American Sign Language to Text - Speech using Background Subtraction using Running Averages

Jyoti Tripathi, Prafull Goel, Raman Bhadauria, Nikhil Yadav, Keshav Gupta

***Abstract**: This Paper Proposes A System Which Converts American Sign Language Hand Gestures Into Text Cum Speech And Helps To Bridge The Communication Gap Between Deaf-Mute People And Rest Of The Society. Any System For This Purpose Generally Has Four Modules: Segmentation, Feature Extraction, Classification And Text-To-Speech. This Paper Focuses On An Improved Method For The Segmentation And The Feature Extraction Processes To Get More Better Resultswhile Using The Standard Techniques On The Other Two Modules. Proposed Algorithm Captures Initial 30 Frames Of The Live Video From The Web Cam Of The System To Construct The Background Model. It Then Finds The Absolute Difference Between The Current Frame And The Background Model In Order To Get The Foreground. Various Features Are Extracted To Classify The Gestures Like Contour, Convexity Hull Etc.. Proposed Algorithm Has Been Tested Under Low And Normal Room Light Conditions. The Overall Performance Of The Proposed Model Will Be Very High And Will Produce Far More Better Resultsdue To Improved Proposed Algorithms For The Initial Two Modules In Comparison To Other Standard Techniques Used Like Hsv, Ycbcr The Above System Can Be Incorporated Into Simple Web Applications, Mobile Applications And Many Other Applications Translating Gestures In The Conversations In Real Time.*

***Keywords:**ASL, Background Subtraction, Running Averages, Segmentation, Feature Extraction, HSV, YCbCr.*

I. INTRODUCTION

Human beings interact with each other to convey their ideas, thoughts, and experiences to the people around them. They use various languages for the communication with each other. But this is not the case for deaf-mute people. Sign language paves the way for deaf-mute people to communicate. Sign language is a visual language that is used by deaf and dumb people as their mother tongue. It is achieved by simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions.

According to a report by World Federation of the Deaf (WFD)[1], there are more than 70 million Deaf-Mute people around the world. Even though the development of sign language has provided a way for the deaf-mute people to interact with the ordinary people but due to lack of proper education and knowledge of sign language in the society, the communication gap within them is still intact due to dissimilarity of communication modes, as most people do not understand the sign language.

This makes it difficult for signers to move around independently in society without an interpreter.Hence, the proposed system tries to bridge this communication gap by acting as an interpreter. The proposed system/application converts Hand Gestures (ASL) to text/speech and act as an effective medium for deaf and mute people to communicate with someone who don’t know sign language and hence bridging the communication gap between them.

II. RELATED WORK

Hand gesture recognition techniques can be categorized into two categories. The first category is wearable glove based hand gesture recognition. The glove includes various sensors which converts the fingers and hand movements into electrical signals. These signals are further used to extract meaningful data for hand gestures recognition. The second category is vision-based hand gesture recognition. It includes capturing the visual frames of hand and extracting relevant features for recognizing the hand gestures. It can be further classified into two categories. The first one includes constructing a 3-D model of signer’s hand using various projections with the help of optical sensors. The second one includes capturing 2-D visual frames. These input frames are then processed to extract the visual characteristics of the hand.

Vijayalakshmi P. et al. [2] proposed a system which uses Flex-sensor to recognize English alphabets and few other words. Flex sensors were mounted on the wearable gloves and fitted all around the fingers. The output data stream from the (degree of bend) sensor, tactile sensor (determine interaction between fingers) and accelerometer are fed to the Arduino microcontroller, where it is processed and then converted to its corresponding digital values. The microcontroller unit will compare these readings with the pre-defined threshold values and the corresponding gestures are recognized. The proposed system achieves the accuracy of 87.5%. But the proposed model is not practical as it is not feasible to carry these gloves everywhere in real life.

Anup Kumar et al. [3] developed an efficient mechanism for real-time hand gesture recognition using visual-based techniques. The proposed approach uses HSV (Hue, Saturation, Value colour space) for skin colour detection. In order to obtain the greatest contour (hand), firstly the face is recognized and subtracted from the input frame using Viola and Jones algorithm.

Revised Manuscript Received on December 30, 2019.

* Correspondence Author
Jyoti Tripathi, Department of Computer Science and Engineering G.B. Pant Government Engineering College, New Delhi, India.
Prafull Goel, Department of Computer Science and Engineering G.B. Pant Government Engineering College, New Delhi, India.
Raman Bhadauria, Department of Computer Science and Engineering G.B. Pant Government Engineering College, New Delhi, India.
Nikhil Yadav, Department of Computer Science and Engineering G.B. Pant Government Engineering College, New Delhi, India.
Keshav Gupta, Department of Computer Science and Engineering G.B. Pant Government Engineering College, New Delhi, India.

The proposed system achieves the accuracy of 93% for static gestures (24 alphabets) and 100% for dynamic gestures (2 alphabets and 2 dynamic words: ‘no’ and ‘bye’). The major drawback of this model is that it works with greater accuracy only in the constrained environment i.e. proper lightening and background. As soon as the lightening decreases or/and the background includes more skin coloured objects, its performance decreases because it uses HSV colour space for the segmentation purpose.

Fariha Nasir et al. [4] worked on sign language recognition using Kinect sensor. The author proposes a model which generates a 3-D video stream of gestures using a sequence of static images along with Kinect sensor to provide depth. Kinect provides the data for 20 skeleton joints of hand out of which they used 6 joints which are used for the majority of the gestures performed. For sample test for words, the model provides the average accuracy detection of 86% with standard deviation of 8.89. The proposed model doesn’t recognize the gestures performed at finger level. Hence, it is difficult to differentiate alphabet.

Satya Prakash et al. [5] designed a system to detect hand gestures using vision-based technique. For the segmentation of hand, it uses HSV (Hue, Saturation, Value) colour space with values (315, 94, 37) for skin colour detection. With the help of palm mask, fingers and palm were segmented. Labelling algorithm is applied for the segmentation of fingers. Thumbs and other fingers are recognized with the help of palm lines and 4 horizontal divisions of coordinates. Gesture recognition on relevant dataset provides an accuracy of 99% but the result highly depends on hand detection. In case of different skin colours and moving objects, result of hand detection degrades and the overall performance of the system is decreased. Also, in the case of low light condition, the system fails to provide the above mentioned accuracy.

Song Yuheng et al. [6] analyzes four commonly used methods for image segmentation. The first method is Region based segmentation. It is of two types. The first one is Threshold segmentation which directly converts the image into different gray scales and process information based on gray value of different targets. The other one is Regional growth segmentation which uses the idea to have similar properties of pixels together to form a region. The second method is Edge detection segmentation. The edge of the object is in the form of discontinuous local features of an image. It uses discontinuities to detect the edge so as to achieve the purpose of image segmentation. The third segmentation method is based on clustering. It uses commonly used clustering algorithm (K-means) to transform the samples into different clusters according to the distance. The last method is based on weakly supervised learning in CNN. It uses Expectation-Maximization algorithm to identify marked and unmarked pixel followed by training images.

Parul Prashar et al. [7] implemented the image classification technique using SURF (speeded up robust features) descriptor and SVM (support vector machine) classifier. The image processing along with extracting SURF descriptors is done by converting each picture of the sequence into gray scale image. SURF awareness point detector is applied to each picture to become aware of points of interests in the picture. The training database consists of 50 images and the test database consists of 25 images. The proposed work provides an accuracy of around 98%.

Donghoon Kim et al. [8] presents a feature - based method to classify salient points as belonging to objects in the face or background classes using SURF descriptor and SVM classifier. For the segmentation of the skin areas in the image, it uses YCbCr (Luminance, Chrominance-blue, Chrominance-red) colour space and a set of experimentally defined thresholds. For detecting features, SURF descriptor is used. In high and low resolution, test data provides extremely good result for eyes and mouth. However, at the lowest resolution, detection results were not very good for all components (below 50%).

III. LANGUAGE SELECTION

The most popular sign languages includes: **ASL** – American Sign Language, **CSL** – Chinese Sign Language, **IPSL** – Indo-Pakistani Sign Language, **JSL** – Japanese Sign Language&**BSL** – Brazilian Sign Language.[9]

Among all these, ASL being one of the standard sign language doesn’t have much variation unlike other sign languages which varies from place to place and it uses single hand only for the gestures. Andalso, English is one of the most popular languages in the world hence, we have chosen ASL for the proposed system.

IV. PROPOSED MODEL

The proposed system converts ASL (hand gestures) to text cum speech with the help of four integrated modules namely- Segmentation, Feature extraction, Classification and Text to Speech as shown in the figure1.

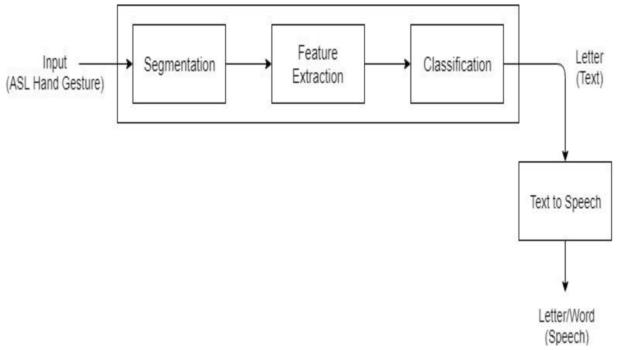


FIG 1: Block diagram of the proposed model

The main aim of the proposed model is to devise an algorithm which improves the efficiency of the system in the first two modules itself. The devised algorithm uses background subtraction using running averages for segmenting the foreground (i.e. hand) from background. Firstly, when the application is started, it captures the background for initial 30 frames and builds the background model. After it, when the hand is placed inside the region of interest (ROI) it uses the background model to segment (separate) the foreground from background. Then, the features of the foreground with the greatest area are extracted like contour, convex hull, defect points, COG etc. Using the extracted features, the hand gestures are classified.

Classified hand gestures are displayed in the form of text on the screen. This text is converted to speech using the fourth module.

The working of the modules is as follows:

A. Segmentation

The very first and most important step of any system which converts hand gestures into speech is Segmentation. It takes image (frames of video) as input, processes it and outputs the foreground (in our case it is "hand") by eliminating the background from the image (frame).

There are several techniques for segmenting the foreground form the background like RGB, HSV, and YCrCb. These are the most commonly used techniquesfor human skin colour detection and uses various colour components for the same. The efficiency of these techniques depends on various factors like lightning condition of the environment, human skin colour (varies w.r.t. geographical locations). It has been observed that the performance of these techniques decreases in low light condition and varying the skin colour. Hence, we have come up with a technique named "Background Subtraction using running averages".

The proposed technique builds the background model for initial 30 frames. Then the absolute difference between the current frame and the background model is calculated to obtain the foreground.

The video is captured in real-time using the web cam. Each frame is separated from the video sequence. The frame is resized and flipped to remove the mirror effect. After this, ROI is displayed on the video so that the hand can be placed in the ROI by the signer. Then the ROI is converted from BGR to GRAY scale. After it, Gaussian blur is applied on the output.

After above pre-processing, background model is build by capturing the background for initial 30 frames using Running Averages. After 30 frames, absolute difference of current frame and the background model is computed.

After obtaining the foreground, thresholding is applied to assign the white colour(1) to foreground and blackcolour(0) to background. After thresholding, smoothening and dilation is done on thresholded image so as to reduce noise.

Segmentation Algorithm:

1. Video Capture
2. Flip & Resize current Frame
3. Extract ROI
4. Convert BGR2GRAY
5. Apply Gaussian Blur
6. Build Background Model (initial 30 Frames) using Running Averages
7. absDiff (Current Frame, Background Model)
8. Thresholding
9. Smoothing
10. Dilation
11. Go to step 2

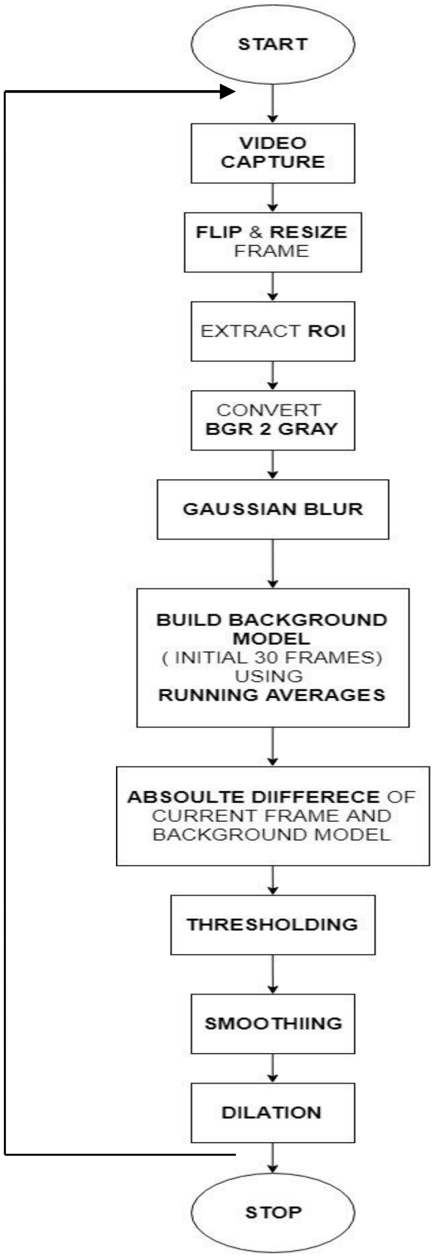

START
VIDEO CAPTURE
FLIP & RESIZE FRAME
EXTRACT ROI
CONVERT BGR 2 GRAY
GAUSSIAN BLUR
BUILD BACKGROUND MODEL (INITIAL 30 FRAMES) USING RUNNING AVERAGES
ABSOULTE DIIFFERECE OF CURRENT FRAME AND BACKGROUND MODEL
THRESHOLDING
SMOOTHIING
DILATION
STOP

FIG 2: Flow chart of the segmentation algorithm

B. Feature Extraction

After successful segmentation, the next step is to extract the features from the output of the previous module. These features will be used in the classification process. Hence, it is important to select the correct features to be extracted so as to get precise classification.

As far now, contour, convex hull, convexity defect points and COG are extracted. With the currently extracted features, we were able to count the number of opened fingers and distinguish between static and dynamic gestures.

- **Contour:** Contours can be explained simply as a curve joining all the continuous points (along the boundary), having same colour or intensity. The contours are a useful tool for shape analysis and object detection and recognition.

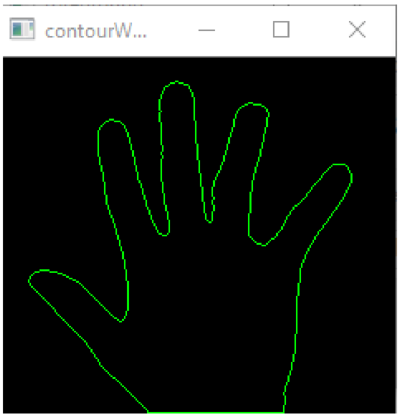


FIG 3: Contours of hand

- **Convex Hull:** A Convex object is one with no interior angles greater than 180 degrees. Hull means the exterior or the shape of the object. Therefore, the Convex Hull of a shape or a group of points is a tight fitting convex boundary around the points or the shape.

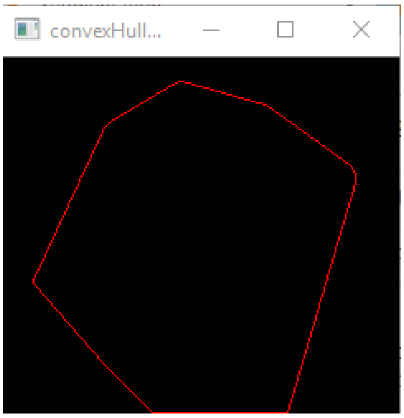


FIG 4: Convex Hull of hand

- **Convexity Defects:** Convexity defect is a cavity in an object (convex hull, contour) segmented out from an image. That means an area that does not belong to the object but located inside of its outer boundary - convex hull.

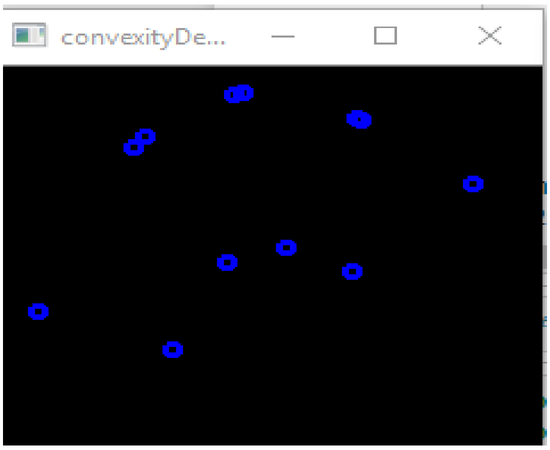


FIG 5: Convexity Defect Points of hand

- **COG:** The centre of gravity (CG) of an object is the point at which weight is evenly dispersed and all sides are in balance.

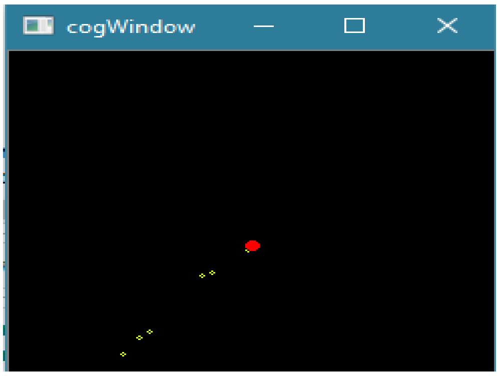


FIG 6: Latest 10 positions of COG

The last two modules C. Classification and D. Text to Speech can use any standard effective techniques and hence, the overall performance of the proposed model will be very high due to improved proposed algorithms for the initial two modules in comparison to other standard techniques like HSV, YCbCr, etc.

V. EXPERIMENTAL RESULT

Background Subtraction using Running Averages in the Segmentation module showed much better results in proper lighting as well as low lighting condition with static background.

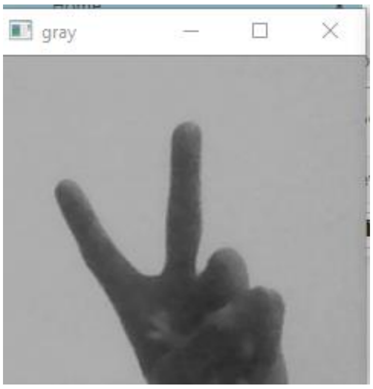


FIG 7: Current Frame in Gray Scale

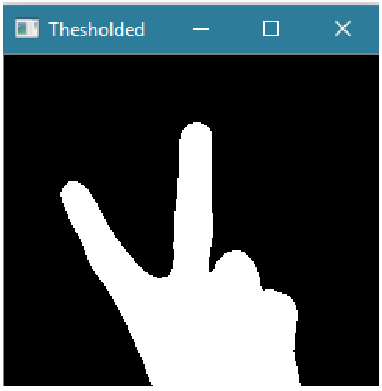


FIG 8: Thresholded image

The features extracted in the Feature Extraction module includes: Contour, Convex Hull, Convexity Defect Points and COG.

Retrieval Number: B3621129219/2019©BEIESP
DOI: 10.35940/ijeat.B3621.129219
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

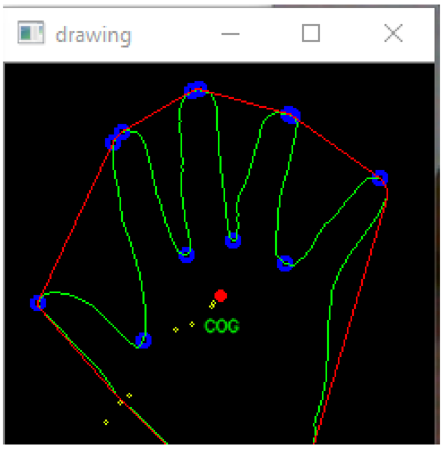


FIG 9: Extracted Features of hand

After successful extraction of all the features from the output of the Segmentation module, all those features are integrated and then used to count the number of fingers opened and to differentiate the gestures as “Static” or “Dynamic”.

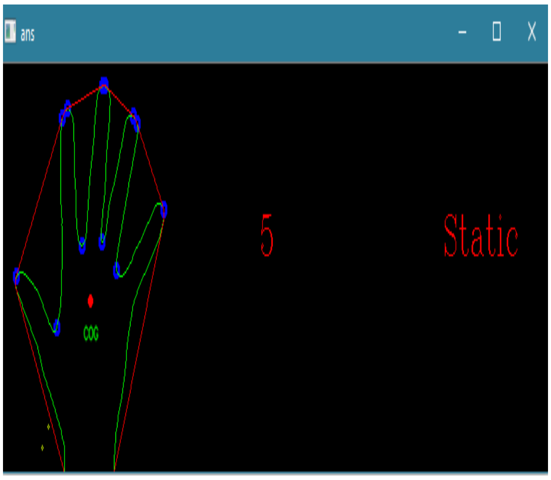


FIG 10: Counting the number of fingers and the type of gesture

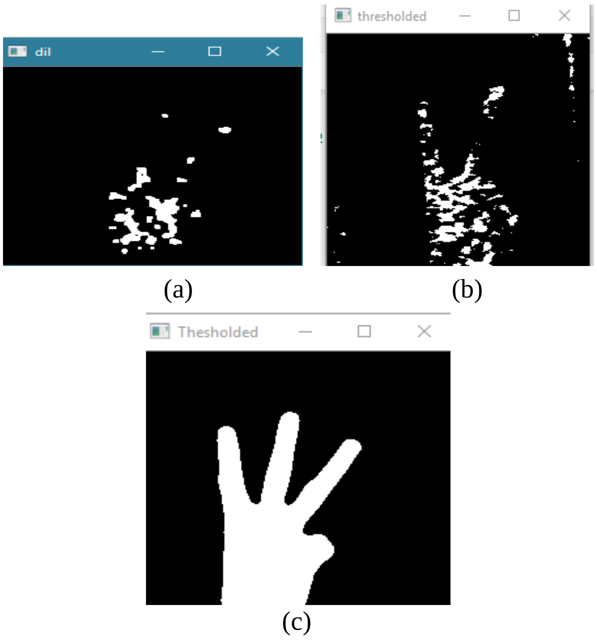
VI. COMPARISON WITH OTHER METHODS

The proposed model uses Background Subtraction using Running Averages for Segmentation module. Some other most commonly used techniques are HSV and YCrCb. All the three above mentioned techniques for Segmentation have been tested under low light and under sufficient light condition (using flash light). The results obtained are:

Under low light condition:

Table 5.1: Comparison of Various Techniques in Low Light

S. No.	Technique	Area of Contour (in pixels)
1.	HSV	3510
2.	YCrCb	11550
3.	Background Subtraction	17990

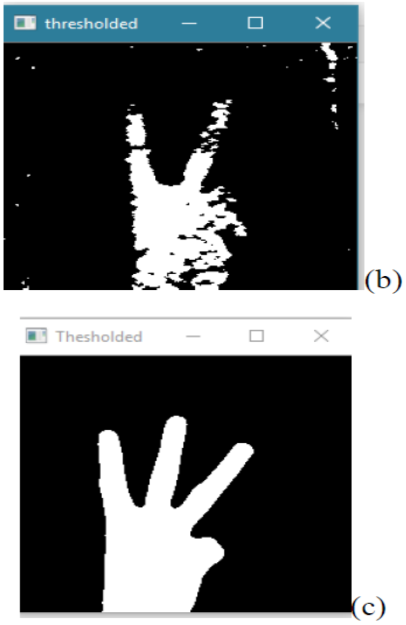


HSV (a) **YCrCb**(b)**Running Averages**(c)
FIG 11: Comparison in Low Light

Under Sufficient light condition:

Table 5.2: Comparison of Various Techniques in Sufficient Light

S. No.	Technique	Area of Contour (in pixels)
1.	HSV	8260
2.	YCrCb	17440
3.	Background Subtraction	18940



HSV(a)YCrCb(b)Running Averages (c)
FIG 12: Comparison of Various Techniques in Low Light

Graphical Representation:

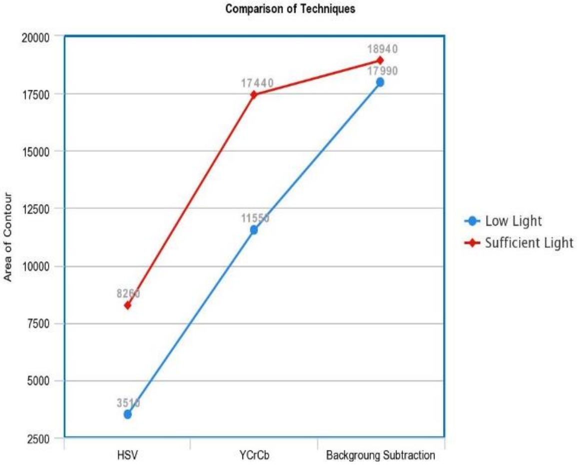


FIG 5.20: Comparison of Various Techniques

As, it can be clearly seen from the results of tested techniques that background subtraction using running averages gives much better results even in the normal room lighting in comparison to the other techniques i.e. HSV and YCbCr in the enough lighting of extra illumination of the flash light of mobile phone in addition to the room lighting.

VII. CONCLUSION

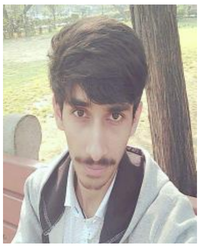
The system is novel approach to bridge the communication gap between the deaf-mute people and the rest of the society.The system includes the four modules: Segmentation, Feature Extraction, Classification and Text-to-Speech. The proposed system uses Background subtraction using running averages for Segmentation module and Contour, Convex hull, Convexity defect points and COG are extracted in the Feature Extraction module for classification. All the tests conducted in different lighting conditions (low light and sufficient light) conclude that the proposed method produces significant results in comparison to the other standard methods like HSV, YCbCr. These results further enhance the quality of the features obtained through the feature extractor module. The overall accuracy of the proposed system increases as the performance of the Segmentation module has been increased.

REFERENCES

1. http://wfdeaf.org/
2. Aarthi M, Vijayalakshmi P "SIGN LANGUAGE TO SPEECH CONVERSION" Fifth International Conference on Recent Trends in Information Technology, 2016
3. Anup Kumar, Karun Thankachan and Mevin M. Dominic "SIGN LANGUAGE RECOGNITION" 3rd InCI Conf. on Recent Advances in Information Technology I RAIT- 2016
4. Fariha Nasir, Umer Farooq, Zunaira Jamil, Maham Sana, Kashif Zafar "AUTOMATED SIGN LANGUAGE TO SPEECH INTERPRETER" 12th International Conference on Frontiers of Information Technology, 2016
5. Satya Prakash, Kapil kumar ahuja, Rahul thakur and Vamsi krishna pendyala "SIGN LANGUAGE TRANSLATOR FOR SPEECH-IMPAIRED", 2016
6. Song Yuheng, Yan Hao "IMAGE SEGMENTATION ALGORITHMS OVERVIEW" SiChuan University, SiChuan, ChengDu, 2017
7. Parul Prashar, Harish Kundra Rayat Institute of Engineering and IT Hybrid Approach for Image Classification using SVM Classifier and SURF Descriptor -, 2015
8. Donghoon Kim & Rozenn Dahyot Trinity College Dublin Face Components Detection using SURF Descriptors and SVMs -, Ireland, 2008
9. https://en.wikipedia.org/wiki/List_of_sign_languages_by_number_of_native_signers
10. P.K. Bora M.K. Bhuyan and D. Ghosh. Trajectory guided recognition of hand gestures having only global motions. International Science Index, 2008.
11. Emil M. Petriu Qing Chen, Nicolas D. Georganas. Feature extraction from 2d gesture trajectory in dynamic hand gesture recognition, 2006.
12. Thad Eugene Starner. Visual recognition of american sign language using hidden markov models. Master's thesis, Massachusetts Institute of Technology, Cambridge MA, 1995.
13. Michael Vorobyov. Shape class i_cation using zernike moments, 2011.
14. Nasser H. Dardas and Nicolas D. Georganas. Real-time handGesture detection and recognition using bag-of-features and support vector machine techniques. IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, 2011.
15. Syed muhammad saqlain shah, Husnain abbas naqvi, Javed i. khan,Muhammad ramzan, zulqarnain,Hikmat ullah khan "SHAPE BASED PAKISTAN SIGN LANGUAGE CATEGORIZATION USING STATISTICAL FEATURES AND SUPPORT VECTOR MACHINES" Digital Object Identifier 10.1109

AUTHORS PROFILE

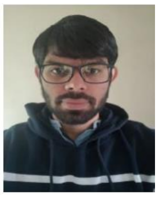
Jyoti Tripathi, is working in the capacity of Assistant Professor in the department of Computer Science and Engineering since December 2019 till date .She is having more than 16 years of teaching experience .She has taught more that 12 subjects .Many B.tech and M.tech projects were supervised.Her area of interest includes Natural Language Processing, Machine Learning, Big Data.She has participated in many international /national conferences, Workshops, FDPs, Swayam and NPTEL courses.



Raman Bhadauria, graduated from Govind Ballabh Pant Government Engineering College in June 2019 with a bachelor's of technology in Computer Science and Engineering. He has been a scholar throughout his academic career and has been awarded with Prime Minister's Scholarship during his B.Tech. He is currently working as a Systems Engineer Specialist in Infosys Limited, Chandigarh. He has cracked the all India hackathon of Infosys, HackWithInfy'18 by making it to the set of top 1009 participants and is currently part of the STG unit of Infosys. He has worked in the fields of android development, web development, computer vision and machine learning. He holds a keen interests in the area of machine learning and computer vision.

Prafull Goel, received his graduation degree in B.Tech Computer Science and Engineering from Govind Ballabh Pant Govt. Engineering College, New Delhi. Presently, he is working as a Systems Engineer Specialist in the department of Software Technology Group in Infosys Ltd. He has been highly active in the field of competitive programming and qualified for ACM-ICPC 2017 regionals level. He has also appeared for Infosys placement hackathon , HackwithInfy'18 and ended up in top 1009 participants. He has worked in the field of android application development, full stack development and holds keen interest in areas of Machine learning, computer vision and Deep Learning.

Keshav Gupta, is a highly motivated professional with a demonstrated history of working in computer software industry with a combination of excellent communication skills and a laser-like focus. He is capable of moving big ideas from design stage to implementation, currently working as a Application developer at Wake'n'Code Technologies Private Limited . Strong engineering professional with a Bachelor of technology focused in Computer Science and Engineering from Govind Ballabh Pant Government Engineering College .He holds a keen interest in the fields of machine learning and has a strong proficiency with javascript and knowledge of Node.js and frameworks available for it with a good understanding of server-side templating language.



Nikhil Yadav, is currently pursuing his masters in Computer Science and Information Security from NIT Warangal and received his graduation in B.Tech Computer Science and Engineering from Govind Ballabh Pant Govt. Engineering College, New Delhi. He has been a recipient of Prime Minister's Scholarship Scheme during his B.Tech. He has worked in the fields of Android Development, Web Development, Machine Learning. He has very keen interest in Machine and Deep learning, Cryptography, Web and Database Security, Digital Video Processing. He is always eager to learn new things.