A Brief Review of Lip Reading Recognition Techniques

Syed Sabeel Abbas Naqvi[1], Gaurav Shrivastav [2]
[1]Computer Science & Engineering, Institute of Technology & Management, Aligarh,
sabeelnaqvi@gmail.com
[2] Associate Professor, Computer Science & Engineering, Institute of Technology & Management, Aligarh

Abstract— **Lip detection has attracted a lot of notice lately in the computer vision group of people. This increasing interest stems from the wide range of application in which visual information is an integral part of, or can improve the performance and healthiness of the overall system. These applications include audio-visual speech acknowledgment, lip organization, artificial talking faces and facial expression analysis. yet, accurate and robust lip detection is a non-trivial task due to large variations caused by the high deformable level of lips, different lip color tone, light conditions, appearance of teeth and tongue, presence of facial hair.**

Keywords- 3D Discrete Wavelet Transformation (3D-DWT), Locality Sensitive Discriminate Analysis (LSDA), And Indian Sign Language (henceforth called ISL)

I. INTRODUCTION

Discourse is the most characteristic method for the correspondence among individuals. Lamentably, the moronic and hearing hindered individuals can't utilize this regular type of correspondence. The technique for word ID displayed to help hearing weakened or stupid individuals to speak with the others actually. It is a visual technique for discourse handling in which just lip developments are utilized to recognize the talked word [1].

1.1 Lip identification has pulled in a considerable measure of consideration recently in the PC vision group. This expanding interest comes from the extensive variety of use in which visual data is a necessary piece of, or can enhance the execution and power of the general framework. These applications incorporate varying media discourse acknowledgment, lip synchronization, engineered talking appearances and outward appearance investigation. All things considered, precise and strong lip identification is a non-minor undertaking because of substantial varieties brought on by the high deformable level of lips, diverse lip shading tone, light conditions, appearance of teeth and tongue, nearness of facial hair.

1.2 Programmed Discourse Acknowledgment is a vital segment without bounds Human-PC Interface; it is intended with the end goal of acknowledging recognizable proof acknowledgment and normal dialect cognizance by methods for human voice. Discourse Acknowledgment innovation has gained huge accomplishments with some effective prevalence and applications. IBM's Via Voice framework, for example, has great exhibitions when the vocabulary pool is little and when the clamor is low. Be that as it may, its execution will be extraordinarily corrupted when utilized as a part of genuine application situations. It is harder to develop a lip-perusing framework than to develop a conventional discourse acknowledgment framework. What we for the most part concentrate on its issues identified with the visual acknowledgment, for example, the accompanying viewpoints: a. Acknowledgment techniques; b. Combination [2].

1.3 Visual data is a vital piece of, or can enhance the execution and power of the general framework. These applications incorporate varying media discourse acknowledgment, lip synchronization, engineered talking countenances and outward appearance investigation [6]. Lip perusing turns out to be more mainstream research, since it has broad applications. The programmed discourse acknowledgment (ASR) framework is generally utilized as a part of mechanical autonomy, PC, and cellphone [7].
Programmed lip perusing is a dynamic research zone. Lip perusing is a discourse perusing process from visual data of lip and territories around the mouth. Programmed acoustic discourse acknowledgment frameworks have a tendency to perform ineffectively in loud condition. Sound data is influenced by acoustic commotion and crosstalk among speakers. While visual methodology gives data which is vigorous and along these lines there

have been various reviews going ahead to evaluate and enhanced the execution of visual discourse acknowledgment [3].
The most vital thing to concentrate lip's data is lip division, which is the partition amongst lip and skin. Numerous lip division calculation have been proposed in the previous few year, the most prevalent is shading change based lip division. Shading change based lip division is comprise of picking the shading space and shading change itself. This current calculation's motivation is to extend the contrast between the lip and non-lip zone [8].

High dimensional list of capabilities can adversely influence the execution of example or picture acknowledgment frameworks. As such, excessively numerous components in some cases decrease the characterization precision of the acknowledgment framework since a portion of the elements might be repetitive and non-instructive [9].

Lip form contains a lot of data. Lip extraction procedure has awesome potential for human-PC connection, and it has gotten much consideration lately. Its application includes in science and social exercises, for example, varying media discourse acknowledgment, speaker recognizable proof, showing hard of hearing youngsters articulation et cetera. Additionally, discourse acknowledgment with visual components can incredibly enhance the exactness in loud condition. Yet, it's confronted with some troublesome issues because of high fluctuation of lip shading and shape.

Another plan going for speed and exactness is proposed to extricate external lip shape in shading pictures. RGB shading space is chosen to section lip from skin. R part is more unmistakable than different segments in the lip district. Another shading change is proposed to highlight the diverse between the lip and the non-lip locales [4].

PC lip-perusing framework is separated into four section. They are video picture obtaining module, lip position module, highlight extraction module and lip acknowledgment module. Video picture securing module is to get a succession of video edges by parsing the video that is caught with the camera. Lip situating module is to find the lip district in the picture grouping. Lip include extraction module is to extricate the lip highlights. Lip distinguishing proof module is to perceive the substance of the discourse through the lip highlights in view of the relating lip includes acknowledgment calculation [5].

1.4 Gesture based communication acknowledgment can be partitioned into disengaged sign acknowledgment and nonstop gesture based communication acknowledgment. Disconnected sign acknowledgment is an extraordinary instance of motion acknowledgment. Signal acknowledgment frameworks are regularly intended to perceive counterfeit motions. The client needs to take in these signals so as to speak with the framework. So as to have a thought regarding the best approach expected to manufacture a communication through signing acknowledgment framework it gets to be distinctly pivotal to study the effectively created frameworks [10]. The main gathering is concentrating on location of lip edges in the mouth area of intrigue (return for money invested). They apply dynamic shape models to the shading changed return on initial capital investment or utilize deformable Layouts [11].

II. LITERATURE REVIEW

2.1 Standard Deviation: An enhanced word division calculation is proposed, which is utilized to discover the segment that speaks to the elocution part from the video outline grouping in the PC lip-perusing framework. In the PC lip-perusing framework, subsequent to extricating the lip includes, the standard deviation of the lip elements is figured. At that point the paper proposes the word division technique in light of the standard deviation and the outcomes demonstrate that this strategy must be utilized to decide the inexact articulation area and that can't precisely decide the area of the start and the consummation of the elocution position. So an enhanced technique is displayed that the area of the start and the completion of the articulation segment is resolved by the circulation of the slope of the standard deviation. The outcomes demonstrate that the enhanced technique can enormously enhances the precision of situating of the start and the completion of the elocution area.

2.2 3D Discrete Wavelet Transformation (3D-DWT) or 3D Discrete Cosine Transform (3D-DCT) or Locality Sensitive Discriminate Analysis (LSDA): Human uses visual data while attempting to comprehend discourse, particularly in boisterous conditions or in circumstances where the sound flag is not accessible. Lip perusing is the procedure of a far reaching understanding the basic discourse by handling on the development of lips. The acknowledgment of lip movement is a troublesome errand since the area of intrigue (return for money invested)

is nonlinear and loud. In proposed technique lip perusing framework we have utilized two phase highlight extraction display which is précised, discriminative and calculation productive. The main stage 3D Discrete Wavelet Change (3D-DWT) or 3D Discrete Cosine Change (3D-DCT) is utilized and the second stage is Region Delicate Discriminate Examination (LSDA) to trim down the element measurements. These elements make a novel lip perusing framework with little element vector measure. Notwithstanding the novel element extraction strategy, the execution of Gullible Bayes and SVM classifier is thought about. CUAVE database of 0 to 9 articulations in English is utilized for experimentation. Consequences of 3 measurement change with LSDA are contrasted and 2 measurement change with LSDA. Test comes about demonstrate that 3D-DWT+LSDA include mining are contrasted and 3D-DWT with PCA or LDA. 3D-DWT+LSDA result is likewise contrasted and 3D-DCT + LSDA.

2.3 Indian Sign Language (henceforth called ISL): Gesture based communication acknowledgment has pulled in much consideration in PC vision. A gesture based communication is a method for passing on the message by utilizing hand, arm, body, and face to pass on musings and implications. Like talked dialects, communications via gestures develop and advance normally inside hearing-hindered groups. Be that as it may, communications via gestures are not all inclusive. There is no globally perceived and institutionalized gesture based communication for all hard of hearing individuals. Similar to the case in talked dialect, each nation has got its own gesture based communication with high level of syntactic varieties. The communication via gestures utilized as a part of India is generally known as Indian Gesture based communication (from this time forward called ISL).

2.4 rg-color histogram: The division of mouth and lips is a crucial issue in facial picture examination. Propose a technique for lip division in light of rg-shading histogram. Measurable examination appears, utilizing the rg-shading space is ideal for this motivation behind an unadulterated shading based division. At first an unpleasant versatile edge chooses a histogram district, which guarantees that all pixels in that locale are skin pixels. In view of that pixels construct a Gaussian model which speaks to the skin pixels appropriation and is used to get a refined, ideal edge. They are not consolidating shape or edge data. In trials they demonstrate the execution of our lip pixel division strategy contrasted with the ground truth of our dataset and a customary watershed calculation.

2.5 Novel color transformation in RGB space instead of complex color models: Different techniques for lip division have been proposed, despite everything it remains a testing and troublesome issue because of high changeability of lip shading and low chromatic difference between the lip and skin. A novel programmed lip division calculation is proposed in light of a novel shading change in RGB space rather than complex shading models. Near review with some current lip division calculations has shown the prevalent execution of the created calculation. The proposed calculation delivers better division quick and effectively.

The following table contains all the techniques used for Lip Recognition

Table 2.1 Lip Recognition techniques

Technique	**Author**	**Year**	**Based on**	**Merits**
Standard Deviation	Xinjun Ma, Xiaohui Jios, Hongjun Zhang*	2016	Standard Deviation	Accuracy of locating of the beginning and the ending of the pronunciation section
3D Discrete Wavelet Transformation(3d-DWT) or 3D Discrete Cosine Transform (3D-DCT) or Locality Sensitive Discriminate Analysis(LSDA)	Sunil S. Morade, Suprava Patnaik	2016	LSDA with DCT or DWT	DWT +LSDA performance is marginally better than DCT +LSDA. The results of LDA are better as compared to PCA with DCT or DWT.
Indian Sign	Shweta Dour,	2015	Use of Depth data	Internationally recognized and

Language (henceforth called ISL)	M.M. Sharma		and Kinect Sensors for the obtaining the features from the hands signs in the videos	standardized sign language for all deaf people
rg-color histogram	Axel Panning, Robert Niese, Ayoub Al-Hamadi and Bernd Michaelis	2009	Lip segmentation based on rg-color histogram	Performs better than conventional Watershed based methods since it can basically adapt better to the various ratios of mouth to skin in the mouth ROI.
Novel color transformation in RGB space instead of complex color models	Jing Pan a, Yepeng Guan a;b;¤, Shuangcheng Wang a	2012	Novel color transformation in RGB space instead of complex color models	Algorithm produces Better segmentation fast and anciently.

III. CONCLUSION

In this paper the various concepts of lip reading, recognition techniques and their merits have been analyzed. These techniques are Standard Deviation, 3D Discrete Wavelet Transformation (3d-DWT) or 3D Discrete Cosine Transform (3D-DCT) or Locality Sensitive Discriminant Analysis (LSDA), Indian Sign Language (henceforth called ISL), rg-color histogram, novel color transformation in RGB space instead of complex color models. Accuracy of locating of the beginning and the ending of the pronunciation section, DWT +LSDA performance is marginally better than DCT +LSDA. The results of LDA are better as compared to PCA with DCT or DWT. Internationally recognized and standardized sign language for all deaf people, performs better than square divide based methods since it can basically adapt better to the various ratio of mouth to skin in the mouth ROI and the algorithm produces better segmentation fast and anciently.

REFERENCES

[1]. Neeru Rathee,"A novel approach for lip Reading based on neural network", IEEE, 2016, pp.1-6.
[2]. Zeliang Zhang and Wenliang Qu and Fumei Liu,"Review of the Lip-reading Recognition", IEEE, 2014, pp.593-596.
[3]. Sunil S. Morade and Suprava Patnaik,"Visual Lip Reading using 3D-DCT and 3D-DWT and LSDA", IEEE, 2016, pp.7-15.
[4]. Jing Pan, Yepeng Guan, Shuangcheng Wang,"A New Color Transformation Based Fast Outer Lip Contour Extraction", IEEE, 2012, pp.2505-2514.
[5]. Xinjun Ma, Xiaohui jiao and Hongjun Zhang,"An Improved Word Segmentation Algorithms for Lip-reading", IEEE, 2016, pp.3099-3104.
[6]. Evangelos Skodras and Nikolaos Fakotakis,"An Unconstrained Method for Lip Detection in Color Images", IEEE, 2011, pp.1013-1017.
[7]. Aris Nasuha,Tri Arief Sardjono, Mauridhi Hery Purnomo, "Lip Segmentation and Tracking Based on Chan-Vese Model", IEEE, 2013, pp.1-4.
[8]. Hadid Tunas Bangsawan and Ronny Mardiyanto, Tri Arief Sardjono, "Six Key Points Lip's Feature Extraction Using Adaptive Threshold Segmentation", IEEE, 2015, pp.261-265.
[9]. Sunil Sudam Morade, Suprava Patnaik, "A Genetic Algorithm-Based 3D Feature Selection for Lip Reading", IEEE, 2015, pp.1-6.
[10]. Shweta Dour, M.M. Sharma,"Review of Literature for the Development of Indian Sign Language Recognition System", IEEE, 2015, pp.27-34.
[11]. Axel panning, Robert Niese, Ayoub Al-Hamadi and Bernd Michaelis,"A New Adaptive Approach for Histogram based Mouth Segmentation", IEEE, 2009, pp.1564-1569.