

# Lip and Head Gesture Recognition Based PC Interface using Image Processing

P. JOHN HUBERT<sup>1</sup> and M.S. SHEEBA<sup>2</sup>

<sup>1</sup>Embedded System, Sathyabama University, India.

<sup>2</sup>Electronics and Communication, Sathyabama University, India.

DOI: <http://dx.doi.org/10.13005/bpj/584>

(Received: February 02, 2015; accepted: March 13, 2015)

## ABSTRACT

The multimodal human-computer interface (HCI) called Lip Mouse is presented, allowing a user to work on a computer using movements and gestures made with his/her mouth only. Algorithms for lip movement tracking and lip gesture recognition are presented in details. User face images are captured with a standard webcam. Face detection is based on a cascade of boosted classifiers using Haar-like features. A mouth region is located in the lower part of the face region. Its position is used to track lip movements that allows a user to control a screen cursor. Three lip gestures are recognized: mouth opening, sticking out the tongue and forming puckered lips. Lip gesture recognition is performed by various image features of the lip region. An accurate lip shape is obtained by the means of lip image segmentation using lip contour analysis.

**Key words:** Matlab, Lip contour, Human computer interface, Image processing.

## INTRODUCTION

Lip boundary extraction is an important problem it has studied to some extent in the literature [1, 2, 3, 4]. Lip segmentation is an important part of audio-visual speech recognition, lip-synching, modeling of talking epitome and facial feature tracking systems. In an audio-visual speech recognition, it have been shown that using lip texture information is more valuable than using the lip boundary information [5, 6]. However, this result may have partly due to inaccurate boundary extraction as well, Therefore lip segmentation performance has not been independently evaluated in earlier studies. It can approach to use lip segmentation information complementary in the texture information. Lip boundary features has been utilized to lip texture features in a multi-stream Hidden Markov model framework with an appropriate weighting scheme Thus, we conjecture it was beneficial to use lip boundary information it can improve accuracy in AVSR. Once the boundary of a lip was found, one may extract geometric or algebraic features. These features has been used

in audio-visual speech recognition systems in complementary features to audio and other visual features. The visual appearance of the human lip kept a lot of information about an individual methods it belongs in the region. It can not be only a specific part of the each person's look at a lip shape also serves of expressing our emotions. However the lips' passage indicates when the person was talking and even allows conclusions defined has being uttered. Localizing a exact lip boundarys an image or video was appeal. Valuable information is in various applications in human computer interaction and an automated control was required in lac commercial applications. In recent years, difference in an automatic speech recognition (ASR) has cropped up and drawn an attention of the researchers [1]-[3]. With the existence of noise as in practical world circumstances, the ASR rate can be dramatically decreased. The ASR system would also be able to afford an appreciable performance only under a certain controlled environment system. With the motivation of lips-analyzing capability from the flawed society and a limitation of the noise hefty techniques, the audio-visual speech recognition

(AVSR) has become in a research direction and it was growing rapidly [4]. Prior to the ASM part.

In this paper, we propose a compelling method for extracting lip contour. The lip shape is expressed as well a set of landmark.

### Existing methods

#### FRINGE Detection

Fringe detection it is the mathematical method is used for defending the fringe of the lip region. And it can be identify the digital pixel image function to change the image brightness strongly and it can discontinuous. The edge of the image region which can come closer to or it can end with sharp edges in curved of the image it can be calculated the fringe of the lip region in 1D signal is called as step detection. Fringe detection is the fundamental tool of the image processing .

#### Fringe properties

The fringe can be extracted by the function of dual dimension image and also can be three dimension image can be classified to the viewpoint dependent otherwise viewpoint independent. The viewpoint independent function can be fringe typically reflects to the inherent properties of the three dimensional subjects, such as surface pointing and also surface figure. The viewpoint dependent fringe can be change belonging to the viewpoint changes, and it can be typically change in the reflects of the geometric of the image. Such

as object occulting one another. The typical fringe might be instance of the border between in a red border colour and also yellow of block. It is a contrast of a line as can be extracted by ridge direction it can be small number of pixel density of a different colour on unchanged background image, it may be therefore usually be one edge on each direction of the space line.

### Proposed work

#### lip region extraction

In the lip region extraction function there are three various function of lip region extraction are available. In the first variant of the lip region extraction is named as (V1) it can be calculated as well as the function of lip region in the eclipse similarity. The lip shape can be constituted in a rectangular box in the mouth region and its contains eclipse. Its is defined that the lip region size is not a constant function and the lip region can be change according to the lip movements. The second variant of the lip region extraction is named as(V2), it is a horizontal, constant shape square is used as in the lip region. In the lip region function the center of the part can is always anchored at the center of the eclipse and length is fixed and it can calculated at the beginning of the calibration process in the width of the eclipse. The third variant of the lip region extraction (V3). Domination of the eclipse of the lip region is very less. And also the lip region can be determined as well in the square form of the lip extraction and it's the fixed point and determined at the beginning of the calibration phase. The center

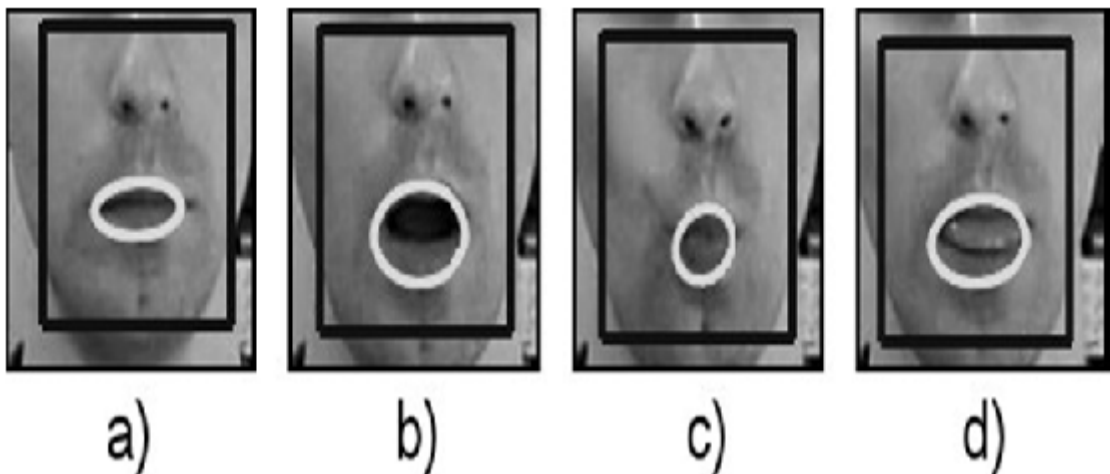


Fig. 1. Lip boundary analyzing

of the eclipse is located in the center of the square function. And the length of the left and right direction is equal to half of the width of the lip region. The third variant is only worked as well as when the first variant is getting failed.

When we extract the visual features, an accurate extraction of the lip shape was needed. The main use of the edge base approach function in the lip region extraction is a problematic function since the edge map obtained it is usually very noisy and also it can determine the false edge of the mouth region, however edges are absent on the lip edge or they have very low magnitude function and also can often be overwhelmed by strong false edges not associated with lip edge region. In view of these difficulties function our approach to find an optimum partition of a given lip image into the lip and non-lip region based on the function of pixel density and color.

**Skin and lip colour analyzing**

The method of the color space is essential because it has a direct influence on the robustness and the veracity of the segmentation is finally achieved. Our goal is to find the color pixel space that enable a good separation of the skin and lip pixels. In [1] the HSV system is the method used for to determine the lip region ‘hue’ has a good discriminative power. Even for different human talkers, hue of lip and skin are relatively constant color and well separated. But hue is generally noisy function because it’s a wrap around nature function (low values can be determine of hue lie close to the high pixel density ) and its has bad reliability for slow saturation pixels.

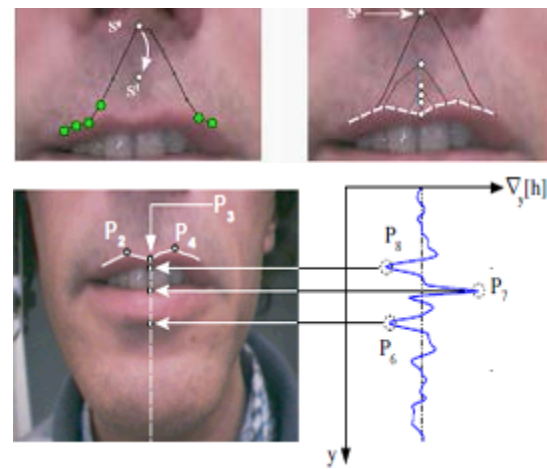
CIELUV and YCrCB spaces can also be used in the face analyzing method. It is shown that the skin color can be cover a small space of the (u,v)planes([12,3]), whatever the distribution of the skin and lips color often overlap and vary for different speakers. This makes spaces unsuitable lip segmentation. In a RBG color space region function, lip and skin have quite different. For both red color is common, whatever there is more green than blue color is in the skin color mixture and for the lip these two components are almost same[14]. Skin appears more yellow color than lip region. Because the difference between the skin and lip, green color is

more in lip. Hulbert and Poggio[15] propose a pseudo hue define that exhibits this difference. It can be computed by,

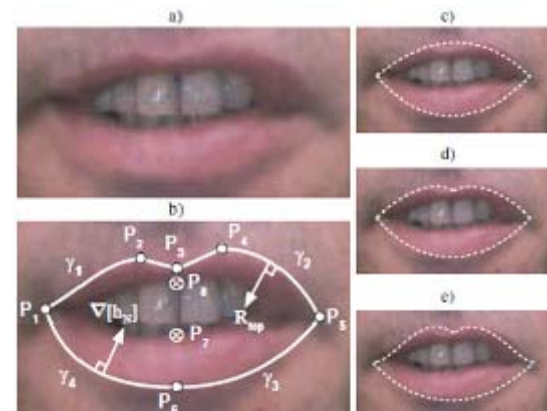
$$h(x, y) = \frac{R(x,y)}{G(x,y)+R(x,y)} \dots(1)$$

where R(x,y) and G(x,y) are respectively mentioned as red and green color of the pixel(x,y). it is a higher pixel for lip than the skin[14]. Luminance also a good cue to be taken into the account. In normal way of system the light can comes along with above the speakers. Then the top frontier of the top lip is illuminated information, we also can used a hyper edge  $R_{top}(x,y)$ , introduced in[16].

$$R_{top}(x, y) = \nabla[Hn(x, y) - Ln(x, y)] \dots(2)$$



**Fig. 2. lip colour analyzing**



**Fig. 3. various lip shape; (a) asymmetric mouth shape (b) composed of 4 cubes (c&d) parametric using parabola (e) using quartics**

**Contour extraction**

There are several parametric models has been developed in the lip boundary function. Tian used a method made for two parabolic function. It is the one easiest method to compute, but it is not too fit with the lip edges with accuracy. Other author propose to use two parabolic instead of to improve accuracy, but the model is still limited by its rigidity. Particularly in the case of asymmetric mouth shape. The model we use is very flexible it is enough to reproduce the specificities of very different lip boundarys shapes. It can be composed by 5 independent curves. Each one of them can be describe a part of the lip boundary. Between P2 and P4 cupidon's bow is drawn with a broken line and the other parts are approximated by 4 cubic polynomial curves yi. Moreover we can conceder that every each cubes is null derivative at key points P2,P4 or P6. For example y1 is null derivative on P2.

**Cursor movements**

Mouth region localization method is compared with a reference of the mouth position, is used to control the screen cursor in PC. Generally the cursor can move faster to given direction. The reference lip position has been saved in the application startup and it can may be altered at any time for user comfortable. Translation between the current mouth position and the screen cursor are determined by three parameters. Threshold t, sensitivity s, and acceleration a, is given with this

equation (In vertical and horizontal direction separately ).\

$$\Delta x = \begin{cases} a(\Delta p - t) * (\Delta p - t) + S(\Delta p - t), & \Delta p > t \\ -a(\Delta p - t) * (\Delta p - t) + s(\Delta p - t), & \Delta p < -t \\ 0 & -t \leq \Delta p \leq t \end{cases} \dots(3)$$

Where "x is denotes the resulting screen cursor movement distance in pixels and delta p is the distance between the mouth position and the reference position. Threshold, sensitivity and acceleration have positive float approach values and are defined for horizontal and vertical direction are implemented. The mouth position shift delta (P) in horizontal and vertical directions are calculated as given in equation,

$$\Delta p_x = \frac{m_x - r_x}{w} \dots(4)$$

$$\Delta p_y = \frac{m_y - r_y}{w} \dots(5)$$

Where (Mx,My) denotes the current mouth position the middle of the lip region upper boundary in video frame pixels density, (Rx,Ry) is the reference lip position and we denotes the current lip region width. Normalization of the mouth position shift by the mouth assure that the screen cursor can moves in the same way independently of the

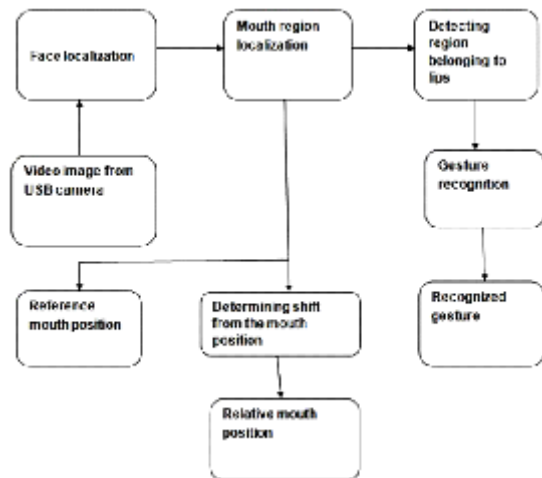


Fig. 4. System architecture

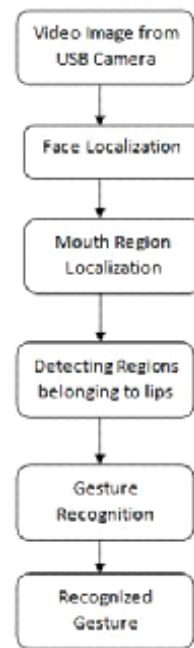


Fig. 5. Flow diagram

user image distance from the web cam. Threshold is the minimal mouth shift from the reference position that is required for the screen to move a cursor. The greater threshold is the more the user head needs to be turned in order to move the cursor. Sensitivity S and acceleration A determine directly how the mouth shift values from its translated into screen cursor movement speed. The greater values of these parameters are, the faster the screen cursor moves at the same mouth shift from the reference position.

**RESULT AND DISCUSSION**

When we run the program in matlab.Its extracts the face region from the input image and its analyzing the lip region of the input image. And then the lip region is analyzed by using lip contour

analysis algorithm ,By analyzing the lip region function, If the lip extracts function is closed, then the mouse cursor can move lift to right direction of the system

The second term function can run by matlab and its extracts the face region from the input image and its analyzing the lip region of the input image. When the Lip region function can be analyzed by lip contour analysis algorithm. By analyzing the lip extraction functioning data, If the lip region is open and also teeth region also visible it can show a pop up window message 'teeth open'. The mouse cursor will be move upward to downward

In this function the web cam can extracts the face region from the input image and its analyzing the lip region of the input image. When Lip region function can be analyzed by lip contour analysis algorithm, By analyzing the lip extraction functioning data, If the lip region is open ,it can show a pop up window message 'lip close 1'.The mouse can optimize the left click function

**CONCLUSIONS**

Results of the experiments carried out show that the effectiveness of the algorithm is sufficient for comfortable and efficient usage of a computer by anyone who does not want or cannot use a traditional computer mouse. We have presented our ALiFE beta version system of visual speech recognition. ALiFE is a system for the extraction of visual speech features and their modeling for visual speech. The system includes three principle parts, lip localization and tracking, lip feature extraction, and the classification and recognition of the viseme.

However, more work should be carried out to improve the efficacy of our lip-reading system. As a perspective of this work, we propose to add other consistent features (for example the appearance rate of tooth and tongue). New parameters will be defined and their extraction method will be tuned. Another research thread will be focused on increasing the number of recognized gestures



**Fig. 6.Lip region in close loop**



**Fig. 7. Teeth open**



**Fig. 8.Mouse close1**

## REFERENCES

1. X. Zhang, R. M. Mersereau, M. A. Clements and C. C. Broun, "VisualSpeech Feature Extraction for Improved Speech Recognition", *In Proc. ICASSP'02*; 1993-1996 (2002).
2. A. Nefi an, L. Liang, X. Pi, L. Xiaoxiang, C. Mao and K. Murphy, "A couple HMM for Audio-Visual Speech Recognition" *In Proc. ICASSP'02*, 2013-2016 (2002).
3. M. Kass, A. Witkin, D. Terzopoulos. "Snakes: Active contour models". *Int. Journal of Computer Vision*, 1(4), 321-331 (1988).
4. D. Terzopoulos and K. Waters, "Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models". *IEEE Trans Pattern Analysis and Machine Intelligence*, 15(6); 569-579 (1993).
5. P. Delmas, P.-Y. Coulon and V. Fristot, "Automatic Snakes For Robust Lip Boundaries Extraction", *In Proc. ICASSP'99*; 3069-3072 (1999).
6. P. S. Aleksic, J. J. Williams, Z. Wu, and A. K. Katsaggelos, "AudioVisual Speech Recognition Using MPEG-4 Compliant Visual Features," *URASIP Journal on Applied Signal Processing*, Spec. Issue on Joint Audio-Visual Speech Processing; 1213-1227(2002).
7. J. Luettin, N.A. Tracker and S.W. Beet, "Active Shape Models for Visual Speech Feature Extraction", University of Sheffi eld, UK, Electronic System Group Report N°95/44, 1995.