Marathi Sign Language Recognition

Swaraj Dahibavkar[1], Jayesh Dhopte[2], Mahesh Patole[3], Sulochana Madachane[4]
[1,2,3]*Student, Dept. of Computer Engineering, K.C. College of Engineering & Management Studies & Research, Thane, India*
[4]*Professor, Dept. of Computer Engineering, K.C. College of Engineering & Management Studies & Research, Thane, India*

***Abstract*—Sign Language is a system of communication which uses visual gestures and signs, with the help of which, a deaf or a mute person communicates with other people. As India is home to many spoken languages, there are also many Indian sign languages. In this paper, we are focusing on the Marathi sign language native to Maharashtra, there is a good number of people in the remote areas and also in the urban areas of Maharashtra who don't have the ability to speak or listen, since there local language is Marathi, most of them use Marathi sign language so that they can speak their mother tongue and express themselves, but not everyone knows the Marathi sign language and also, there cannot be interpreters of sign language at every time at every place, this gap of the sign language interpreter is filled by our product, as it converts the Marathi sign language to Marathi alphabets. There are over 43 alphabets in Marathi language (Swar & Vyanjan), each alphabet has a different sign gesture and our product can identify the gesture made from both right and left hand. Our product captures the real time image of the hand gestures, recognizes the gesture and displays the recognized Marathi alphabet in real time.**

***Keywords*—Marathi sign language recognition, skin color detection, segmentation, pattern prediction;**

I. INTRODUCTION

From the very beginning, Communication has been significant in every aspect, whether the communication is between two individuals or in a group, Verbal or Non-verbal, the exchange of information is very vital, because that's how we express ourselves and can tell what's going on in our mind. Whenever we say 'Communication', the idea we picture in our mind is that, someone is speaking and the other person is hearing, which is very common. But unfortunately, there are people who are differently abled, who cannot express themselves by speaking or who can't hear. The Language in which they communicate is "The Sign Language", they express themselves using hand gestures, facial expressions and body postures. The problem over here is that, the differently abled people can only communicate with the other person who knows The Sign Language, and that's very rare. India is home to 12 main spoken languages and every spoken language has its respective Sign Language. In the state of Maharashtra, the native language is Marathi, there are over 9 lakh people who are deaf and mute, most of them use "The Marathi Sign Language" to communicate.

The evolution in the Information technology and the image processing sector has made it possible to develop recognition systems e.g. text and gesture recognition systems[1]. Our proposed system is a real time Marathi sign language recognition system which converts the Marathi sign language to Marathi alphabets in real time, allowing user to communicate easily with other people.

II. LITERATURE SURVEY

Image Recognition Method based on Deep Learning, the author uses Convolutional Neural Networks, Restricted Boltzmann Machines, Autoencoder & Sparse Coding methodologies, The state-of-the-art approaches of the four classes are discussed and analyzed in detail.The CNN architectures can be optimized toward improving desirable properties such as invariance and class discrimination.

Security using Image Processing & Deep Convolutional Neural Networks, the use of computer vision and Image processing helps identify objects frame by frame using improved version of the open source OpenCV's computer vision algorithms to achieve higher accuracy when compared to the trivial systems.Improvement to this system can be done using the OpenFace and the classifier it offers. OpenFace helps us to get the 128 measurements of the face and that is sent as an input to the classifier.

Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning, the accuracy of the system is about 94.7 percent.Due to the low processing speed of CPU, the trained classifier requires about three seconds to classify the image. Adding more images in the dataset and training the classifier via Deep Learning method requires a lot of time.

Skin Color Detection, this paper proposes an improved skin detection method that combines RGB color space & YCgCb color space, then use this texture into detect skin area.Result show that this method can remove a large number of non-skin color under the complex background which enhanced the resolution between skin color & non-skin color.The experimental results still have some mistakes for the false judgement which was caused by the interference under the complex background in the process of skin detection

III. PRODUCT OVERVIEW

The input gesture image is given to the system using the camera, this input image is processed and skin color is detected. The process of differentiating the foreground with the background is called segmentation. The method used for segmentation of image is the edge-based segmentation

technique. In edge-based segmentation method, first of all, the edges are detected and then are connected together to form the object boundaries to segment the required regions. The basic two edge-based segmentation methods are: Gray histograms and Gradient based methods. To detect the edges, one of the basic edge detection techniques like Sobel operator, Canny operator and Robert's operator etc. can be used. Result of these methods is basically a binary image. A fairly large dataset is formed using these binary images for each Marathi alphabet. The training or the machine learning algorithm used is the Convolutional Neural Network (ConvNet/CNN), which is a deep learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. This was the training part, now for the sign recognition, the user gives the gesture input through the camera, this image is processed using Keras to predict the sign, if the probability of the predicted sign is more than 80% then, the system fetches the name of the gesture from the database.

IV. PROPOSED SYSTEM

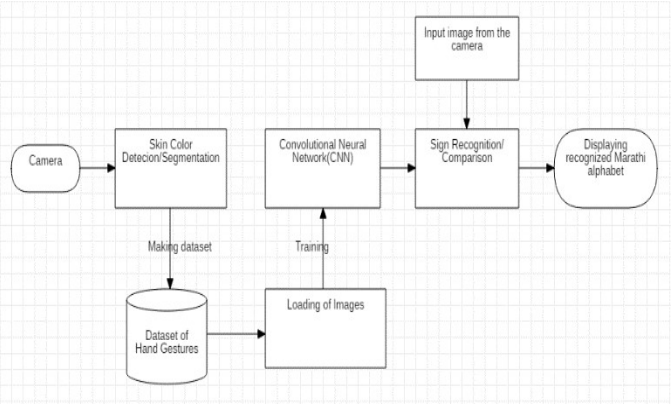

Fig. 1. Product Design

A. Setting hand histogram

In skin color detection, the image is converted from RGB to HSV color model and the hand is detected using OpenCV. As mentioned, hand shown in the 300x100 grid and the rest of the image will be converted into a hsv binary image, making the part in the 300x100 grid as black and rest of the image (background) as white, making it easier in the start itself to classify it much better. Using Gaussian blur on the captured image in the above step and then passed it to the median blur for making the binary image borders more distinct, giving us the required histogram. This process has to be done again only when the lighting conditions or background is changed, as this histogram is the threshold for our system to detect the hand gestures distinctly from the background for the required processing.

Fig. 2. Color image

Fig. 3. Binary image

B. Creating gestures:

This is used to create the dataset for the system and uses a video stream to capture the frames and store it into the folder. For each gesture the video will stream until it captures 1200 frames. The histogram from the initial step will act as a threshold and will help us to store these 1200 frames as binary images.Creates folder with its name as gesture id in "gesture" folder in the root directory, where the images of the gestures with respect to the gesture ids will be sorted, also creates a database of g_id and g_name i.e. name(letter) and Id of the gesture. The images of gesture are kept in the respective folder name, which is named after the g_id corresponding to the g_id in the database and which refers to the name of the gesture in the database. This g_id helps us to connect the images stored in the gesture folder with the name (letter) of the gesture. The gestures can be added and even overwritten using this module but the system has to be retrained whenever gesture is added or deleted or overwritten.

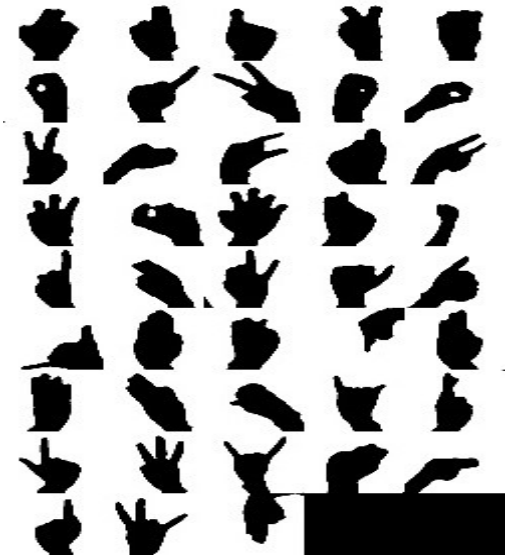

Fig. 4. Sign Gestures

C. Creating dataset for left/righ thand

In this 43 gesture samples using OpenCV. For each gesture 1200 images are captured which are of 50x50 pixels. All these

images are in grayscale which are stored in the gestures/ folder. This script flips every image along the vertical axis. Hence each gesture has 2400 images. This helps us to get the gesture for both left and the right hand.

Fig. 5. Right hand gesture Fig. 6. Left hand gesture

D. Loading images for training

This creates the labels and the validations for the training and also bifurcated the dataset into the training and testing images, labels and validations into a file with FILE type. Total of 103200 images of 43 gestures which are bifurcated as 86000 for training,8600 as validations to be passed with training as validation data and 8600 for testing.

E. Training the system

For sign recognition, the training or the machine learning algorithm used is the Convolutional Neural Network(CNN), A Convolutional Neural Network(ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The model used for training this system is CNN Sequential model. In the initial stage the images are fetched from the folder where the gestures are stored then thenumber of the classes are calculated (43 in our case). The CNN layers are generated as:

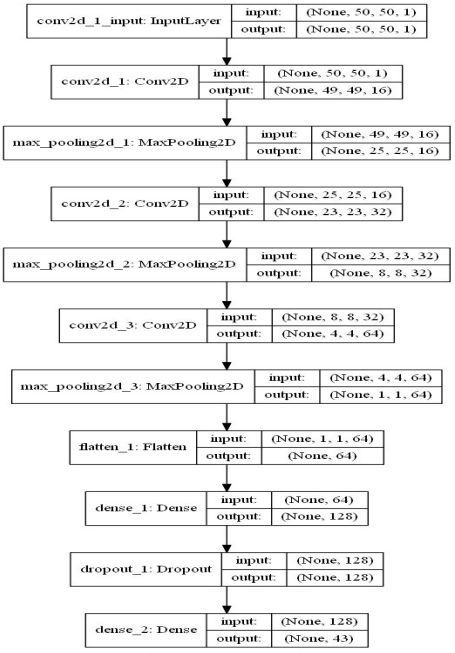

Fig. 7. CNN model

The activation function used in this model is Rectified linear unit activation function(Relu) for all the layers except the last Dense layer of the model where Softmax activation function is used as this softmax activation function gives output of softmax as all values are non-negative and sum to 1,which is useful for the loss function(Categorical Crossentropy) that is chosen in this model which computes the crossentropy loss between the labels and predictions. This crossentropy loss function is used as we have more than one classes to classify.

The optimizer used in this model is Stochastic Gradient Decent(SGD) is a variant of gradient descent. Instead of performing computations on the whole dataset, which is redundant and inefficient. This optimizer only computes on a small subset or random selection of data examples. The optimizer produces the same performance as regular gradient descent when the learning rate is low. The data of the training is stored in a h5 file which is referred when the recognition has to be done. The model is then stored in an image which will be stored in the root directory.

V. IMPLEMENTATION AND RESULT

In this module of the system, the output is displayed on the screen of the recognized gesture. The image issent to keras_predict() to predict the sign, where the sign corresponding to the gesture id is fetched , then if the probability of the predicted sign is more than 80%,then the system fetches the name of the gesture from database with the help of the g_id of the sign.As OpenCV does not support Marathi font by default, we use “Mangal truetypefont” in python using of pilow library which allows us to draw the Marathi letters on the screen with our desired font size and color.

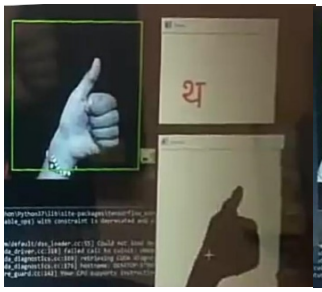

Fig. 8

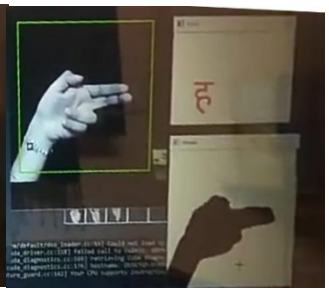

Fig. 9

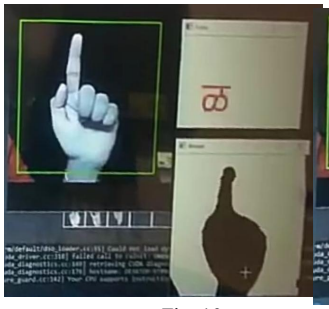

Fig. 10

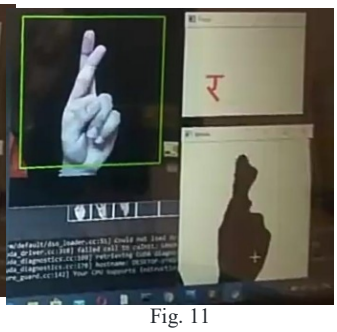

Fig. 11

VI. CONCLUSION

The proposed system is designed to translate Marathi sign language-to-alphabet. It is capable of handling the different input sign images and translates them into alphabets. The proposed system works for both right and left hand gestures, it consumes less time in recognizing the gesture as everything is happening in real time. The accuracy of the system is between 95-100%. There are few measures to be considered while giving input to the camera, the signer should keep in mind that they should wear non-skin colored or dark coloredclothes, thebackground should not be same as the skin color, if these conditions aresatisfied, further the image can be sent for skin color detection. The proposed system is trained on predefined dataset.

In future work, we are looking to deploy it on Raspberry Pi which will be attached to a 7-inch touchscreen and a Pi camera making it a fully portable device that can be carried anywhere, this will be very cost effective, because there won't be any requirement of computer system, we are also looking forward to develop a system which would convert Sign language to words which would help in forming sentences, also we are planning for a dynamic gesture recognition system which will identify the dynamic actions or signs done by the deaf or mute sign user.

REFERENCES

[1] Amitkumar A. Shinde and Ramesh M. Kagalkar, "Sign Language Recognition System in Marathi for Deaf Sign User"
[2] Xin Jia, "Image Recognition Method Based on Deep Learning", 2017
[3] Shadman Shahriar, Ashraf Siddiquee, Tanveerul Islam, Abesh Ghosh, Rajat Chakraborty, Asir Intisar Khan, Celia Shahnaz, Shaikh Anowarul Fattah, "Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning", 2018
[4] Gaurav Kumar, Pradeep Kumar Bhatia, "A Detailed Review of Feature Extraction in Image Processing Systems", 2014
[5] Goutham Reddy Kotapalle, SachinKotni, "Security using Image Processing and Deep Convolutional Neural Networks", 2018