Lip Segmentation based on Automated Edge Detection based on P-SVM

Kharak Singh[1], Ramanjot Kaur[2]
M.Tech (Scholar), Assistant Professor
Department of Computer Science Engineering/IT, DIET , Kharar, Punjab

Abstract- Image Segmentation is the process where there is the segmentation of the image in different sets of the regions. The main objective of the different regions is the organisation of the regions such as crops, forests and urban areas. With the advancement in the artificial intelligence and increasing usage of the technical devices, human-computer communication technology became a common technology .Hence, there are different methods used for the human –computer interaction and languages are used by humans to interact with the machines as an efficient method. Lip segmentation plays a significant role in a photographic lip reading classification because the segmentation establishes an accurate value. In the existing work, lip region is segmented from face image using Markov random field (MAP-MRF) framework for the recognition local and texture by image segmentation. In traditional MRF method, the whole image is processed with constant scale and input image may have different scale images. In proposed work, Digital images are used to recognise the quality of the image. The various methods used to determine the quality of the images are machine learning, classification and feature extraction method. The image data set of the lip images is observed and then the region of the noise free image is detected using segmentation method. In addition, the Principal Component Analysis algorithm used for image as feature extraction method for the extraction of the features from the digital images. After that, the classification of images is done in training and testing section based on support vector method .The experimental analysis is done to compute the parameters which are overlap error, segmented error and accuracy. The different parameters are compared and errors are reduced as well as accuracy is increased in the proposed approach.

Keywords- Lip Segmentation, Feature Extraction Process, Training section and Testing Section.

I. INTRODUCTION

The image segmentation referred as the partitioning an image into set of regions. The primary goal in several tasks of region is to represent meaningful region of image like crops, urban areas, and forests of a satellite image. In different analysis tasks, the areas may be sets of border pixels assembled into structures such as circular arc segments and line segments in 3D images of industrial objects. Regions can be characterized as gathering of pixels having both a border and specific shape like circle or polygon or ellipse[3]. At a point the intriguing regions do not cover the entire picture, presently discussion topic is about segmentation, to foreground regions of interest and foundation regions to be overlooked [1].

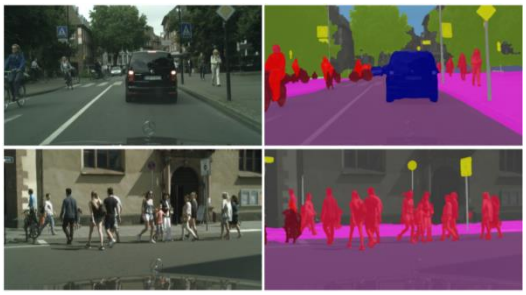


Fig.1: Image Segmentation of Objects [2]

The visual information has aroused the interest of many researchers. The visual information of lip movement can help enhancing the robustness of automatic speech recognition systems especially in noisy environments [4]. Useful geometric information about lip movement, such as the temporal variation of mouth width and height, can be obtained easily from a segmented lip. However, accurate lip segmentation has proved to be difficult due to the weak color contrast and the significant overlap in color features between the lip and the face regions. Color lip segmentation is treated as a two-class clustering and segmentation problem. Clustering-based method allows the segmentation to be based on color difference between lip and non-lip regions, without assuming a particular hue distribution for the lip. The original lip images are in the RGB color format. It is desirable to work in a uniform color space, where the distance between two points in the color space is directly proportional to the perceived color difference. The lip image is first transformed into the color spaces. A feature vector is constructed for each pixel in the image. Then, the intensity non uniformity (For instance smooth variation in intensity value) in the component due to uneven illumination is reduced by the following procedure.

i. Estimate the intensity non uniformity along the column direction on the upper and lower border of the image using a small window.
ii. Compute the mean value by averaging.
iii. Modify the luminance value for each pixel along each column.

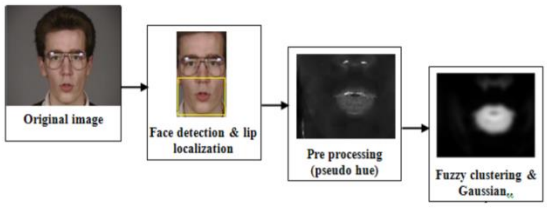


Fig.2: Process of Lip Segmentation [5]

In fig 2. The lip segmentation process is described and its phases are face detection, lip localization, pre-processing and various techniques for the detection [5].

Segmentation has two objectives as follows: Firstly, to decompose the image into parts for further analysis. In simple cases, the environment might be well enough controlled so that the segmentation process reliably extracts only the parts that need to be analysed further. For example, in the chapter on color, an algorithm was presented for segmenting a human face from a color video image. The division is reliable, provided given that the individual's dress or room foundation does not have similar shading parts as a human face. In complex cases, for example, removing an entire street organize from a grayscale ethereal picture, the division issue can be extremely troublesome and might require utilization of a lot of space a building learning [6].

Secondly, segmentation is to perform a change of representation. The image pixels must be organized to higher-level units that are either more meaningful or more efficient for further analysis (or both). A critical issue is whether or not segmentation can be performed for many different domains using general bottom-up methods that do not use any special domain knowledge. Segmentation techniques have potential use in many different domains. Both region-based and curve-based units are discussed in the several sections. The prospects of having a single segmentation system work well for all problems appear to be dim. Experience has shown that an implementer of machine vision applications must be able to choose from a toolset of methods and perhaps tailor a solution using knowledge of the application.

II. LITERATURE SURVEY

Lu, Y., et al., (2018) [7] proposed a work on lip segmentation on the basis of an automatic initial contour. The planned contours relied on the location of active contour approach

. With the rapid development of artificial intelligence and the increasing popularity of smart devices, human-computer interaction technology had become a multimedia and multimode technology from being computer-focused to people centred. Among all ways of human-computer interactions, using language to interact with machines was the most convenient and efficient one. However, the performance of audio speech recognition systems was not satisfied in a noisy environment. Lip segmentation played an important role in a visual lip reading system, since the segmentation result was crucial to the final recognition accuracy. The planned proposal was a localized active contour model-based method using two initial contours in a combined color space. The illumination equalization to original RGB images was to decrease the interference of uneven illumination. A combined color space consists of the U component in CIE-LUV color space and the sum of C2 and C3 components of the image after discrete Hartley transform. From the experiment, it was concluded that this method could get better segmentation results compared with the method using a circle as the initial contour to segment gray images and images in combined color space, especially for open mouth. An extremely obvious advantage of this method was the results of open mouth excluding internal information of mouth such as teeth, black holes, and tongue, because of the introduction of the inner initial contour.

Thein, T., et al., (2018) [8] proposed a research on lip movement recognition. Lip reading system was supportive technology to human being especially for hearing im-paired, or elderly people. Lip reading was a process where visual information was extracted by watching lip movements of the speaker with or without sound. So, reliable lip movements were required to extract visual information. This was the first work for lip movement recognition for Myanmar consonants. So, the major challenge was to recognize lip movements because of many possible lip motions and lip shapes. The accuracy and reliability of speech recognition systems could be improved by using visual information from the movements of the lips, and the need for lip-reading systems continues to grow for every language. Therefore, this paper presents Myanmar consonant recognition based on lip movements toward lip reading by using CIEL color transformation, Moore Neighborhood Tracing Algorithm and linear SVM classifier. The purpose of this study was to develop a visual training technique to accurately identify the characteristics of the movement of the lips for hearing impairment.

Lu, Y., et al., (2018) [9] describes the explanation of an automatic lip reading approach. As a significant component of the Human Computer Interface (HCI), automatic lip reading was designed for the purpose of understanding the content of speech by interpreting the movement of the lips. Although performance of automatic lip reading system was easily affected by challenging conditions such as noise, illumination and low resolution, enormous advancements in the relevant fields accompanied with enhancement in computer capability had improved the robustness of the system, making it more adaptable to the real environment. In this paper, the field and it give a detailed discussion on the actuality and the developing level of automatic lip reading. The introduction about the feature extraction and recognition model algorithms was described. Consequently, a comparison and analysis of the various visual speech databases for their characteristics and functions in speech recognition systems were given. In addition, it was demonstrated that the challenges and offer our insights into future research direction of automatic lip reading.

Ju, Z., et al., (2018) [10] proposed a deep research on the lip segmentation and multi scale features. Recently, lip image segmentation had attracted much attention because the lip image segmentation result provides much visual information for recognition of visual speech. In this paper, a new approach for lip segmentation based on the deep neural network was presented. The deep neural network was firstly introduced to lip image segmentation task while the current methods using the hand-crafted feature. In this method, multi-scaled information was integrated to do lip segmentation by using various size convolution kernels. Two neural networks were combined in this method which reduces the bad effect of dirty annotation. Also, a new loss function was employed, which take full advantage of the knowledge between successive frames, so that it was more robust to the dirty annotation and unbalanced data. The intensive approach achieved good

performance using the low-quality annotation. Experimental results show that this method achieves the state-of-art performance.

Ahmad B.A. Hassan at et al., (2016) [11] introduced a color based approach for segmentation of digital images. Initially, some color spaces were used to segment pixels using artificial neural networks (ANN). A novel method for fusion of color spaces produces improved results than individual color spaces. Hands, faces, lips, fingers were several objects for segmentation. Several databases were used to represent the problem and ANN was trained on color of the pixel and its surrounding 8 neighbors. In testing phase the trained set segments 9 pixels of test image. Feature vector was utilized to extract data from several color pixels for training and testing outputs from fusion of colors. Various experiments were imposed on the different database to calculate the method proposed. Valid results were recorded that show the expression of color and texture information of object segmentation.

Olaf Ronneberger, et al., (2015) [12] presented a network and training strategy based on strong use of data augmentation to use available annotated samples more efficiently. A huge consent for successful training of deep networks need unlimited annotated samples training. The design comprises of contracting way to catch setting and extending in symmetric manner that empowers exact confinement. They demonstrated a system could be prepared end-to-end from not very many pictures and outflanks the earlier best strategy on the ISBI challenge for division of neuronal structures in electron tiny stacks. Utilizing a similar system prepared on transmitted light microscopy pictures they won the ISBI cell following test 2015 in such classifications by huge edge. Besides, the system is quick. Division of a 512x512 picture takes not as much as a moment on a current GPU.

III. ISSUES IN LIP SEGMENTATION

A number of state-of-the-art segmentation methods have been proposed to segment the lip region from a given face image in the past decades. Nevertheless, lip segmentation is still a challenging problem under which there are at least two major issues:

i. Segment number dependency problem, and
ii. Local scale determination 5 problem. [13]

The following sub-sections give the details:-

i. **Segment number dependency problem:** Maximum a posteriori - Markov random field (MAP-MRF) framework is widely used in image segmentation because they yield a local and economical texture description. However, the methods require pre-assigning the number of segments (also called number of clusters interchangeably) appropriately in advance. Otherwise, the performance may result in over- or under-segmentation.[14]

ii. **Local scale determination 5 problem:** In the traditional MRF based image segmentation methods, the entire given image is processed under a constant scale in usual. However, the ability to describe the variable scale behaviours is limited. That is, their performance may become deteriorated as long as the input image is composed of the structures with different scales which are common in real world. [15]

However, for most of the existing multi-scale MRF based segmentation methods, there are two main shortcomings:

i. The scale map fixing and segmentation procedures are separated, which may induce redundant computational complexity, and
ii. The segment number must be pre-assigned in advance; otherwise the estimation bias of the segment number may cause over- or under-segmentation.[16,17]

IV. PROPOSED METHODOLOGY

In current year, Digital images are vital to searching quality of the image. We will implement the machine learning classification method and feature extraction methods to identify the unique properties of the image.

To attain these objectives, the proposed methodology will be as follows:

i. Image dataset is downloaded in the UCI MACHINE LEARNING REPOSITORY site using lip images (open mouth) in .jpg and .png.
ii. Segmentation used to detect the region of the noise free image and Principal Component Analysis algorithm is used for image based feature vector extraction form the collected data set.
iii. Apply the classification Methods (SVM) on the image dataset to classify images into two parts like training and testing section.
iv. Perform the study and calculate various performance parameters in terms of accuracy, Segmented Error and Overlap Error.
v. Compare the performance metrics with the exiting methods like as contour model.

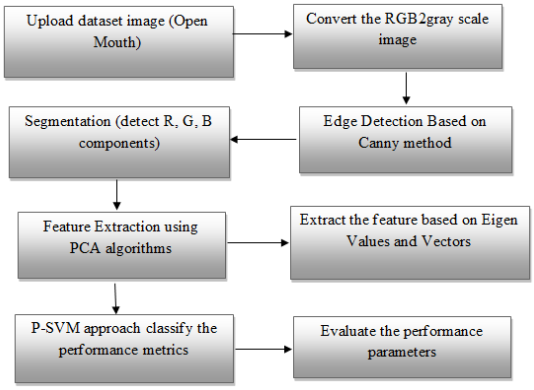


Fig.3: Proposed Work Flow Chart

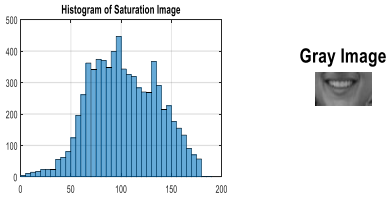
V. RESULT AND DISCUSSIONS

In this research work, we implement a lip segmentation using P-SVM algorithm. We used the simulation tool in MATLAB 2016 version. Lip segmentation plays a significant role in VLRS (Visual Lip Reading System).Because the segmentation consequence is vital to the final recognition accuracy rate.

(i) (ii)

Fig.4: (i) Original Image and (ii) Cropped Image

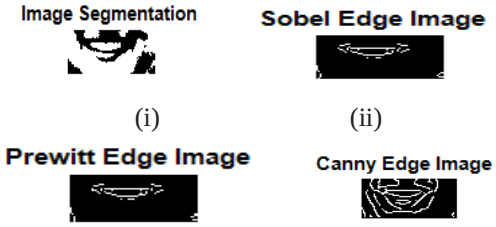
The above figure 4(i) shown that the upload the original image from the main dataset. Fig 4(ii) Uploading image from the dataset, then crop the lip area from the face image. In research work is done by lip segmentation.



(i) (ii)

Fig.5: (i) Histogram (ii) Gray level Image

Fig 5(i) defined that an image histogram is a type of histogram that acts as a graphical representation of the tonal distribution in a digital image. It plots the number of pixels for each tonal value. By looking at the histogram for a specific image a viewer will be able to judge the entire tonal distribution at a glance. Fig 5(ii) defined in digital photography, computer-generated imagery, and colorimetric, a grayscale or grayscale image is one in which the value of each pixel is a single sample representing only an amount of light, that is, it carries only intensity information.

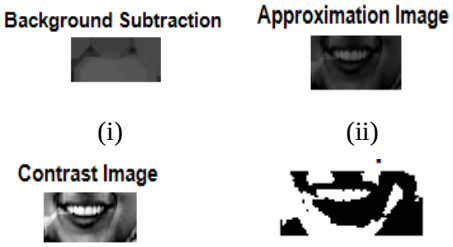


(i) (ii)

(iii) (iv)

Fig.6: (i) Image Segmented Image (ii) Sobel (iii) Prewitt and (iv) Canny Image

Fig 6(i) shown image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.Fig 6(ii) Defined that the edge detection based on various edge detector applying. In fig 5.6 defined that the sobel operator computing an approximation of the gradient of the image intensity function. At each point in the image, the result of the Sobel–Feldman operator is either the corresponding gradient vector or the norm of this vector. The Sobel–Feldman operator is based on convolving the image with a small, separable, and integer-valued filter in the horizontal and vertical directions and is therefore relatively inexpensive in terms of computations. On the other hand, the gradient approximation that it produces is relatively crude, in particular for high-frequency variations in the image. Fig 5 (iii) defined that the prewitt operator is based on convolving the image with a small, separable, and integer valued filter in horizontal and vertical directions and is therefore relatively inexpensive in terms of computations like Sobel operators. On the other hand, the gradient approximation which it produces is relatively crude, in particular for high frequency variations in the image. Fig 5(iv) defined that the canny edge detector is a technique to extract useful structural information from different vision objects and dramatically reduce the amount of data to be processed. It has been widely applied in various computer vision systems. Canny has found that the requirements for the application of edge detection on diverse vision systems are relatively similar. Thus, an edge detection solution to address these requirements can be implemented in a wide range of situations.



(i) (ii)

(iii) (iv)

Fig.7: (i) Background Subtraction (ii) Appro. Image (iii) Contrast Image (iv) Binary Version

Fig 7 (i) defined that the Background subtraction is a major pre-processing steps in many vision based applications. For example, consider the cases like visitor counter where a static camera takes the number of visitors entering or leaving the room, or a traffic camera extracting information about the vehicles etc. 7(ii) defined that the approx.. Image with FB (Fourier Basis) is not effective since it is too global and also it treats the region of the lip image using periodic situations. Fig 6(iii) shown that the contrast image is the dissimilar luminance that creates an object distinguishable the maximum image contrast of an image is the contrast and dynamic range. Fig (iv) defined that the binary image is a DI that has only twice possible values for individual image pixels. Typically, the two colors used for a binary image are black and white. The color used for the object(s) in the image is the foreground color while the rest of the image is the background color.

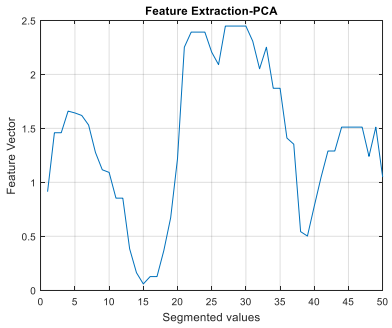


Fig.8: Feature Extraction (PCA)

Fig 8 defined that the Principle Component Analysis (PCA) is a common feature extraction method in data science.That is, it reduces the number of features by constructing a new; smaller number variables which capture a significant portion of the information found in the original features.

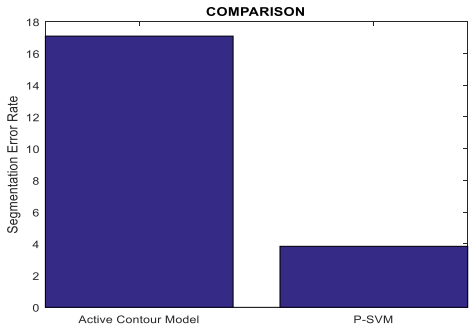


Fig.9: Comparisons – SE

Fig.9: defined that the comparison between proposed and existing work with segmented Error Rate. It SE is reduces rather than other parameters.

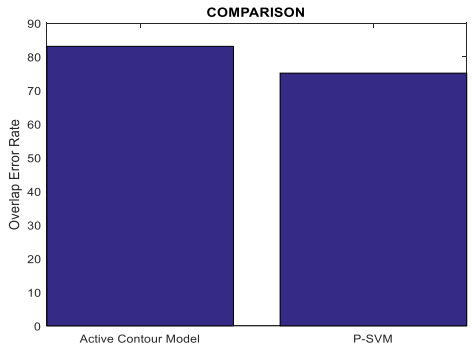


Fig.10: Overlap Error Rate Comparison

Fig 10 defined that the comparison between proposed and existing works with Overlap Error Rate. It OER is reduces rather than other parameters.

Table 1:- Comparison between proposed and existing work (Overlap and Segmented Error)

Parameter Metrics	Proposed Work	Existing Work
Overlap Error	74.43	83.10
Segmentation Error	4.4306	17.10

Table 2. Proposed Performance Metrics

Parameter Metrics	Values
Overlap Error	74.4
Segmentation Error	4.43
Accuracy Rate	74.4
MSE	5.99

Table 1 and 2 described that the performance metric in proposed work and comparison between proposed and existing work in Principle Component Analysis with Support Vector Machine.

VI. CONCLUSION AND FUTURE SCOPE

It is concluded that the lip image segmentation is done by acquiring the noise free images from the data set of the uploaded digital images. The image segmentation method based on the process of the segmentation of the digital image in to multiple segments with set of the super pixels. The main objectives of the segmentation are for changing and simplify the image in to meaningful and easy approach. The segmentation of image is for positioning the objects and boundaries. The main goal of the proposed work is the recognition of the quality of the digital images. The machine learning, classification and feature extraction method are different methods used to determine the quality of the images. The image data set of the lip images is observed and then the region of the noise free image is detected using segmentation method. In addition, the Principal Component Analysis algorithm used for image as feature extraction method for the extraction of the features from the digital images.After that, the classification of images is done in training and testing section based on support vector method .The experimental analysis is done to compute the parameters which are overlap error, segmented error and accuracy. The different parameters are compared with the exiting methods like as contour model. The errors are reduced as well as accuracy is increased.

In Future Scope, the method can give more efficient shape and region control approach using other than machine learning approach. The new machine learning audio-visual method can be established in to audio- visual speech detection method. This method is used for the detection of the face or the face detection system.

VII. REFERENCES

[1]. Cheng, H. D., Jiang, X. H., Sun, Y., and Wang, J. (2001). Color image segmentation: advances and prospects. Pattern recognition, 34(12), 2259-2281.

[2]. Kundu, A., Vineet, V., and Koltun, V. (2016). Feature space optimization for semantic video segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3168-3175). 2222

[3]. Kaur, D., and Kaur, Y. (2014). Various image segmentation techniques: a review. International Journal of Computer Science and Mobile Computing, 3(5), 809-814. 1111

[4]. Wang, S. L., Lau, W. H., Liew, A. W. C., and Leung, S. H. (2007). Robust lip region segmentation for lip images with complex background. Pattern Recognition, 40(12), 3481-3491.

[5]. Rohani, R., Alizadeh, S., Sobhanmanesh, F., and Boostani, R. (2008, December). Lip segmentation in color images. In Innovations in Information Technology, 2008. IIT 2008. International Conference on (pp. 747-750). IEEE.
[6]. Mortensen, E. N., and Barrett, W. A. (1998). Interactive segmentation with intelligent scissors. Graphical models and image processing, 60(5), 349-384.
[7]. Lu, Y., and Liu, Q. (2018). Lip segmentation using automatic selected initial contours based on localized active contour model. EURASIP Journal on Image and Video Processing, 2018(1), 7.
[8]. Thein, T., and San, K. M. (2018, May). Lip movements recognition towards an automatic lip reading system for Myanmar consonants. In 2018 12th International Conference on Research Challenges in Information Science (RCIS). IEEE.
[9]. Lu, Y., Yan, J., and Gu, K. (2018). Review on Automatic Lip Reading Techniques. International Journal of Pattern Recognition and Artificial Intelligence, 32(07), 1856007.
[10]. Ju, Z., Lin, X., Li, F., and Wang, S. (2018, June). Lip Segmentation with Multi-scale Features Based on Fully Convolution Network. In 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC). IEEE.
[11]. Hassanat, A. B., Alkasassbeh, M., Al-Awadi, M., and Esra'a, A. A. (2016). Color-based object segmentation method using artificial neural network. Simulation Modelling Practice and Theory, 64, 3-17.
[12]. Ronneberger, O., Fischer, P., and Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
[13]. Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.
[14]. Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., and Shen, D. (2015). Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. NeuroImage, 108, 214-224.
[15]. Kaur A. and Kaur N. (2015). Image Segmentation Techniques. International Research Journal of Engineering and Technology, vol. 2, no. 2, pp. 944-947.
[16]. Haitao, L., and Shengpu, L. (2016). An Algorithm and Implementation for Image Segmentation. International Journal of Signal Processing, Image Processing and Pattern Recognition, 9(3), 125-132.
[17]. Pare, S., Kumar, A., Bajaj, V., and Singh, G. K. (2016). A multilevel color image segmentation technique based on cuckoo search algorithm and energy curve. Applied Soft Computing, 47, 76-102.