OPERATION OF PC THROUGH A LIP MOUSE: A REVIEW

Ritu Chouhan[1], Shreya Sharma[2] and Anuroop Gaur[3]

[1,2,3] Department of Electronics and Communication Engineering,
B.K Birla Institute of Engineering and Technology, Pilani, Rajasthan-333031, (India)

ABSTRACT

A Lip mouse system is a non-obtrusive method that helps a user to work on a computer using movements and gestures made with his/her mouth only, especially for handicapped people; it is a great source of interaction with a Computer. The system tracks the computer user's movements with a video camera and translates them into the movements of the mouse pointer on the screen. The entire procedure includes detection, gesture recognition, region extraction, localization, shape approximation and cursor movement. This paper presents a review of the work done on a Lip Mouse for a computer and also presents the advantages and disadvantages of various approaches.

Keywords: Extraction, Face Detection, Human Computer Interface, Localization, Segmentation.

I INTRODUCTION

A Human Computer Interface (HCI) system called "The Lip Mouse" is evaluated in this paper. It tracks a user's movements with a video camera and translates them to the movements of the mouse pointer on the screen. Human-Computer Interface (HCI) systems are key components in allowing persons with severe disabilities to communicate with family, friends and caregivers. In many cases, persons with severe disabilities are non-verbal and have limited voluntary movement that greatly hinders their ability to share their needs and desires with others. An assistive HCI device like the "Camera Mouse" [1, 3] not only allows severely disabled people to communicate their wants, but also allows them to use the computer for educational and recreational and other purposes.

Lip image segmentation and lip movement tracking is a cumbersome task, mainly because of a very small contrast between lips and a face skin. The earlier approaches required user preparation, such as placing marks on a user face or particular make-up, but the newer approaches don't.

Lip image is usually segmented by the means of transforming RGB color space into desired space such as-CIE-LUV, HSV or YCbCr. Authors have also proposed a new transformation called a chromatic curve map. An automatic lip segmentation algorithm has been described based on the wavelet multi-scale edge detection across the discrete Hartley transform [4].

Another method for lip segmentation utilizes creating a lip shape model and fitting it to a lip image. Lip shape models may be based on deformable templates, active contour models or active shape models and they generally use a set of feature points to approximate the lip contours [5].

Color dissimilarity between lip and skin and a spatial distance from an ellipse approximating lip shape in order to facilitate lip segmentation has been combined by another method. In this method various mouth shapes are used for lip reading. Another system is able to control the mouse based on eye and mouth movements.

Lip Mouse is based on the principle of any usual camera mouse. A camera mouse system is usually composed a video camera for capturing video frames and a processing unit like a computer which uses image processing algorithm to convert the motion events in video frames to mouse operations. The algorithm is usually formed from two different modules- a visual tracking module and a mouse control module. The visual tracking module retrieves motion information from the video, and the mouse control module specifies the rules of control [1]. Face detection proposed by Viola and Jones based on statistic methods is most popular among the face detection approaches. This face detection is a variant of the AdaBoost algorithm which achieves rapid and robust face detection. They proposed a face detection method based on the AdaBoost learning algorithm using Haar features that detected the face successfully with high accuracy. However the accuracy of the method is still not enough when this method is used to detect facial feature [14].

A Lip Mouse is thus a great means of Human Computer Interface and helps the handicapped people to use computer easily.

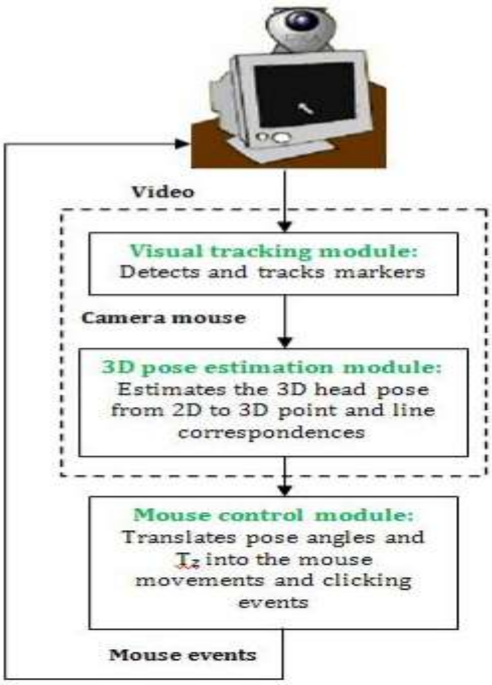

Figure 1: The block diagram of a camera mouse implementation [6].

II HUMAN COMPUTER INTERFACE ELUCIDATION

As technology is getting to play an important role in enhancing the quality of life, more and more research is being directed towards natural Human-Computer Interaction. People usually hope to use the most natural and convenient ways to express their intentions and interact with the environment. The most traditional means of giving commands to household appliances is button pressing, and is used in the remote control panel.. Such kind of operation, however, is not easy to use, especially for elders or visually disabled people who are not able to distinguish the buttons on the device. In this regard, gesture-based interaction offers an easy and alternative way in a smart environment for such people [2].

Lip Mouse is a novel, patent-pending, contactless, human-computer interface that allows a user to work on a computer using movements and gestures made with his or her mouth only. Lip Mouse is an application running on a standard PC computer. It requires only one hardware component: a display-mounted, standard web camera that captures images of the user face. The main task of Lip Mouse is to detect and analyze images of user's mouth region in a video stream acquired from a web-camera. All movements of mouth (or head) are converted to movements of the screen cursor. Various parameters regarding speed of the cursor movement may be set according to user preferences [2]. Lip Mouse also detects different mouth gestures. Each gesture may be associated with an action, which may be freely chosen by a user. Possible actions include clicking or double-clicking various mouse buttons, moving mouse wheel and others. Many actions may be defined as single or continuous ones. The single actions are executed only once, in the very moment when a new gesture is detected; continuous actions are executed as long, as a gesture is kept. For example, opening mouth gesture may be connected with an action executing left mouse button click in the moment or single click of the mouse, when the mouth is opened, or with an action that keeps left mouse button pressed as long, as a user keeps his/her mouth open. Additionally, based on the mouth (head) movement speed, Lip Mouse detects two other gestures (head shaking "Yes" and "No") that consist in shaking a head energetically in vertical or horizontal direction [7].

It allows a user to configure Lip Mouse according to his preferences. In the right part of the window, vertically-flipped video frames from the camera are displayed (a user sees his mirror-like reflection). In the frames, the mouth region is denoted with a rectangle, and the lip shape is denoted with an ellipse. Before a user starts working with Lip Mouse, a short calibration lasting about 30 seconds needs to be executed. During the calibration, the user is asked to perform some head movement and gestures according to the instructions seen on the screen. The purpose of the calibration is to tune Lip Mouse to detect gestures made by the user in the current lighting conditions. The target users for the tool are people who, for any reason, cannot or do not want to use traditional input devices. Therefore Lip Mouse is a solution enabling severely disabled and paralyzed people to use a computer and communicate with the surrounding world. No user adaptation, such as placing marks on the face, is required in order to successfully work with Lip Mouse [8].

Figure 2: Lip Mouse application main window [1]

III AN OVERVIEW OF LIP MOUSE METHODOLOGY

Figure 3 presents a scheme of the algorithm used in Lip Mouse. First, a user's face is detected in every image frame captured by a web camera. Further stages of the algorithm are restricted to the ROI containing the user's face. Then, a mouth region is localized and its shift from the reference mouth position is calculated. This shift is directly used to move a screen cursor. Simultaneously, a small region (blob) placed on user lips are found in mouth region. This region is used as a starting condition for an iterative method for lip shape extraction. Lip shape and lip region image features are used by an intelligent decision system utilizing an artificial neural network to classify gestures made by a user [9].

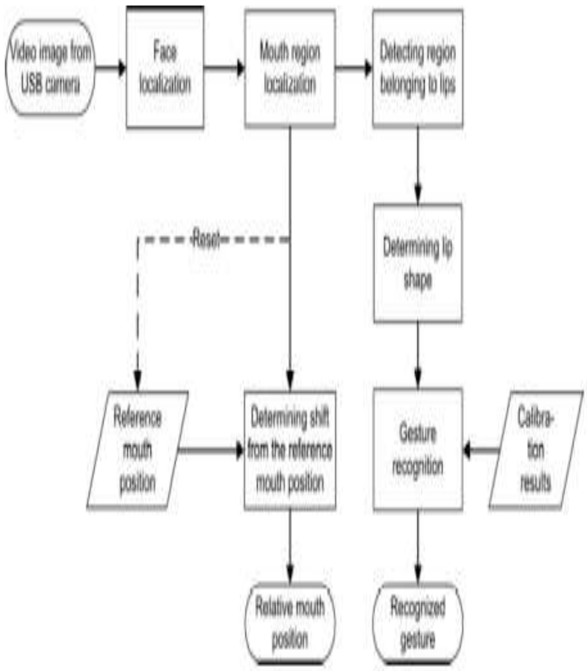

Figure 3: Block diagram of Lip mouse methodology [1].

IV DETECTION, LOCALISATION, EXTRACTION AND RECOGNITION

A cascade of boosted classifiers working with Haar-like features is used to detect a user's face in images captured by a web camera. It is a very efficient and effective algorithm for visual object detection. Each classifier in the cascade consists of a set of weak classifiers based on one image feature each. Features used for face detection are grey-level differences between sums of pixel values in different, rectangle regions in an image window. The window slides over the image and changes its scale. Image features may be computed rapidly for any scale and location in a video frame using integral images [1]. For each window, the decision is made whether the window contains a face; all classifiers in the cascade must detect a face for the classification result to be positive. If any classifier fails to detect a face, the classification process is halted and the final result is negative. Classifiers in the cascade are trained with Ada Boost algorithm that is tuned to minimize false negatives error ratio. Classifiers in the cascade are combined in the order of increased complexity; initial classifies are based on a few features only [8]. This makes possible for the algorithm to work in the real time because it allows background regions of the image to be quickly discarded while spending more computation on promising regions. Face detection algorithm finds location of all faces in every video frame. It is assumed, that only one person is present in the camera field of view therefore only the first face location is used for further processing. In order to increase speed of the face detection and to make sure that the face is large enough to recognize lip gestures, the minimal width of a face was set to the half of the image frame width. Sample results of face detection and mouth region. The mouth region is localized arbitrary in the lower part of the face region detected. It is defined by the half-ellipse horizontally centered in the lower half of the face region. The width and the height of the half-ellipse is equal to the half of the height and half of the width of the face region, respectively. Only the mouth region of each video frame is used for lip gesture recognition [6].

Figure 4: Face detection and mouth region finding [2].

The Lip gesture is recognized by using any of the techniques such as Artificial Neural Network etc. A feature vector for the ANN contains parameters describing image region containing lips only. A feed-forward ANN with one hidden layer is used to detect lip gestures. Each image frame is classified independently [3].

The lip region is then extracted using ellipse phenomenon. An ellipse is formed around the lips which can be varied according to the shape of the lip. . The centre of the square is located at the centre of the ellipse, and the length of sides is equal to the half of the width of the whole mouth region [5].

Figure 5: Lip shape detection [5].

In order to facilitate lip gesture recognition by ANN, an algorithm for determining region of the image containing lips only must be very precise and has to be robust against head movements in the vertical and horizontal directions. In order to locate lips, a series of face image transformations is performed. First, an image is smoothed with Gaussian filter and converted from RGB color space to the CIE LUV space. The U component of LUV space and the third component C3 of DHT transform are used for further processing, because they provide distinct, linear separation of lip and non-lip areas. These two components are multiplied in order to increase the differences.

The lip image features are thus extracted using image extraction and the cursor movements are made. A reference position is set and the distance in reference to this position is responsible for how far the mouse cursor moves [10].

V LIP CONTROL SYSTEM USING A JOYSTICK

Another method can be using a Lip Joystick. A great advantage of the joystick is the possibility of soft and free movements in any direction. Chin control and mouth joystick depend on neck move human–computer interface with a headset and a joystick positioned in front of the lower lip. The studies to develop the prototype showed that the lip control must be head mounted in order to capture the lower lip muscles movements. The joystick, as an interaction method, was chosen because it is easy to use, provides an intuitive control, is compatible with the lips movements; the body must be fixed and the head must be able to move freely [5].

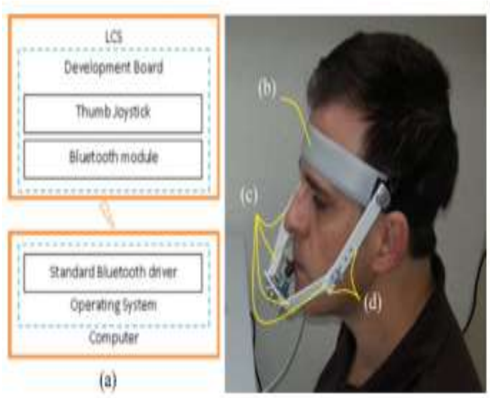

Figure 5: (a) Architecture (b) Head support (c) Joystick support (d) Calibration holes [7]

VI ADVANTAGES AND DISADVANTAGES

Lip mouse is a very feasible method for the physically disabled people to communicate with a computer. It is a novel, patent-pending, and contactless human-computer interface that allows a user to work on a computer using movements and gestures made with his or her mouth only [1].

Thus it becomes simpler for the people with any of the disabilities, especially the users with Tetraplegia. But there are certain limitations also for the same. Vibration during the drive and body spasms (common in spastic tetraplegia) can generate false commands. Outside the wheelchair, the user has no control, due to the dependence of the apparatus on the wheelchair structure [3].

VII CONCLUSION AND FUTURE WORK

Various techniques for the lip gesture that have been studied here prove to be very useful and easy for use. In future, the work can be done on improving each of the techniques and new ideas can be discovered to make the methods easy. Various methods include Viola Jones, skin pixel detection etc. and each of the method have their own advantages and limitations. The system can be made faster and accurate, different search algorithm techniques can be used [12] and also can design the library software to auto-generate a folder for most used hand gesture by the user and discarding the least used one. This would make the search process faster and better for the user. Future work includes not only improvement of the designed strategy but also taking into account more challenges such as dynamic gestures. The final objective involves a system does not generate false commands and detects the lip gestures efficiently.

REFERENCES

[1] Piotr Dalka, Andrzej Czyzewski-Lip movement and gesture recognition for a multimodal human-computer interface, Proceedings of the International Multiconference on Computer Science and Information Technology, pp. 451 – 455.

[2] Hlaing Htake Khaung Tin-FACIAL EXTRACTION AND LIP TRACKING USING FACIAL POINTS, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.1, No.1, April 2011.

[3] Margrit Betke-The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access for People With Severe Disabilities, IEEE Transactions on Neural Systems and Rehabilitation Engineering, VOL 10, NO 1, MARCH 2002.

[4] Piotr Dalka-Human-Computer Interface based on Visual Lip Movement and Gesture Recognition, International Journal of Computer Science and Applications, Technomathematics Research Foundation Vol. 7 No. 3, pp. 124 - 139, 2010.

[5] Andrzej Czyzewski- Human-Computer Interface based on Visual lip movement and gesture recognition, International Journal of Computer Science and Applications, Technomathematics Research Foundation Vol. 7 No. 3, pp. 124 - 139, 2010.

[6] Masoomeh Nabati- Camera Mouse Implementation Using 3D Head Pose Estimation by Monocular Video Camera and 2D to 3D Point and Line Correspondences,, 2010 5th International Symposium on Telecommunications (IST'2010).

[7] Alireza Behrad- Camera Mouse Implementation Using 3D Head Pose Estimation by Monocular Video Camera and 2D to 3D Point and Line Correspondences,, 2010 5th International Symposium on Telecommunications (IST'2010).

[8] E. Perini, S. Soria, A. Prati, and R. Cucchiara- FaceMouse: A Human-Computer Interface for Tetraplegic People.

[9] Robyn Cloud, Margrit Betke, and James Gips-Experiments with a Camera-Based Human-Computer Interface System .

[10] Marcelo Archajo Jos-Human–Computer Interface Controlled by the Lip, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 19, NO. 1, JANUARY 2015.

[11] Roseli de Deus Lopes - Human–Computer Interface Controlled by the Lip, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 19, NO. 1, JANUARY 2015.

[12] A.W.M Smeulders and M. Worring, Content-based image retrieval at the end of the early years, in Proc. Of IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 22, Issue 12, pp. 1349-1380, 2000.

[13] Ijaz Khan, Hadi Abdullah and Mohd Shamian Bin Zainal, Efficient Eyes and Mouth Detection Algorithm using Combination of Viola Jones and Skin Colour Pixel Detection, International Journal of Engineering and Applied Sciences, June 2013. Vol. 3, No. 4.

[14] Wen Feng, A Novel Lips Detection Method Combined Adaboost Algorithm and Camshift Algorithm, 2nd International Conference on Computer Application and System Modeling (2012).