Dynamic Tool for American Sign Language Finger Spelling Interpreter

Prateek SG
Department of Information Science and Engg
BVBCET
Hubli, India
prateeksg21041996@gmail.com

Jagadeesh J
Department of Information Science and Engg
BVBCET
Hubli, India
jagadish.j03@gmail.com

Siddarth R
Department of Information Science and Engg
BVBCET
Hubli, India
siddharthrayaraddi@gmail.com

Smitha Y
Department of Information Science and Engg
BVBCET
Hubli, India
ymsmita@gmail.com

P. G. Sunitha Hiremath
Department of Information Science and Engg
BVBCET
Hubli, India
pgshiremath@bvb.edu

Neha Tarannum Pendari
Department of Information Science and Engg
BVBCET
Hubli, India
neha_ip@bvb.edu

***Abstract*— Sign language is a type of language that uses manual communication to convey meaningful messages to other people. This includes simultaneous employing of hand gestures, movement, orientation of the fingers, arms or body, and facial expressions to convey a speaker's ideas. American Sign Language is one of the popular sign language used by most of deaf and dumb people to communicate with each other. American Sign Language is also referred to as ASL. A real-time sign language translator is required for facilitating communication between the deaf community and the general public. We propose a system called Dynamic tool for American Sign Language (ASL) finger spelling interpreter which can consistently classify the letters a-z. The dataset consists of a set of American Sign Language videos. Our approach first converts the videos into frames and then pre-processes the frames to convert them into greyscale images. Then the Convolutional Neural Network (CNN) classifier is used for building the classification model which classifies the frames into 26 different classes representing 26 English alphabets. Finally, the evaluation of the classification model is carried out with test data providing the output in the form of text or voice. The cross-validation accuracy results of 98.66% is achieved from our approach.**

Keywords*— *American Sign Language (ASL); Finger spelling interpreter; Sign classification; CNN classifier;

I. INTRODUCTION

According to statistics, one person in every 1480 of the population is either deaf or dumb. They communicate with each other by making hand gestures, which is called as Sign language. This language is normally called as American Sign Language (ASL). Hand gestures have both static and dynamic elements. A normal person cannot understand these gestures when a person communicates with him using sign language. Therefore, there is a need to have an ASL conversion system, which can convert gestures made using ASL into normal text or voice.

There are about 70 million deaf/dumb people who use sign language as their first language or mother tongue. Only close family members and friends of deaf/dumb people understand ASL. These people rely on family members to translate for them to other people. Due to which everyday communication is cut down from these people. Therefore, sign language translation finds great importance as it focuses on facilitating communication between the deaf community and the general public.

In this paper, an efficient method is proposed that focuses on building a model to translate ASL into English text or voice which can be understood by anyone who do not know ASL and want to communicate with deaf and dumb. The proposed model provides advantages to the deaf community in order to overcome the difficulty they face in life thereby hoping that with better understanding of sign

ISBN: 978-1-5386-4119-4/18/$31.00 ©2018 IEEE

language, deaf or dumb people will find themselves on an equal footing in the society.

II. RELATED WORK

Many of the researchers have used different methods of translating sign language, such as Convolutional Neural Network in [1], where authors proposed a pipeline that takes video of a user signing a word as the input. Individual frames of the video are extracted and letter probabilities for each are generated using Convolution Neural Networks. The limitation in the proposed paper is the lack of variation in datasets as a result of which, the validation accuracies are not directly reproducible upon testing. In future it is expected that the models would be able to generalize with higher efficiency and would produce a robust model for all letters.

Sign language refers to motion of hands and recognizing meaningful expressions of these motions is gesture recognition. The major tools surveyed for this purpose include Hidden Markov Models (HMMs), particle filtering and condensation algorithm and Finite State Machines (FSMs) [2]. There are many ways to recognize these gestures, such as Hand gesture, Face gesture and Body gesture [3]. Speaking about hand gesture, which is used to identify specific human gesture to convey information, Bhushan Bhokse and Jagadish L used Neural Network analysis (NN) [4] - [5].

In [6], Image Processing Module algorithm is used for real time processing of Sign language. The results obtained shows successful extraction of foreground object from background. The system accuracy is obtained by testing each character 50 times.

Some of the researchers employed a method consisting of three main phases of processing viz., Edge detection, Clipping, Boundary Tracing [7]. Edge detection plays an important role in identifying and locating discontinuities in an image. It helps in optimizing network bandwidth and to extract useful features for pattern recognition [8]. There have been many researches carried out in improving the canny edge detection algorithm as in [9]. Canny edge detection algorithm can also be applied in digital image processing [10]. There are numerous edge detection techniques available such as Sobel Operator, Robert's cross Operator, Prewitt's operator and canny edge detection [11]. Four different edge detection techniques have been used for pre-processing of images in [12]. On comparison, the best result was obtained by Canny Edge Detection [13]. In [14], the authors have used Support Vector Machine (SVM) classifier and has obtained accuracy of 96.347%. In [15], the authors have used Multi Scale Mode Filtering (MSMF) algorithm for the recognition of hand gestures and they have achieved a global recognition rate of 95.66%.

As far as existing methods are concerned, many papers have been published making use of various classification algorithms. Among them Convolutional Neural Network classifier algorithm as explained in [16] - [18] is framed as the best algorithm that can be used for the classification of American Sign Language.

III. DATA DESCRIPTION

The dataset consists of videos of people talking in American Sign Language. These videos are converted into frames and then used as a training data for building a Convolutional Neural Network classification model.

The Fig. 1 shows a sample dataset used for building an American Sign Language Interpreter model. We have a Training dataset of size 80 MB which includes 32,400 training images. Dataset consists of 26 different English alphabets and for each alphabet a unique label is assigned. Thus 26 classes are created while building a classification model.

Fig. 1: Training data for building a classifier

IV. PROPOSED SYSTEM MODEL

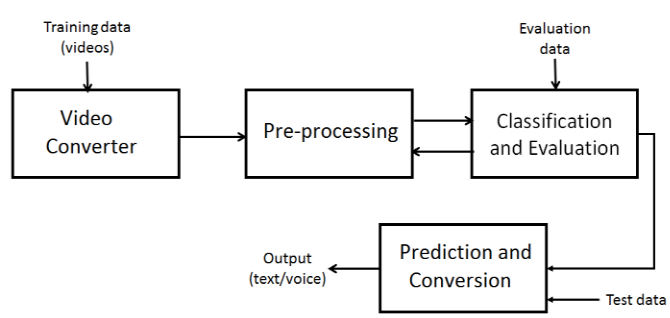

Fig. 2: System model showing the steps involved in the proposed system

The system model describes how the translation of American Sign Language can be done dynamically into text or voice as described in Fig. 2. Firstly, the videos of people communicating in American Sign Language (ASL) is converted into frames using ffmpeg API, where frames extracted are 1 frame per second. Then each frame is stored with a unique frame ID. Further, the extracted frames are pre-processed in order to convert frames into greyscale images by using pre-processing techniques such as Skin Filtering. Skin Filtering technique is performed to the input frames for detection of hand gestures. It is done so that the required hand gesture can be extracted from the background. Skin Filtering is a technique used for separating the skin colored regions from the non-skin colored regions. Skin filtering is performed by using canny edge detection algorithm. Further, for extracting the local visual descriptors from each image, Scale Invariant Feature Transform (SIFT) method is used. The method works in two steps. First step is detection of feature point whereas the second step is finding the feature description. Then the Convolutional Neural Network (CNN) classification model is built using the Scale Invariant Feature Transform (SIFT) features of different American Sign Language alphabets consisting of 26 classes. Evaluation of the classification model is carried out by providing the test data as the input at this step. Finally, the output is provided in the form of text/voice depending on the gesture provided as the input.

V. RESULTS AND DISCUSSION

The Dynamic tool for American Sign Language finger spelling interpreter system is demonstrated on videos of person communicating in American Sign Language. The algorithm for Pre-processing, edge detection and classification are implemented on Intel Core i5-5200U processor @2.20GHz*4 and 8GB RAM with NVIDIA GeForce 840M graphics.

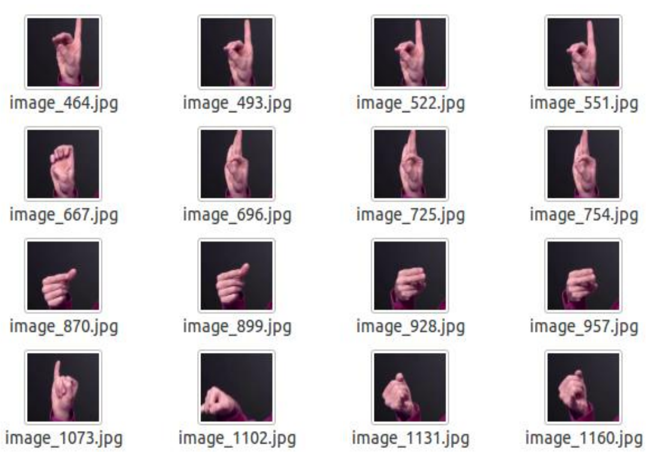

Fig. 3: Conversion of videos into Frames

Firstly, the video of a person communicating in American Sign Language is captured dynamically with different signs. In order to classify those signs into text/voice we need to convert these videos into frames as shown in Fig. 3. The conversion of videos into frames is done at the rate of one frame per second while assigning the unique frame ID.

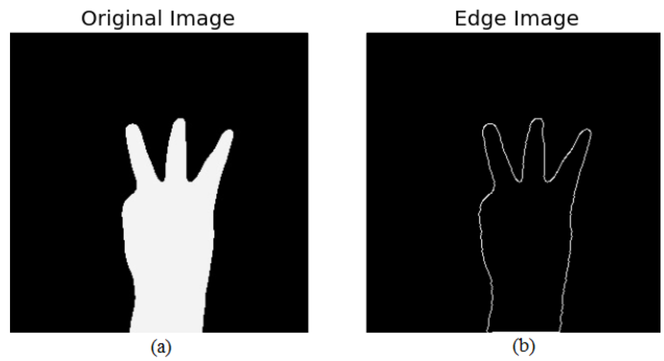

Fig. 4: Canny Edge Detection for pre-processing of frames

Then, pre-processing of frames is done in order to convert frames into grey scale images and to detect the edges of hand gestures. There are different kinds of pre-processing techniques like Canny Edge Detection, Sobel Operator, Robert's Cross Operator, and Prewitt's Operator etc. Each technique will give different accuracy rate. Out of these techniques, Canny Edge Detection will give the best results as shown in Fig. 4.

```
Training...
Output model: ../databasetrainingdata.svm.model
--------------------
## outputting results
model file: ../database../databasetrainingdata.svm.model
codebook file: ../databasecodebook.file
category      ==>  label
a             ==>     0
c             ==>     2
b             ==>     1
e             ==>     4
d             ==>     3
g             ==>     6
f             ==>     5
i             ==>     8
h             ==>     7
k             ==>    10
j             ==>     9
m             ==>    12
l             ==>    11
o             ==>    14
n             ==>    13
q             ==>    16
p             ==>    15
s             ==>    18
r             ==>    17
u             ==>    20
t             ==>    19
w             ==>    22
v             ==>    21
y             ==>    24
x             ==>    23
z             ==>    25
```

Fig. 5: Training of the Classification Model.

Further, the pre-processed frames act as the training data for building the Convolutional Neural Network Classification model. The Training data consists of 32,400 images of 26 different sign gestures. Then the training data is fed as input for classification model. The classifier is built and it assigns unique label for all 26 different classes as shown in Fig. 5.

The Classifier model is then evaluated using n-fold Cross-validation technique. The Test data consists of 5000 sample images. The test data is fed as input for classifier model. The results obtained after prediction of classes is shown in Fig. 6. The cross-validation accuracy of 98.66% is achieved for the CNN classifier model.

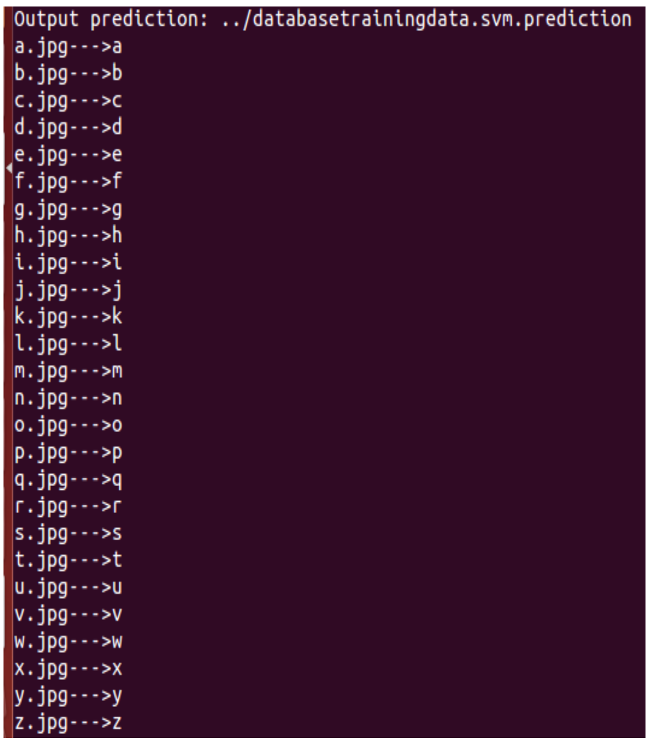

Fig. 6: Evaluation of Classification Model

The Fig. 7 and Fig. 8 are the examples of live dynamic input captured and that input is correctly classified into sentence.

Fig. 7: Prediction of American Sign Language for sentence I WANT TEA

Fig. 8: Prediction of American Sign Language for sentence HELP ME

In Fig. 9, Fig. 10 and Fig. 11, the Convolutional Neural Network Classification model is fed with the dynamic live input of a person communicating in American Sign Language and CNN Classifier detects the sign correctly.

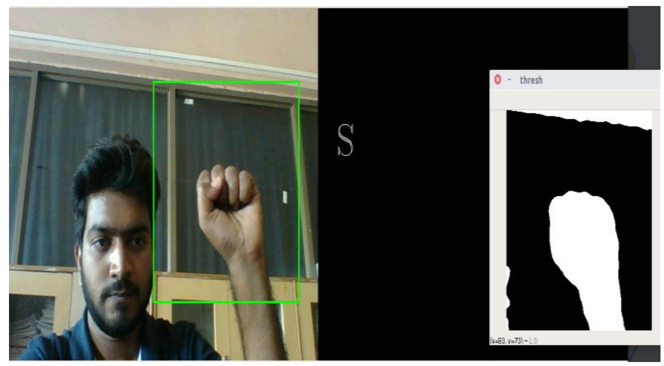

Fig. 9: Prediction of American Sign Language Letter S.

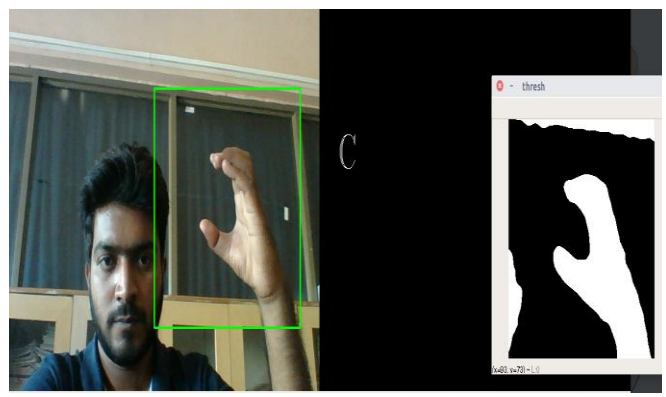

Fig. 10: Prediction of American Sign Language Letter C.

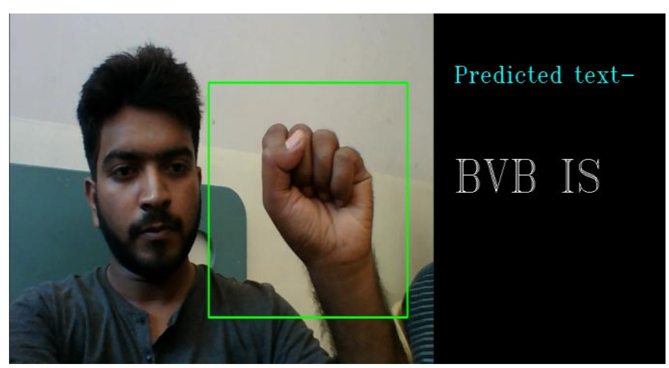

Fig. 11: Prediction of American Sign Language sentence.

VI. CONCLUSION

American Sign Language is the only source of communication for deaf or dumb people. For common people, it is difficult to understand their language of communication. Hence our tool will provide the opportunity for common people to understand their mode of communication. The American Sign Language finger spelling tool will capture the ASL gestures made by deaf or dumb people in real time and classify those gestures into text or voice. For classification of ASL gestures, we have used CNN classifier and achieved a cross validation accuracy of 98.66%. In future, we want to develop an android application which would dynamically translate the American Sign Language into text or voice.

REFERENCES

[1] Sigberto Alarcon Viesca, Brandon Garcia. “Real-time American Sign Language Recognition with Convolutional Neural Networks". Stanford University, Stanford, CA, 2015.

[2] IEEE Sushmita Mitra, Senior Member and IEEE Tinku Acharya, Senior Member. “Gesture Recognition: A Survey". IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 37, NO. 3, May 2007.

[3] Prof. V. S. Ubale Miss. Kawade Sonam P. “Gesture Recognition-A Review". IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), ISSN: 2278-2834,ISBN: 2278-8735:pp: 19–26.

[4] Dr. A.R.Karwankar Bhushan Bhokse. “Hand Gesture recognition Using Neural Network". IJISET International Journal of Innovative Science, Engineering and Technology, Vol. 2 Issue 1, January 2015.

[5] Sadab Jagdish L. Raheja, A. Singhal. “Android based Portable Hand Sign Recognition System". CSIR-CEERI., Vol. 2, ISSN: 2278-6252, 2014.

[6] Ariruna Dasgupta Dibyabiva Seth, Anindita Ghosh and Asoke Nath. “Real Time Sign Language Processing System". Springer Nature Singapore Pte Ltd. 2016 A. Unal et al. (Eds.): SmartCom, pages pp. 11–18, 2016.

[7] Suhas Mahishi Dheeraj R Sudheender S Nitin V Pujari Ravikiran J, Kavi Mahesh. “Finger Detection for Sign Language Recognition". Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol. 1:Pages: 18–20, March 2009.

[8] Debosmit Ray, “Edge Detection in Digital Image Processing”, Thursday, June 06, 2013.

[9] Wojciech Mokrzycki, Marek Samko, “New version of Canny edge detection algorithm”, Faculty of Mathematics and Informatics, University of Warmia and Mazury Olsztyn, Poland, ICCVG, Chapter: I, Publisher: Springer, Editors: Bolc et all, pp.533-540, January 2012.

[10] Indrajeet Kumar, Jyoti Rawat, Dr. H.S. Bhadauria, “A CONVENTIONAL STUDY OF EDGE DETECTION TECHNIQUE IN DIGITAL IMAGE PROCESSING”, International Journal of Computer Science and Mobile Computing, Vol.3 Issue.4, April 2014, pg. 328-334.

[11] Md. Ashikur Rahman. “Computer Vision Based Human Detection". International Journal of Engineering and Information Systems (IJEAIS), Vol. 1 Issue 5:Pages: 62–85, july 2017.

[12] Raman Maini and Dr. Himanshu Aggarwal. “Study and Comparison of Various Image Edge Detection Techniques". International Journal of Image Processing (IJIP), Volume (3): issue (1):Pages: 62–85.

[13] Monica Avlash and Dr. Lakhwinder Kaur. “PERFORMANCES ANALYSIS OF DIFFERENT EDGE DETECTION METHODS ON ROAD IMAGES". International Journal of Advanced Research in Engineering and Applied Sciences, Vol. 2, ISSN: 2278-6252:No. 6, June 2013.

[14] Shama, Mr. Karun Verma, “Sign Language Recognition using SVM”, Computer Science and Engineering Department, Thapar University, Patiala-147004, July 2014.

[15] Georgiana Simion, Ciprian David, Vasile Gui, Catalin-Daniel Caleanu, “Fingertip-based Real Time Tracking and Gesture Recognition for Natural User Interfaces”, Politehnica University of Timişoara, Faculty of Electronics and Telecommunications, 2 Parvan Av., 300223, Timisoara, Romania, Vol. 13, No. 5, 2016.

[16] Le Kang, Peng Ye, Yi Li, and David Doermann, “Convolutional Neural Networks for No-Reference Image Quality Assessment”, University of Maryland, College Park, MD, USA, NICTA and ANU, Canberra, Australia.

[17] Keiron O’Shea and Ryan Nash, “An Introduction to Convolutional Neural Networks”, Department of Computer Science, Aberystwyth University, Ceredigion, SY23 3DB, School of Computing and Communications, Lancaster University, Lancashire, LA1, Dec 2015.

[18] Xiaofeng Han and Yan Li, “The Application of Convolution Neural Networks in Handwritten Numeral Recognition”, College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, 266590, China, International Journal of Database Theory and Application, Vol.8, No.3 (2015), pp. 367-376, 2015.