IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

SIGN LANGAUGE TRANSLATOR

Sana.A.Ansari
Computer engineering,Shree LR Tiwari college of Engineering,Thane
Mumbai,India

Prof.Aabha Patil
Assistant Professor,Computer engineering,Shree L R Tiwari college of Engineering,Thane Mumbai,India

***Abstract*— Communication plays a vital part for mortal beings. Human beings interact with each other either using a natural language channel similar as words, jotting, or by body language (gestures)e.g. hand gestures, head gestures, facial expression, lip stir and so on. The sign language is the introductory communication system within hail disable people. The deaf community suffer a lot to interact with the community. Subscribe language is the way through which the people communicate with each other. In order to give commerce with normal people there's a system which can convert the sign languages to the mortal accessible language. The purpose of this work is to give a real- time system which can convert Subscribe Language to the textbook form. For this particular perpetration we're moving forward with American Subscribe language. Since Subscribe Language Restatement is grueling, but largely important exploration field for several computer vision systems that essay to grease the communication among the deaf and hail disabled people.**

Keywords—Convolution Neural Network, Tenserflow, Open CV, Keras

I. INTRODUCTION

American sign language is a predominant sign language Since the only disability D&M people have is communication related and they cannot use spoken languages hence the only way for them to communicate is through sign language. Communication is the process of exchange of thoughts and messages in various ways such as speech, signals, behavior and visuals. Deaf and dumb(D&M) people make use of their hands to express different gestures to express their ideas with other people. Gestures are the nonverbally exchanged messages and these gestures are understood with vision. This nonverbal communication of deaf and dumb people is called sign language..

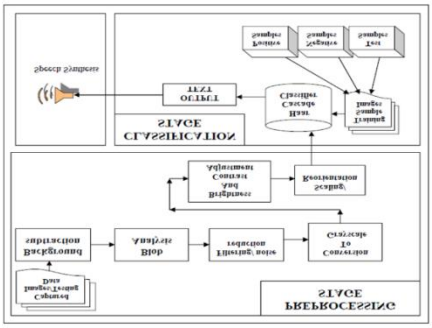
II. EASE OF USE

A. Motivation: For interaction between normal people and Deaf and Dumb people a language barrier is created as sign language structure which is different from normal text. So they depend on vision based communication for interaction.If there is a common interface that converts the sign language to text the gestures can be easily understood by the other people. So research has been made for a vision based interface system where Deaf and Dumb people can enjoy communication without really knowing each other's language.The aim is to develop a user friendly human computer interfaces (HCI) where the computer understands the human sign language. There are various sign languages all over the world, namely American Sign Language (ASL), French Sign Language, British Sign Language (BSL), Indian Sign language, Japanese Sign Language and work has been done on other languages all around the world.

B. Objectives: Sign language is used by deaf and hard hearing people to exchange information between their own community and with other people. The data acquisition, data preprocessing and transformation, classification are going to implement. The focus of this work is to create a vision-based application which offers sign language translation to text thus aiding communication between signers and non-signers. We use a CNN (Convolutional Neural Network) for recognizing spatial features. The dataset used is the American Sign Language Dataset.

III. EXISTING SYSTEM

The existing which was present in "Machine Learning Model for Sign Language Interpretation using Webcam Images [4]" paper written by Kanchan Dabre in 2014. In this paper, they had proposed In this paper, we have proposed a marker-free, visual Indian. The Sign Language Interpretation System works in two stages as shown in Figure. The first is the preprocessing phase i.e. image processing phase, where the hand shape and other distinguishable features are extracted from the image using background subtraction, blob analysis, filtering and noise removal, grayscale conversion, brightness and contrast normalization, scaling and several other image processing techniques. The second stage involves the classification of an image into given many different possible gestures using Haar Cascade Classifier, where this classifier is trained on a given training set that contains samples of the different gestures. This training sample images are taken from several different angle and captured in different lightening conditions. Training dataset consist of positive, negative as well as test sample database. Positive samples are those image samples which contain perfect hand gesture where as in negative sample images the required gesture is absent or only background details are available, no hand moment is present. These datasets are mostly used in training part of classification phase. The test sample dataset can be used in testing part of classification phase. After the training of setup is done, the system is now ready to interpret input images from the videos. A database of haar cascade classifier which denotes different signs is then observed. The classifier which produces the highest probability is then chosen as the most possible interpretation of the sign. Classification or ANN phase which follows the text to speech conversion this phase is known as speech synthesis phase. There was delayed response and I was not identified facial expression and whole body gesture. In future, it Will work on large vocabulary and also will work on different platforms like mobile

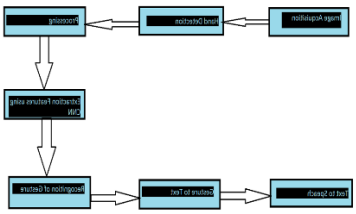


A. Sign Language recognition system using image processing, computer vision and neural network methodologies, to identify the characteristics of the hand in images taken from a video trough web camera. This approach will convert video of daily frequently used full sentences gesture into a text and then convert it into audio. Identification of hand shape from continuous frames will be done by using series of image processing operations. Interpretation of signs and corresponding meaning will be identified by using Haar Cascade Classifier. Finally displayed text will be converted into speech using speech synthesizer.

IV. PROBLEM STATEMENT

Identify applicable funding agency here. If none, delete this text box.

Communication provide interaction among the people to exchange the feelings and ideas. Sign language is the way through which the people communicate with each other. In order to provide interaction with normal people there has to be a system which can convert the sign languages to the understandable form without extra preprocessing.



A. Proposed Solution for the above Problem

The purpose of this work is to provide a real-time system which can convert American Sign Language (ASL) to the text. Most of the existing work is based on handcrafted feature. In this we are introducing a deep learning approach which can classify the sign using the convolutional neural network. In the first phase we make a classifier model using the alpha numeral signs using the Keras implementation of convolutional neural network using python. In phase two another real-time system will be implemented which will take output from phase one and converts to speech.

B. Proposed Model

In the world different languages are used among the people to provide communication, while talking about physically impaired people both deaf and dump community also use different sign languages. The different languages are American Sign Language, Chinese sign language, Indian sign language etc. In each case symbols are vary with the involvement of motion, single handed and double handed representations. A real time system among these communities will allows the communication barrier among them. Once it has been converted to using the Computer Vison approach then it can be converted to any language. There are many researches undergone by in this are to build an efficient and accurate system. The previous works done by the researchers using the hand crafted feature but having limitation and used special conditions. The most works are based on the pattern recognition, feature extraction based on HOG, SIFT, LBP, etc. But the system using a single feature is not sufficient in most of the cases and the Hybrid approach are introduced to solve this problem. But for a real time system we need faster methods to solve our problems. Now a days our computers are improved with the speed of processing using parallel implementation. In most of the time our system utilizes a single core for solving a problem. Using the GPU system, the problems can be solved by parallel computing and the number of cores is higher than the CPU system. Using the Deep Learning approach, we can model a self-learning system for our needs. Convolutional neural network is one of the trending deep learning system which is capable of solving any kind of computer vision problem. In our method we used a Region of interest – Convolutional neural network for the real time implementation.

Convolution Neural Network :Unlike regular Neural Networks, in the layers of CNN, the neurons are arranged in 3 dimensions: width, height, depth. The neurons in a layer will only be connected to a small region of the layer (window size) before it, instead of all of the neurons in a fully-connected manner. Moreover, the final output layer would have dimensions (number of classes), because by the end of the CNN architecture we will reduce the full image into a single

vector of class scores.

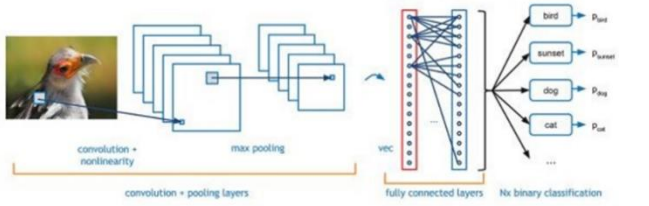


Figure 5.2: Convolution neural networks

1. Convolution Layer : In convolution layer we take a small window size that extends to the depth of the input matrix. The layer consist of learnable filters of window size. During every iteration we slid the window by stride size [typically 1], and compute the dot product of filter entries and input values at a given position. As we continue this process well create a 2-Dimensional activation matrix that gives the response of that matrix at every spatial position. That is, the network will learn filters that activate when they see some type of visual feature such as an edge of some orientation or a blotch of some color

2. Pooling Layer : We use pooling layer to decrease the size of activation matrix and ultimately reduce the learnable parameters. There are two type of pooling :

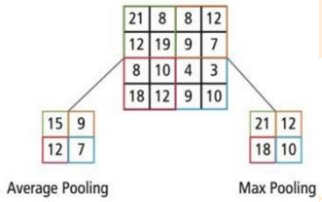


Figure 5.3: Types of pooling

3) Max Pooling : In max pooling we take a window size and only take the maximum of 4 values. Well lid this window and continue this process, so well finally get a activation matrix half of its original Size.

4) Average Pooling : In average pooling we take average of all values in a window.

5) Fully Connected Layer : In convolution layer neurons are connected only to a local region, while in a fully connected region, well connect the all the inputs to neurons.

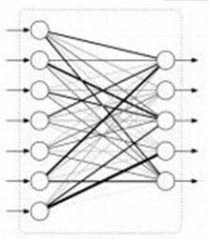


Figure 5.4: Fully Connected Layer

6) Final Output Layer : After getting values from fully connected layer, well connect them to final layer of neurons that will predict the probability of each image to be in different classes.

V. Literature Review

1) The paper "Sign Language Recognition in Image based Hand Gesture Recognition Techniques [1]" was introduced software which presented a system prototype that was able to recognize sign language to help deaf and dumb people. Pattern recognition and Gesture recognition are the developing fields of research. Basically there were two approaches for sign recognition and sign detection. Sign detection was an extracting feature of certain object with respect to certain parameters. Sign recognition was recognizing a certain shape that differentiates the object from the remaining shapes. The technical point of view characteristic features of sign language communication was: its social direction and meaning; technical and technological convenience and easy to use. The system had used a webcam for the capturing images and pre-processing of the signs was done by using Microsoft Visual Studio as an IDE and OpenCv library[20]. Initially it captured the live stream images with the help of webcam. After that applying image preprocessing steps to remove unwanted noise and adjust the brightness of images. Then perform image analysis and apply convexity algorithm to extract contour of hand position. Convexity hull algorithm[8] implemented just for finger point detection and number recognition. A similar study can be done on English alphabets as well as Marathi alphabets by using contour analysis process. Because it contained methods for preliminary handling of the images, contours extraction, their filtrations and a Recognition. Also, it contains tools for automatic generation of templates for recognition of printing symbols. Recognition was done by matching templates of hand by considering Contour curve[11] shapes. In contour analysis by considering vectors value it is ease to recognize hand.

2) In the study conducted by Murat Taskiran "A Real-Time System For Recognition Of American Sign Language By Using Deep Learning [5]" developed in 2018 The sign language used in this paper was American sign language. They had used convolutional neural network [16]was trained by using dataset collected in 2011 by Massey University, Institute of Information and Mathematical Sciences. After network training is completed, the network model and network weights are recorded for the real-time system. In the real-time system, the skin color [17] was determined for a certain frame for hand use, and the hand gesture was determined using the convex hull algorithm[8], and the hand gesture was defined in real-time using the registered neural network model and network weights. In future it will be implemented on different words and different sign languages will be used.

3) In 2018, Kartik Shenoy had implemented, "Real time Indian Sign Language Recognition [3]". It was captured from a smartphone camera and its frames were transmitted to a remote server for processing. The use of any external hardware (such as gloves or the Microsoft Kinect sensor) was avoided. Techniques such as Face detection, Object stabilisation and Skin Colour Segmentation[21] are used for hand detection and tracking. The image is further subjected to a Grid-based Feature Extraction technique which represents the hand's pose in the form of a feature Vector. Hand poses were then classified using the k-Nearest Neighbours algorithm[19]. On the other hand, for gesture classification, the motion and intermediate hand poses observation sequences were fed to Hidden Markov Model[19] chains corresponding to the 12 pre-selected gestures defined in ISL. ISL had used gestures for representing complex words and sentences. It had contained 33 hand poses.The solution described can be easily extended to two-handed gestures. In the next section of this paper, the related work pertaining to sign language translation. The pre-processing first was involved face removal, stabilisation and skin colour segmentation[21] to remove background details and later morphology operations to reduce noise. The hand of the person is extracted and tracked in each frame.

For recognition of hand poses, features are extracted from the hand and fed into a classifier. The gesture whose HMM chain had given the highest score with forward-backward algorithm was determined to be the recognized gesture for this pattern. Currently, only single-handed gestures in ISL were considered for research. With the use of advanced hand extraction algorithms, this approach can be extended to two handed gestures as well. Also, with the use of Natural Language Processing algorithms.

4) The Sign Language Interpretation System worked in two stages as showed in this paper. The first was the preprocessing phase i.e. image processing phase, where the hand shape and other distinguishable features were extracted from the image using background subtraction, blob analysis, filtering and noise removal, grayscale conversion, brightness and contrast normalization, scaling and several other image processing techniques. The second stage involved the classification of an image into given many different possible gestures using Haar Cascade Classifier, where this classifier was trained on a given training set that contains samples of the different gestures. There was delayed response and I was not identified facial expression and whole body gesture. In future, it Will work on large vocabulary and also will work on different platforms like mobile phones.

5) Many Researchers have worked on bangla sign language recognition but no researches have been carried out using HOG as feature vector. Muttaki Hasan had developed in 2016 "A Machine Learning Based Approach for the Detection and Recognition of Bangla Sign Language [5]". In Bangladesh a formal sign language has been established only recently. In the year 2000, Centre for Disability in Development (CDD) took the initiative to standardize

6) communication with sign languages in this country. Before this step, there were different local variants and no national dialect existed in their training centre. People in Bangladesh still are ignorant of this mode of communication and thus the deaf children still cannot lead an uncomplicated life here. In this paper they presented a vision-based method that detects hand gestures in different illuminations. For this purpose, they had employed machine learning approach which includes training the classifier using HOG (Histogram of Oriented Gradients) features. Classification has been done using k-NN[12] (k-Nearest Neighbour). They had used OpenCV[2] as image processing library and python as programming language. An USB webcam has been used for capturing the gesture images. This research approach used 16 signs of expressions. The signs have been chosen from "বাংলা ইশারা ভাষার অভিধান" published by BdSL. Signs have been chosen in such way that it includes single handed gesture. The training dataset was prepared using different background. Each of these signs takes 20 samples for data training in the SVM classifier[24]. For sign of expression they had used 16× 20=320 images for training and applied k-Nearest Neighbour algorithm[20]. They used SVM for decision making. Many previous works had been done for detection and recognition of hand gestures but HOG feature along with audio output enables the speech impaired to communicate more efficiently. They firstly constructed HOG feature based hand detector for bangla sign language detection. However, this method was suitable for static gesture recognition. In future, they will work on two hands gestures, also will work on meaningful images and different sign languages as well.

7) "Machine Learning Model for Sign Language Interpretation using Webcam Images [4]" was introduced by Kanchan Dabre in 2014. In this paper, they had proposed a marker-free, visual Indian Sign Language recognition system using image processing, computer vision and neural network methodologies, to identify the characteristics of the hand in images taken from a video trough web camera. This approach was to be converted video of daily frequently used full sentences gesture into a text and then convert it into audio. Identification of hand shape from continuous frames was done by using series of image processing operations. Interpretation of signs and corresponding meaning was identified by using Haar Cascade Classifier.

The system is a vision based approach. All the signs are represented with bare hands and so it eliminates the problem of using any artificial devices for interaction.

VI. DATA GENERATION

For the project we tried to find already made datasets but we couldn't find dataset in the form of raw images that matched our requirements. All we could find were the datasets in the form of RGB values. Hence we decided to create our own data set. Steps we followed to create our data set are as follows.

We used Open computer vision(OpenCV) library in order to produce our dataset.Firstly we captured around 800 images of each of the symbol in ASL for training purposes.

First we capture each frame shown by the webcam of our machine and then we apply our gaussian blur filter to our image which helps us extracting various features of our image. The image after applying gaussian blur looks like below.

Algorithm Layer 1:

1. Apply gaussian blur filter and threshold to the frame taken with opencv to get the processed image after feature extraction.

2. This processed image is passed to the CNN model for prediction and if a letter is detected for more than 50 frames then the letter is printed and taken into consideration for forming the word.

3. Space between the words are considered using the blank symbol.

Algorithm Layer 2:

1. We detect various sets of symbols which show similar results on getting detected.

2. We then classify between those sets using classifiers made for those sets only.to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Training and Testing :

We convert our input images(RGB) into grayscale and apply gaussian blur to remove unnecessary noise.We apply adaptive threshold to extract our hand from the background.

a) Challenges Faced :There were many challenges faced by us during the project. The very first issue we faced was of dataset. We wanted to deal with raw images and that too square images as CNN in Keras as it was a lot more convenient working with only square images. We couldn't find any existing dataset for that hence we decided to make our own dataset. Second issue was to select a filter which we could apply on our images so that proper features of the images could be obtained and hence then we could provided that image as input for CNN model. We tried various filter including binary threshold, canny edge detection, gaussian blur etc. but finally we settled with gaussian blur filter. More issues were faced relating to the accuracy of the model we trained in earlier phases which we eventually improved by increasing the input image size and also by improving the dataset.

VII. RESULTS

The American Sign Language Translator, the real time system is been developed for alphabets as well as numbers. The model will be created by implementing Deep Learning system using Convolutional Neural Network. After training our model against sign language hand expression dataset we are expecting to get is around 90-95%. In future, we will be going to implement this model on different meaningful words and sentences as well.

VIII. ACKNOWLEDGEMENT

I wish to express my sincere thanks and deep sense of gratitude to respected guide Prof.Aabha Patil, Associate Professor, SLRTCE who has always been optimistic and a constant source of motivation for me throughout. Her support and cooperation has really helped me to sail through tough times.I would also like to thank Prof.Neelam Phadnis, Head of Department, Computer Engineering, SLRTCE, for her support. Her valuable suggestions from time to time have been a great help to me.My gratitude extends to my family and friends who have always stood by me. Without their moral support and encouragement this dissertation would not have been possible.

IX. REFERENCES

[1] https://ieeexplore.ieee.org document/7916786
[2] https://ieeexplore.ieee.org/document/8441304
[3] https://ieeexplore.ieee.org/ document/8493808
[4] https://ieeexplore.ieee.org/ document/6839279
[5] https://ieeexplore.ieee.org/abstract/document/7835387
[6] Jorge Badenas, Josee Miguel Sanchiz, Filiberto Pla, Motion-based Segmentation and Region Tracking in Image Sequences, Patternrecognition 2001, 34, pp. 661-670
[7] P. T., Thinh, H. C., Van Phuc, B., & Kha, H. H. (2018). Automatic feature extraction for Vietnamese sign language recognition using support vector machine. 2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom).
[8] https://en.wikipedia.org/wiki/Convex_hull.
[9] M. Panwar and P. Singh Mehra, "Hand gesture recognition for human computer interaction," IEEE Int. Conf Image Inf Process., 2011 no. Iciip, pp. 1-7
[10] Yang, et al. "A Kinect based gesture recognition algorithm using GMM and HMM." Biomedical Engineering and Informatics (BMEJ),20 13 6th International Conference on.IEEE,2013.
[11] https://debrief.github.io/tutorial/contouring_algorithm.html.
[12] Anuja V. Nair, Bindu V., A Review on Indian Sign Language Recognition, International Journal of Computer Applications (0975 – 8887) July 2013.
[13] [1]T. Yang, Y. Xu, and "A. , Hidden Markov Model for Gesture Recognition", CMU-RI-TR-94 10, Robotics Institute, Carnegie Mellon Univ.,Pittsburgh,PA, May 1994.
[14] Pujan Ziaie, Thomas M¨uller , Mary Ellen Foster , and Alois Knoll"A Na¨ive Bayes Munich,Dept. of Informatics VI, Robotics and Embedded Systems,Boltzmannstr. 3, DE-85748 Garching, Germany.
[15] https://docs.opencv.org/2.4/doc/tutorials/imgproc/gausian_median_blur_b ilateral_filter/gausian_median_blur_bilateral_filter.html
[16] Mohammed Waleed Kalous, Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language.
[17] aeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convol utional-Neural-Networks-Part-2/
[18] http://www-i6.informatik.rwth-aachen.de/~dreuw/database.php
[19] Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925.
[20] Zaki, M.M., Shaheen, S.I.: Sign language recognition using a combination of new vision based features. Pattern Recognition Letters 32(4), 572–577 (2011)
[21] Byeongkeun Kang , Subarna Tripathi , Truong Q. Nguyen "Real-time sign language fingerspelling recognition using convolutional neural networks from depth map" 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR).