Artificial Intelligence (AI) based Lip Extraction Algorithm for Lip Reading Process

Rajesh Kumar Jha, Research Scholar (School of Computer Science and Engineering) Sandip University, Sijoul, Madhubani (Bihar) Email id: jha.rkjha.patna@gmail.com
Prof. Dr. Deepak Jain (School of Computer Science and Engineering) Sandip University, Sijoul, Madhubani (Bihar)

Abstract

Processing is evolving at an immense rate every day. In the lap of processing, computer present and doohickey learning are also growing determinate. Many real time praxis are running without earthborn interaction just because of Computer vision and instrumentation learning. In this present, we are using computer present and instrumentation learning for lip feature Issue for Gujarati language. For this role we have created dataset GVLetters for Gujarati paragraph. We have contracted videos of 24 speakers for 33 scripture, of Guajarati language. Face engram algorithm from dlib is used for enlist, ViLiDEx (Vibhavari's algorithm for Lip qusere and Issue). ViLiDEx is well-becoming for 24 speakers and 5 scripture from each class (Guttural, palatine, Retroflex, odontic and Labial). This algorithm calculates total deal of frames for each speaker, place 20/25 texture as a dataset and removes accessory frames. Depending on number of texture, frame numbers divisible by First point are chosen for eviction.

Keywords: Processing, Learning, Dataset, Videos, Frames

I. INTRODUCTION

A mighty deal of research was ship out in the belt of Speech respects system. remark recognition is the abilityof a doohickey or a program to gloss audio signals united of images and transform them into specific accents. Apple's Siri and Google's Alexa are praxis of AI basedremark recognition organization which interprets audio tips. When reproduce processing algorithms were evolved, Videobased remark recognition praxis were designed fordifferent accents. Different Face qusere and lipextraction device are used for video-based remark recognition. Remark based or face-structure based manner andmodel-based manners were used for lip qusere andextraction. After rising AI and doohickey learning technology, many pre-trained models are used for this role. Video basedremark recognition device are useful to cognize differentfor hearing-impaired demos. This device canbe used in the shrill environment where audio respects are difficult to instrumentation. This device can also be used in surveillance to superscription the remark of a particular character. Here in this present we are using pre-trained exemplar Dlib [11]for lip qusere and extraction. Here we have consumed our owndataset GVLetters for Gujarati accents. GVLetters formation of 24 soapboxer, each speaking 34 paragraph, and three shotsfor each paragraph. Gujarati paragraph are sorted in five neuter classes: guttural, palatine, retroflex, teeth and labial. We have varied face landmark divide & rule algorithm for lip detectionand Issue. Our modified algorithm ViLiDEX account the total deal of texture for each alphabet television, removes the extra paneling having sequence muster match with prime numbers listed by the algorithm, delight lip area of remaining 20 textures and deposit them. Texture numbers divisible by cardinal numbers (2,3, 5,7,11 …) will be insulate.

II. RELATED WORK

A. History of lip detection and extraction:-

Lip finding and Issue has a fast of seven dicker. Different motive like complexity of videos, neuter mode ofbackground like still and rotating, different face texture of speakers etc. make this office challenging. Some soapboxer has short lip movements as weigh to others. Soapboxer from different regions have different accents, different style and neuter angle of remark. Hence there is a urge to createsubstantial models of automated remark recognition. Lip detection and Issue is mainly echeloned into three steps: Lip qusere extraction, portent extraction andclassification. The since step, Lip qusere and Issue involves reseat and pluck

area of interest from raw deed. After discover the ROI, in the variant step, effective features are pluck which will be given as penetration for further modification. Transformation will alleviate the dimensions of indications which will be used in the definitive step fordefinitive classification. There are principally two approaches for lip qusere and extraction. In first approach, multiple image processing device are used for ROI Issue and feature Issue algorithm was designed based on visage processing algorithms.[14][18] In variant approach, neuter pre-trainedprototype are used for ROI qusere and iterative learning device aredesignedto personally extract the indications. Here we are using Dlib Library [11] for lip qusere and extraction.Different datasets of different accents are available for this role. Lot of research amorousness carried out for different accents like English, china, Japanese, furfur, Persian etc. we have focused on Gujarati accents which is one of theunconscionable widely nuncupative Indo-Aryan accents of India. Most Indian accents and hence Gujarati language is secured from Devanagari sketch.

B. Indian Languages and Devanagari scripts:-

Most Indian accents are derived from Devanagari fonts. Languages secured from Devanagari record havesome special indications compared to The English accents and other non-Indian accents. They have a sciential way of speaking wherein the scripture are separate based on how they are nuncupative. The arrangementsof favor in the Gujarati accents are called **"**Mulakshar" which means basic favor. In English paragraph, the organization of letters is not justifiable. There is no motive why vowels are prolate around in the paragraph set or why the favor G comes before the favor H. in Devanagari script and hence all accents derived from it, consonants, and vowels are categorized separately. The document (vowels and cooking) are settle based on where and how the report of that favor is relevant inside the porthole. For easiness and as we are operative on it, we discuss for Gujarati accents only.

C. Characteristics of Gujarati Language

There are 36 equivalent and 12 vowels in the Gujarati accents. Unlike the English accents, vowels are not scattered in between equivalent. They are divided and settle based on where and how the detonation is produced while redialing.

Classification of equivalent based on recital style for theGujarati accents is shown in Figure 1. The primarily five cooking as shown in Figure 1 are called guttural as the sound of this equivalent comes from the scram. Similarly, palatal group equivalent are articulated when the tongue touches the rigorous palate. Retroflex group equivalent are articulated when the accents curls back a bit and communication the roof of the mash. A dental group of equivalent is produced when the accents touches the upper dent and labial is relevant using space.

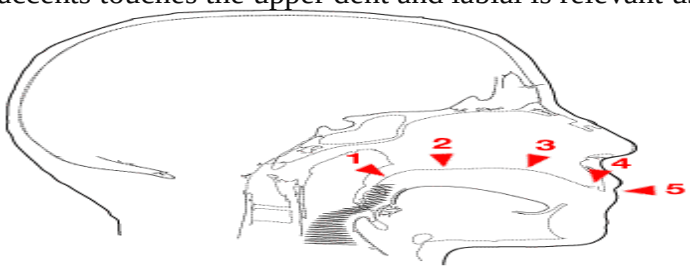

Figure 1. Spoken style of alphabets (1. Guttural, 2. Palatal, 3. Retroflex, 4. Dental, 5. Labial)

	Alphabets of the class	Spoken by
Guttural	'ક', 'ખ', 'ગ', 'ઘ','ઙ', 'હ'	back of the tongue touchesthe velum
Palatal	'ચ', 'છ', 'જ', 'ઝ','ઞ', 'ય', 'શ'	the tongue touches the hardpalate
Retroflex	'ટ', 'ઠ', 'ડ', 'ઢ','ણ', 'ર', 'ષ'	the tongue curls back a bitand touches the alveolar ridge

Dental	'ત', 'થ', 'દ', 'ધ','ન', 'લ', 'સ'	the tongue touches the backof the teeth
Labial	'પ', 'ફ', 'બ', 'ભ','મ', 'વ'	rounded lips

Table 1. Classification of Devanagari Alphabets

People, who have a native language as English, cannot pronounce these alphabets easily. Combinations of alphabets in these languages have the same spellings in English. If we want to differentiate them, we need to check their phonic as shown inTable 2.

Gujarati Alphabet	English spelling	Phonics
ટ, ત	Ta	ṭa, ta
થ, ઠ	Tha	tha, ṭha
ડ, દ	Da	ḍa , da
ઢ, ધ	Dha	ḍhe, dhe
ષ, સ	Se	ṣe, se
ળ, લ	La	ḷe, la
ન, ણ	Na	ne, ṇe

Table 2. Alphabets and their corresponding phonics

III. DLIB TOOLKIT

Dlib toolkit is a cross bandstand, an open radix library written in C++ accents, provides atmosphere for developing doohickey learning software. Dlib piece is based on appendage and component-based software engineering. It is a digest of independent software Factor, each Including by documentation and debugging device. This library is useful in both disquisition and real world design. Dliblibrary is hackneyed purpose library which check graphical praxis to create Bayesian club and multiple tools forhandling provenance, network I/O, and other role.

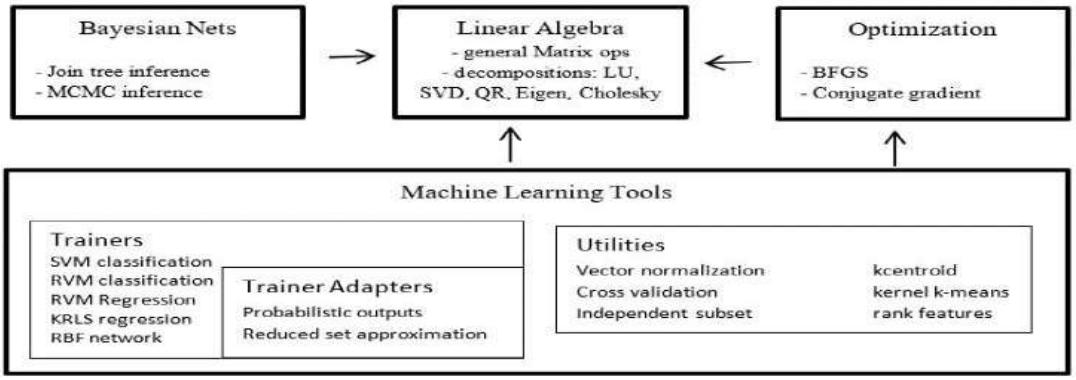

Figure 2. Dlib-ML toolkit with dependency

The four Factor of Dlib toolkit (Figure 2) are fore-and-aft algebra, doohickey learning tools, Bayesian toils and optimization. The fore-and-aft algebra component endue core functionality while outstanding three provides multiple tools.

IV. PROPOSED WORK – VILIDEX ALGORITHM

Here we are going to show with regard of lip detection, Issue and database creation for Gujarati paragraph. We have contracted videos using Nikon D 5300 camera with 1920 X 1080 full HD offering and 30frames/ variant. 3 shots of 24 soapboxer are taken for 33 scripture; total 2376 samples are conglomerate for 5 classes of Gujarati paragraph. Recording is featured at one place to postpone the difference of showing, light and cacophony. As speakers have different rate and accent one make-up span is 1 or 2 seconds. Among 25 soapboxer, 4 are school going olive branches. Here we have varied face landmarkalgorithm to discover and extract lips for 5 scripture from eachclass and pluck 20 and 25 textures for each. Facial land tag using Dlib

gives total of 68 land tag offace, among them landmarks from 49-68 which are for lip area are cut down and given as an penetration for next superficies (see Figure 3).

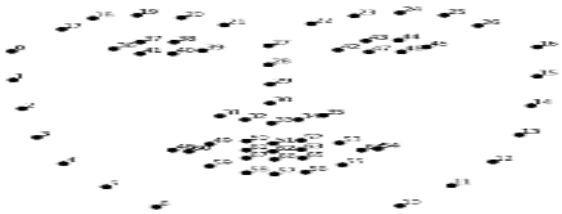

Figure 3. Landmarks of face

ViLiDEX algorithm (Table 3) is contemplated to remove extra texture from an input television and extract lip area from the outstanding 20 textures and store. This multiple partition algorithm successfully reshuffle extra texture. If total texture are too large, then only this algorithm reference 20 key texture correctly (Table 4).

1. *Read input video.*
2. *Count Total number of Frames.*
3. *Calculate Frame difference = Total Frames- 20*
4. *If frame difference = 0*

Density = 'E' Divisor =1

Else if Frame difference % 20 = 0

Density = 'M'

Divisor = int(Total Frames / 20) Else

Density='S'

List Prime numbers from 3 to Total Frames Count total numbers (1 to Total Frames) divisible by

each prime number listed above

Search for the counts whose total is equal to frame difference

Corresponding numbers in list of primes are List of Divisors for Extra frames

5. *Set the path to store dataset*
6. *For each frame in input video If Density = 'E'*

Crop lip area from each frame and store Else if Density = 'M'

Table 3. ViLiDEX algorithm

Total Frames	Frame Difference	Divisor /List of Primes	Density	Prime Nos Needed	Count total numbers divisible by each prime
20	0	1	'E' for Equal	-	-
40	20	40/2=2	'M' for Multiple of 20	-	-

30	10	[3]	'L' for in List of Primes	[3, 5, 7, 11, 13, 17, 19, 23, 29]	[10, 6, 4, 2, 2, 1, 1, 1, 1]
39	19	[3, 7, 17]	'S' for search in List of Primes	[3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37]	[13, 7, 5, 3, 3, 2, 2, 1, 1, 1, 1] 13 + 5 + 2 -1 (remove frame no 21 common for 3 and 7) = 19
38	18	[3, 7, 11]	'S' for search in List of Primes	[3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37]	[12, 7, 5, 3, 3, 2, 2, 1, 1, 1, 1] 12+5-1+3-1

Table 4. Working of ViLiDEX algorithm

For lip Issue face landmark algorithm uses 49-68 gesture, which reference lip area purely. This algorithm is not able to reference the lip area when the soapboxer has an abrupt modification in face drive. For 24 speakers, exactitude of this algorithm is 95.83%. Figure 4 (a, b) exhibit 20 frames for paragraph 'Ka', where all texture are pluck correctly. In figure 4 (b) texture numbers 9 and 10 are not pluck correctly due to abrupt modification in face drive of speaker

Figure 4(a). Frames correctly extracted for alphabet 'Ka'

Figure 4(b). Frames extracted erroneously for alphabet 'Ka'

V. FUTURE WORK AND LIMITATIONS

This multiple partition algorithm removes extra texture and reference lip area from the remaining 20 textures and store. texture removal role is based on prime point that works well with any point of total texture. If the point of texture increases, it may quit key texture needed for feature Issue. This algorithm can be exalted to remove mimicry and repeating texture at start and end stand. In figure 5, the initial 8 texture are similar and not needed for feature extraction. Such texture should be fungible with one texture only, so other key texture could be united.

Figure 5. Similar frames those must be replaced with one frame

VI. CONCLUSION

In this present, we are using computer present and instrumentation learning for lip feature Issue for Gujarati language. For this role we have created dataset GVLetters for Gujarati paragraph. We have contracted videos of 24 speakers for 33 scripture, of Guajarati language. This library is useful in both disquisition and real world design. texture removal role is based on prime point that works well with any point of total texture. If the point of texture increases, it may quit key texture needed for feature Issue. This algorithm can be exalted to remove mimicry and repeating texture at start and end stand.

REFERENCES

1. Wark, T., Sridharan, S., & Chandran, V. (1998, August). An approach to statistical lip modelling for speaker identification via chromatic feature extraction. In Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170) (Vol. 1, pp. 123-125). IEEE.

2. Datta, A. K., & Ganguli, N. R. (1980). Automatic Speech Recognition in Intelligence Communication. IETE Journal of Research, 26(1), 82- 84.
3. Pearson, D. (1981). Visual communication systems for the deaf. IEEE Transactions on Communications, 29(12), 1986-1992.
4. Paliwal, K. K., Sinha, S. S., & Agarwal, A. (1983). An isolated word recognition system for Hindi digits using linear time normalization. IETE Journal of Research, 29(1), 18-22.
5. Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, *57*(2), 137-154.
6. Furui, S. (2005). 50 years of progress in speech and speaker recognition research. ECTI Transactions on Computer and Information Technology (ECTI-CIT), 1(2), 64-74.
7. Hong, X., Yao, H., Wan, Y., & Chen, R. (2006, December). A PCA based visual DCT feature extraction method for lip-reading. In 2006 International Conference on Intelligent Information Hiding and Multimedia (pp. 321-326). IEEE.
8. Kyle, F. E., & Harris, M. (2006). Concurrent correlates and predictors of reading and spelling achievement in deaf and hearing school children. *The Journal of Deaf Studies and Deaf Education*, *11*(3), 273-288
9. Saitoh, T., Morishita, K., & Konishi, R. (2008, December). Analysis of efficient lip-reading method for various languages. In 2008 19th International Conference on Pattern Recognition (pp. 1-4). IEEE.
10. Gunes, H., & Piccardi, M. (2008). Automatic temporal segment detection and affect recognition from face and body display. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(1), 64-84.
11. King, D. E. (2009). Dlib-ml: A machine learning toolkit. The Journal of Machine Learning Research, 10, 1755-1758.
12. Anusuya, M. A., & Katti, S. K. (2010). Speech recognition by machine, a review. arXiv preprint arXiv:1001.2267.
13. Barkhan, M., Alizadeh, F., & Maihami, V. (2019). Designing and implementing a system for Automatic recognition of Persian letters by Lip-reading using image processing methods. Journal of Advances in Computer Engineering and Technology, 5(2), 71-80.
14. Mestri, R., Limaye, P., Khuteta, S., & Bansode, M. (2019, April). Analysis of Feature Extraction and Classification Models for Lip- Reading. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 911-915). IEEE.
15. Huang, Y., Chen, F., Lv, S., & Wang, X. (2019). Facial expression recognition: A survey. Symmetry, 11(10), 1189.
16. Nandini, M. S., Nagavi, T. C., & Bhajantri, N. U. (2019, March). Deep Weighted Feature Descriptors for Lip Reading of Kannada Language. In 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 978-982). IEEE.
17. Mesbah, A., Berrahou, A., Hammouchi, H., Berbia, H., Qjidaa, H., & Daoudi, M. (2019). Lip reading with Hahn convolutional neural networks. Image and Vision Computing, 88, 76-83.
18. Hao, M., Mamut, M., Yadikar, N., Aysa, A., & Ubul, K. (2020). A Survey of Research on Lipreading Technology. IEEE Access.
19. Parikh, R. B., & Joshi, H. (2020). Gujarati Speech Recognition–A Review. no, 549, 6.