International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 03, March-2015

Eye & Speech Fusion Technique to Interact With Computer

Payal Shivhare
Student of Computer Technology,
R.G.C.E.R., Nagpur - 441110,
Maharashtra, India

Shubhangi Patil
Lecturer of Computer Technology,
R.G.C.E.R., Nagpur - 441110,
Maharashtra, India

Lokesh Jindal
Student of Computer Technology,
R.G.C.E.R., Nagpur - 441110,
Maharashtra, India

Abstract **- In this world of technology everything is being computerized. But there are many people with several disabilities like handicapped people who are unable to take advantage of this technology. This system proposed an efficient approach to handle the computer. The first approach is to use the webcam to control the cursor, the second approach researches speech recognition as an input and the third approach deals with the Fusion of both Eye and Speech. As most of the new Computer System comes with built in Webcams and Microphones, it will be very beneficial and cost effective.**

Keywords- Face Capturing, Image Extraction, Harris Corner Detector, Contours, Cursor Controlling, Speech Recognition.

I. INTRODUCTION

Computer is most common device which is used by individuals all over the world. It is the most intelligent machine available for the computational purpose and lots of input output devices are developed to interact with it (like mouse, keyboard, joystick, scanner, printer and many other).The main motivation behind designing every new system is to make interactions in more humanly fashion. Every new device can be seen as an attempt to make the computer more intelligent and making humans able to perform more complicated communication with the computer.

In the world of humans most of the communications are done by using vision and sound. So this interface will prove better than any other available systems. The most important factor using this system is that the user does not have to be physically in contact with contact with the system. And if we consider noisy environment it will prove better than that.

In this paper we are proposing feature based detection and image localization algorithms [1], [2], [3]. The advantage using this algorithm is that it performs faster but it requires high quality images. Rather we can use combined algorithms such as neural network, genetic algorithm [4], [5], [7] which makes the use of low quality images. Here we are considering steady state fixes system which will continuously detect head movements of the user and capture images of eye. The eye features are extracted in real time using the images captured .From extracted eye features eye gaze is detected to control the cursor. It can be divided into two forms viz. Head mounted [6], [8], [3] and steady state fixed. Both will have their advantages and disadvantage depending on the environment they are used.

II. BACKGROUND

A. Human Eye Structure

The eye is not shaped like a perfect sphere; rather it is fused two- piece unit. The smaller frontal unit, more curved called the Cornea is linked to the larger unit called SCLERA black centre, the pupil are seen instead of the cornea due to cornea's transparency. The cornea segment is typically about 8 mm in radius. The sclerotic chamber constitutes the remaining 5/6, its radius is about 12 mm. The cornea and sclera are connected by a ring called limbus.

The eye is made up of three coats, enclosing three transparent structures. The outermost layer, known as fibrous tunic, is composed of cornea and sclera. The middle layer, known as the vascular tunic or uveas, consists of the choroid, cilliary body, and iris. The innermost is the retina, which gets its circulation from the vessels of the choroid as well as the retina vessels which can be seen in an ophthalmoscope.

B. Corner Detection

1. What Are Corners?

- Intuitively, junctions of contours.
- Generally more stable features over changes of viewpoint.
- Intuitively, large variations in the neighbourhood of the point in all direction.
- They are good features to match.

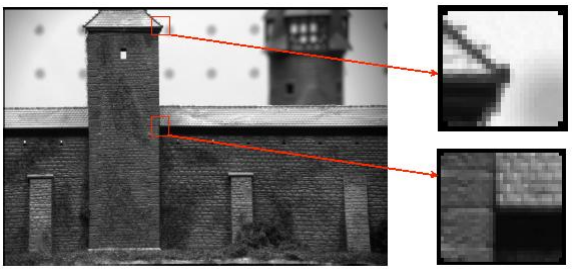


Fig 1: Corners (Junction of Contours)

2. Corner Points: Basic Idea

- We should easily recognize the point by looking at intensity values within a small window
- Shifting the window in any direction should yield a large change in appearance.

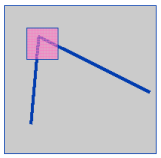


Fig 2: Detection of Corner Points.

3. Harris Corner Detector: Basic Idea

The most common idea of corner is defined by Harris [9]. The Harris corner detector [10] is based on local auto-correlation function of signal.

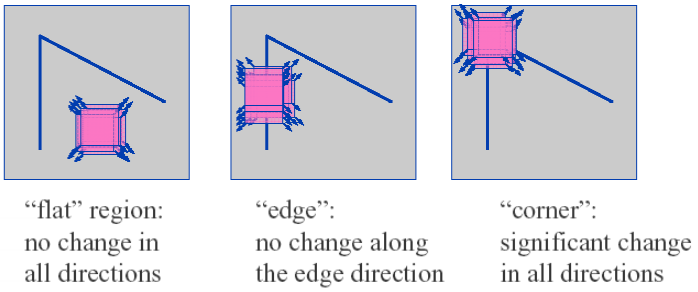
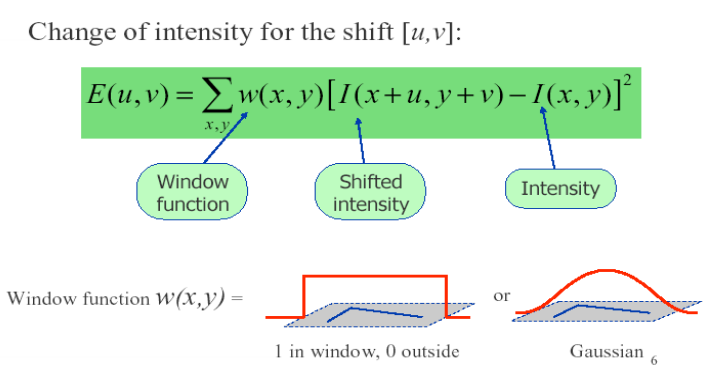


Fig 3: Mathematical approach for determining which case holds.

4. Harris Detector: Mathematics



III. DEVELOP THE ALGORITHM TO DETECT IRIS POSITION AND CALCULATE EYE-GAZE DIRECTION

In this section, we describe the proposed algorithm for position's iris detection and eye-gaze estimation. A single web camera is used for eye movement detection. This web camera can continuously capture images of user's eye and find out the information of eye movement.

In this paper, matching between the iris boundary model and the limbus are performed as an approach to detect the eye movements. The part between dark iris of eye and the white sclera is called as Limbus.

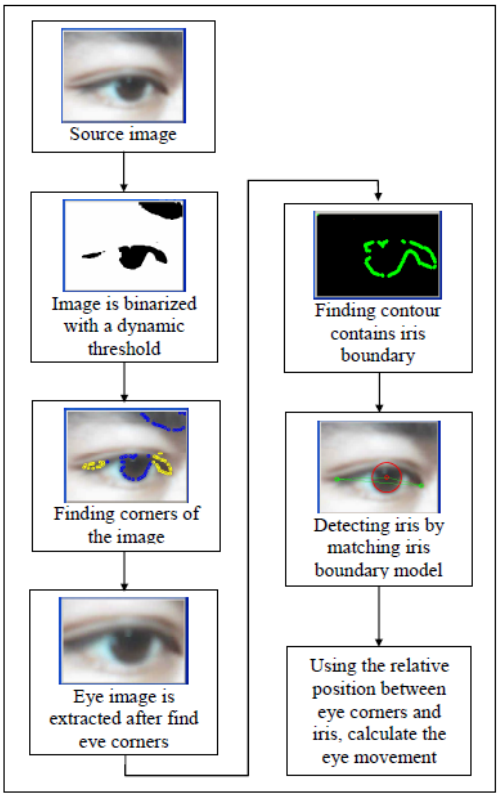


Fig 4: Structure of eye movement tracking algorithm

Above figure shows the flow chart of proposed algorithm:

- A dynamic threshold is used for binarized the source image.
- On the basis of two eye corners selected from corners of binary image, an eye image is extracted.
- The contour that contains iris boundary is selected after finding the contours of the eye image.
- To detect iris position, the contour obtained from previous step and the iris boundary model is matched.
- Calculate the eye movement using the relative position between iris and eye corners.

A. Thresholding the Source Image

The image which is captured from web camera is converted from RGB to Grayscale. After that, by using Peak and Valley method a dynamic threshold is found. At last, the gray scale is threshed with this threshold.

Peak and Valley method is described as follows:

- Let the number of pixel which have gray level (g) is denoted as h(g).
- Now, two most prominent peaks of histogram's image are found with constrain: g is a peak if
- $h(g) > h(g \pm \Delta g)$, $\Delta g = 1, \ldots, k$
- Suppose g_1 and g_2 be two highest peak, with $g_1 < g_2$.
- Find g as the deepest valley between g_1 and g_2: g is the valley if
- $h(g) < h(g')$, $g, g' \in [g_1, g_2]$
- g is used as the threshold.

Fig 5: Process threshold image.

B. Detecting Position of Eye Corners.

First, Harris corner detector is used to extract the corners from binary image. Then, based on Geometric structure of eye, the corners that belong to the eye are selected. Then, detect the position of two corners of eye. At the end, we extract a part of binary image that is limited by two eye corners and the remaining parts are rejected. Finally, the image of eye is found.

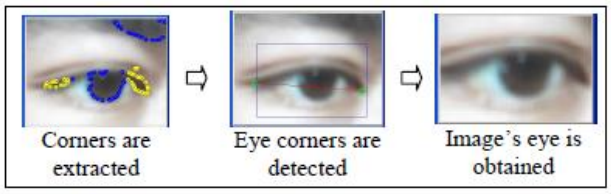


Fig 6: Detection of eye corner and extracting eye image.

C. Finding Contours

In this process, the contours that probably contain the iris boundary are found. For this, from the boundary of the binary eye image contours are detected. Based on the number of elements in this contours we choose one that have this number higher a certain amount. The contours which are found above contains iris's boundary which is extracted from original contours.

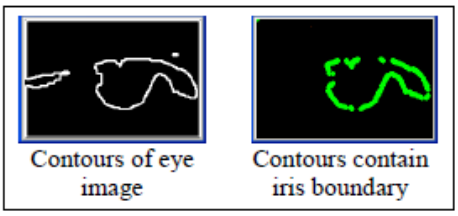


Fig 7: Results of finding contours.

D. Modelling Iris Boundary

We know that, the boundary of the iris is generally circular. While Haslwanter et al. was researching the robust method to find the eye movement and the iris boundary [4], they observed that the error in modelling the iris boundary as a circle is very small.

Therefore, three parameters are found after modelling the iris boundary as a circle which are, centre of iris (x_c , y_c) and radius $r_{min} < r < r_{max}$ which present the circle model.

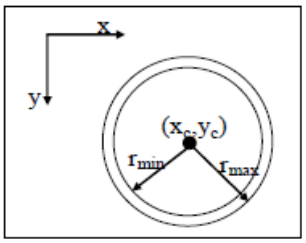


Fig 8: The iris boundary model

E. Matching Model

Position of iris centre is detected by matching between the iris boundary model and contours found in the previous section. Results of detecting iris position using model matching method are shown in Fig.

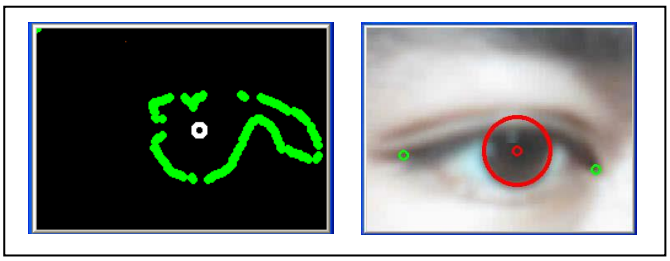


Fig 9: Results of iris position detection

F. Calculating the Eye-Gaze Direction

Three coordinates and their relationships are determined to compute eye-gaze direction.

These are listed as coordinates of:

- Image captured from web camera.
- Eye image based on positions of two eye corners.
- Monitor.

After detecting the position of iris centre, the position that the user is looking at monitor is calculated based on relationship of these coordinates.

1) Relationship between captured image and eye image coordinates

The central point of the line segment connects two eye corners is the origin of the eye image coordinate. Fig. depicts captured image and eye image coordinates and their relationship.

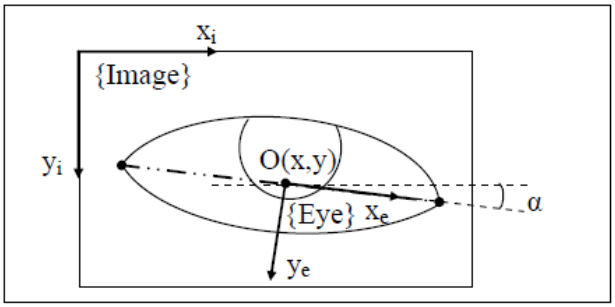


Fig 10: Captured image and eye image coordinates.

Relationship between the two coordinates is as follows,

$$\begin{cases} x_e = (x_i - O(x))\cos\alpha \\ y_e = (y_i - O(y))\cos\alpha \end{cases}$$

where O(x, y) is origin of eye image coordinate.

2) Relationship between eye image and monitor coordinates

Horizontal axes of monitor and eye image coordinates are opposite to each other because captured images incidentally obtains horizontal movements which are opposite to actual movements of the eye.

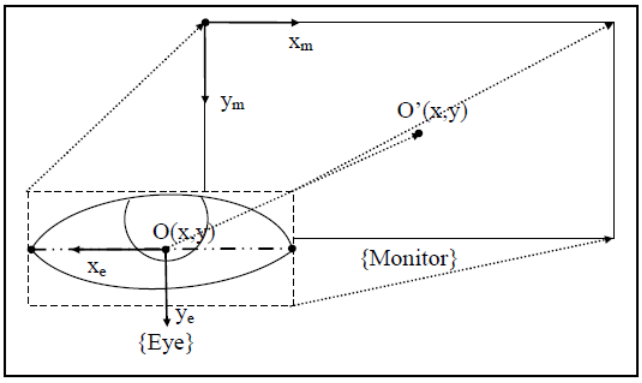


Fig 11: Eye image and monitor coordinates

IV. SPEECH RECOGNITION MODULE

For people main and natural mode of communication is speech and we have learned all related skills in our childhood & we do it without any instruction and continue to use it throughout our life.

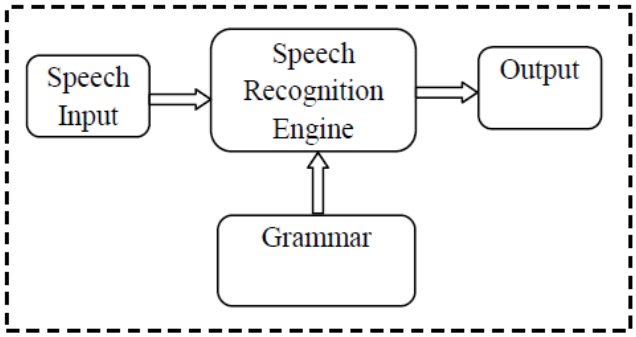


Fig 12: Speech Recognition Module

The accuracy of system can be measured along the following points:

A. Vocabulary

It is easy to differentiate between small set of words but as the size of words set increases error rate grows naturally. For Ex: Let us consider a set of 10 words. In this we can easily differentiate between this 10 words but if we take a set of 10000 words it will be very difficult and error rate may rise to 7 -45 % even if it will be difficult to recognize small set of word if it contains confusable words. For Ex: 26 letters of English alphabets can be pronounced with different toning but some alphabets are very difficult to discriminate like the E-set “B, C, D, E, G, P, T, V, Y”.

B. Speaker Independence vs. Dependence

System may be speaker dependent or independent. Speaker Independent system is difficult to achieve because parameters defined to particular speaker are highly specific. Therefore, single speaker system is good for use.

C. Types of Speech

Isolated speech, discontinuous speech, continuous speeches are the 3 types of speech. Isolate speech contains single words. Discontinuous speech contains full sentence in which word are artificially separated by silence. And naturally spoken sentences are continuous speech. In Isolated and discontinuous speech recognition, the words tend to be cleanly pronounced and word boundaries are detectable which makes it relatively easy.

D. Read vs. Spontaneous Speech

Speech that is either read from prepared scripts or spoken spontaneously is evaluated by Systems. Spontaneous speech tends to be peppered with influent speech like "uh" and "um", false starts, incomplete sentences, stuttering, coughing, and laughter which makes it more difficult and moreover the system must be able to deal intelligently with unknown words as the vocabulary is essentially unlimited.

E. Adverse Conditions

Range of adverse conditions can also degrade the performance of system. These include environmental noise (e.g., noise in a car or a factory); acoustical distortions (e. g, echoes, room acoustics); different microphones (e.g., close-speaking, or telephone); limited frequency bandwidth (in telephone transmission); and altered speaking manner (shouting, whining, speaking quickly, etc.).

V. FUSION OF EYE GAZE POINT AND SPEECH RECOGNITION

Fusion is a process of combining two or more things together to form a single entity.

Fusion processes are stated as follows:

A. Levels of Fusion.

One of the earliest considerations is to decide what strategy to follow when fusing multiple modalities. The most widely used strategy is to fuse the information at the feature level, which is also known as early fusion. The other approach is decision level fusion or late fusion [11] which fuses multiple modalities in the semantic space. A combination of these approaches is also practiced as the hybrid fusion approach [11].

B. How to Fuse?

There are several methods that are used in fusing different modalities. These methods are particularly suitable under different settings. The discussion also includes how the fusion process utilizes the feature and decision level correlation among the modalities, and how the contextual and the confidence information influence the overall fusion process [11].

C. When to Fuse?

The time when the fusion should take place is an important consideration in the multimodal fusion process. Certain characteristics of media, such as varying data capture rates and processing time of the media, poses challenges on how to synchronize the overall process of fusion. Often this has been addressed by performing the multimedia analysis tasks (such as event detection) over a timeline [11]. A timeline refers to a measurable span of time with information denoted at designated points. The timeline-based accomplishment of a task requires identification of designated points at which fusion of data or information should take place. Due to the asynchrony and diversity among streams and due to the fact that different analysis tasks are performed at different granularity levels in time, the identification of these designated points, i.e. when the fusion should take place, is a challenging issue [11].

D. What to Fuse?

The different modalities used in a fusion process may provide complementary or contradictory information and therefore knowing which modalities are contributing towards accomplishing an analysis task needs to be understood. This is also related to finding the optimal number of media streams [11] or feature sets required to accomplish an analysis task under the specified constraints. If the most suitable subset is unavailable, can one use alternate streams without much loss of cost-effectiveness and confidence?

VI. SYSTEM ARCHITECTURE

The system architecture consist different modules as shown in figure. Each module performs the specific task, The Eye tracking module uses various image processing techniques and the tracking algorithm for estimating the direction of the pupil. The Speech Recognition module recognizes the spoken words and compares them according to the specified grammar.

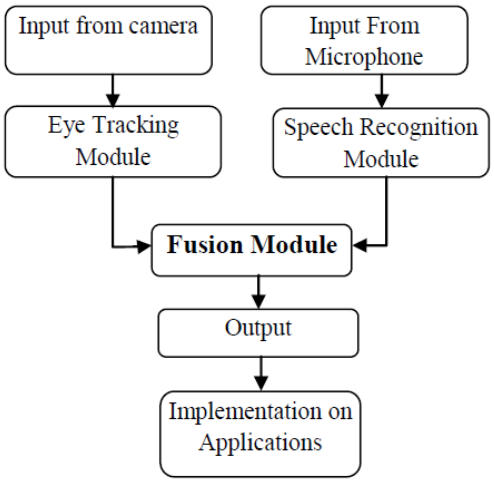


Fig 13: Fusion of Eye Tracking and Speech Recognition.

VII. EXPERIMENTAL RESULT

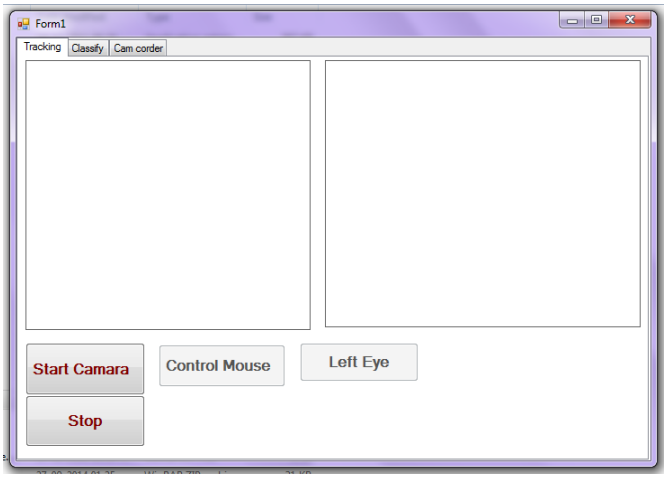


Fig 14: Main Form

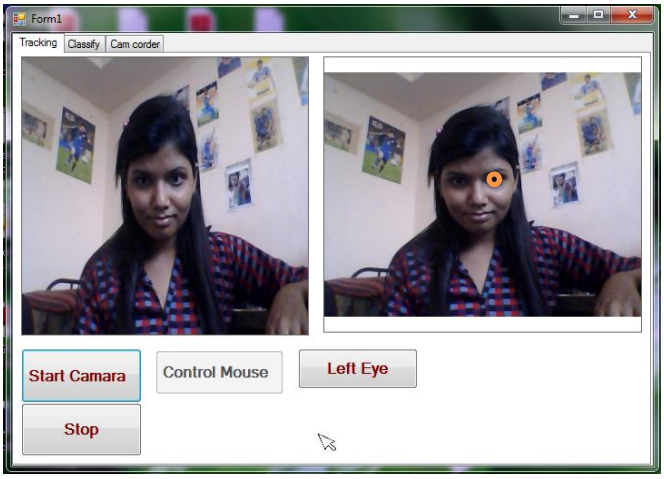


Fig 15: Captured Image

VIII. CONCLUSION

In this paper, a real time eye-gaze detection system is presented. We proposed an eye-gaze detection algorithm using images captured by a single web camera. The user with several disabilities can use this system for handling computer. A real time eye motion detection technique is presented. We verified the system accuracy by performing the experiments. Although the proposed algorithm may look rather simple, but it is able to detect the eye gaze with high successful rate.

IX. FUTURE WORK

Our future work will mainly concentrate on improving the accuracy of the proposed eye-gaze detection algorithm, making the system quicker and robust, and considering other eye movements such as the user blinking. The proposed system will be verified by numerous experiments with different users.

X. REFERENCES

[1] J. Zhu and J. Yang, "Sub pixel eye gaze tracking", IEEE Conference on Automatic Face and Gesture Recognition, pages 124-129, May 2002.

[2] S. Kawato and N. Tetsutani, "Detection and tracking of eyes for gaze camera control", Proceedings of the 15th International Conference on Vision Interface, 2002.

[3] X. Long, O. K. Tonguz, and A. Kiderman, "A high speed eye tracking system with robust pupil centre estimation algorithm", The 29th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, August 2007, pp. 3331-3334.

[4] R. Stiefelhagen, J. Yang and A. Waibel, "Tracking eyes and monitoring eye gaze", Proceedings of the Workshop on Perceptual User Interfaces, 1997 .

[5] S. Amarnag, R. S. Kumaran and J. N. Gowdy, "Real time eye tracking for human computer interfaces" , Proceedings of the International Conference on Multimedia and Expo, Volume 3, July 2003, pp. 57-60.

[6] S. Ramadan, W. A. Almageed and C. E. Smith, "Eye tracking using active deformable models", Processing of the Conference on Computer Vision, Graphics and Image , India, December 2002.

[7] T. Akashi, Y. Wakasa and K. Tanaka, "Using genetic algorithm for eye detection and tracking in video sequence", Journal of Systemics, Cybernetics and Informatics, Volume 5, No. 2, pp. 72-78, 2007 .

[8] R. Argue, M. Boardman, J. Doyle and G. Hickey, "Building a low-cost device to track eye movement", December 2004.

[9] C. Harris and M. Stephens, "A combined corner and edge detector", Proceedings of the 4th Alvey Vision Conference,1988, pp. 147-151.

[10] G. Derpanis, "The Harris corner detector", York University, October 2004.

[11] Pradeep K. Atrey, M. Anwar Hossain, Abdulmotaleb El Saddik, Mohan S. Kankanhalli, " Multimodal fusion for multimedia analysis: a survey", Multimedia Systems, Springer-Verlag 2010.