2012 International Conference on Communication, Information & Computing Technology (ICCICT), Oct. 19-20, Mumbai, India

Facial-Feature Based Human-Computer Interface For Disabled People.

[1] Kaushik Parmar,[2] Bhavin Mehta,[3]Prof. Rupali Sawant

Dept. of Information Technology
Sardar Patel Institute of Technology
Mumbai. Maharashtra, India
[1]kaushikparmar24@gmail.com,
[2]bhavin.m91@gmail.com,
[3]rupali_sawant@spit.ac.in

Abstract- **In recent years there has been an increased interest in Human-Computer Interaction Systems allowing for more natural communication with machines. Such systems are especially important for elderly and disabled persons. Face detection has always been a vast research field in the computer vision world, considering that it is the backbone of any application that deals with the human face (e.g. biometric systems). The paper presents a vision & feature-based system for detection of long voluntary eye blinks and interpretation of blink patterns for communication between man and machine. Supplemented by the mechanism for detecting multiple eye blinks, this paper provides a complete solution for building intelligent hands-free input devices. Due to recent increase of computer power and decrease of camera cost, it became very common to see a camera on top of a computer monitor. The described technique uses off-the self cameras that allow one for tracking nose features, eyebrows and head position robustly and precisely in both 2D and 3D coordinates. This tracking and monitoring allows user to give input to the computer machine and access the entire system in a hands free manner. The detailed theory behind the technology is presented in the paper and the results from running several perceptual user interfaces built with this technology are shown.**

Keywords-HCI; face detection; SSR; BTE; ROI; SVM;

I. INTRODUCTION

HCI can be described as the point of communication between the human user and computer. Typical input devices used nowadays for communication with the machine are: keyboard, mouse, trackball. Touch pad and a touch screen. All these interfaces require manual control and can't be used by person impaired in movement capacity. This fact induces the need for development of the alternative method of communication between human and computer that would be suitable for the disabled. Therefore the work on the development of innovative HCI attracts so much the attention of researchers all over the world [1]. For severely paralyzed persons, whose ability of movement is limited o the muscles around the eyes most suitable are systems controlled by eye-blinks since blinking is the last voluntary action the disabled person loses control of[2]. Eye blinks can be classified into 3 types: voluntary, reflexive and spontaneous. Spontaneous eye blinks are those with no external stimuli specified and they are associated with the psycho-physiological state of the person [3]. Voluntary eye blinks are results of the person's decision to blink and can be used as a method for communication. The eye-movement or eye blink controlled HCI systems are very useful for persons who cannot speak or use hands to communicate (hemiparesis, quadriplegia, ALS). The systems use techniques based mainly on infrared light reflection or electrooculography.

The example of gaze-communication device is Visionboard system [4]. The infrared diodes located in corners of the monitor allow for detection and tracking of the user's eyes employing the "bright pupil" effect. The system replaces the mouse and keyboard of a standard computer and provides access to many applications, such as writing messages, drawing, remote control, Internet browsing or e-mail. However, majority of users were not fully satisfied with this solution and suggested improvements.

More efficient system, based on passive eye and blink detection techniques, was proposed in [5]. The system enables communication using eye blink patterns. The spelling program and two interactive games were proposed for the users of this system.

The application of vision-based analysis of eye blink biometrics was demonstrated in [6]. The authors also examined the influence of ambiguous blink behaviors on specific HCI scenarios. In [7] the authors developed text editor, which is operated using EOG signals. Eye movements are responsible for the movement of the cursor and decision making is simulated by successive blinks. Electrooculography was also used in [8] to develop a system allowing physically impaired patients to control a computer as assisted communication. The application allowing for entering alphanumeric signs to text editor was controlled by eye blinks. Double eye blinking was used as a decision making signal to avoid errors resulting from physiological blinking. The results confirmed, that eye-movement interface can be used to properly control computer functions and to assist communication of movement-impaired patients.

The vision-based system presented is designed for the disabled users who are capable of blinking voluntarily. The proposed algorithm allows for eye blink detection, estimation of the eye blink duration and interpretation of the sequence of blinks in the real time to control the non-intrusive HCI. The employed image processing techniques are used to measure the area of the visible part of an eye and in this way analyze the eye blink dynamics.

The paper is organized as follows: the next section describes the Face monitoring methods, the detailed description of the proposed system and the applications designed for Human Computer Interaction using the system is given in Section 3 and 4.Experimental results are presented in Section 5 and Section 6 concludes the paper.

II. FACE MONITORING

Face detection has a number of applications, like eye blink dynamics analysis, eye detection and tracking or face detection. For this reason a number of eye blink and face detection techniques have been developed. They can be divided into contact methods requiring direct connection of the measuring equipment to the monitored person, and non-contact methods. Another classification divides eye blink detection techniques into non-vision and vision-based. Non-vision methods include electrooculography (EOG) and high-frequency transceivers. EOG uses the recordings of the electric skin potential differences collected from the electrodes placed around the eyes. The blink s are identified as the spike waveforms with amplitudes of 0.5/1 mV [9]. Another non-vision-based method for eye blink detection employs the high-frequency (~30GHz) transceivers. The head is illuminated by the transceiver and the analysis on the Doppler components of the reflected signal allow for identification of the eye blinks [10].

Camera based eye blink detection techniques are in general non-contact ones. They make use of some properties or characteristics of an eye that can be detected by the camera. Vision-based methods can be classified into two groups: active and passive. In active eye blink detection additional light sources are required in order to take advantage of the reflective properties of human eye. In contrast passive eye detection methods do not need special illumination. However, more advanced image processing methods need to be applied on the images taken in natural conditions.

Eye blink detection systems make use of number of combined image processing methods, usually including active infrared (IR) illumination. The system proposed by Grauman et.al [5] employs motion analysis and template matching for eye blink detection and analysis.

The passive vision based system proposed in this paper is constructed in such a way that it is as reliable, safe and user-friendly as possible. The assumptions are:

- Nonintrusive system
- Avoid specialized hardware and infrared illumination
- Real-time performance;
- Use only part of the processing power of the computer;
- Run on a consumer grade computer.

III. PROPOSED SYSTEM DESCRIPTION

The passive vision based system proposed in this paper is constructed in such a way that it is as reliable, safe and user-friendly as possible.

A. Face Detection

These methods however have improved over time and become more robust to lighting conditions, face orientation, and scale. Despite the large number of face detection methods, they can be organized in two main categories: Feature-based methods, and image-based methods. The first involves finding facial features (e.g. nose trills, eye brows, lips, eye pupils etc.) and in order to verify their authenticity performs geometrical analysis of their locations, areas, and distances from each other. This feature-based analysis will eventually lead to the localization of the face and the features that it contains.

Some of the most famous methods that are applied in this category are skin models, and motion cues which are effective in image segmentation and face extraction.

On one hand feature-based analysis is known for its pixel-accuracy features localization, and speed, on the other hand its lack of robustness against head rotation and scale has been a drawback of its application in computer vision. The second is based on scanning the image of interest with a window that looks for faces at all scales and locations.

B. Six Segment Rectangular Filter

Six Segmented Rectangular filters (SSR) divides the detected face frames into six rectangular pixel frames. At the beginning, a rectangle is scanned throughout the input image. This rectangle is segmented into six segments.SR filter is scanned on the image and the average gray level of each segment is calculated from integral image. Then, the bright-dark relations between each segment are tested to extract its center as a candidate point for Between-the-Eye.

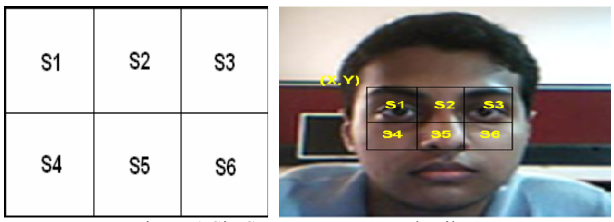


Figure 1 Six Segment Rectangualr Filter

C. Integral Image

In order to facilitate the use of SSR filters an intermediate image representation called integral image will be used. In this representation the integral image at location x, y contains the sum of pixels which are above and to the left of the pixel x, y. With this representation calculating the sectors of the SSR filter becomes fast and easy. No matter how big the sector is, we will need only 3 arithmetic operations to calculate the sum of pixels which belong to it. So each SSR filter requires 6*3 operations to calculate it.

We will be using feature based face detection methods to reduce the area in which we are looking for the face, so we can decrease the execution time. To find face candidates the SSR filter will calculate the integral image by making one pass over the video frame using these equations:

$$s(x, y) = s(x, y-1) + i(x, y) \qquad (1)$$

$$ii(x, y) = ii(x-1, y) + s(x, y) \qquad (2)$$

Where $s(x, y)$ is the cumulative row sum, $s(x,-1) = 0$, and $ii(-1, y) = 0$.

D. Support Vector Machine(SVM)

SVM stands for: Support Vector Machines, which are a new type of maximum margin classifiers. In 'learning theory' there is a theorem stating that in order to achieve minimal classification error the hyper plane which separates positive samples from negative ones should be with the maximal margin of the training samples, and this is what the SVM is all about.SVM takes as an input training data samples, where each sample consists of attributes and a class label (positive or negative).

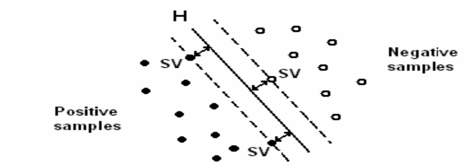
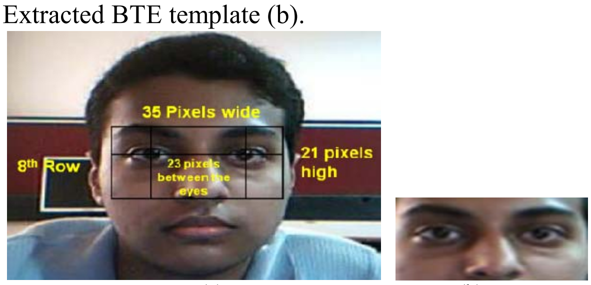


Figure 2 Support Vector Machines-Hyper planes

The data samples that are closest to the hyper plane are called support vectors. The hyper plane is defined by balancing its distance between positive and negative support vectors in order to get the maximal margin of the training data set. We will use the SVM to verify the 'between the eyes' template. Each template is defined as follows:

Extract a 35 wide by 21 high template where the distance between the eyes is 23 pixels, and they are located on the 8th row. Fig 3 Illustrates of how to extract a training template (a) and Extracted BTE template (b).

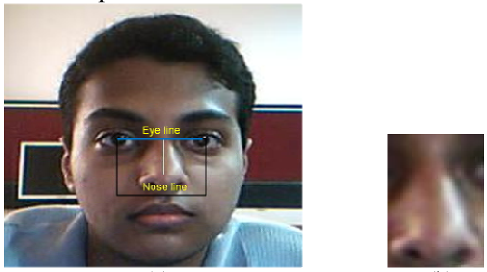


(a) (b)

Figure 3 Extraction of Template

IV. FIND NOSE TIP

Now that we located the eyes, the final step is to find the nose. We can see in the Fig 4(a) that the blue line defines a perfect square of the pupils and outside corners of the mouth; the nose tip should fall inside this square, so as shown in Fig 4(b), this square becomes our region of interest (ROI) in finding the nose tip.



(a) (b)

Figure 4 Extraction of Region of Interest

So the first step is to extract the ROI in case the face was rotated we need to rotate the ROI back to a horizontal alignment of the eyes

V. PREOCESS EXECUTION

Now that we found the facial features that we need, we will be tracking them in the video stream. The nose tip is tracked to use its movement. The eyes are tracked to detect their blinks, where the blink becomes the mouse. The flow of process is outlined in the Fig 5 shown below.

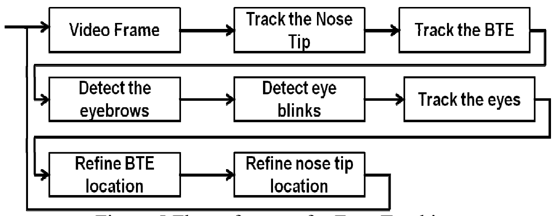


Figure 5 Flow of events for Face Tracking

A. Tracking the nose tip

Tracking the nose tip will be achieved by template matching inside the ROI.

B. Track the BTE

The BTE now represents a different concept from the one in the detection mode; it refers to a 15*15 template which is located between the eyes. The use of BTE tracking will be explained in eyes tracking and nose tip refinement processes. The BTE is tracked like the nose tip.

C. Motion Detection

To detect motion in a certain region we subtract the pixels in that region from the same pixels of the previous frame, and at a given location (x, y). If the absolute value of the subtraction was larger than a certain threshold, we consider a motion at that pixel.

D. Blink Detection

The blink detection process is run only if the eye is not moving (we consider the eye not moving if it didn't change its location in the past two frames). Because when a person uses the mouse and wants to click, he moves the pointer to the desired location, stops, and then clicks; so basically the same for using the face: the user moves the pointer with the tip of the nose, stops, and then blinks.

E. Eyes Tracking

If a blink was detected, the tracking process of the eye will be skipped and its location will be considered as the same one from the previous frame (because blink detection is applied only when the eye is still). Eyes are tracked in a bit different way from tracking the nose tip and the BTE, because these features have a steady state while the eyes are not. To achieve better eyes tracking results we will be using the BTE (a steady feature that is well tracked) at each frame after locating the BTE and the eyes, we calculate the relative positions of the eyes to the BTE. In the next frame after locating the BTE we assume that the eyes have kept their relative locations to it, so we place the eyes' ROIs at the same relative positions to the new BTE (of the current frame).

F. Refine BTE location

Despite the fact that the BTE is a stable tracking feature, we found by experiment that refining its location gives more accurate results. As we remember, we used the BTE to track the eyes; now it's time to use the eyes in refining the location of the BTE. Since we chose the BTE as a 15*15 template in the middle of the line that connects the eyes, we will center the BTE's ROI on that point and apply template matching to find the final BTE. The BTE's ROI is placed in the middle distance between the tracked eyes. The small square inside the ROI is the final BTE template (after refining).

G. Refine nose tip location

We will use two criteria to decide whether to refine the nose tip location or not. From we noticed that the blue line (the eyes line) defines the nose tip ROI. (So basically the length of the line that connects the BTE and the nose tip (we call it the nose line) should be smaller than the length of the eyes line); if not, nose refinement is applied. If the face plane is parallel to the capturing device; the nose line should be perpendicular on the eyes line). If the face plane was not exactly parallel or the face was rotated, the angle will not be exactly 90 degrees. We calculate the angle between the nose line and the eyes line; if it does not fall inside 90 ± 30 degrees nose refinement is applied. If refinement is needed, we center the nose tip ROI at the end of a line that is perpendicular on the eyes line in the BTE location, and which length is half of the eyes line length.

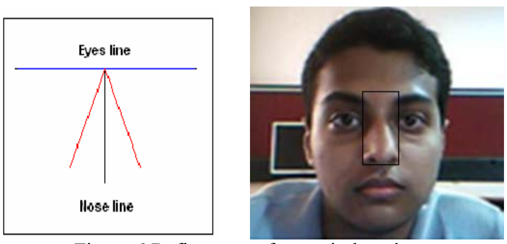


Figure 6 Refinement of nose tip location

VI. Experimental Results

A. In Detection mode:

The eyes and nose tip were located accurately when the following conditions were fulfilled**:**

- The face is not rotated more than 5° around the axis that passes from the nose tip (as long as the eyes fall in sectors S1 and S3 of the SSR filter).
- The face is not rotated more than 30° around the axis that passes from the neck (profile view).
- Wearing glasses does not affect our detection process.
- As for different scales it is best to get about 35 cm close to the webcam, because when the face is a bit far from the screen the program may detect a false positive especially when the background is jammed (crowded).

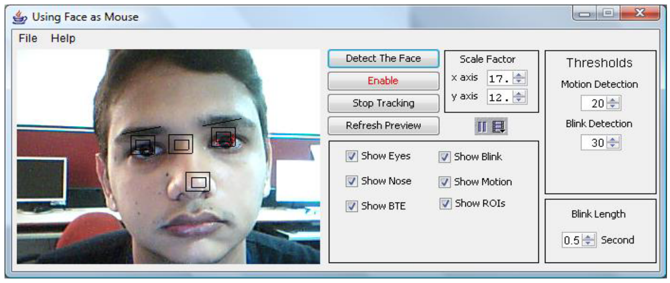


Figure 7 Result under detection mode

B. In Tracking mode

The results were very robust when:

- The frame rate is 20 fps and higher; the user can move very fast without having the program losing his facial features.
- The glasses reflect light and cause bright spots that sometimes force our program to lose track of the eyes.
- For the detection and tracking to be accurate and robust the lighting conditions must be set so the light is frontal in a way that it will spread evenly on the face, because side light will cause false face detection and will affect eventually the tracking process.

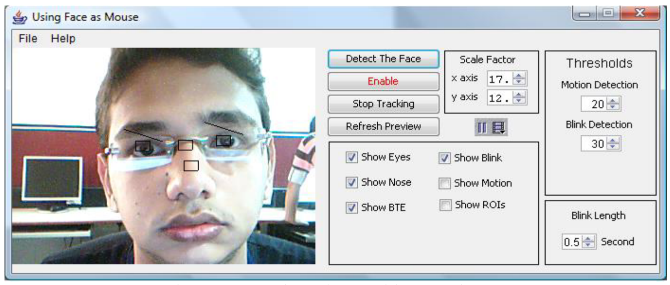


Figure 8 Result under tracking mode

C. System Performance

The developed system was tested on 10 different healthy people, aged between 18-25 years. Each person was asked to use the system under assistance and perform their operations. The assessment of the effectiveness of the system was based on three measures: precision, recall and accuracy calculated according to formulas (3), (4) and (5).

$$precision = \frac{TP}{TP+FP} \quad (3)$$

$$recall \frac{TP}{TP+FN} \quad (4)$$

$$accuracy \frac{TP}{TP+FP+FN} \quad (5)$$

Where:

TP (True Positive) - no. of detected eye blinks when the blink actually appeared.

FP (False Positive) – no. of detected eye blinks when the eye blink did not appeared.

FN (False Negative) – no. of eye blinks that appear but were not detected by the system.

The possibility of head movement while using the system was also tested. The maximum head turn for correct eye blink detection was equal to ±40º.

VII. CONCLUSION

Obtained results show that the developed algorithm and its software implementation is viable alternative communication technique suitable for disabled persons. Performed tests demonstrate that the system is able to accurately distinguish between voluntary and involuntary blinks. This is an important aspect as it is used as an interface controlled by facial gestures. The important advantage of the proposed system is that the fact that it does not need prior knowledge of face location or skin color is not required, nor any special lighting. Moreover the system is passive (no IR used) and works in the real time. In many human-computer interfaces for the disabled additional hardware is required to be worn by the user, such as special head gear, sensors, or markers. Since our proposed system uses only the web camera, it is completely non-intrusive, and therefore more user-friendly and easier to configure.

VIII. REFERENCES

[1] D. Gorodnichy. On importance of nose for face tracking. Involume 3,pages 514–518, Singapore, 1995.Proc. Intern. Conf. on Automatic Face and Gesture Recognition (FG'2002), Washington, D.C., May 2002).

[2] R. Ruddarraju, et al, Perceptual user interfaces using vision-based eye tracking, in: the 5th international conference on Multimodal interfaces, Vancouver, British Columbia, Canada, 2003, pp. 227 – 233.

[3] M.Chau, M.Betke:"Real time eye tracking and blink detection using USB camera", Boston University Computer Science Technical Report No.2005-12, May 2005.

[4] Carolann Cormier, MS, CCC-SLP, ATP. *Establishing an Encoded Eye GazeCommunication ystem.* www.connsensebulletin.com/eyegaze.html on 6/3/04.

[5] T.Morris, P.Blenkhorn, F.Zaidi:"Blink Detection for Real-time Eye Tracking" Journal of Network and Computer Application, Vol.25, num. 2, pp 129-143, April 2002.

[6] Akamatsu, S., Sasaki, T., Fukamachi, H., and Suenaga,Y.(1993). Automatic extraction of target images for face identification using sub-space classification IEICE Transactions on Information and Systems (D), E76-D(10):1190-1198.

[7] R. Brunelli and T.Poggio. Face recognition: Features versus face and facial feature tracking and applications. In Proc. templates. IEEE Trans. on Pattern Analysis and Machine AVSP'98,pages 79–84, Terrigal, Australia, 1998. Intelligence, 15(10),pp. 1042-1052, 1993.

[8] G. Loy, E. Holden, R. Owens, 3D head tracker for an automatic lipreading system, in: Proceedings of Australian Conference on Robotics and automation (ACRA 2000), Melbourne, Australia, August,2000

[9] "Talking Face Video", Face and Gesture Recognition working group, IST-2000-26434,http://www.prima.inriaple.fr/FGnet/data/01-TalkingFace/talking_face.html

[10] http://www.cameramouse.com/

[11] Stringa,L.(1993).Eyes detection for face recognition Applied Artificial Intelligence, 7(4):365-382

[12] Nouse 'Use Your Nose as a Mouse' – a New Technology for Hands-free Games and Interfaces by D.O. Gorodnichy, S. Malik, Computational Video Group, IIT, National Research Council, Ottawa, Canada K1A 0R6, http://www.cv.iit.nrc.ca/research/Nouse

[13] www.mousevision.com

[14] Lee.C.H.,Kim,J.S., and Park,K.H.(1996). Automatic Human Face Location in a Complex Background Using Motion and Color Information Pattern Recognition, 29(11):1877-1889.

[15] *Velez, J.[Jose], Borah, J.D.[Joshua D.]*, Visor and camera providing a parallax-free field-of-view image for a head-mounted eye movement measurement system, *US_Patent*4,852,988,Aug1,1989

[16] *Chung, A.J.[Adrian J.]*, *Deligianni, F.[Fani]*, *Hu, X.P.[Xiao-eng]*, Extraction of visual features with eye tracking for saliency driven 2D/3D registration, *IVC(23)*, No. 11, 1 October 2005, pp. 999-1008.

[17] Zdzislaw S. Hippe and Juliusz L. Kulikowski, Eye-Blink Controlled Human-Comouter Interface for the Disabled, Human-Computer system Backgroundandapplication, http://www.springerlink.com/content/9148657LW5LTH651