JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Sign Gesture Recognition using MachineLearning Platform

H.Y. Devang[1] , S.J.Sharma[2] , A.R.Wande[3] , S.K.Jagtap[4] , K.A.Pujari[5] , A.S.Pandit[6]

Department of E&TC Engg., Smt.Kashibai Navale College of Engineering , SPPU, Pune, Maharashtra.

[1]harshdewang@gmail.com
[2]shreyassharma6651gmail@.com
[3]ashraywande1422000@gmail.com
[4]skjagtap.skncoe@sinhgad.edu
[5]kapujari.skncoe@sinhgad.edu
[6]anuradha.pandit_skncoe@sinhgad.edu

Abstract— Ability to visualize, understand, interpret and react to environment is a blessing which is necessary, it sometimes either not fully developed or even worse totally absent in some individuals so thus individual have to depend on sign language to be able communicate without words. Understanding sign language is the first step in helping users of sign language to communicate with the rest of the society. A real time sign language translator is an important aspect in facilitating communication with the hearing impaired and the general public. Hence, the purpose is to make this translator using Convolutional Neural Network (CNN) and to achieve high accuracy. The input to the model will be hand gestures in real-time. Here focus is to develop sign language translation system to identify hand gestures into text form. Communication barrier of mute and deaf people is overcome.

Keywords— Sign Language, Communication, Convolutional Neural Network (CNN), Hand Gesture, Translator.

I. INTRODUCTION

A sign language is a way of communicating by using the hands and other parts of the body. According to the world Federation of the deaf, there are around 70 million people in the world who use sign language to communicate. Sign languages differ from country to country. Sign language does not only use signs to communicate. It also uses facial expressions, gestures, body language hand movements and position to communicate. Recently there has been a lot of research through which the sign language can be translated automatically into spoken language or text with the help of advancements in technology. With the help of complex neural network architecture we can detect, classify and translate the sign languageinto spoken language.The Indian Sign Language is not a language understood by everyone. The signs also differ region to region. This further, complicates the communication between the local people and people with hearing or speaking disability. There are other ways like written communication, but the hearing-impaired community is not much skilled in writing a spoken language. Also, such type of communication is slow in real time conversations. In emergency situations, such type of communication is not helpful. The purpose of our project is to contribute to the field of sign recognition. The project will mainly detect and classify the hand gestures or signs. Firstly, extraction of features from the frame sequences will be done with the help of image processing and CNN. After this, classification will be done.Hand gesture recognition is important and challenging task. Hand tracking is a critical part for many applications related to hand gesture recognition. Hand gesture taking out from the screen is most important task in this project, and that will be achieved by Open-CV library. Model will be trained by different hand gesture data.

II. LITERATURE SURVEY

Dixit, Karishma, and A. S. Jalal. suggested Automatic Indian sign language recognition system.A method for automatic recognition of signs on the basis of shape based features is presented. For segmentation of hand region from the images, Otsu's thresholding algorithm was used, that chooses an optimal threshold to minimize the within-class variance of thresholded black and white pixels. Features of segmented hand region were calculated using Hu's invariant moments that were fed to Artificial Neural Network for classification. Performance of the system was evaluated on the basis of Accuracy ,Sensitivity and Specificity.[1] Ahuja, M.Kaur, and A.Singh. proposed "Hand gesture recognition using PCA.A scheme using a database driven hand gesture recognition based upon skin color model approach and thresholding approach along with an effective template matching with can be effectively used for human robotics applications and similar other applications was discussed . Initially,hand region was segmented byapplyingskin color model in YCbCr color space. In the next stage thresholding was applied to separate foreground and background. Finally, template based matching technique was developed using Principal Component Analysis (PCA) for recognition.[2] Sethi, Ashish, S. Hemanth, K.Kumar, N. Bhaskara Rao, and R. Krishnan. proposed SignPro-An Application Suite for Deaf and Dumb. Application that helps the deaf and dumb person to communicate with the rest of the world using sign language was discussed. The key feature in this system was the real time gesture to text conversion. The processing steps include: gesture extraction, gesture matching and conversion to speech. Gesture extraction involves use of various image processing techniques such as histogram matching, bounding box computation, skin colour segmentation and region growing. Techniques applicable for Gesture matching include feature point matching and correlation based matching. The other features in the application include voicing out of text and text to gesture conversion.[3] Rajamohan, Anbarasi, R. Hemavathy, and M. Dhanalakshmi. proposes Deaf-mute communication interpreter. Various prevailing methods of deaf-mute communication interpreter system was discussed. The two broad classification of the communication methodologies used by the deaf mute people are Wearable Communication Device and Online Learning System. Under Wearable communication method, there are Glove based system, Keypad method and Handicom Touch-screen. All the above mentioned three sub- divided methods make use of various sensors, accelerometer, a suitable micro controller, a text to speech conversion module, a keypad and a touch-screen. The need for an external device to interpret the message between a deaf mute and non-deaf mute people can be overcome by the second method i.e online learning system. The Online Learning System has different methods. The five subdivided methods are- SLIM module, TESSA, Wi-See Technology, SWI_PELE System and Web-Sign Technology[4]

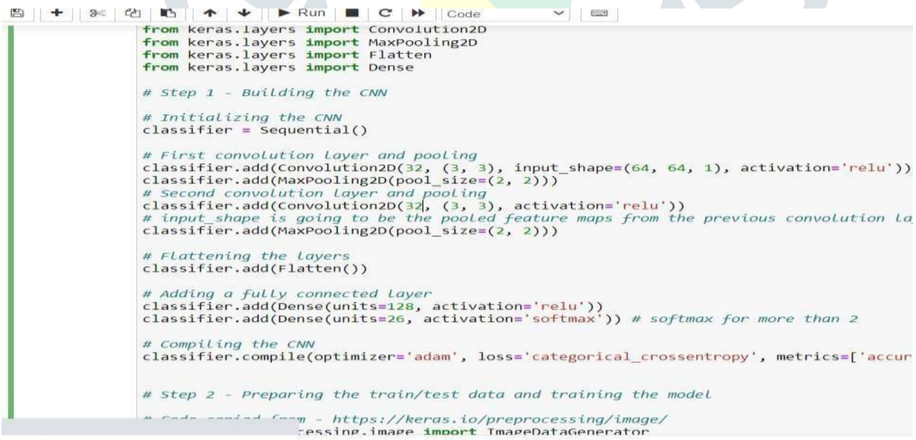
III. METHODOLOGY

Data set of different hand gestures is already provided to system for training and testing purpose, now system will take hand gesture as an input from laptop's camera for that user need to put their hand in region of interest,in image pre-processing system will remove all the unnecessary noise, smoothing of edges, resizing of images and segmentation images will be take place.Feature Extraction aims to reduce the number of feature in dataset by creating new feature form existing ones and then discarding the original feature .These new reduced set of features should then be able to summarize most of the information contained in the original set of feature

Fig:3.1: Block diagram of hand gesture sign recognition

Image classification is the process of predicting a specific class, or label, for something that is defined by a set of data points. Image classificationis a subset of the classification problem, where an entire image is assigned a label.In the output state our system will firstly perform all the above operation in that it will compare gesture which was acquired by system with data set and it will try to predict appropriate hand gesture And it willget printed on out screen.

Sequential classifier: The model type is Sequential. Sequential is the easiest wayto build a model in Keras. It allows you to build a model layerby layer. We use the 'add()' function to add layers to our model. Our first 2 layers are Conv2Dlayers. These are convolution layers that will deal withour input images, which are seen as 2-dimensional matrices.64 in the first layer and 32 in the second layer are the number of nodes in each layer.



```
from keras.layers import Convolution2D
from keras.layers import MaxPooling2D
from keras.layers import Flatten
from keras.layers import Dense

# Step 1 - Building the CNN

# Initializing the CNN
classifier = Sequential()

# First convolution layer and pooling
classifier.add(Convolution2D(32, (3, 3), input_shape=(64, 64, 1), activation='relu'))
classifier.add(MaxPooling2D(pool_size=(2, 2)))
# Second convolution layer and pooling
classifier.add(Convolution2D(32, (3, 3), activation='relu'))
# input_shape is going to be the pooled feature maps from the previous convolution la
classifier.add(MaxPooling2D(pool_size=(2, 2)))

# Flattening the layers
classifier.add(Flatten())

# Adding a fully connected layer
classifier.add(Dense(units=128, activation='relu'))
classifier.add(Dense(units=26, activation='softmax')) # softmax for more than 2

# Compiling the CNN
classifier.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accur


# Step 2 - Preparing the train/test data and training the model
```

Fig:3.2: Sequential Classifier

Number can be adjusted to be higher or lower, depending on the size of the data set. In our case, 64 and 32 work well,so we will stick with this for now. Kernel size is the size of the filter matrix for our convolution. So a kernel size of 3 means we will have a 3x3 filter matrix. Refer back to the introduction and the first image for a refresher on this. Activation is the activation function for the layer. The activation function we will be using for our first 2 layers is the ReLU, or Rectified Linear Activation. This activation function has been proven to work well in neural networks. Our first layer also takes in an input shape. This is the shape of each input image, 28,28,1 as seenearlier on, with the 1 signifying that the images are greyscale. In between the Conv2Dlayers and the dense layer, there is a 'Flatten' layer. Flatten serves as a connection between the convolution and dense layers. 'Dense' is the layer type we will use in for our output layer. Dense is a standard layer type that is used in many cases for neural networks. We will have 10 nodes in our output layer, one for each possible outcome (0–9).The activation is 'softmax'. Softmax makes the output sum up to 1 so the output can be interpreted as probabilities.The model will then make its prediction based on which option has the highest probability.

IV. EXPERIMENTAL RESULTS

DATA SET FOR EXPERIMENTATION:

In the communicative hand/arm gesture taxonomies, modern sign language is considered as the most organized and structured form out of various gesture categories and is an important means of communication amongthe hearing-impaired and deaf community. Sign language involves the usage of different parts of the body, such as fingers, hand, arm, head, body, and facial expression. One class of sign languages, known also as finger spelling, is limited to a collection of manual signs.

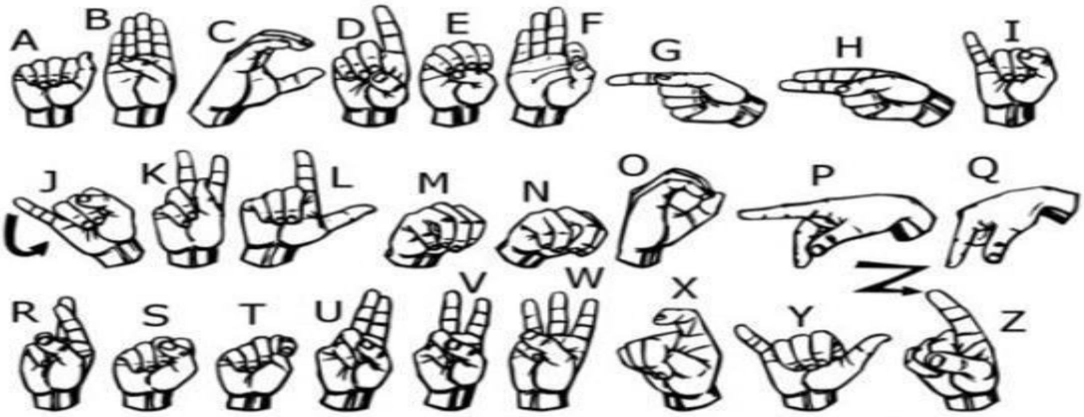


Fig:4.1: Data set of hand gesture

There are many challenges associated with the accuracy and usefulness of gesture recognition software. For image-based gesture recognition there are limitations on the equipment used and image noise. Images or video may not be under consistent lighting, or in the same location. Gesture based interfaces allow human computer interaction to be in a natural and intuitive manner.

ACCURACY FOR ALPHABETS GESTURE:

```
Epoch 1/5
12380/12380 [==============================] - 251s 20ms/step - loss: 0.5077 - accuracy: 0.8462 - val_loss: 0.0777 - val_accura
cy: 0.9791
Epoch 2/5
12380/12380 [==============================] - 228s 18ms/step - loss: 0.0885 - accuracy: 0.9728 - val_loss: 0.0089 - val_accura
cy: 0.9971
Epoch 3/5
12380/12380 [==============================] - 243s 20ms/step - loss: 0.0560 - accuracy: 0.9838 - val_loss: 0.0292 - val_accura
cy: 0.9915
Epoch 4/5
12380/12380 [==============================] - 360s 29ms/step - loss: 0.0450 - accuracy: 0.9876 - val_loss: 0.0072 - val_accura
cy: 0.9981
Epoch 5/5
12380/12380 [==============================] - 248s 20ms/step - loss: 0.0497 - accuracy: 0.9873 - val_loss: 0.0034 - val_accura
cy: 0.9988
```

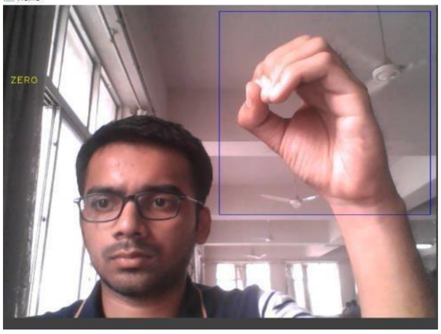
Fig: 5.1:Accuracy for alphabets gesture.

ACCURACY FOR NUMBER'S GESTURE

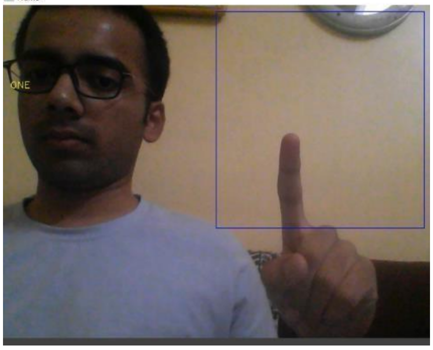
```
Epoch 1/10
600/600 [==============================] - 13s 19ms/step - loss: 0.8364 - accuracy: 0.7167 - val_loss: 1.3615 - val_accuracy:
0.5000
Epoch 2/10
600/600 [==============================] - 5s 9ms/step - loss: 0.2753 - accuracy: 0.9267 - val_loss: 1.1316 - val_accuracy: 0.6
667
Epoch 3/10
600/600 [==============================] - 7s 12ms/step - loss: 0.1921 - accuracy: 0.9400 - val_loss: 1.4716 - val_accuracy: 0.
5333
Epoch 4/10
600/600 [==============================] - 7s 11ms/step - loss: 0.1120 - accuracy: 0.9617 - val_loss: 1.5380 - val_accuracy: 0.
6000
Epoch 5/10
600/600 [==============================] - 6s 10ms/step - loss: 0.0770 - accuracy: 0.9783 - val_loss: 1.3809 - val_accuracy: 0.
6667
Epoch 6/10
600/600 [==============================] - 6s 11ms/step - loss: 0.1092 - accuracy: 0.9650 - val_loss: 0.8701 - val_accuracy: 0.
7333
Epoch 7/10
600/600 [==============================] - 7s 12ms/step - loss: 0.0505 - accuracy: 0.9883 - val_loss: 1.1684 - val_accuracy: 0.
7000
Epoch 8/10
600/600 [==============================] - 6s 10ms/step - loss: 0.0466 - accuracy: 0.9883 - val_loss: 1.1279 - val_accuracy: 0.
6000
Epoch 9/10
600/600 [==============================] - 7s 12ms/step - loss: 0.0571 - accuracy: 0.9817 - val_loss: 3.5057 - val_accuracy: 0.
3667
Epoch 10/10
600/600 [==============================] - 8s 13ms/step - loss: 0.0731 - accuracy: 0.9700 - val_loss: 1.4951 - val_accuracy: 0.
7000
```

Fig:5.2:Accuracy for number gesture.

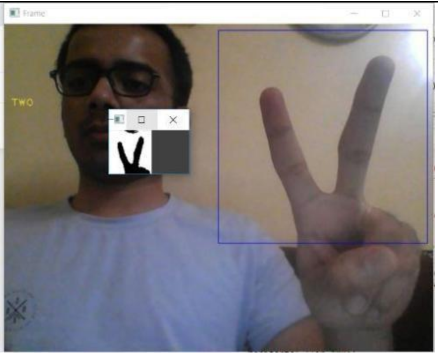
GESTURE RECOGNITION



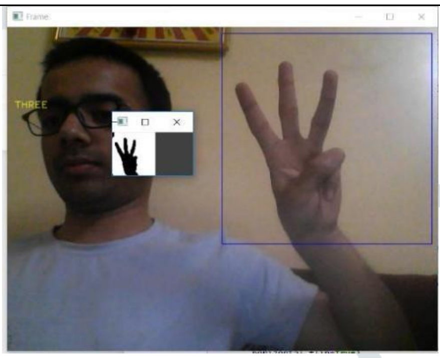
a



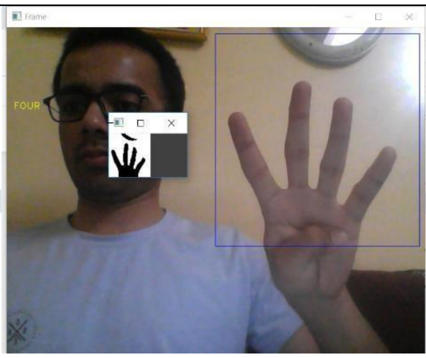
b



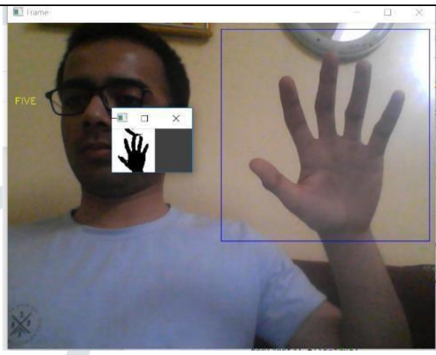
c



e



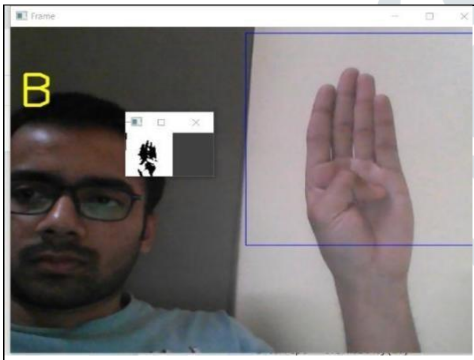
f



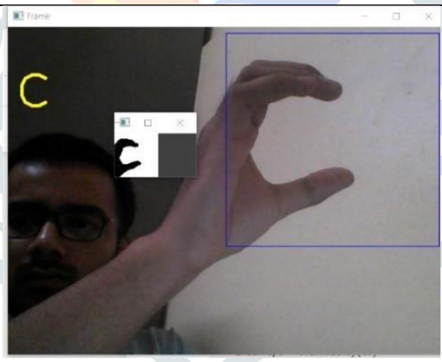
g

Fig 5.3 Results with identification number gestures form 0 to 5

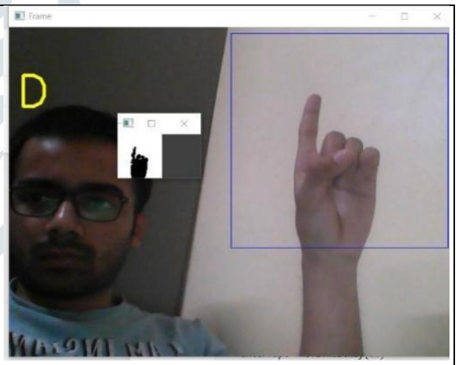
Fig a is representing gesture of 0, Fig b is representing gesture of 1 , Fig c is representing gesture of 2Fig d is representing gesture of 3, Fig e is representing gesture of 4, Fig f is representing gesture of 5



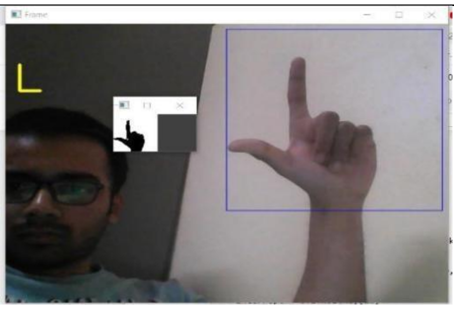
a



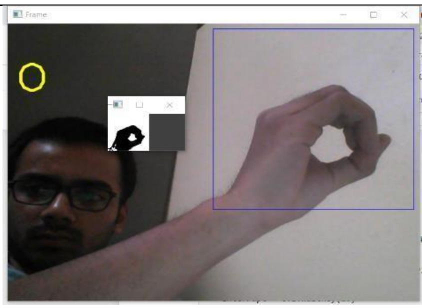
b



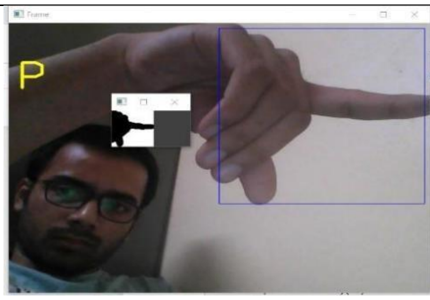
c



d



e



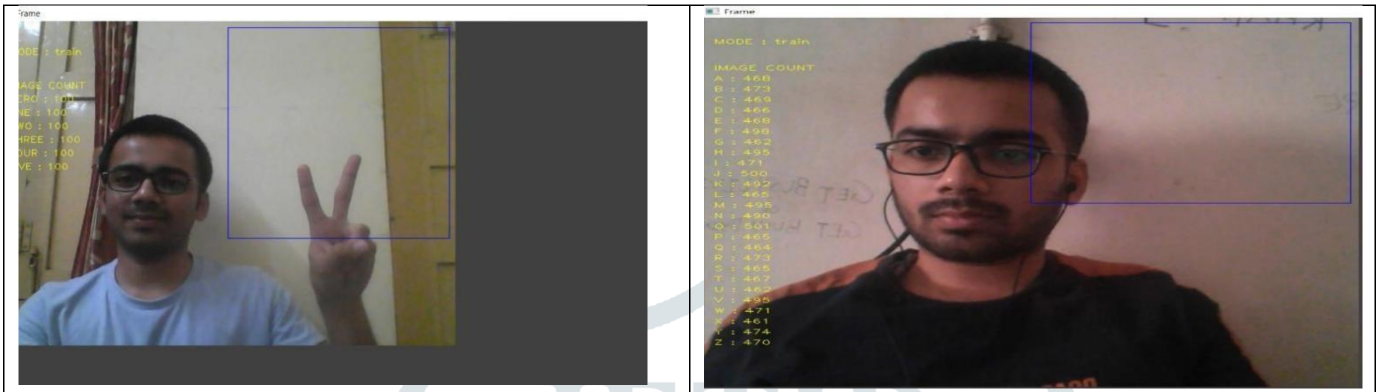
f

Training Mode:-

Fig5.4: Above Images shows the identification of alphabets gestures Fig: a shows gesture B , Fig: b shows gesture C , Fig: c shows gesture D,Fig: d shows gesture L,Fig: e shows gesture O , Fig: f shows gesture P,

In training mode one can give images/data from keyboard by pressing that specific key for gesture. for example:- As shown in below figureone can create data or add data for training purpose of gesture TWO 2 .

Testing Mode:-



(a) Count of number data (b) Count of alphabet data

Fig 4.5 Model in traning mode

In testing mode user need to put their hand in blue region of interest box and user will get expected output printed on left side.

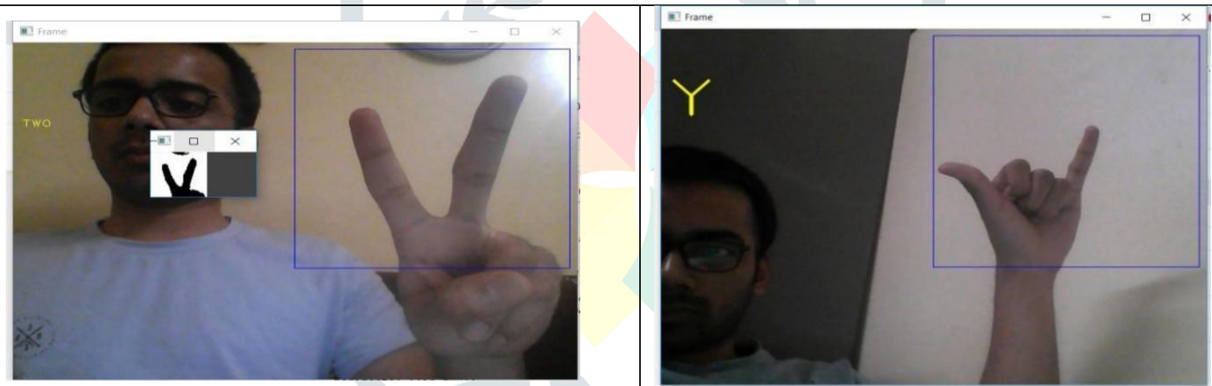


Fig5.5:Model in testing mode

ADVANTAGES

Easy to operate
Gesture patterns are not critical
User doesn’t required any advanced training to

useitUser Friendly

LIMITATIONS

Large training data is required
Many gestures recognition system do not read motions accurately or optimally due to factors like insufficient background light.

APPLICATIONS

In home automation hand gesture can be used to switch ON/OFF
lights.It can also be used to control robot.
It is widely used by deaf and mute people for communication purpose.

VI. CONCLUSIONS AND FUTURE SCOPE

A Real time vision based sign language recognition system for deaf mute people have developed using neural network.Model is created with low cost.This high speed model creates meaningful word or Sentence using different hand gesture.The created model shows high accuracy.

Gesture recognition, along with voice recognition, facial recognition, lip movement recognition and eye tracking combined can be used to create something called a perceptual user interface (PUI), a completely different way to interact with computer systems which will improve usability and creativity by leaps and bounds.Most video games today are played either on game consoles, arcade units or PCs, and all require a combination of input devices. Gesture recognition can be used to truly immerse a players in the game world like never before. In homes, offices, transport vehicles and more, gesture recognition can be incorporated to greatly increase usability and reduce the resources necessary to create primary or secondary input systems like remote controls, car entertainment systems with buttons or similar.One of the biggest challenges faced today is providing separate and equally non cumbersome services to the differently abled and handicapped. While there are special provisions around the world, there's still huge room for improvement to bring all lives on equal footing. Gesture recognition technology can eliminate a lot of manual labor and make life much easier for those who aren't as fortunate as most of us are.

REFERENCES

[1] Dixit, Karishma, and A.S.Jalal. "Automatic Indian sign language recognition system." *2013 3rd IEEE International Advance Computing Conference (IACC)*. IEEE, 2013.

[2] Ahuja, M.Kaur, and A.Singh. "Hand gesture recognition using PCA." *International Journal of Computer Science Engineering and Technology (IJCSET)* 5.7 (2015): 267-27.

[3] Sethi, Ashish, S. Hemanth, K.Kumar, N. Bhaskara Rao, and R. Krishnan. "SignPro-An Application Suite for Deaf and Dumb." *IJCSET* 2, no. 5 (2012): 1203-1206.

[4] Rajamohan, Anbarasi, R. Hemavathy, and M. Dhanalakshmi. "Deaf-mute communication interpreter." *International Journal of Scientific Engineering and Technology* 2.5 (2013): 336-341.

[5] Ren, Z., Yuan, J., Meng, J., & Zhang, Z. (2013). Robust part-based hand gesture recognition using kinect sensor. IEEE transactions on multimedia, 15(5), 1110-1120.

[6] Chen, Feng-Sheng, Chih-Ming Fu, and Chung-Lin Huang. "Hand gesture recognition using a real-time tracking method and hidden Markov models." Image and vision computing 21, no. 8 (2003): 745-758.