CS365 Project Report
Joint Eye Tracking and Head Pose Estimation for Gaze Estimation

Ankan Bansal(10095)
Salman Mohammad(10630)
Advisor: Prof. Amitabha Mukerjee

1 Motivation

Head pose estimation, eye location and gaze estimation from images is an important field of research. They have many applications in commercial, entertainment and strategic sectors. Their uses vary from Human-Computer interaction (e.g. games, pointer control and other vision bases applications) to advertisement research. Driver attention analysis, human behaviour information and control for disabled people are some of the uses of gaze estimation. Some people have used several kinds of restricting hardwares for the purpose of head pose and eye location estimation. So there was a need to develop single static camera based algorithms for this purpose.
Head pose and eye location have been separately used for estimation of visual gaze. But these are prone to certain errors and inaccuracies. A combined model of both head pose and eye location estimates can give better results than both.

2 Past Work

Our work is mostly based on [1], which uses both head pose and eye location data for robust and accurate gaze estimation.
A cylindrical model of the head was used in [3] to recover the motion of the head. This presented a robust estimate of the motion of the head. But head pose estimates alone fail to finely estimate the gaze direction. In [2], the authors used isophote curvatures for accurate loacation and tracking of the center of the eye. Eye locations are sensitive to head pose. Trying to estimate the gaze using just the eye locations puts a restriction on the orientation of the head with respect to the camera. The work in [1] combines these two robust methods to obtain an even better estimate of eye locatiion, head pose and gaze.

Results on robustness of eye location obtained after the presentation have been included in the results section

3 Algorithm

Gaze estimation can either be done using head pose or eye locations. But [1] combines both these and introduces a synergic relation between the two i.e. one supports the other.
The method for eye center localization proposed in [2] has been shown to produce a very accurate estimate of the eye center. The centres of circular patterns have been estimated using isophotes. This method requires low computation and is invariant to linear changes in illumination. Under the assumptions of cylindrical head model (CHM), a static camera, pin hole camera model and a given camera calibration [3] gives very good results for head-pose estimation. This work proposes the early integration of the two, i.e, the eye center is located using head pose information which itself is then adjusted using the eye centers.

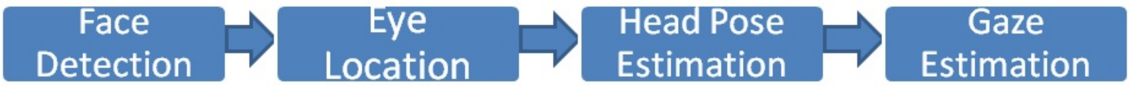

Figure 1: Outline of the Algorithm

Figure 1 gives an outline of the algorithm for gaze estimation using eye locations and head pose estimates.

3.1 Face Detection

The Viola-Jones Algorithm [6] has been used for detecting the face in a live video. It uses a complex combination of haar features for object detection. Some examples of such features are shown in Figure 2. OpenCV has cascades for faces whih have been used for detcting faces in live videos.

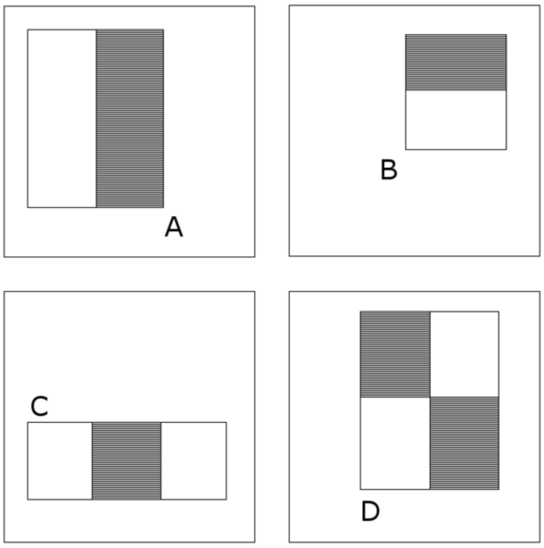

Figure 2: Haar Features [1]

[1] Image from en.wikipedia.org

3.2 Eye Location

A robust method for estimating the eye locations is given in [2]. The eye region is found in the face using the Haar cascades available in OpenCV. Isophotes (i.e. curves of equal intensity) are then found in this region. "The eyes are characterised by radially symmetric brightness patterns; hence, this method looks for the center of the curved isophotes in the eye region" [1]. Each pixel gives and estimate of the distance to the center of its isophote and the direction. All these centers are accumulated and weighted. The peaks of these accumulated regions give estimates of locations of eye centers. This method requires low complexity and can be applied in real time [2].

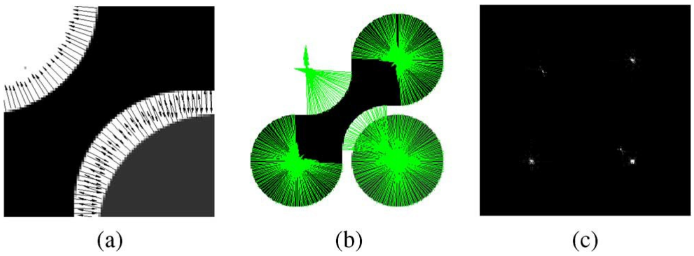

Figure 3: (a) Direction of the gradient, (b) displacement vectors pointing to the isophote centers, and (c) the centermap. [1]

3.3 Head Pose Estimation and Tracking

We use the method proposed in [3] for head pose estimation. Most of the earlier methods were restricitve in the sense that they either required stereo data or did not work in real time due to complex models of the head. The work in [3] uses a cylindrical head model (CHM). In this method the head is approximated as a cylinder. It has been shown to be very robust and can be implemented in real time.

The pose vector comprises of the rotation parameters, ω_x, ω_y, and ω_z, and translation parameters, t_x, t_y, and t_z. ω_x, ω_y, and ω_z corrspond to the pitch, yaw and roll angles of the head respectively and t_x, t_y, and t_z give the dispacements of the head from the camera center. The cylindrical head model is initialised using the eye locations. Perspective projection is used to obtain the 3D points from 2D coordinates in the image and the initial pose. Eye region is obtained in the obtained head region and eye centers are again located in this region. In this way, both eye location estimation and head pose estimation complement each other to give a better estimate of the pose.

Lukas Kanade [7] tracking algorithm is used for tracking the head in the images.

3.4 Visual Gaze Estimation

It has been shown that the visual field of view is characterised by the head pose. The human field of view can be approximated as a pyramid with its center at the center of the eyes as shown in Figure 4. "The displacement vectors between the resting position of the eyes (reference points) and the estimated eye location are used to obtain joint visual gaze estimation, constrained within the visual field of view defined by the head pose" [1].

[1] Image from [1]

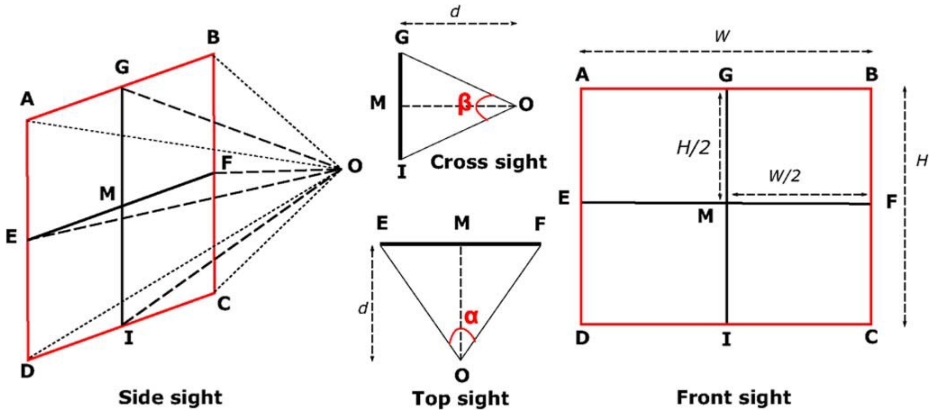

Figure 4: Representation of the visual field of view at distance d [1]

4 Data-set and Resources

Live data is obtained from the webcam. The Boston University head pose database [4] can also be used for performance estimation for both eye localization using head pose and head pose adjustment using eye center information. The authors of [1] have added eye annotations to the dataset [5]. The eye location estimation is done using EyeAPI [5].

5 Results

The face detection is done using the Haar cascades available in OpenCV. Eye detection is done using the EyeAPI. Eye detection can be said to be quite robust to rotations of the head. Eye tracking is robust to pitch angles upto about 20 degrees. Figure 5 shows eye tracking with a pitch angle of approximately 20 degrees. Similarly it is robust to a yaw angle of about 15 degrees. Figure 6 shows eye tracking with a yaw angle of approximately 20 degrees. The sampled points on the face in the image are shown in Figure 7. Face tracking in an image is shown in Figure 8. The black lines in the image show the motion of the sampled points from one frame to the next.

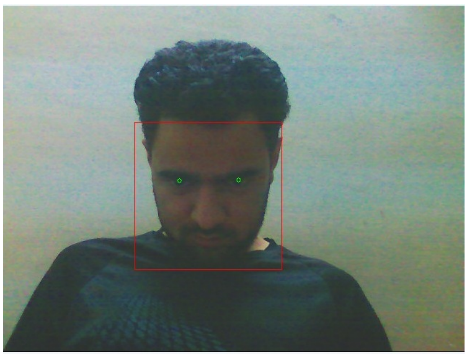

Figure 5: Eye tracking

[1] Image from [1]

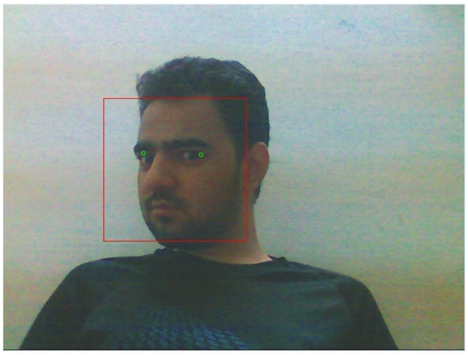

Figure 6: Eye tracking at a yaw angle of about 20 degrees

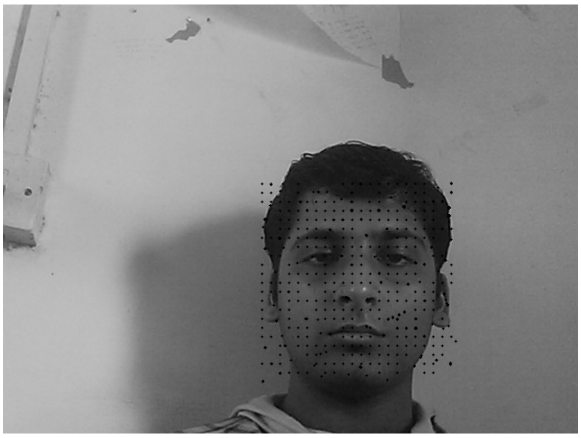

Figure 7: Sampled points on the face

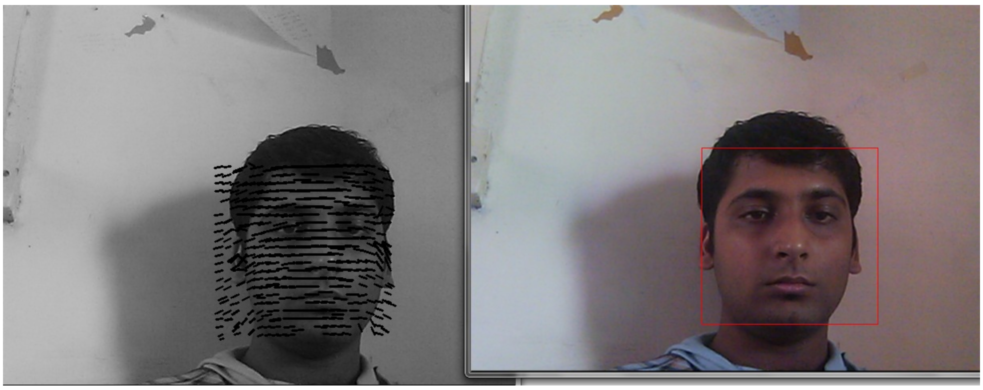

Figure 8: Flow vectors of the sampled points

6 Work to be done

The problem of estimating the gaze of a person accurately is considered to be a tough one. Due to lack of time estimation of the gaze has not been completed. Some work needs to be done in this direction. A basic model of the cylindrical head model has been written. A lot of improvement is needed in it. Using this gaze estimation can be implemented.

References

[1] R. Valenti, N. Sebe and T. Gevers, "Combining head pose and eye location information for gaze estimation," *IEEE Transactions on Image Processing, Volume 21 (2), pp. 802 - 815,* 2012

[2] R. Valenti and T. Gevers, "Accurate eye center location and tracking using isophote curvature," *IEEE Conference on Computer Vision and Pattern Recognition,* 2008.

[3] J. Xiao, T. Kanade and J. Cohn, "Robust full motion recovery of head by dynamic templates and re-registration techniques," *Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FG'02), pp. 156 - 162,* 2002.

[4] M. La Cascia, S. Sclaroff and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models," *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 22(4):322336,* 2000.

[5] R. Valenti, "UvAEyes, eye annotations for the Boston University head pose dataset [Online]," Available at:
http://staff.science.uva.nl/ rvalenti/index.php?content=UvAEyes

[6] Paul Viola and Michael Jones "Robust real-time object detection," *IJCV,* 2001

[7] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proceedings of Imaging Understanding Workshop, pp. 121 - 130,* 1981

[8] Roberto Valenti, Zeynep Ycel and Theo Gevers, "Robustifying eye center localization by head pose cues," *CVPR 2009: 612-618*

[9] Jun-Su Jang and Takeo Kanade, "Robust 3D Head Tracking by Online Feature Registration," *IEEE International Conference on Automatic Face and Gesture Recognition,* September, 2008