Lip Movement Features Point Extraction For Lip Reading System

M. R. Suresh [1], Dr. Nagappa U. Bhajantri[2]

[1] Associate Professor, [2] Professor
[1]Department of Information Science & Engineering
[1]PES College of Engineering, Mandya, India

Abstract - **Lip reading is a technique of communication used by a hard hearing person in their conversation between themselves or with the normal person. Sometime the word they understand is not the same as what the other speaker talk. Computer-based lip reading system may help them to track those words based on the movement of the lips. When speak, lip make a movement that may differ between several words. For the computer to recognize the spoken word, feature from the lip need to be extracted and is stored in the database. A surface area of the lip is proposed as the feature of the lip movement. The horizontal and vertical distances of the lip are extracted to determine the surface area. Data from the lip feature then been resampled to estimate some parameter and their reliability. Result from the resampled then will be initialized to reduce memory usage in the database. In the experiments, several spoken words at the hospital have been chosen. The experimental results show that the ellipse feature could be employed to train the computer understands the spoken word from the human.**

Index Terms **- Lip reading, Movement, extraction, classifiers.**

I. INTRODUCTION

Computer-based lip reading system is a system that may help deaf and hard hearing person to learn how to speak with the correct lip movement. Lip has a significant shape that may help in recognizing word speaking. In the conventional lip reading learning, instructor needs face to face with the hard hearing person to correct their lip movement for word been spoke.

When we speak our lips will make several horizontal and vertical movements. Many researches find out that detection of lip activities and movements may help to classified talking faces in television-content [1].

Piotr Dalka [2] in his research had developed a LipMouse to help paralyze people controlling mouse curser by using shape of the lips. Accurately tracking facial features requires coping with the large variation in appearance across subjects and the combination of a rigid and a non-rigid motion [3]. Lip region may sometime difficult to extract due to variation color of the lip. Lip tracking not a trivial task, since there are varieties in people in skin color, lip width and lip movement during speech [4].

In this work, the research objective is to detect the movement of lips based on the changes of the surface area of the ellipse. In [5] the researcher try to locate the lip region based on the whole face detection. The surface area is calculated based on the width and height of the tracking lip.

This paper is structured as follows: In the methodologies, the fundamental of image processing employs in the research is briefly explained. In the experiments, the obtained results from ellipse tracking are discussed.

II. LITERATURE REVIEW

In the research of image processing, several researches had proposed various techniques for the purpose of vision sensing.

Jamal Ahmad Dargham et al. [9] proposed a method for lips detection in color image in the normalized RGB color scheme. They proposed a new method called the maximum intensity normalization. The method successfully reduces the error in image segmentation compared to the pixel intensity normalization.

Ying-li Tian et al. [10] described a method of tracking a feature of facial expression; in particular, lip contours, by using a multi-state mouth model and combining lip color, shape and motion information. They the lip and skin color by using Gaussian mixture.

Yun-Long Lay [11] used principal component analysis and mouth changing rate for getting the feature of the moving lip. The author also states that in the pre-processing, several steps need to ensure first for getting the robust lip recognition. These steps include brightness adjustment, color space transforms, face color cutting, facial operation region and lip localization.

Yonghong Xie [12] in his paper demonstrates a new method for ellipse detection. In his paper, the ellipse detected is based on the one-dimensional accumulator array to accumulate the length information for minor axis of the ellipse. This method does not require the evaluation of the tangent or curvatures of the edge contours.

III. METHODOLOGY

A. *Flow chart of the proposed works*

Before recognizing of the spoken words used in the hospital, the word database needs to be developed. Image processing technique is employed to extract lip features. Figure 1 shows the flow chart of the proposed image processing technique.

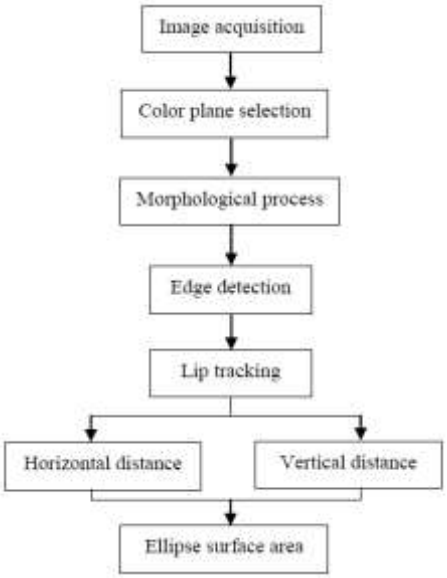

Fig.1 This figure shows the flow chart of the image processing technique

B. *Lip region detection*

Detection of the lip region required several process of image preprocessing technique. Before subtract the lip region, the image information of the lip is in RGB color space. To ensure the robust lip detection, the RGB color space image is converted to HSL color space. This is because of avoiding effect from lighting glare, by converting to HSL color space. After converting to HSL, the image then will be processed in grayscale image. The grayscale image then is converted to the binary image. Before converting to the binary image, the threshold value for the lip region is identified. The lip region produced in the binary image is not smooth because contains noises that distract the image. To manage capturing lip region without noises, the morphological process [6] needs to be done to smoothen it.

Moment threshold

This technique is suited for images that have the poor contrast. The moment method is based on the hypothesis that the observed image is a blurred version of the theoretically binary original. The blurring that is produced from the acquisition process, caused by electronic noise or slight defocalization, is treated as if the statistical moments of average and variance were the same for both the blurred image and the original image. This function recalculates a theoretical binary image.

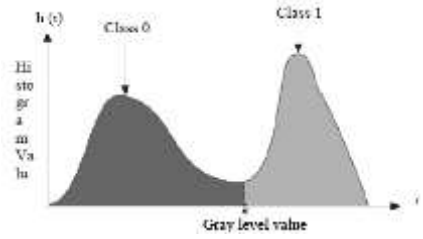

Fig.2 Graph of the threshold value selection

i = represents the gray level value

k = represents the gray level value chosen as threshold value

$$m_k = \frac{1}{n}\sum_{i=0}^{i=n-1} i^k h(i) \qquad (1)$$

Where n is the total number of pixels in the image.

m_k is the value for moment threshold

Morphology

Morphological image processing relies on the ordering of pixels in an image and many times is applied to binary and grayscale images. Through processes such as erosion, dilation, opening and closing, binary images can be modified

Dilation/Erosion

First, define *A* as the reference image and *B* is the structure image used to process *A*.
Dilation is defined by the equation:

$$A \oplus B = \left\{ Z \mid \left[\left(\hat{B} \right)_Z \cap A \right] \subseteq A \right\} \qquad (2)$$

Where $\hat{B}$ is *B* rotated about the origin. Dilation has many uses but a major one is bridging gaps in an image due to the fact that *B* is expanding the features of *A*. Dilation on the other hand can be considered a narrowing of features on an image. Again defining *A* as the reference image and *B* as the structure image:

$$A \Theta B = \left\{ Z \mid \left[(\hat{B})_Z \cap A \right] \subseteq A \right\} \qquad (3)$$

Many times dilation can be used for removing irrelevant data from an image.

Opening/Closing

By utilizing the processes of erosion and dilation, opening and closing is simply and extension of these applications. The process of "opening" an image will likely smooth the edges, break narrow block connectors and remove small protrusions from a reference image. "Closing" an image will also smooth edges but will fuse narrow blocks and fill in holes.

Opening $$A \circ B = (A \ominus B) \oplus B \qquad (4)$$

Closing $$A \cdot B = (A \oplus B) \ominus B \qquad (5)$$

By these definitions, the opening of *A* is the erosion of *A* by *B* and then that image dilated by *B*. The closing of *A* is the dilation of *A* by *B* and then eroded by *B*.
By knowing that dilation and erosion are duals of each other:

$$(A \ominus B)^c = A^c \oplus \hat{B} \qquad (6)$$

We can conclude that with respect to set complementation and reflection, that opening and closing are complements of each other:

$$(A \cdot B)^c = A^c \circ \hat{B} \qquad (7)$$

C. ***Lip Tracking***

Distance measurement

When the edge of the lip has been detected, the next step is to tracking the movement of the lip. For the robust lip tracking, the horizontal and vertical distances of the lip are measured as shown in figure 3. The horizontal and vertical distances are denoted by D_H and D_V, respectively. This value can be gotten after using LABView function, which are *Imaq Horizontal clamp* and *Imaq Vertical clamp* [8].
The distance value from *Imaq function* is in number of pixel distance between two reference points. For horizontal distance, left and right oral commissure is the reference point. For vertical distance, the maximum upper lip the center of cupid bow and the lower lip detect will be lower lip vermillion. When speaking, this four reference points move.

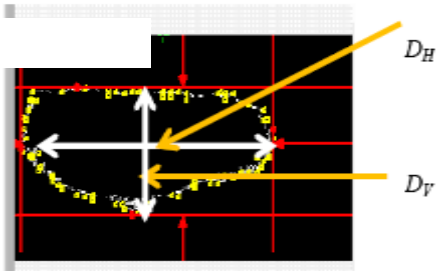

Fig.3 D_H and D_V and of the lip

The pixel value obtains from the distance measurement is not the best value to measure the ellipse surface area. The pixel value is converted to the mathematical unit. Formulation for converting to millimeter is used as stated below:

$$d_p^{\,2} = W_p^{\,2} + h_p^{\,2} \qquad (8)$$

$$PPI = \frac{d_p}{d_i} \quad (9)$$

Where,

d_p is the diagonal resolution

w_p is the width resolution

h_p is the horizontal resolution

d_i is the size of screen

PPI is the pixel per inch

Ellipse surface area

Ellipse is the best feature represents the lip area. From experiments that have been conducted, this ellipse area covers all surface of the lip. For measuring the ellipse surface area, the formulation is as below:

$$Area\ ellipse = pi \times \frac{D_H}{2} \times \frac{D_V}{2} \quad (10)$$

Where, D_H is the horizontal distance of lip from D_V is the vertical distance of lip.

D. ***Development***

Resampling

Movement data that acquired from the spoken word is time based and has a different characteristic each time it is recorded. One of the reasons is due to the emotional factor of the subjects. Data need to be converted to another form. The simplest example is to measure the average and the distribution of data such as the standard deviation. However, this kind of statistical parameters reduce the information contain in the raw data. This research proposes the use of resampling algorithm and the raw data are converted to several points called features point.

The formulation use to convert the raw data to 10 features point is shown in eq. 11. First the size of the array is defined as the size of the partition for every ten point of features point.

$$partition\ size = \left\lfloor \frac{n}{10} \right\rfloor \quad (11)$$

Where n = array size

The first element in the array subtracts second element. This process is continuously happens until the last value contains in the array as shown in eq. 12 until 14.

$$\begin{bmatrix} m_1 \\ \dots \\ m_n \end{bmatrix} = \begin{bmatrix} e_1 \\ \dots \\ e_n \end{bmatrix} - \begin{bmatrix} e_2 \\ \dots \\ e_n \end{bmatrix} \quad (12)$$

where

$m_1 \sim m_2$ = movement data

$e_1 \sim e_2$ = element data of array

$$rotate\ array_n = -(ps \times n) \quad (13)$$

$$pd = (ps)\ data\ from\ upper\ array \quad (14)$$

ps = partition size

n = number of respoint

pd = partition data

In each patrician, the mean of data are measured and stored as features point.

Initialize

Data from the features point then is initialized to the value ranging from 0 to 1 by using the eq. 15.

$$initialize\ data = \frac{respo\,\text{int}}{\max respo\,\text{int}} \quad (15)$$

IV. EXPERIMENTS

The experiments were conducted in the laboratory with the optimum lighting condition. The subject needs to sit in front of the web camera with approximately 8 to 10 cm in distance. The white screen was placed at the back of the subject.

The subject was asked to speak out 10 words that were commonly used in the hospital by the repetition 10 times each. Every word will be repeated 10 times.

V. CONCLUSION

In this paper, the methods to extract features of lip movements for the purpose of the tracking and recognizing of lip movements had been explained. The lip region that had been extracted was smoothen by the morphological process. The lip was tracked based on the horizontal and vertical distances of the edge. The surface area of ellipse was produced from the obtained values of the horizontal and vertical distances of the edge in the tracking process. Method for developing the database had been explain in this paper and data from database had been used in the recognition process. The experimental results show that the ellipse could be used as the feature of the lip movements. In the future, the system to rehabilitate speech therapy could be developed to be used by hard-hearing person.

VI. REFERENCES

[1] Meriem Bendris, Delphine Charlet and Gerard Chollet, "*Lip activity detection for talking faces classification in tv-content*," in 3rd International Conference on Machine Vision (ICMV), 2010, pp. 187-190.

[2] Piotr Dalka and Andrzej Czyzewski, "Lip movement and gesture recognition for a multimodal human-computer interaction" Proceeding of the International Multiconference on Computer Science and Information Technology, 2009, pp. 451-455.

[3] Ying-li Tian, Takeo Kanade and Jeffrey F. Cohn, "Robust lip tracking by combining shape, color and motion" Proceeding of the 4th Asian Conference on Computer Vision (ACCV'00), 2000.

[4] Hamed Talea and Kashayar Yaghmaie "Automatic combined lip segmentation in color image" Proceeding of IEEE, 2011, pp. 109-112.

[5] S. C. Chen et al. "A text input system developed by using lips image recognition based on labview for the serious disabled" Proceeding of the 26th Annual International Conference of the IEEE EMBS, 2004, pp. 4940-4943.

[6] Rafael C. Gonzalez, "Morphological Image Processing (3rd Edition)," in Digital Image Processing, Pearson Prentice Hall, 2008, pp. 630-676.

[7] Rafael C. Gonzalez and R. Woods in Digital Image Processing, Addison Wesley, 1992, pp. 414-428.

[8] Thomas Klinger, "Image Analysis: Edge detection," in Image Processing with LABView and Imaq Vision, Prentice Hall Professional, 2003, pp. 223-231.

[9] Jamal Ahamad Dargham and Ali Chekima, "lips detection in the normalized RGB color scheme" Proceeding of IEEE, 2006, pp. 1546-1551.

[10] Ying-li tian, Takeo Kanade and Jeffrey F. Cohn "robust lip tracking by combining shape, color and motion" Proceeding of the 4th Asian Conference on Computer Vision (ACCV'00), 2000.

[11] Yun-Long Lay et al. "Lip language recognition for specific word" Indian Journal of Science and Technology, 2012, pp. 3565-3572.

[12] Yonghong Xie and Qiang Ji "A new efficient ellipse detection method" Proceeding of IEEE, 2002, pp. 957-960.