



CONVERSION OF SIGN LANGUAGE TO TEXT/SPEECH

MINOR PROJECT(4IT31)

INFORMATION TECHNOLOGY DEPARTMENT

BIRLA VISHVAKARMA MAHAVIDYALAYA ENGINEERING COLLEGE

(AN AUTONOMOUS INSTITUTION)

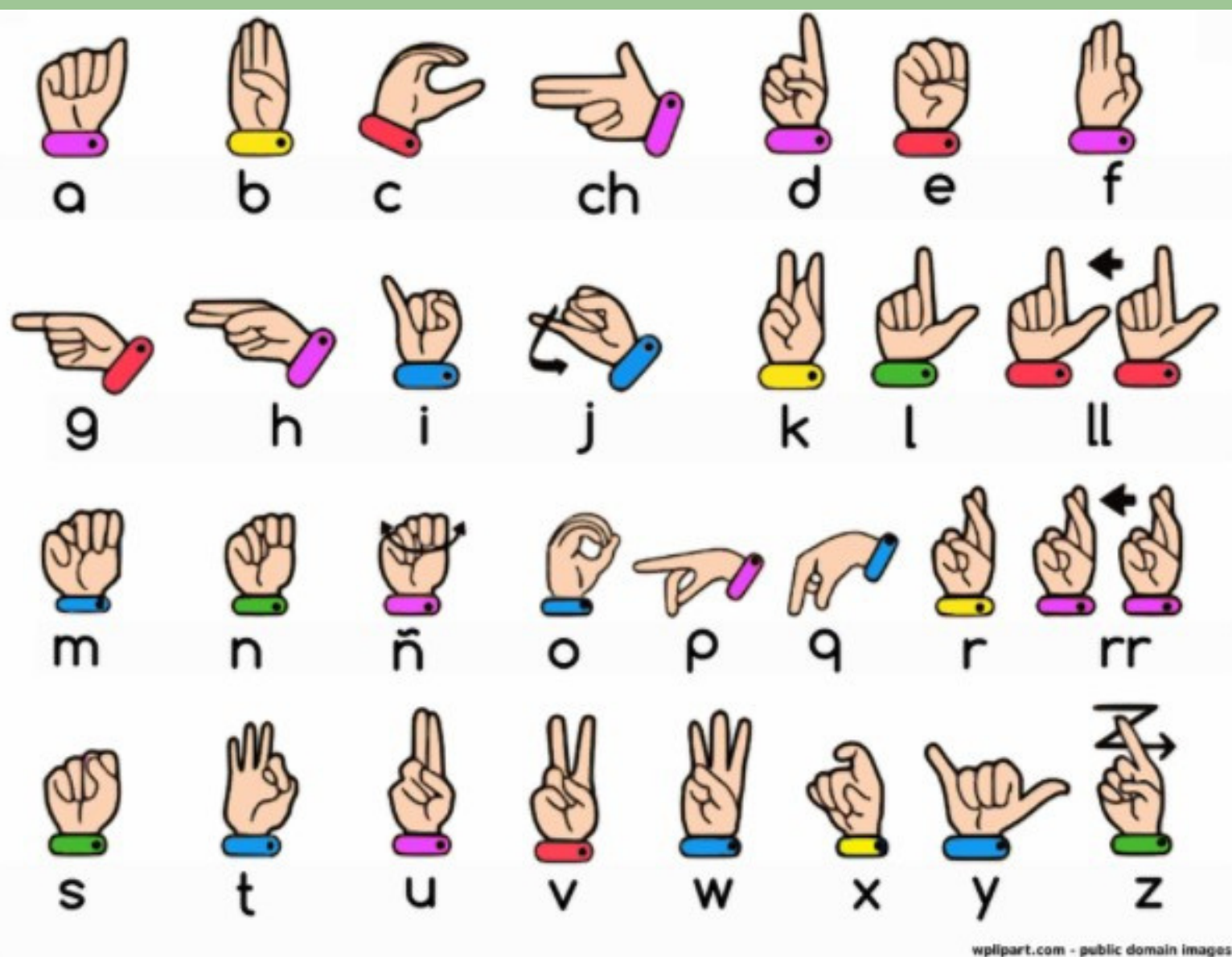


ABSTRACT

SIGN LANGUAGE IS ONE OF THE OLDEST AND MOST NATURAL FORM OF LANGUAGE FOR COMMUNICATION, HENCE WE HAVE COME UP WITH A REAL TIME METHOD USING NEURAL NETWORKS FOR FINGER SPELLING BASED AMERICAN SIGN LANGUAGE. WE PROPOSE A CONVOLUTION NEURAL NETWORK (CNN) METHOD WITH THE HELP OF MEDIA PIPE LIBRARY TO RECOGNIZE HAND GESTURES OF HUMAN ACTIONS FROM AN IMAGE CAPTURED BY CAMERA.

INTRODUCTION

MORE THAN 70 MILLION DEAF PEOPLE AROUND THE WORLD USE SIGN LANGUAGES TO COMMUNICATE. SIGN LANGUAGE ALLOWS THEM TO LEARN, WORK, ACCESS SERVICES, AND BE INCLUDED IN THE COMMUNITIES. IT IS HARD TO MAKE EVERYBODY LEARN THE USE OF SIGN LANGUAGE WITH THE GOAL OF ENSURING THAT PEOPLE WITH DISABILITIES CAN ENJOY THEIR RIGHTS ON AN EQUAL BASIS WITH OTHERS. SO, THE AIM IS TO DEVELOP A USER-FRIENDLY HUMAN COMPUTER INTERFACE (HCI) WHERE THE COMPUTER UNDERSTANDS THE AMERICAN SIGN LANGUAGE THIS PROJECT WILL HELP THE DUMB AND DEAF PEOPLE BY MAKING THEIR LIFE EASY. THEY JUST NEED TO DO THAT SIGN LANGUAGE GESTURES AND THIS SYSTEM WILL IDENTIFY WHAT HE/SHE IS TRYING TO SAY AFTER IDENTIFICATION IT GIVES THE OUTPUT IN THE FORM OF TEXT AS WELL AS SPEECH FORMAT.



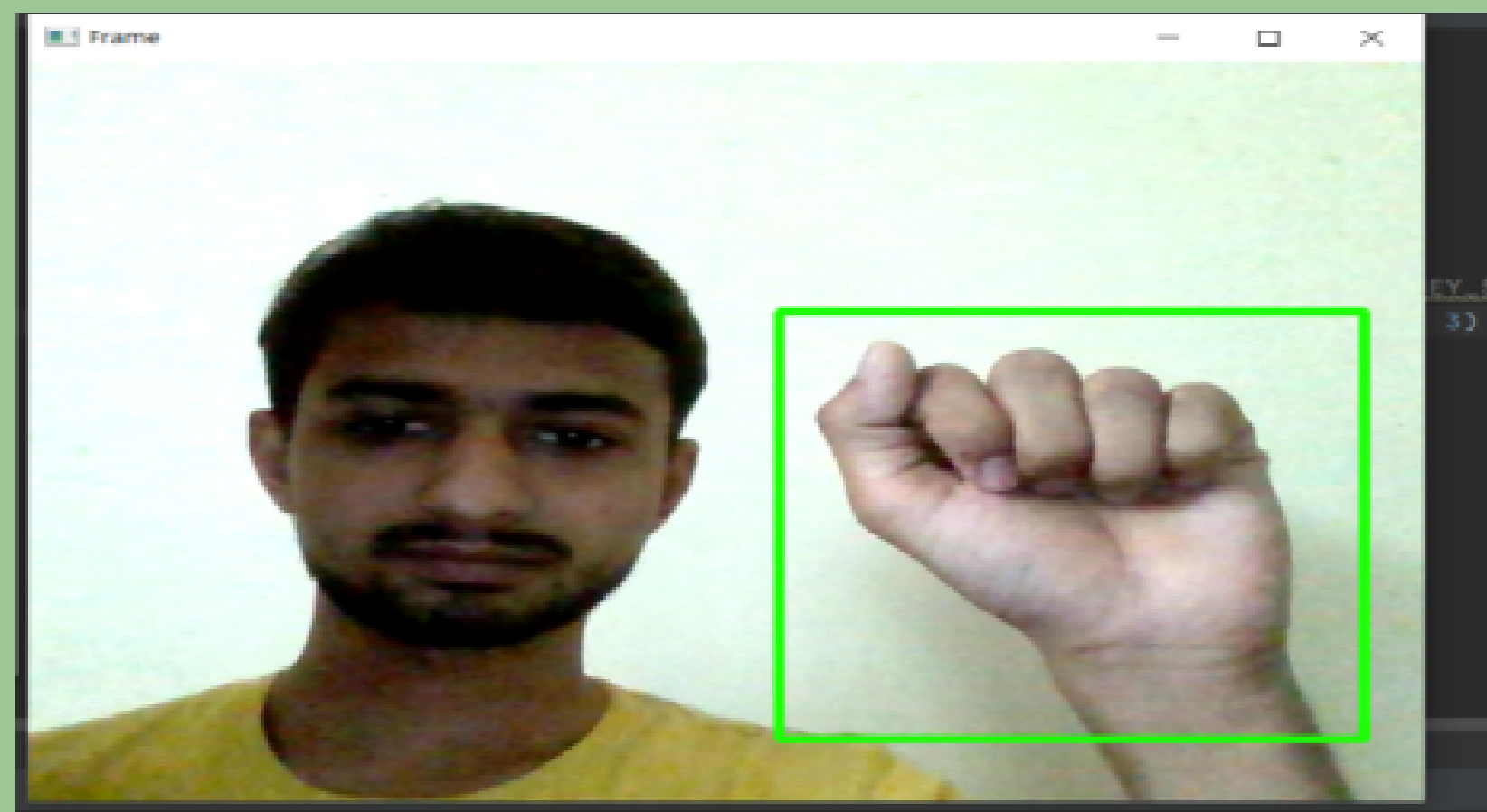
LITERATURE SURVEY

Comparison Table:

Author name	Mahesh Kumar	Krishna Modi	Bikash K. Yadav	Ayush Pandey	Victor Adebimpe Akano	Rakesh Kumar
Algorithm	LDA	Blob Analysis	CNN	CNN	KNN	contour measurement
Accuracy	80%	93%	95.8%	95%	92%	86%
Year	2018	2013	2020	2020	2018	2021

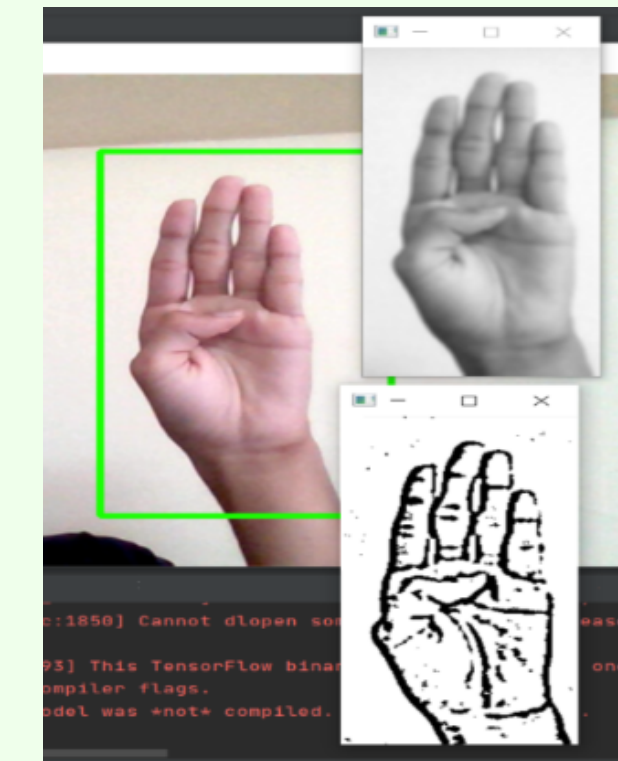
DATA ACQUISITION

IN VISION-BASED METHODS, THE COMPUTER WEBCAM IS THE INPUT DEVICE FOR OBSERVING THE INFORMATION OF HANDS AND/OR FINGERS. THE VISION BASED METHODS REQUIRE ONLY A CAMERA, THUS REALIZING A NATURAL INTERACTION BETWEEN HUMANS AND COMPUTERS WITHOUT THE USE OF ANY EXTRA DEVICES, THEREBY REDUCING COSTS.



DATA PREPROCESSING AND FEATURE EXTRACTION

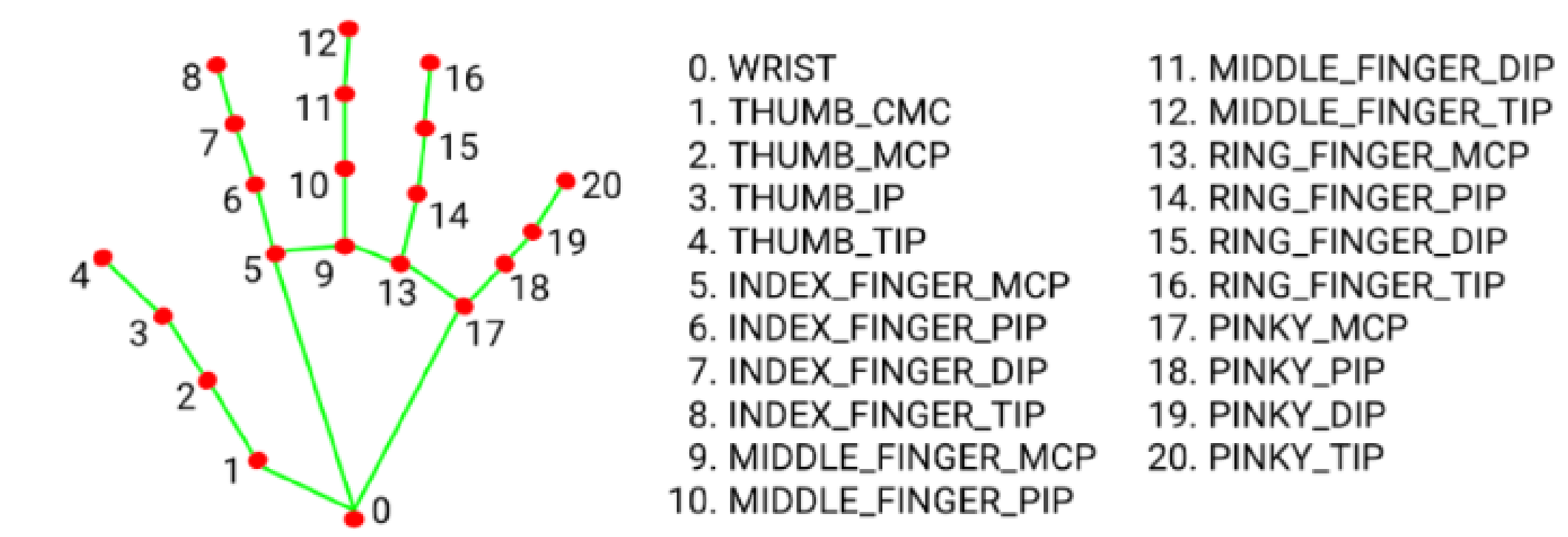
IN THIS APPROACH FOR HAND DETECTION, FIRSTLY WE DETECT HAND FROM IMAGE THAT IS ACQUIRED BY WEBCAM AND FOR DETECTING A HAND WE USED MEDIA PIPE LIBRARY WHICH IS USED FOR IMAGE PROCESSING. SO, AFTER FINDING THE HAND FROM IMAGE WE GET THE REGION OF INTEREST (ROI) THEN WE CROPPED THAT IMAGE AND CONVERT THE IMAGE TO GRAY IMAGE USING OPENCV LIBRARY AFTER WE APPLIED THE GAUSSIAN BLUR .THE FILTER CAN BE EASILY APPLIED USING OPEN COMPUTER VISION LIBRARY ALSO KNOWN AS OPENCV[1]. THEN WE CONVERTED THE GRAY IMAGE TO BINARY IMAGE USING THRESHOLD ANDADAPTIVE THRESHOLD METHODS..



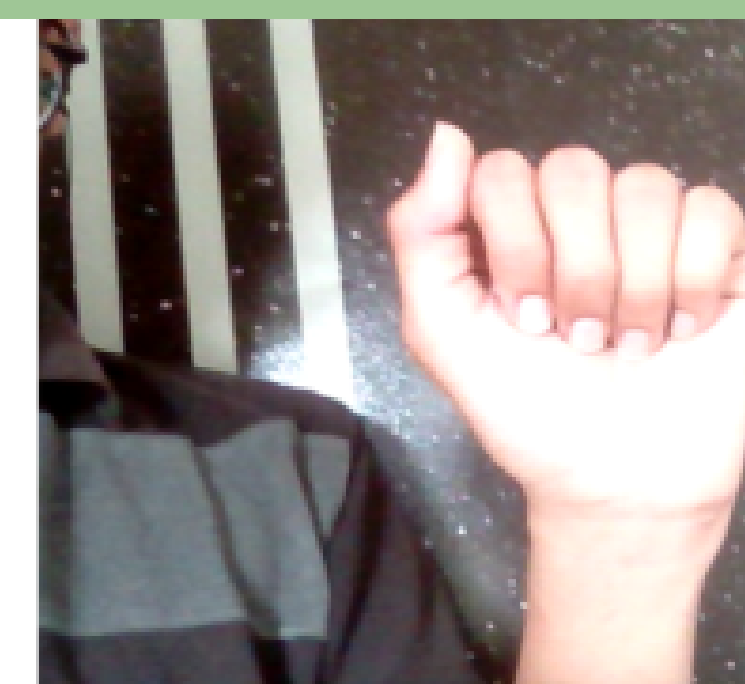
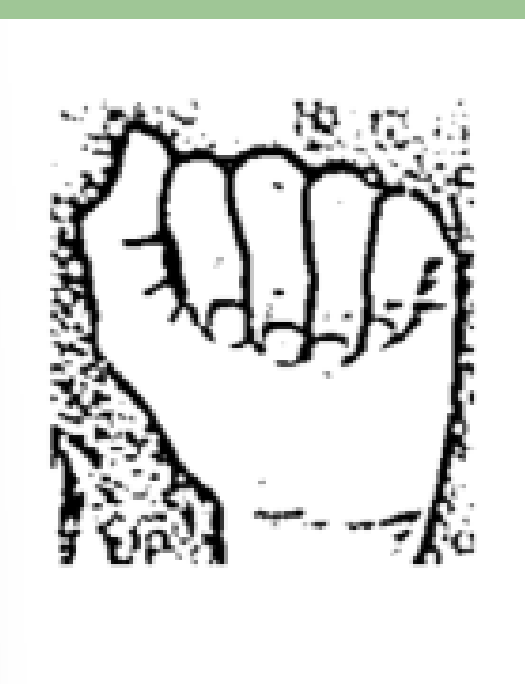
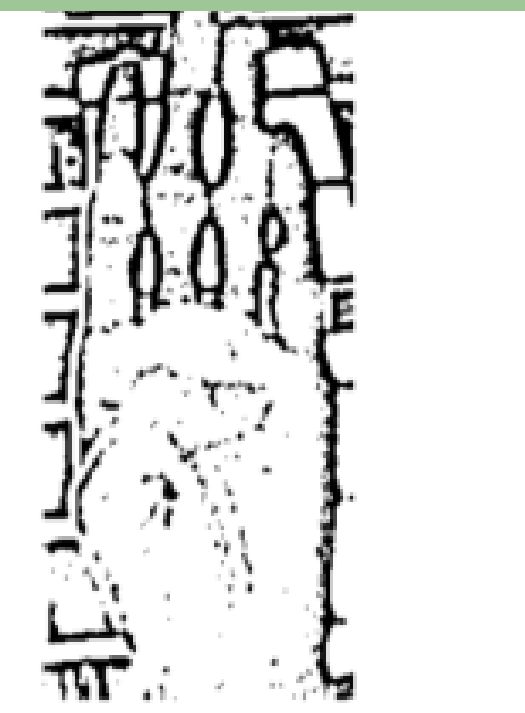
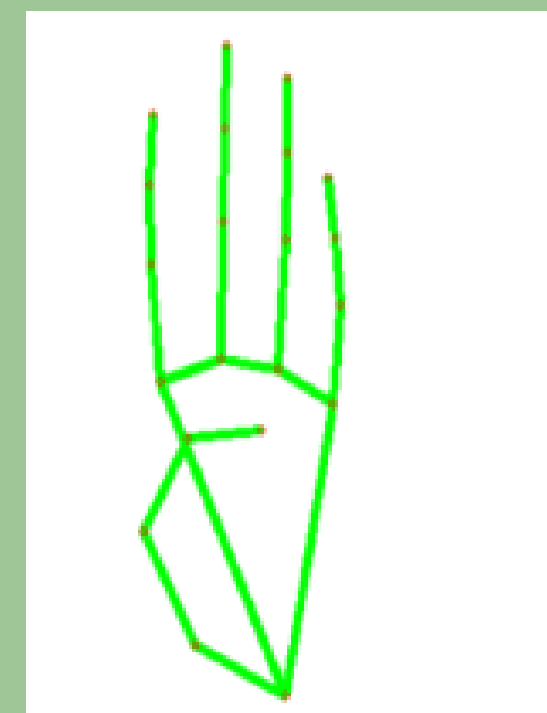
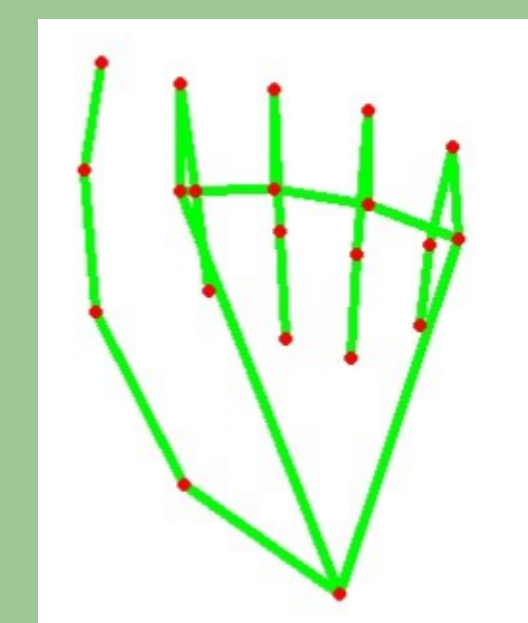
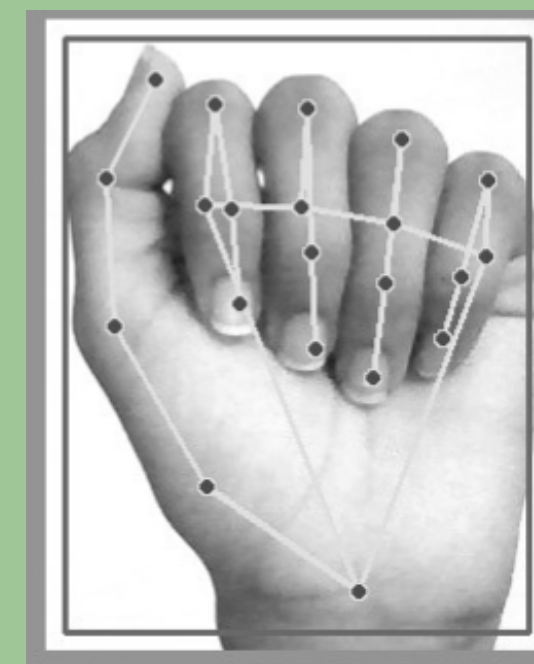
IN THIS METHOD THERE ARE MANY LOOP HOLES LIKE YOUR HAND MUST BE AHEAD OF CLEAN SOFT BACKGROUND AND THAT IS IN PROPER LIGHTNING CONDITION THEN ONLY THIS METHOD WILL GIVE GOOD ACCURATE RESULTS BUT IN REAL WORLD WE DONT GET GOOD BACKGROUND EVERYWHERE AND WE DON'T GET GOOD LIGHTNING CONDITIONS TOO.

SO TO OVERCOME THIS SITUATION WE TRIED DIFFERENT APPROACHES THEN WE REACHED AT ONE INTERESTING SOLUTION IN WHICH FIRSTLY WE DETECT HAND FROM FRAME USING MEDIAPIPE[2] AND GET THE HAND LANDMARKS[3] OF HAND PRESENT IN THAT IMAGE THEN WE DRAW AND CONNECT THOSE LANDMARKS IN SIMPLE WHITE IMAGE

Mediapipe Landmark System:



NOW WE WILL GET THIS LANDMARK POINTS AND DRAW IT IN PLAIN WHITE BACKGROUND USING OPENCV LIBRARY



GESTURE CLASSIFICATION

THE PREPROCESSED 180 IMAGES/ALPHABET WILL FEED THE KERAS CNN MODEL[3]. BECAUSE WE GOT BAD ACCURACY IN 26 DIFFERENT CLASSES THUS, WE DIVIDED WHOLE 26 DIFFERENT ALPHABETS INTO 8 CLASSES IN WHICH EVERY CLASS CONTAINS SIMILAR ALPHABETS: [AEMNST][BDFIKRUVW][CO][YJ][CO][PQZ][L][X]

ALL THE GESTURE LABELS WILL BE ASSIGNED WITH A PROBABILITY. THE LABEL WITH THE HIGHEST PROBABILITY WILL TREATED TO BE THE PREDICTED LABEL. SO WHEN MODEL WILL CLASSIFY [AEMNST] IN ONE SINGLE CLASS USING MATHEMATICAL OPERATION ON HAND LANDMARKS WE WILL CLASSIFY FURTHER INTO SINGLE ALPHABET A OR E OR M OR N OR S OR T.

RESULTS & CONCLUSION

Sign Language To Text Conversion



Character : O

Sentence : HELL

Clear Speak

Suggestions :

HELL

JELL

ELL

HELLS

-FINALLY, WE GOT 97% ACCURACY (WITH AND WITHOUT CLEAN BACKGROUND AND PROPER LIGHTNING CONDITIONS) THROUGH OUR METHOD. AND IF THE BACKGROUND IS CLEAR AND THERE IS GOOD LIGHTNING CONDITION THEN WE GOT EVEN 99% ACCURATE RESULTS

TEAM

GUIDE : DR. NILESH B. PRAJAPATI
TEAM MEMBERS : DEVANSH RAVAL (19IT470)
KELVIN PARMAR (19IT473)

REFERENCES

- [1]-[HTTPS://DOCS.OPENCV.ORG/4.X/INDEX.HTML](https://docs.opencv.org/4.x/index.html)
- [2]-[HTTPS://GOOGLE.GITHUB.IO/MEDIAPIPE/SOLUTIONS/HANDS](https://github.com/mmediapipe/solutions/hands)
- [3]-[HTTPS://WWW.TENSORFLOW.ORG/TUTORIALS/IMAGES/CNN.](https://www.tensorflow.org/tutorials/images/cnn)