A NOVEL APPROACH TO LIP SEGMENTATION FOR VISUAL SPEECH RECOGNITION

H DEEPA LAXMI & B SUJATHA

Department of Computer Science, Meenakshi College for Women, Chennai, Tamil Nadu, India

ABSTRACT

Lip area detection and extraction is essential for many applications like facial expression, lip reading and visual speech recognition. In general, the lip segmentation in color image is a tedious task because the color difference between lip and skin region is not so apparent sometimes, which is a major issue in lip reading system. The objective of this work is to provide the solution to the problem of lip segmentation from the video using Image Processing techniques. In this work, Otsu based thresholding and mid-point based lip segmentation is carried out. The performance of the proposed technique is evaluated by testing it on 20 different videos which are taken under standard condition. From the results it is observed that, the proposed method has achieved 98% of success rate for lip segmentation.

KEYWORDS: Face Detection, Image Segmentation, Otsu Method, Thresholding, Inverted Image

I. INTRODUCTION

Recently, lip detection and segmentation has received extensive attention in the field of research because of its potential application in areas such as lip reading, audio visual speech recognition and facial expression analysis. The automatic speech recognition (ASR) system is widely used in robotics, personal computer, and cell phone. However, in noisy circumstances, ASR performance will drop. With the help of a lip reading system, this problem can be addressed [4]. Lip reading system consists of several steps, and lip segmentation is one of the most important and critical pre-processing step, because it provides basic information to be processed in the next step. There are many techniques used by different researchers in various phases of lip segmentation methods.

A large category of techniques are model-based. In these techniques, at the first, a model of the face is built and then the structure of lip area is described by a set of model parameters. These techniques include snakes, active contour models and several other parametric models. The advantage of these systems is that the important features are represented in a low-dimensional parameter space and the calculation complexity is decreased. Also, this system exhibits good performance in conditions such as rotation, scaling and illumination variation. These models also have few drawbacks, for example sometimes large training set is needed to cover the high variability range of lips [5]. For lip detection several gray scale or color image segmentation techniques are used. For example, Ishvari S. Patel [1] used edge detection technique for lip segmentation. Edge detection was performed on gray scale images using canny method, cascade classifier approach was used to detect the face part from the image and dilation approach was used to find the lip position. After finding the lip region lip segmentation was done using Region props with Bounding Box property.

The goal of this proposed work is to segment the lip of the speaker from the video. In this work video is converted into frames, and then face detection technique is applied to extract the face of the speaker. The extracted face is then segmented based on automatic threshold segmentation called Otsu's method and the lip region is detected from the

segmented image using proposed method. The paper is organized as follows, Section II gives related works, Section III gives the proposed work in detail, section IV gives the simulation results and discussion and finally section V gives the conclusion and future work.

II. RELATED WORK

In [1] Lip Segmentation is done based on Edge Detection Technique. This model works more effectively and gives 98.68% result for image sequences. In [2] they tried to perform color lips segmentation using thresholding method based on a new and efficient color space IHLS it has good performance in different color images and gives 93% Accuracy. In [3] they present a novel active contour guided geometrical feature extraction approach for lip reading and different active contour methods are used for lip segmentation. In [4] Chan-Vese model is a region-based segmentation algorithm, which can also be used as tracking method and investigates about lip segmentation and tracking based on Chan-Vese model, preceded by color segmentation using k-means and a*-b* component of CIEL*a*b* color space. In [5] for each part they calculated statistical information such as standard deviation and based on them they detected the lip area in face image. For separate lip pixels from skin pixels, they use YCbCr and HSI color spaces at this work. In [7] a new feature extraction technique called Active Lip Model (ALM) is used for Visual only isolated digit fuzzy ART (f-ART) classifier is presented.

III. PROPOSED WORK

In this paper, an effective Lip Segmentation method is proposed to overcome the complexity and limitations of traditional Lip Segmentation techniques. This work has three phases, as follows.
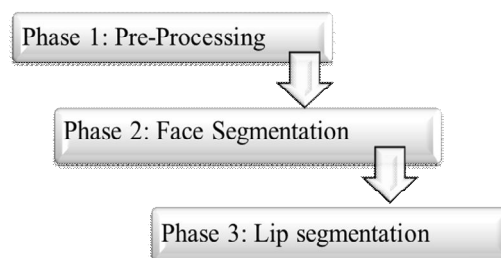

Figure 1: Block Diagram for Proposed System

Pre-Processing

In pre-processing phase, frames are extracted from the video file of the individual speaker which is given as the input. A video can be considered as a series of images shown in rapid succession and the number of frames captured within a second depends on the frame rate of the camera used. After extracting the frames face detection is carried out from which the lip region has to be segmented. Lips comprise a very small part which is volatile and constantly in motion for a speaker. Therefore, to limit the area in which lips need to be searched, it is a popular practice to perform face detection first. Once the face is detected, crop the face from the given image for further processing. Image Processing is a technique by which one can process any image to extract the real information i.e. useful information from the image. Sometimes the specific part of the image will be required for the process, and therefore the particular portion of the image is selected. In this work, conversion from RGB to Gray-Scale is necessary for further implementation. Therefore the color transformation of RGB to Gray-Scale is performed and the segmentation accuracy is improved in the following image processing stages due to a proper color space.

Face Segmentation

There are three steps in this phase; the first step is to convert the gray scale image into inverted image. After getting the inverted image, thresholding is done automatically, using Otsu's method to obtain the binary image. Each step is explained in detail as follows.

The image is inverted when the objects are darker than the background in the original image or the negative of the image is required. In inverted image, the background pixels and object pixels represents edges and non-edge pixels respectively. Thresholding is carried out in the inverted image using Otsu method. It is necessary to select an optimum threshold of gray level for extracting objects from the background. The threshold value can be determined manually or calculated automatically. The automatic thresholding methods are widely used because of their advantages such as simple to implement and less time consumption. The Otsu method is one of the Automatic thresholding methods which is frequently used in various fields. It is easy to implement and fast. However, it is important to consider background brightness variations when trying to segment an image using this method. Background variations can be caused by inhomogeneous illumination or by reflection [6]. Pixels with gray values greater than threshold are mapped to logical true whereas pixels with gray values lower than or equal to threshold get mapped to logical false.

A Threshold image (x, y) is defined as:

$$g(x, y) = \begin{cases} 1 & \text{If } f(x, y) > T \\ 0 & \text{If } f(x, y) \leq T \end{cases}$$

Where, T is a Threshold value.

Lip Segmentation

The final phase of the proposed work is to detect and segment the lip region from the logical face image. In this work, a new method is proposed to detect and segment the lip region based on the midpoint of the segmented face.

Steps involved in this phase are as follows:

Step 1: Convert the Boolean value of the logical image into zero-one matrix

Step 2: Find the total no of rows and columns of the given image.

Step 3: Find the midpoint of the image i.e. mid-row and mid-column of the given image.

Step 4: Round the mid-row and mid-column value to the nearest integer for accuracy.

Step 5: Sum the 0's and 1's in each row, starting from the mid-row to the last row.

Step 6: Sort those values in descending order and save the equivalent row position of that value.

Step 7: Select appropriate number of largest values and find the actual row position of those values using the below equation

$$O_R = M_R + (N_R - 1)$$

Where, OR – Original row value, MR – Mid-row value, NR – New row value

Step 8: Sort the calculated actual row values in ascending order

Step 9: Subtract a value from the left side of the mid-column value and add a value to the right side of the mid-column value.

Step 10: Finally the lip region alone is extracted from the face.

IV. SIMULATION RESULTS AND DISCUSSIONS

In this work, real time video sequences are captured according to the specific requirements of this project. Video database is created in Indian English accent, recording distance is kept constant, and no head movement is allowed, but the pictures have quite different backgrounds and different lighting conditions. In addition, this database includes only females for video recording. The video is stored in Audio Video Interleave (AVI) format and has a frame rate of 30 fps. For each person a video is recorded in normal lighting condition without any particular makeup; volunteers are selected from the age group of 22-25. The accuracy of the face detection and Lip Segmentation was tested using this database. This model was tested on 10 videos recorded using web cam and for implementation purpose 20 frames with an interval of 0.45 seconds is considered per video.

Results of the proposed method in various categories like face detection and lip segmentation are shown in table 1, where lip segmentation has two different categories like lip segmentation on inverted and lip segmentation on non-inverted image

Table 1: Result of the Proposed System

	Success	Failure	Percentage
Face Detection	200	-	100
Lip segmentation with non-inverted image	140	60	70
Lip segmentation with inverted image	196	4	98

To get the high accuracy rate, the image is inverted before segmentation. Figure 1, 2 and 3 shows successful images for Pre-processing phase, segmentation phase and lip segmentation phase result respectively.
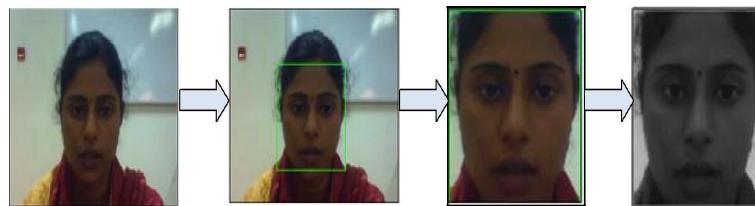

Figure 2: Results of Pre-Processing Phase

Figure 3: Results of Segmentation Phase

Figure 4: Results of Lip Segmentation Phase

V. CONCLUSIONS

In recent days, speech recognition systems are becoming more important. Lip segmentation is one of the important pre-processing steps in the lip reading system. However, detecting and segmenting the lip accurately and robustly from the face is very difficult because lips are highly deformable and they vary in shape, size and color. It is difficult to detect the lip, when there is a shadow area near the lip; or the lip color is similar to that of the face. The proposed work addresses these issues of Lip Segmentation in color space. This work detects the face of the speaker from the video, segments it using Otsu's method, and then the lip region is detected and segmented based on midpoint.

From the results it is concluded that the proposed method works effectively and achieves good accuracy for video sequences. In case of videos with extremely bright or dark lighting conditions, the system sometimes shows less detection accuracy. The future work is to improve the system, so as to adapt to all lighting conditions and to extend the work for male speakers.

REFERENCES

1. Ishvari S. Patel, Apurva A. Desai. "Lip Segmentation Based on Edge Detection Technique" IJSR - International Journal of Scientific Research Volume: 2, Issue: 5, ISSN No 2277 – 8179, May (2013). Information".

2. Mohamad Amin Bakhshali, Mousa Shamsi. "Segmentation of color lip Images by Optimal Thresholding using Bacterial Foraging Optimization (BFO)" Journal of Computational Science - JOCS Vol- 5 Page No: 251–257 (2013).

3. Sunil S.Morade, B.S Uprava Patnaik. "Automatic Lip Tracking and Extraction of Lip Geometric Features for Lip Reading" International Journal of Machine Learning and Computing (IJMLC), Vol. 3, No. 2, April (2013).

4. Aris Nasuha, Tri Arief Sardjono, Mauridhi Hery Purnomo. "Lip Segmentation and Tracking Based on Chan – Vese Model" (2013). In proceeding of: ICITEE 2013, At Yogyakarta, Indonesia.

5. Hashem Kalbkhani, Mehdi Chehel Amirani. "An Efficient Algorithm for Lip Segmentation in Color Face Images Based on Local Information" Journal of World's Electrical

6. file:///C:/Users/User/Downloads/Tutorial%20(5).pdf Image processing with Scilab and Image processing design Toolbox

7. Sujatha B, Santhanam T. "A novel f-ART Model for Visual Lipreading based on Active Lip Model" Proceedings of IEEE International Conference on Signal Acquisition Processing (ICSAP 2011).

8.   http://www.sciencedirect.com/science/article/pii/S2212671612002338 An Improved Median-based Otsu Image Thresholding Algorithm

9.   http://www.cse.unr.edu/~bebis/CS791E/Notes/Thresholding.pdf Thresholding

10. http://www.cs.bgu.ac.il/~klara/ATCS111/gonzales_10.1_10.2.pdf Image Segmentation