Gaze Tracking in Visible Spectrum Extending Head Pose Estimation

[1] Rohit P. Gaur, [2]Krupa N. Jariwala
[1]M.E. - GTU PG School, [2]Assistant Professor – SVNIT - Surat
[1]CO - Wireless and Mobile Computing
[1]GTU, Ahmedabad, India

Abstract- Gaze tracking in visible spectrum is a widely researched area. Many researches have been carried out up till now combining various methods in different scenarios. Even though the results are quite accurate, most researches either restrict the head movements or does not include the eye movements. This limits the feasibility of application in the gaze tracking domain. In this paper we have extended the head pose estimation and incorporated eye estimation to track gaze by using simple methods and approach. The approach is to estimate gaze from a simple webcam without restricting the head movements. The experimental results suggest that the accuracy is feasible enough in the x plane and needs an additional work to perfect the y plane as the eye movements are less visible here compared to the x plane.

Keyword- Gaze estimation, Head Pose estimation, Pupil tracking, Fusion, Visible Spectrum.

1. INTRODUCTION

Gaze estimation is the process of obtaining the point where the person is currently looking at. Gaze estimation has many applications in various domains. It can be used to see where a user is looking on the computer screen to estimate and develop an efficient display of information to the areas where the user tends to look more and the lesser visited areas can be prioritized accordingly. Another application of gaze estimation can be to read the behavioral reaction of a person under a specific condition. For example, where the person looks particularly at the time of accident. There are many methods to estimate the gaze of a person as this area is under research since many decades. Gaze tracking generally lie under two domains, namely, Visible Spectrum and Infrared Spectrum. The basic difference is of using an infrared source to produce a glint in the eye to estimate the pupil location and using that information to estimate the gaze. Visible spectrum is further classified in intrusive and non-intrusive systems. Intrusive systems are the ones in which a person has to wear a headgear which contains the camera to estimate the eye movements. Non-intrusive systems on the other hand does not have any such requirement. There is always a trade-off between restrictions and accuracy. If you remove the restrictions and provide the ease of use to the user end, you have to deal with the conditions leading to inaccurate results. A person, in general, tends to move his head during a normal interaction with the computer. Restricting the head movements has limited the usability of gaze estimation in the visible domain, however, it yields good accuracy. There are many methods that use eye tracking alone to estimate the gaze. Related work has been done by Laura Sesma et. al. [1] where they have used the eye corner and pupil center vector. The approach yields good results, however, as the eye corners move along with the eye movements, the accuracy is less. Head pose estimation is the estimation of the plane of head with respect to the camera. It consists of 6 parameters that are to be considered while estimating. These parameters are X, Y, Z, Yaw, Pitch and Roll. X, Y and Z are the 3 planes whereas Pitch, Yaw and Roll are the rotation about those planes respectively. Related work has been done by Michael S. et. al. [2] where they have taken the eye to mouth location and considered it as the facial plane to which the nose base to nose tip vector is the normal. Other related work has been done by P. Padeleris. et. al. [8] where they have used a depth camera to obtain head pose. We also refer the reader to [7] for going through a survey on models and methods of eye tracking, head pose and gaze estimation.

1.1 Motivation and Justification of proposed approach

The existing techniques and methods of estimating the gaze generally limit the head movements as the estimation is based of pupil tracking itself. Without the head movements, the application of the approach is restricted for the end user. On the other hand, head pose alone cannot justify the gaze as it works on the head movements and not on the gaze of the person. Our approach is to extend the existing head pose estimation by incorporating the pupil tracking information and exploit that fusion to estimate gaze. By this incorporation we are able to take away the restriction of head movements as well track the gaze in a non-intrusive domain using a simple web camera. The approach is laid down keeping it simple so that it can work for real time dealing with the complexities.

Our approach considers the offset between the vector normal and the eyes which is added to the gaze estimator in the beginning. Eye movements are then tracked using an image gradient algorithm for pupil estimation by F. Timm et. al. [3] which uses the dot products between the normalized displacement vectors to obtain the optimal pupil center. This pupil estimation information is then passed to the pose estimator to improve stability. Finally the kalman filter is used to obtain a linear estimate of the gaze from the incorporated algorithm.

1.2 Outline of the approach

Figure 1 shows the flow and outline of our approach. We are extending the work done by Michael S. et. al. [2] in our approach. First the face is detected from the input video of webcam using Viola Jones method. Facial features are then extracted by using haar cascades to identify them. Pupil detection and pupil tracking is carried out parallel to features extraction so as to save the computing time. Pupil tracking information is passed to the feature tracking block. Feature tracking works on the template matching of the features extraction by using Normalizes Sum of Squared Distances (NSSD) method. Head Pose is then estimated by calculating Yaw, Pitch and Roll values as mentioned in [2]. The output values of head pose and pupil estimation are then passed in kalman filter and gaze is estimated based on our proposed algorithm.

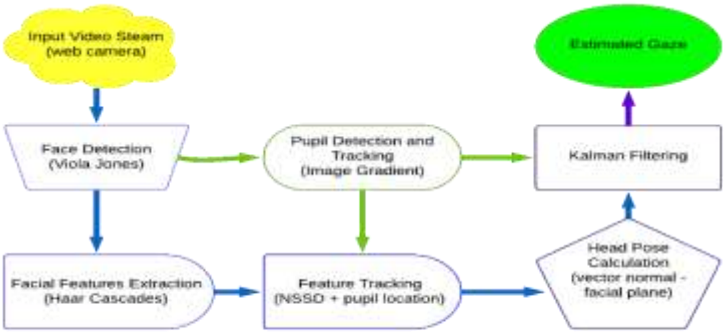

Figure 1: Flow and outline of the approach

2. INCORPORATING HEAD POSE AND PUPIL ESTIMATION TO TRACK GAZE

The existing head pose estimation algorithm works in the following manner:

- Detect the face, then the features such as eyes, nose and mouth until it is stable
- Track the features by using template matching, NSSD, in the next frames
- Calculate the position of cursor according to the face position and the rotation of eyes and nose.

The existing pupil estimation algorithm works in the following manner:

- Detect the face using Viola Jones method
- Locate the approximate box region of the eyes.
- ROI is defined for the next step
- Center of the image’s most prominent circle is then defined by using image gradients and dot products

In our fusion algorithm, we incorporated the above two independent algorithms in such a way that they support each other to produce a much more stable result that also extends the functionality of the overall system. Similar work has been carried out by R. Valenti [5] where they have used isophote method for pupil estimation and the head pose estimation is done with the help of a calibrated depth camera.

3. ALGORITHM FOR GAZE ESTIMATION IN VISIBLE SPECTRUM

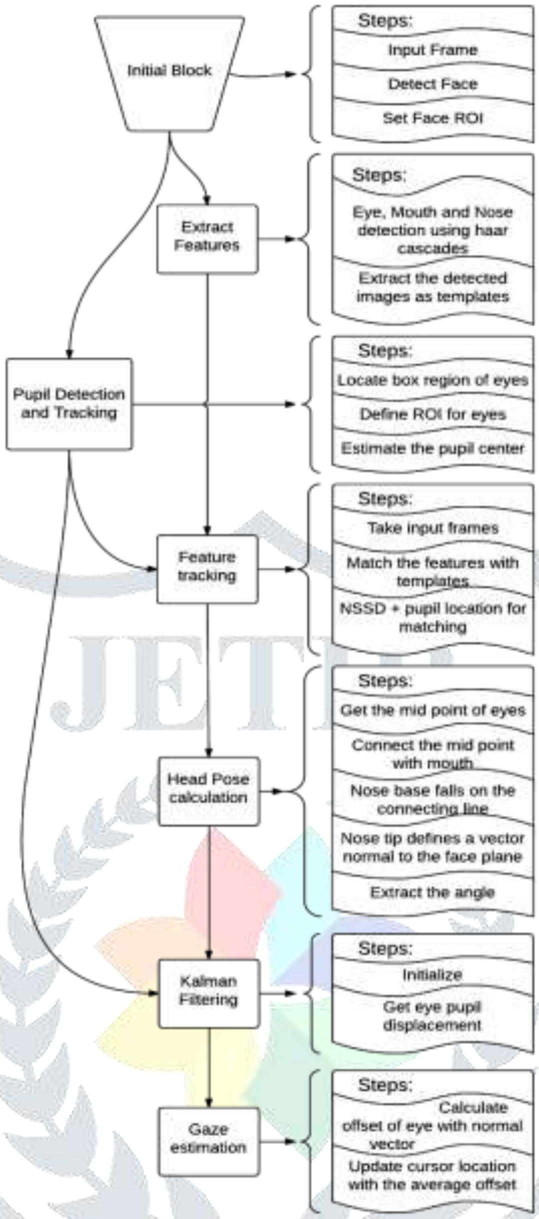

Figure 2 shows the algorithm used for estimating the gaze in visible spectrum extending the head pose estimation.

Detailed description of the algorithm:

INITIAL BLOCK
Frame is taken as input
Face is detected using Viola Jones method
ROI is set for the detected face to save the further computational time

EXTRACT FEATURES
Haar cascades are initialized on the face ROI to detect eye, mouth and nose from the face region
The detected features are extracted and saved for future use as a template

PUPIL DETECTION AND TRACKING
Box region is located for eyes using haar in the face ROI
ROI for eyes is defined
Pupil center is estimated using a function working on image gradients and dot products

FEATURE TRACKING
Frames are taken as input
Features are estimated using template matching. Previously extracted templates are used here.
Pupil center information is added to the eye feature tracking to improve the stability in the system

Template matching is carried out by using Normalized Sum of Squared Differences method

Figure 2: Algorithm for proposed System.

HEAD POSE CALCULATION

- The mid-point of eyes is calculated
- Mid-point is connected to the mouth
- Base of nose falls on this line and it forms the facial plane
- Tip of the nose forms a vector with the base of the nose. This vector is normal to the facial plane
- The vector normal forms an angle σ with a vector d which the normal vector of the image plane. Related mathematical derivation can be referred at [2] and the supporting theory can be referred at [6].

KALMAN FILTERING

- Kalman filter contains two step, predict and update.
- The predict step produces the state of current frame by using the state information of previous frame
- The update step gets the state information of current frame and combines it with the state produced in predict step and the refined state is finally produced.
- Eye pupil displacement is calculated by using predict and update steps.

GAZE ESTIMATION

- Offset of eye with the vector normal to the facial plane is computed and added to the cursor location.
- First the initial position of the pupil is defined.
- To stabilize its position, it is averaged with previous frames
- Suppose the initial eye position is a_n, then it is updated as:

$$a_n = (a_{n-1} * (n-1) + b_n)/n \quad (1)$$

  b_n, here is the detected pupil position in the nth frame.

- Pupil displacement is then calculated by:

$$d_p = b_n - a_n \quad (2)$$

- Kalman filter is used to obtain a stable and more precise movement of eye pupil. Suppose the displacement after kalman filter is d_p', then the cursor is calculated by:

$$c = h + \omega_p * d'_p \quad (3)$$

$$(c.x, c.y) = (h.x, h.y) + \left(\omega_p.x, \omega_p.y\right) * (d'_p.x, d'_p.y)$$

  Here c is the coordinates of the cursor and h is calculated head movement. ω_p is the weight to control how the eye pupil movement will affect the coordinates of the cursor.

4. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

4.1 Performance Parameters for Gaze estimation

In order to evaluate the feasibility and effectiveness of the proposed model, covering gaze estimation that extends a head pose estimation, the experiments were conducted in real time and the estimated gaze was compared with the ground truth values. To obtain the ground truth gaze estimate we asked the user to click in a full screen window and to look at the point of click while doing so. At every click, the gaze coordinates estimated by our algorithm were recorded along with the location of the click. The recording system is a 15.6-inch laptop with 1366 x 768 screen resolutions. The camera used is an integrated webcam working at a standard 26 fps. Gaze estimation is calculated for a 3x3 grid divided across the screen. The distance of face from the screen is roughly around 20 cms and the environment is well lighted.

4.2 Experimental Results

Figure 3 shows the difference between the actual positions and the estimated gaze positions.

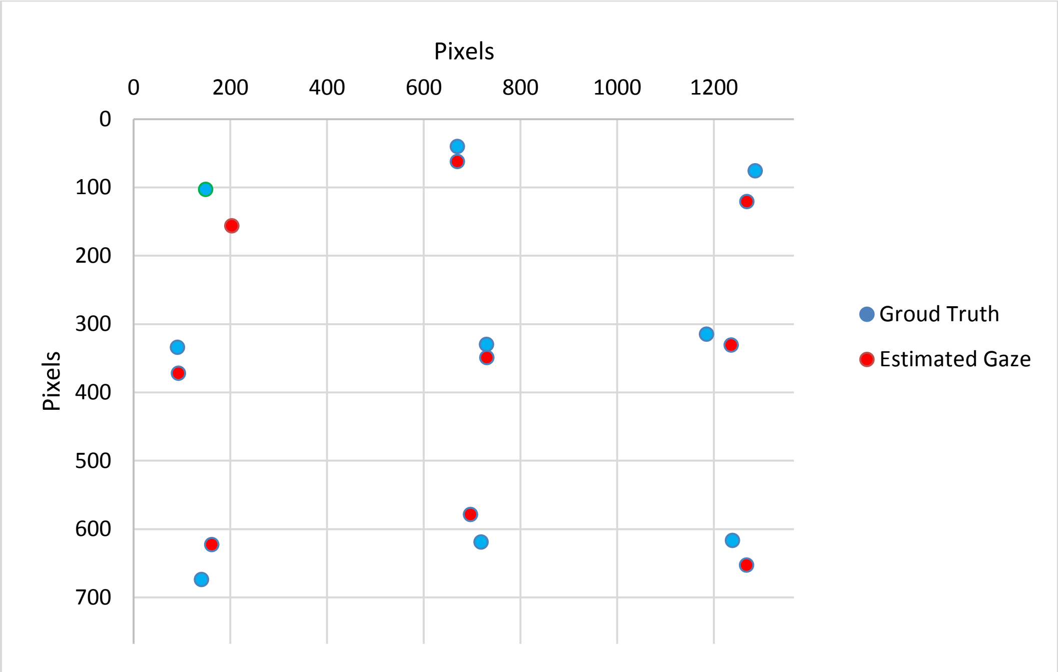

Figure 3: Ground truth and estimated gaze position.

4.3 Performance Analysis

Table 1 shows the mean and the standard deviation for the error in the distance between the actual and estimated gaze position for the experimental results.

X - AXIS		Y - AXIS	
MEAN	SD	MEAN	SD
21.88	20.2	35.55	13.68

Table 1: Average error and standard deviation in pixels

5. CONCLUSION AND FUTURE SCOPE

In this Paper, we proposed an effective approach of incorporating head pose with pupil estimation to estimate the gaze of a person. The system performs well in a well-lighted environment with a feasible delay in the cursor position. However, we observed much variance in the y axis compared to the x axis values. This is because the movements of eye pupil, in the y direction are not as visible as compared to the movements in x direction. Future work can be done to enhance the accuracy for y positions by tracking an additional feature such as sclera or eye lashes. Future work can also be done in applying adaptive weightage to the affecting factor of eye and head pose on the cursor. Our system is currently not able to deliver much accuracy when a user is wearing reflective spectacles or when the background is very bright compared to the user.

6. AKNOWLEDGEMENT

I take this opportunity to express my sincere gratitude to my guide Mrs. Krupa N. Jariwala, Assistant Professor, SVNIT, Surat for the invaluable guidance and moral support rendered by her in the fulfillment of my paper work. I am thankful to her for her constant constructive criticism and invaluable suggestions, which benefited me a lot while doing my research. Without her encouragement and guidance this research work would not have been materialized. I am also thankful to my family and friends who supported me during the preparation of my work.

REFERENCES

[1] Sesma L, Villanueva A. and Cabeza R., (2012) Evaluation of Pupil Center-Eye Corner Vector for Gaze Estimation Using a Web Cam, ETRA 2012, Santa Barbara, CA

[2] Michael Sapienza, Kenneth P Camilleri, "Fasthpe: A recipe for quick head pose estimation", Department of Systems & Control Engineering, University of Malta, Technical Report, 2014

[3] Fabian Timm and Erhardt Barth, "Accurate Eye Centre Localisation By Means of Gradients". In Proceedings of the Int. Conference on Computer Theory and Applications (VISAPP), volume 1, pages 125-130, Algarve, Portugal, 2011. INSTICC

[4] P. Viola and M. Jones, "Robust Real-Time Face Detection", International Journal of Computer Vision, May 2004, Volume 57, Issue 2, pp 137-154

[5] Roberto Valenti, Nicu Sebe and Theo Gevers "Combining Head Pose and Eye Location Information for Gaze Estimation", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 21, NO. 2, FEBRUARY 2012

[6] A. Gee and R. Cipolla, “Determining the gaze of faces in images,” Image and Vision Computing, vol. 12, no. 10, pp. 639–647, 1994

[7] Rohit P. Gaur, Krupa N. Jariwala, "A Survey on Methods and Models of Eye Tracking, Head Pose and Gaze Estimation", JETIR1405004, Volume 1 Issue 5, October-2014, eISSN: 2349-5162

[8] P. Padeleris, X. Zabulis, and A.A. Argyros, "Head pose estimation on depth data based on Particle Swarm Optimization", in Proc. CVPR Workshops, 2012, pp.42-49