Available online at http://www.ijims.com

ISSN: 2348 – 0343

Lip reading for Malayalam Text Entry

Jahfar C

Dept of Computer Science, MCAS Vengara

Abstract
Voice recognition systems [2][5] are used for text entry in many situations. But they are not useful in highly noisy environments or even in areas where the sound is not a preferable option. Also it is not useful for physically challenged people like deaf and dump. Facial gesture interfaces [3][4] can be used for text entry in these environments and also for those people. Facial gesture interfaces, which respond to deliberate facial actions, have received comparatively little attention and also they are in its new born stage. Still researches are going on in these subjects and a comparatively good sign of success is obtained in Japanese language. In this paper we propose an interface which uses a coordinated action of hand and mouth [1], for a user to do Malayalam text entry.

Keywords: Text Entry, human computer interaction, lip movement, lip contour points, lip extractor, text generator

INTRODUCTION

In the proposed system, the system captures the lip movement corresponding to each character using a video capturing device. Most syllables take the form CV (C =consonant, V= vowel). Here, the vowels are identified from the captured video and consonants from key press.

Before using the captured image directly, the image is pre-processed in order to make it suitable for efficient lip identification. The preprocessing operation includes filtering for the noise reduction and segmentation for the purpose of localization of the lip area.

From the lip area, the lip contour points [9] (left, right, top and bottom) are identified. And then calculate the vertical and horizontal pixel distances between these points. It is obvious that for each character these values are unique. These values along with the key press will produce the character.

SYSTEM ARCHITECTURE

The schematic diagram (Figure 1) of the proposed system and detailed description of each module is given below.

A. Video Capturing Module:

This module captures the lip movements, and produces a video clip with frame rate of 30 frames per second or more.

B. Frame Extractor and Image filter:

Frame extractor split the videos in to static images in common format such as jpeg/jpg. The frame extractor will generate number of images and we will select only few of them for further processing to reduce the complexity.

Instead of random selection, we use images on frequent intervals, to ensure the reliability of our system. The selected images then go to the next module. Image filter improves the quality of the image by removing the noises.

Figure 1

C. Lip Detector:

RGB colour scheme of the image is not suitable for immediate processing as it contains a lot of mixed information about lightness. Another color scheme YC_BC_R [6] is used because it separates luminance, blue chrominance and red chrominance. These colors are very convenient as the mouth is considered to contain high red and low blue components in comparison with other face regions. The YC_BC_R colours can be computed from the RGB as follows[8]:

$Y = 0.229*R+0.587*G+0.144*B$
$C_B = -0.168*R-0.3313*G+0.5*B+128$
$C_R = 0.5*R-0.4187*G-0.0813*B+128$

With this colour scheme it is possible to start lip detection. To be more precise the pixels of the lips can be described by the probability, that current pixel is mouth pixel. Because lips pixels contain high red and low blue components the lips detection can be correlated to the red chrominance (C_R):

$$\text{Separator}(x,y)=C_R(x,y)^2 * (C_R(x,y)2 - K* C_R/C_B)^2$$

Where Separator returns the probability of the lip color

$$\mathbf{K} = 0.95 * \frac{\sum_{(x,y)} C_R(x,y)^2}{\sum_{(x,y)} \frac{C_R(x,y)}{C_B(x,y)}}$$

Where constant K fits final value in range 0….255.

The Separator function returns values in range of <0,255> where 0 represents values, which are not similar to lip color and 255 for colors very much similar to lip color.

A threshold value is calculated and we use this value to construct a matrix(M) that contains only 0's and 1's. If the original image have PxQ coordinates, the matrix also have a size of PxQ. If the Separator(x,y) returns a value greater than the threshold value, the M(x,y) will be 1, otherwise the M(x,y) will be 0.

Some other part of the face may have high concentration of red components which will give a value for the separator function greater than threshold value. So the part other than lip also will be represented by 1s. In order to eliminate these scattered 1s, the erosion operation is performed upon this matrix with suitable structuring element.

Figure 2

This matrix is given as the input to the image value calculator module.

D. Image Value Calculator: The image value calculator module finds the following values for each images using the algorithms shown below.

Horizontal distance between outer lip points	(HD)
Vertical distance between inner lip points	(VD)

At the learning phase we will calculate the value for each letter and store them in the data base, and then at the testing phase we calculate the value and compare it with the values in the data base for selecting a more closed value.

1) Algorithm to find left and right outer lip contour points:

```
Step 1:  Start
Step 2:  Scan each column from left to right
Step 3:  If 0 followed by 1 found at M(x,y)
            Right outer lip contour point=(x,y)
         End if
Step4:   Scan row x from right to left
Step5:     If a 0 followed by 1 found at M(x,l)
              Left outer lip contour point=(x,l)
           End if
Step 6:  Stop
```

2)Algorithm to find top and bottom inner lip contour points:

```
Step 1:  Start
Step 2:  n=(y+l)/2
Step 3:  Midpoint=(x,n)
Step 4:  Scan the column from Midpoint(x,n) to the
         top
Step 5:  If 0 followed by 1 found at M(m,n)
            Top inner lip contour point=(m,n)
         End if
Step 6:  Scan the column from Midpoint(x,n) to bottom
Step 7: If a 1 followed by 0 found at M(k,n)
            Bottom inner lip contour point=(k,n)
        End if
Step 8: Stop
```

E. Database:

Database for text entry have one table with the following fields.

Text_entry (HD,VD, letter)

Here HD and VD are horizontal and vertical distances respectively.

F. Database Comparator:

This module compares the image value generated by the image value calculator with the data base values, and selects a more closed result from the database.

G. Text Generator:

The text generator will produce the text corresponding to the lip movement. If lip movement is only taken as the input, it may not be reliable because of that, different words/letters may have the same sequence of lip movements. So a GUI based keypad is used for the selection of consonants and the lip movements represent the vowels. Both of these are combined together to generate the text.

Figure 3

CONCLUSION

In this paper we have proposed a system which allows text entry by action of hand and mouth for a phonetic based language, Malayalam. The principal advantages of this system are 1) It allows single keystroke text entry 2) Incorporation of this system with voice recognition application will increase it's reliability, especially in noisy environment.

REFERENCES

1. M.J. Lyons, C.H. Chan, and N. Tetsutani, "MouthType: Text Entry by Hand and Mouth", In proceedings of CHI 2004, pp. 1383-1386.
2. T.Chen, "Audiovisual speech processing. Lip reading and synchronization," IEEE Signal Processing Mag., vol. 18, pp. 9-21, January 2001.
3. Lei Gao, Y. Mukigawa and Y. Ohta, "Automatic Face and Gesture Recognition" , In proceedings of third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 181-186, 1998.
4. H. Shirgahi, S. Shamshirband, H. Motameni and P. Valipour, "A New Approach for Detection by Movement of Lips Base on Image Processing and Fuzzy Decision", World apllied science Journal 3(2), pp. 323-329,2008.
5.B.Shneiderman, "The limits of speech recognition",Communication of the ACM 42(9), pp.63-65,2000.
6. H. Noda, N.Takao and M. Niimi, "Colorization in YCbCr Space and its Application to Improve Quality of JPEG Color Images ", In proceedings of IEEE International Conference ICIP 2007 on image processing, pp. 385-388,2007.
7. I.S. MacKenzie, and R.W. Soukoroff, "Text Entry for mobile computing: Models and methods, theory and practice." Human-Computer Interaction 17, pp.147 -198, 2002.
8.Yang Yang, Peng Yuhua,Liu Zhaoguang," A Fast Algorithm for YCbCr to RGB Conversion", In proceedings of IEEE International Conference Rosemont, IL, USA.,pp.1490-1493,2007.
9. Abhay Bagai, Harsh Gandhi, Rahul Goyal, Ms. Maitrei Kohli and Dr. T.V.Prasad, . "Lipreading using neural networks. IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.4, April 2009.

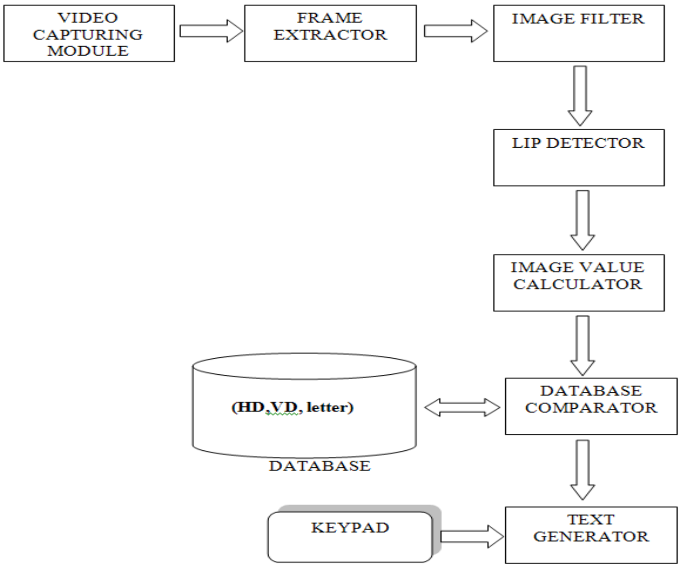

Figure 1: System Architecture

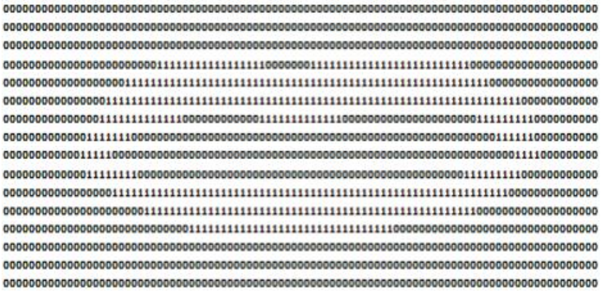

Figure 2: Sample Matrix (M)

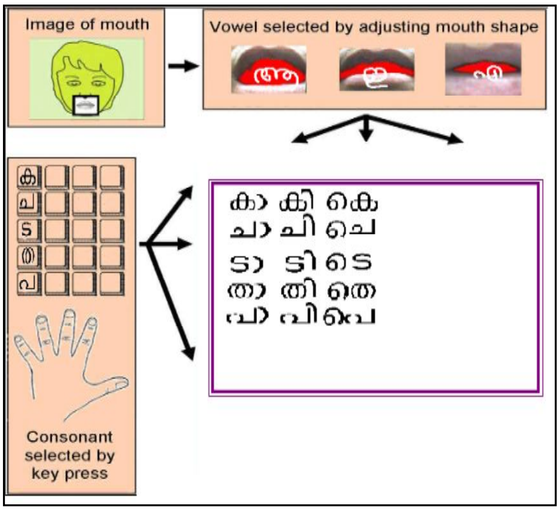
Image of mouth
Vowel selected by adjusting mouth shape
ആ
ഇ
എ
ക
ച
ട
ത
പ
കാ കി കെ
ചാ ചി ചെ
ടാ ടി ടെ
താ തി തെ
പാ പി പെ
Consonant selected by key press

Figure 3: Snapshot of the user interface