array, and then consider the number of items in the largest array element. Construct the Vectorized Statement model, where 1 is the current class for binaries and 0 is all other classes.

The vector support searches the nearest path called the "vector support." The CNN draws a line for the link until it finds the closest point. The vector supporting machine declares the best line which bisects to connect lines and is perpendicular. By finding the hyperplane to distinguish between two groups we perform classification. The vector support system is just a single observation coordinate. CNN classifies each Sample data path, using a data matrix, using the information generated using the CNN train function in a support vector classifier structure, as CNN struct. Sample is a matrix where every row corresponds to an observation or replica and each column corresponds to a feature or variable like training data used to generate CNN Struct. Sample must therefore have the same number of columns as the data for training.

Text to speech

We aim to convert the image to a text string, save it to a file and hear through audio what is written in the image.

We will import some libraries for this purpose.

Pytesseract (Python-tesseract): This is Google-sponsored python visual pattern recognition (OCR). pyttsx3: This is a text-to - speech library offline multiplatform Python Imaging Library (PIL): This provides the Python interpreter with image recognition capability. Google trans: It is a publicly accessible library using the Google Translate API.

Display result

The hand movements recognized are shown as a numerical image, a character image or audio. The product of sign one, two and five format is shown in Figure 6. There are 10 different numbers for experimentation. The accuracy of this device is more than 90%.

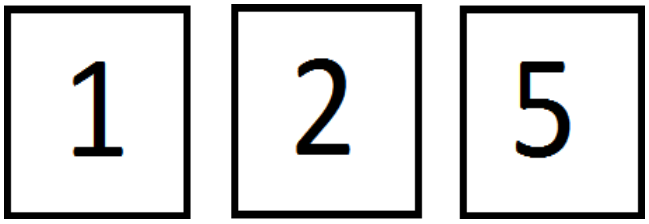

Fig5. Shows the result of input image for the number of one, two and five

IV. CONCLUSION

For performance analysis of the proposed systems, the general performance metrics like False Accept Rate, False Reject Rate are selected. Such systems are below 2 percent at fake acceptance and fake rejection levels. The Real Time (10 or 10 times per class) dynamic recognition approach requires 10 frames. For the database we get 100 images. Expect 5 signs to show the correct result for the test time. The rate of recognition of signs is 95%. Similar test samples give the recognition rate of specific tests. This program is able to use CNN classifier to generate numeric signs with high precision. The coding is done with uniform lighting and simple, white background in real time environments.

REFERENCES

[1] *Balakrishna, G. and P. S. Rajam.. "Recognition of Tamil Sign Language Alphabet using image processing to aid Deaf- Dumb People", International Conference on Communication Technology and System Design. 30: 861-868. 2012.*

[2] *Rekha J., J. Bhattacharya and S. Majumder. 2011."Shape, texture and local movement hand gesture features for Indian Sign Language recognition". 3rd International Conference on Trendz in Information Sciences and Computing (TISC).: 30-35, 2012.*

[3] *Divya Deora and Nikesh Bajaj, "Indian sign language recognition", 1st International Conference on Emerging Technology Trends in Electronics Communication and Networking (ET2ECN), pp.1-5. 2012.*

[4] M. Geetha and U. C. Manjusha, , "A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation", International Journal on Computer Science and Engineering (IJCSE), vol. 4, no. 3, pp. 406-415. 2012.
[5] Rohit Sharma, Yash Nemani, Sumit Kumar, Lalit Kane, Pritee Khanna, Member, IAENG WCE - 5, 2013, London,U.K. "Recognition of Single Handed Sign Language Gestures using Contour Tracing Descriptor". 2013, July 3
[6] Archana S. Ghotkar and Dr. Gajanan K. Kharate., "Study Of Vision Based Hand Gesture Recognition Using Indian Sign Language". International Journal On Smart Sensing And Intelligent Systems VOL. 7, NO. 1, MARCH 2014.
[7] Joyeeta Singha,Karen Das., "Recognition of Indian Sign Language in Live Video". International Journal of Computer Applications (0975 – 8887) Volume 70– No.19) 2013.
[8] Madhuri Sharma, Ranjna Pal and Ashok Kumar Sahoo , "Indian Sign Language Recognition Using Neural Networks And Knn Classifiers". ARPN Journal of Engineering VOL. 9, NO. 8, AUGUST 2014,
[9] Ashok K Sahoo, Gouri Sankar Mishra and Kiran Kumar Ravulakollu, , "Sign Language Recognition": State of the Art, ARPN Journal of Engineering and Applied Sciences, vol. 9, no. 2, pp. 116-134. 2014.
[10] Ashok Kumar Sahoo, Kiran Kumar Ravulakollu, , "Vision Based Indian Sign Language Character Recognition" JATIT & LLS. Vol. 67 No.3. 30 th September 2014.
[11] Zeshan U., M. M. Vasishta and M. Sethna. "Implementation of Indian Sign Language inEducational Settings". Asia Pacific Disability Rehabiliation Journal. (1): 16-40.
[12] Sakshi Goyal, Ishita Sharma and Shanu Sharma, , "Sign Language Recognition System For Deaf And Dumb People", International Journal of Engineering,vol.2, no. 4, pp. 382-387. 2013.
[13] [Sahoo, A. K., G. S. Mishra and K. K. Ravulakollu. "Sign Language Recognition: State Of The Art". ARPN Journal of Engineering and Applied Sciences.9(2): 116-134.2014.

