Mouth Gestures as Human-Computer Interface

K.Meenakshi[1], Dr. A.Geetha[2], A.Raja Prabu[3]
[1, 3] Dept of Computer Science and Engineering
[2]Associate Professor, Dept of Computer Science and Engineering
[1, 2, 3] Annamalai University, Chidambaram, India

Abstract- *In this proposed work, a research work implementing mouth interface between human and computer . The purpose of this work is to compensate the people, especially who have disabilities in their hands which prevent them from using the mouse. The system is developed to make the people interact with the computer to operate all real time computer operations such as internet browsing, audio and video player, GPS, surgical operations etc. The proposed interface allows people to move cursor and perform mouse click operations using mouth gestures without any special hardware. Hence this work leads to the implementation of low cost human-computer interface(HCI) system. Haar-cascade classifier is used to detect faces and mouth regions. The mouth region is tracked for mouse cursor movement. The mouth-open and mouth-closed gestures are detected using a binarization approach for mouse click events. The work has been implemented on openCV, python programming environment with operating system as windows 10 and the system performs better than the existing systems.*

Keywords- Cursor Movement, Face Detection, Haar Cascade Classifier, Human-Computer interface, Mouth Detection, Mouth Gestures, Mouse Click Events.,.

I. INTRODUCTION

The main objective of the proposed work is to implement a hands-free mouse control system using mouth gestures to interact with the system without any special hardware with the computer. The proposed system is designed to interact with a computer using human mouth gestures for carrying out operations like window open and close, navigation between tabs, zooming etc.

1.1 HANDS-FREE HUMAN-COMPUTER INTERFACE

Human-computer interface(HCI) takes a live video from a camera pointed at a user and interpret the user's actions for the purpose of directing computation. The tracked pose of user's mouth gestures allows positioning of the cursor without the use of mouse, joystick, or keyboard.
The application of this technology is as a hands-free mouse substitute for disabled access. Users who have difficulty using a standard mouse could manipulate an on-screen cursor by changing their mouth gestures.

1.2 MOUTH GESTURE TRACKING FOR HCI

Human computer interfaces utilize facial measures that allow users to explicitly control the interface though the use of mouth gestures. These systems usually use the mouth gestures to control the cursor position and use mouth open state to trigger mouse clicks or other events. The aim of this work is to find new forms of interactions using mouth gestures suitable for users.

1.3 APPLICATIONS OF THE PROPOSED SYSTEM

The proposed system can be used for applications like:

- Education
- Medical
- Security
- Industry
- Mobile Operations

II. LITERATURE SURVEY

Todkar Mrunal S, Shah Sarang A., Kore Kaustubh R, Babar Divyata D, Shevade Snehal B,Prof. Gurunath G. Machhae[1] proposed a work extracting eye and nose feature. This technology is used for an application that it is capable of swapping mouse with human face for interactionwith PC. Facial features such as nose tip, eyes are traced and tracked to use their movements for performing mouse events. Co-ordinates of the nose tip in the video feed are interpreted to develop co-ordinates of the mouse on screen. The left/right blink control left/right events for click. The external device will be webcam for the video stream.

Piotr Dalka, Andrzej Czyzewski [2] Implemented a multimodal human-computer interface (HCI) called LipMouse, 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 over 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 an artificial neural network and utilizes various image features of the lip region. An accurate lip shape is obtained by the means of lip image segmentation using fuzzy clustering.

Shadman Sakib Khan, Md. Samiul Haque Sunny, Shifat Hossain, Eklas Hossain [3] developed Human Computer Interface(HCI) control personal computers with high efficiency. The main three features of this interface are nose tracking cursor control, auto brightness control and display control based on the presence anddetection of valid human face. The proposed system is low cost and exhibits inherent security and power saving capabilities. MATLAB environment is used to perform the experiments.

P. John Hubert and M.S. Sheeba [4] developed multimodal human-computer interface (HCI) called Lip Mouse, allowing a user to work on a computer using movements and gestures made with his/her mouth only. User face images are captured with a standard webcam. 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.

Anitha C, M K Venkatesha, B Suryanarayana Adiga, [5] **p**roposed a real-time application for detection and tracking of facial features. The detection and tracking algorithms have to be very robust and efficient with least or zero false positives and false negatives. This work provides a novel combination for detection and tracking purpose. In this paper a robust mouth region extraction and tracking algorithm is proposed that works in real-time. The region of interest for our application requires the face and mouth regions. a novel technique for extracting the mouth region automatically. The proposed technique detects and tracks the mouth in either closed or opened state. the color components for skin tone and lips extraction.

Marcelo Archajo Jos, Roseli de Deus Lopess [6] presented a Lip control system of innovative human–computer interface specially designed for people with tetraplegia. This paper presents an evaluation of the lower lip potential to control an input device, according to Fitts' law (ISO/TS 9241-411:2012 standard). The results show that the lower lip throughput is comparable with the thumb throughput using the same input device under the same conditions. These results establish the baseline for future of research studies about the lower lip capacity to operate a computer input device.

Neha Nikhade1, Maninee Chaki2, Nazneen Shaikh3, A.B.Lamgunde4 [7], developed a HCI that aims to improve the interactions between users and computers by making computers more user friendly .This paper presents a system for eye blink detection and head movements for communication between man and machine. Facial features (eyes and nose) are detected based on which actions such as head movements and eye blinks are performed so that cursor can carry out appropriate functions on screen. Our system provides a comprehensive solution to control computer mouse cursor movement with eye-blinks and head movement for people with special needs. The user may perform the clicking of mouse through eye-blinks. For detection of face candidates our basic strategy is use of Six-Segmented Rectangular Filter (SSR) along with a support vector machine (SVM) for face verification. The patterns of between-the-eyes are tracked with update template matching in face tracking

Piotr Dalka, Andrzej Czyzewski [8], presented a lip movement tracking and lip gesture recognition based human computer interaction (HCI). This system is developed to control the system using lip movements. The mouth region is calculated in the lower part of the face region and is used to track lip movements. Three lip gestures are tracked mouth open, sticking out tongue and "O" shaped lip.

III. PROPOSED SYSTEM

In the proposed work, a human-computer interface is implemented using mouth gestures, which allows user to work on computer using movement and gestures made with user mouth. The workflow of the proposed work is shown in Figure 1.

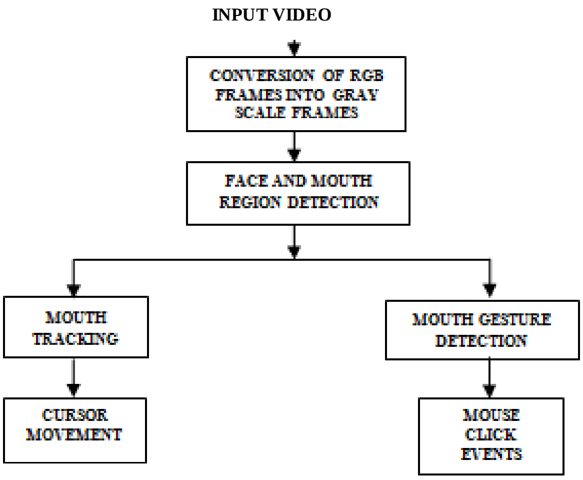

Fig.1 Work Flow of the Proposed Methodology

3.1 FACE DETECTION

The Haar Cascade classifier algorithm is used to detect faces. The classifier is configured with a minimum size of 150 x 150 pixels and a scaling factor of 1.1. The biggest face in the image is chosen as a region of interest.

3.1.1 FEATURES USED BY HAAR CASCADE CLASSIFIER

In Haar-cascade classifier, the system is provided with several numbers of positive images (like faces of different persons at different backgrounds) and negative images (images that are not faces but can be anything else like chair, table, wall, etc.), and the feature selection is done along with the classifier training using Adaboost and Integral images.

Fig.2 Haar Features Applied to a Image

In general, three kinds of features are used in which the value of a two rectangular feature is the difference sum of the pixels within two rectangular regions. These regions have same shape and size and are horizontally or vertically adjacent as shown in Figure 2. In the case of three rectangular features values are computed by taking the sum of two outside rectangles and then subtracted with the sum in a center rectangle. Moreover, in the four rectangle feature value is computed by taking the difference between diagonal pairs of the rectangles as shown in Figure 3.

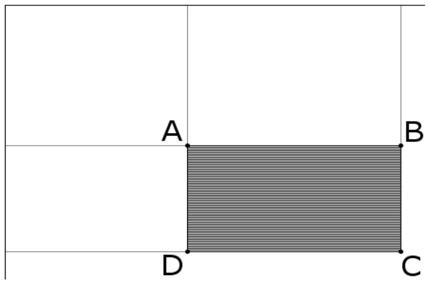

Fig.3 Four Rectangle Feature

3.1.2 *INTEGRAL IMAGE*

The speed of feature evaluation is also a very important aspect since almost all object detection algorithms slide a fixed-size window at all scales over the input image. The features can be computed at any position and any scale in the same constant time as below

$$SAT(X,Y) = (X + A)n \sum_{x' \le X, Y' \le y} I(X',Y') \quad ....(2)$$

All the features can be computed very fast and in constant time for any size by means of two auxiliary images. For upright rectangles the auxiliary image is the Summed Area Table SAT(x, y). SAT(x, y) is defined as the sum of the pixels of the upright rectangle ranging from the top left corner at rectangle to the bottom right corner.

For 45° rotated rectangles the auxiliary image is defined as the Rotated Summed Area Table RSAT (x, y). It gives the sum of the pixels of the rectangle rotated by 45° with the right most corner at (x, y) and extending till the boundaries of the image.

So if SAT[x,y] is the original image and RSAT[x,y] is the rotated integral image then the integral image is computed as shown below.

$$RSAT(X,Y) = \sum_{x' \le x x' \le x - |y-y|} I(X',Y) \quad ....(3)$$

It only takes two passes to compute both integral image arrays, one for each array. Using the appropriate integral image and taking the difference between six to eight array elements forming two or three connected rectangles, a feature of any scale can be computed. Thus calculating a feature is extremely fast and efficient. It also means calculating features of various sizes requires the same effort as a feature of only two or three pixels. The detection of various sizes of the same object requires the same amount of effort and time as objects of similar sizes since scaling requires no additional.

3.1.3 CASCADE CLASSIFIER

The face detection can be performed by cascade classifier as shown in Figure 4 using haar like features. In training a boosted cascade of weak classifiers includes two major stages: the training and the detection stage. The detection stage uses the Haar based models.

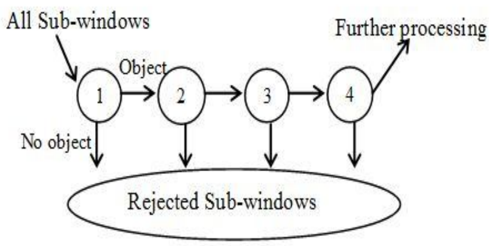

Fig.4 Haar Cascade Classifier

Each stage in the cascade reduces the false positive rate and decreases the detection rate. A target is selected for the minimum reduction in false positives and the maximum decrease in detection. Each stage is trained by adding features until the target detection and false positivesrates are met (these rates are determined by testing the detector on a validation set). Stages are added until the overall target for false positive and detection rate is met.

3.2 MOUTH DETECTION

To identify the location of mouth, the detected face region is used for estimating region of mouth. The mouth will be in the bottom region of the face and so the bottom region of face portion is taken as reference, From that portion there a four regions detected for drawing a rectangle over the mouth, the height and width of the corners from the bottom row of the segmented face, Based upon the height and width of the four corners of the rectangle mouth region is estimated.

3.3 MOUTH TRACKING FOR MOUSE CURSOR MOVEMENT

The proposed system is aimed at creating an interface between user and the computer using mouth gestures. The mouth region is tracked for moving the mouse cursor over the computer screen.

The algorithm for cursor movement is given below:

- Cursor movement is calculated based on the centroid movement from face and mouth regions occurred in images
- The average of the centroid of point between face and mouth is used as a reference to calculate the cursor's movement over the screen.
- The centroid point detected is used as a reference to move the cursor .If face and mouth are not detected the cursor stays at the last position.
- Screen and cursor coordinates are mapped for the cursor movement according to the length of the windows.
- Upper and Lower window coordinates are calculated to map the cursor movement according to the monitor as given below:

Mx=(mx,[lower_mx,upper_mx]) ---------(4)
My=(my,[lower_my,upper_my])

- After mapping, the coordinates point between the face and mouth region points are given as input to the cursor library.

3.4 MOUTH GESTURES FOR MOUSE CLICK EVENTS

The activation of the mouse buttons use the changes that occur in the mouth region like mouth-open and mouth-close states which are detected as mouth gestures. A histogram equalization is applied on the mouth region to reduce changes in lighting and shadows production and subsequently a binarization technique is applied using a threshold of 40 to mouth region. The image with the mouth open will have a greater number of black pixels than the mouth closed state. When the click events are invoked the mouth is in opened state.

The largest number of black pixels is identified by calculating contours of the mouth region, if the mouth is in open state for more than 1.5 seconds it is classified as a mouth open gesture.

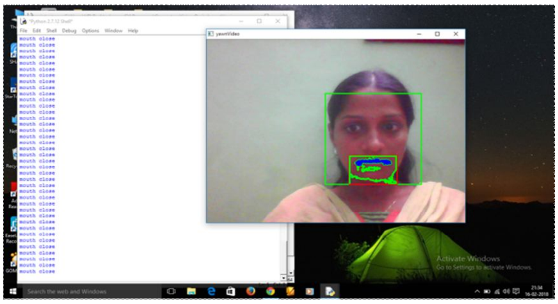

Fig. 5 Mouth Closed State

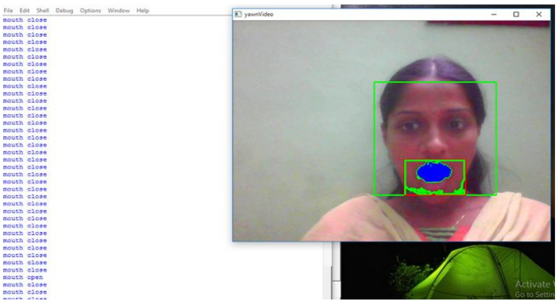

Fig. 6 Mouth Open State

The modules of the proposed work are implemented on OpenCV and PYTHON. The implementation details are given below.

IV. IMPLEMENTATION

4.1 FACE DETECTION

The Haar Cascade Classifier is loaded for detecting face from input video sequence and the faces are detected using ROI by drawing a rectangle over the detected regions. The face is detected as shown in Figure 7.

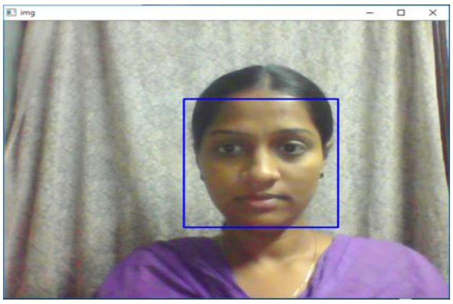

Fig.7 Face Detection

4.2 MOUTH DETECTION

The mouth region is estimated based upon the height and width of the corners from the bottom row of face detection as shown in Figure 8.

.

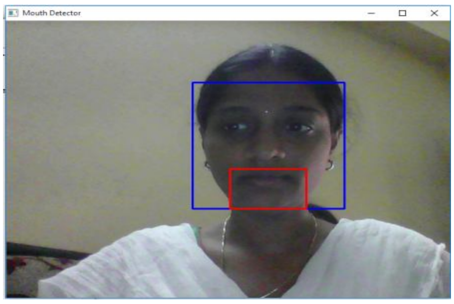

Fig.8 Mouth Region Detection

4.3 MOUSE CURSOR MOVEMENT

The movement of the cursor is implemented using the centre position as reference from nose and mouth,the input values and screen coordinates are calculated to move the cursor over the screen as shown in Figure 9 to Figure 11.

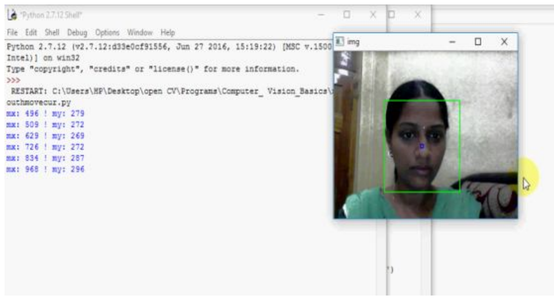

Fig.9 Cursor Movement Using Mouth Tracking

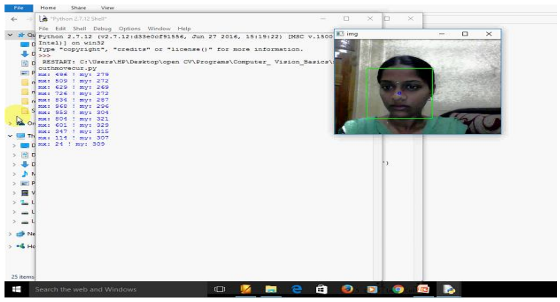

Fig.10 Moving Cursor Towards Right Side Of Screen

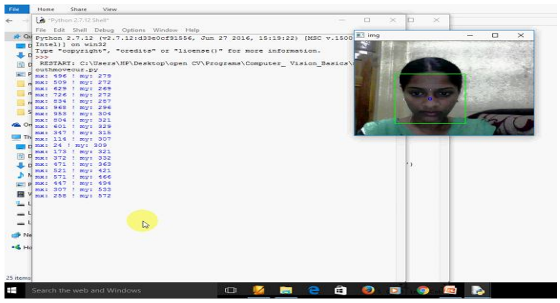

Fig.11 Moving Cursor To Down Side Of The Screen

4.4 MOUSE CLICK EVENTS

Mouth gestures are classified into mouth open and close states. The mouse operations are implemented by opening the mouth. The number of black pixels in the mouth region is calculated and the mouth is stated as open when the black pixels are in higher values. By calculating these values the mouse click actions are performed.The output screenshots are shown in Figures 12 thru 16.

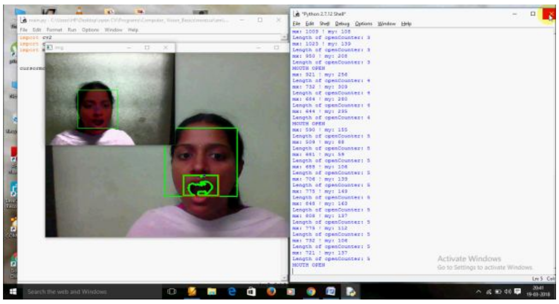

Fig.12 Closing the Program by Mouth-Open

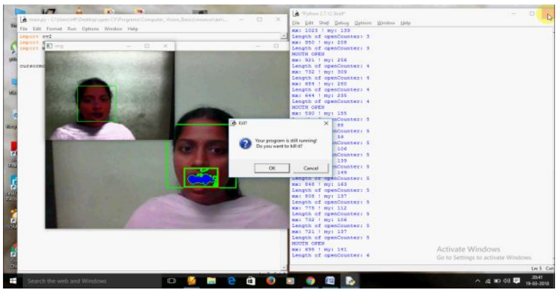

Fig.13 Left Click By Mouth-Open

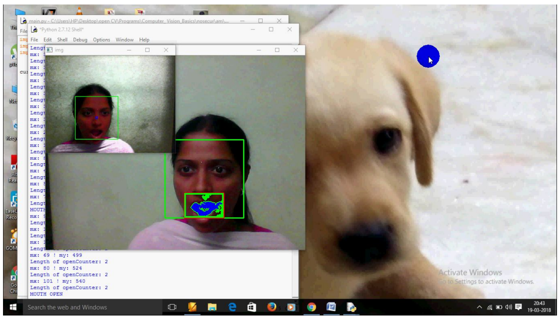

Fig.14 Right Click by Mouth-Open

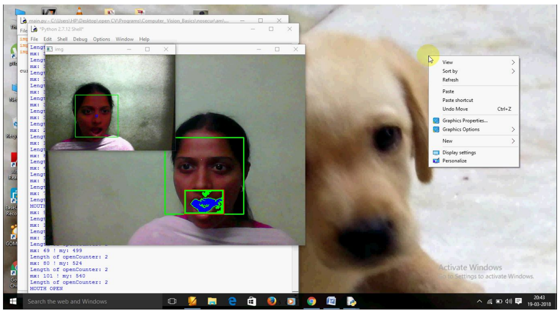

Fig.15 Right Click by Mouth-Open

V. EXPERIMENTAL RESULTS

In this section the experimental results obtained for the proposed approach is presented. The experiments are carried out in PYTHON SHELL in Windows 10 operating system in a computer with Intel Celeron processor with 2.16 GHz with 2GB RAM.

A live video is captured using in-built Web Camera from 20 different subjects. To carry out mouse click events, the subjects are made to open their mouth 50 times /video for just 1 sec during the capture of their images.

The performance of the proposed system is measured based on the accuracy as shown in Table 1 and Figure 16.

Table 1: Accuracy of the Proposed Work

	SUBJECTS	No. of Left Clicks	No. of Right Clicks	Accuracy(%)	
				Left Click	Right Click
1	Subject 1	49	50	98	100
2	Subject 2	46	48	92	96
3	Subject 3	47	48	94	96
4	Subject 4	42	49	84	98
5	Subject 5	45	44	90	88
6	Subject 6	45	46	90	92
7	Subject 7	44	46	88	92
8	Subject 8	43	44	86	88
9	Subject 9	46	44	92	88
10	Subject 10	39	42	78	84
11	Subject 11	45	48	90	96
12	Subject 12	43	46	86	92
13	Subject 13	46	43	92	86
14	Subject 14	42	46	84	92
15	Subject 15	48	48	96	96
16	Subject 16	40	47	80	94
17	Subject 17	42	45	84	90
18	Subject 18	45	46	92	82
19	Subject 19	44	48	88	96
20	Subject 20	46	44	92	88
Average Accuracy for Left and Right Clicks				88.8	91.7

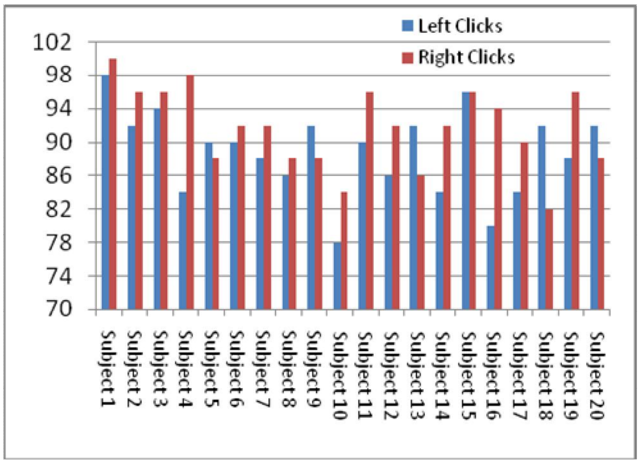

Fig.16 Accuracy for Different Subjects

From the above experimental results, It is proved that the proposed system provides an average accuracy 88.8 % for Left click and 91.7 for Right click events

VI. CONCLUSION

This proposed system has been developed to enable people to interact with computer using mouth gestures considering operations like window open and close, navigate between Tabs, zooming, etc. The algorithm has been implemented to detect the user's face and mouth in an office-type environment. On the other hand, the algorithm is developed to move the position of the cursor on the screen relative to camera from the real-time video captured. The left and right mouse click operations are performed by opening the mouth. This application has been tested with 10 subjects interacting with computer applications. This system performs better in a good lighting office-type environment. The system can be extended for more mouth gestures for different applications.

VII. ACKNOWLEDGEMENT

I express my deep sense of gratitude to my guide Dr. A. GEETHA, M.E., Ph.D., Associate Professor, Department of Computer Science and Engineering for her persistent interest in this work and her guidance and constant encouragement during this work without which the work would not have been materialized.

REFERENCES

[1] Todkar Mrunal S, Shah Sarang A., Kore Kaustubh R, Babar Divyata D, Shevade Snehal B,Prof. Gurunath G. Machhae, “Gesture Based Desktop Control”, International Journal of Innovative Research in Science Engineering and Technology, vol. 5, no. 2, February 2017.

[2] Piotr Dalka, Andrzej Czyzewski, “human-computer interface based on visual lip movement and gesture recognition”, International Journal of Computer Science and Applications, vol. 7, no.3, pp. 124-139, 2010.

[3] Shadman Sakib Khan, Md. Samiul Haque Sunny, Shifat Hossain, Eklas Hossain, “Nose tracking cursor control for the people with disabilities: An improved HCI”, International Conference on Electrical Information and Communication Technology (EICT), DOI: 10.1109, December 2017.

[4] P. John Hubert and M.S. Sheeba, “Lip and Head Gesture Recognition Based PC Interface using Image Processing”, Biomedical & Pharmacology Journal, vol. 8, pp. 77-82, March- 2015.

[5] Anitha C, M K Venkatesha, B Suryanarayana Adiga, “Real Time Detection and Tracking of Mouth Region of Single Human Face”, International Conference on Artificial Intelligence, Modelling and Simulation, pp. 297-303, 2015.

[6] Marcelo Archajo Jos, Roseli de Deus Lopes, “Human–Computer Interface Controlled by the Lip”, IEEE journal of biomedical and health informatics, vol. 19, no.1, pp. 302 – 308, January 2015.

[7] Neha Nikhade1, Maninee Chaki2, Nazneen Shaikh3, A.B.Lamgunde4, “Mouse Handling Using Facial Features”, International Journal of Advanced Research in Computer and Communication Engineering, vol.4, no.2, pp. 176-178, February2015.

[8] Piotr Dalka, Andrzej Czyzewski, ”Lip Movement and Gesture Recognition for a Multimodal Human-Computer Interface”, Institute of Electrical and Electronics Engineers, ISSN: 1896-7094, vol. 3, pp. 451-455, March 2009

[9] Miyoung Nam, Minhaz Uddin Ahmed, Yan Shen, and Phill Kyu Rhee, “Mouth Tracking for Hands-free Robot Control Systems”, International Journal of Control, Automation, and Systems, ISSN:1598-6446, DOI 10.1007/s12555-012-0473-7, pp. 628-636, 2014.

[10] Deepshikha Bhargava1 , L. Solanki2, Satish Kumar Rai, “Real Time Gesture Based Control: A Prototype Development”, Advancement in Science and Technology, DOI. 10.1063/1.4942708, pp 020026-1–020026-6, 2016.

[11] Yo-Jen Tu, Chung-Chieh Kao, Huei-Yung Lin. Chin-Chen Chang, “Face and Gesture Based Human Computer Interaction”, International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 8, no.9, pp. 219-228, 2015.

[12] Mrunalinee Patole , Pooja Athalye, “Controlling Computer with Human Lip: HCI”, Journal of Software Engineering & Software Testing, vol. 2, pp. 11-21, 2017.