02
03
04
45%

IET Computer Vision

Research Article

ISSN 1751-9632
Received on 12th August 2015
Revised 25th February 2016
Accepted on 23rd March 2016
E-First on 3rd May 2016
doi: 10.1049/iet-cvi.2015.0316
www.ietdl.org

Fast and accurate algorithm for eye localisation for gaze tracking in low-resolution images

Anjith George[1] ✉, *Aurobinda Routray*[1]
[1]*Department of Electrical Engineering, IIT Kharagpur, Kharagpur 721302, West Bengal, India*
✉ *E-mail: anjith2006@gmail.com*

Abstract: Iris centre (IC) localisation in low-resolution visible images is a challenging problem in computer vision community due to noise, shadows, occlusions, pose variations, eye blinks etc. This study proposes an efficient method for determining IC in low-resolution images in the visible spectrum. Even low-cost consumer-grade webcams can be used for gaze tracking without any additional hardware. A two-stage algorithm is proposed for IC localisation. The proposed method uses geometrical characteristics of the eye. In the first stage, a fast convolution-based approach is used for obtaining the coarse location of IC). The IC location is further refined in the second stage using boundary tracing and ellipse fitting. The algorithm has been evaluated in public databases such as BioID, Gi4E and is found to outperform the state-of-the-art methods.

1 Introduction

Real-time detection and tracking of the eye is an active area of research in computer vision community. Localisation and tracking of the eye can be useful in face alignment, gaze tracking, human–computer interaction [1] etc. The methods available to find iris position are based on (i) scleral coil (ii) electro-oculography and (iii) video oculography [2]. Video or photo-oculography uses image processing techniques to measure the iris position non-invasively and without contact. The majority of the commercially available eye trackers use active infra red (IR) illumination. The bright pupil–dark pupil method [3] is used for finding the accurate location of the pupil. However, IR-based methods need extra hardware and specifically zoomed cameras that limit the movement of the head. Furthermore, the accuracy of IR-based method falls drastically in uncontrolled illumination conditions. This paper proposes an image-based algorithm for localising and tracking the eye in the visible spectrum. The main advantage of such a method is that it does not require any additional hardware and can work with regular low-cost webcams.

Several approaches have been reported in the literature for the detection of iris centre (IC) in low-resolution images. These methods can be broadly classified into four categories: (i) model-based methods, (ii) feature-based methods, (iii) hybrid methods and (iv) learning-based methods.

Model-based approaches generally approximate iris as a circle. The accuracy of such methods may fall when model assumptions are violated. In feature-based methods [1], local features such as gradient information, pixel values, corners, isophote properties etc. are used for the localisation of IC. Hybrid methods combine both local and global information for higher accuracy than individual methods alone. Learning-based methods [4] try to learn representations from labelled data rather than using the heuristic assumptions.

A hybrid approach for the detection and tracking of IC accurately in low-resolution images is presented here. A two-stage algorithm is proposed for localising the IC. A novel convolution operator is derived from circular Hough transform (CHT) for IC localisation. The new operator is efficient in the detection of IC even in partially occluded conditions and extreme corner positions. Additionally, an edge-based refinement and ellipse fitting are carried out to estimate the IC parameters accurately. IC and eye corners (ECs) are used in a regression framework to determine the point of gaze (PoG).

The important contributions of this paper are:

- A novel hybrid convolution operator for the fast localisation of IC.
- An efficient algorithm that can estimate the iris boundary in low-resolution grey-scale images.
- A framework for the eye gaze tracking in low-resolution image sequences.

The rest of this paper is organised as follows. Section 2 presents representative methods available in the literature for IC localisation and gaze tracking in low-resolution images. Section 3 presents the proposed algorithm. Section 4 discusses the experiments and results and Section 5 concludes this paper.

2 Related works

The localisation of iris or pupil is an important stage in gaze tracking. Once the IC has been successfully localised, regression-based methods can be used for finding the corresponding gaze points on the screen. Most of the passive image-based methods treat iris localisation as a circle detection problem. Circular Hough transform (CHT) is a standard method used for detection of circles [5]. Young *et al.* [6] reported a method for the detection of iris with specialised Hough transform and tracking with active contour algorithm. However, this method requires high-quality images obtained from a head mounted camera.

Smereka and Dulęba [7] presented a modified method for the detection of circular objects. They used the votes from each sector along with the gradient direction to detect circle locations. Atherton and Kerbyson [8] proposed phase combined orientation annulus method for the detection of circles with convolutional operators. The annulus is convolved with the edge image to detect the peaks. Yang *et al.* [9] presented an algorithm for first localising the eye region with Gabor filters and then localising the pupil with a radial symmetry measure. However, the accuracy of the method falls when the iris moves to corners. Valenti and Gevers [10] proposed an isophote property-based IC localisation algorithm. The illumination invariance of isophote curves along with gradient voting is used for the accurate detection of ICs. This method is further extended in [11] for scale invariance using scale-space pyramids. The face pose and IC obtained are combined to determine the PoG achieving an average accuracy of 2°–5° in unconstrained situations. The accuracy of the method falls when iris moves toward the corners resulting in false detection of eyebrows and ECs as ICs. Timm and Barth [12] proposed a method using gradients of the eye region. An extensive search is carried

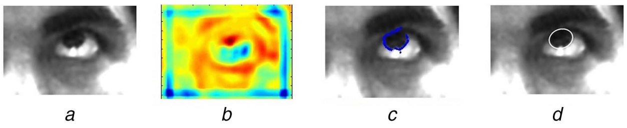

Fig. 1 *Stages in ellipse fitting*
a Cropped eye region
b Correlation surface from the proposed operator
c Selected candidate boundary points
d Fitted ellipse

out in all pixels maximising the inner product of the normalised gradient and normalised distance vector. IC is obtained as the maximum of weighted function in the region of interest (ROI). The time taken for search increases with increase in the search area. The performance of the algorithm degrades in noisy and low-resolution images where the edge detection method fails.

D'Orazio *et al.* [13] have reported a method for detection of IC using convolution kernels. The kernels are convolved with the gradient of images and peak points are selected as candidate points. The mean absolute error (MAE) similarity measure is used to reject false positive cases. Daugman [14] proposed an integro differential operator (IDO) for the accurate localisation of iris in IR images. Curve integral of gradient magnitudes is computed to extract the iris boundary. Recently, Baek *et al.* [15] presented an eyeball model-based method for gaze tracking. Elliptical shapes for eye model is saved in the database and used at the detection time for finding the ICs. A combined IDO and a weighted combination of features are used for the localisation of IC. Polynomial regression methods were used for training the system. They obtained average accuracy of 2.42° visual angles. Sewell and Komogortsev [16] developed an artificial neural network-based method for gaze estimation from low-resolution webcam images. They trained the neural network directly with the pixel values of the detected eye region. They obtained an average accuracy of 3.68°. Zhou and Geng [17] proposed a generalised projection function (GPF) that uses various projection functions and a special case hybrid projection function in localising the IC. The peak positions of vertical and horizontal GPF are used to localise the eye. Bhaskar *et al.* [18] proposed a method for identifying and tracking blinks in video sequences. Candidate eye regions are identified using frame differencing and are subsequently tracked using optical flow. The direction and magnitude of the flow are used to determine the presence of blinks. They obtained an accuracy of 97% in blink detection. Wang *et al.* [19] proposed one circle method where the detected iris boundary contours are fitted with an ellipse and back projected to find the gaze points. Recently, many learning-based methods have been proposed for IC localisation and gaze tracking. Markuš *et al.* [4] proposed a method for localising pupil in images using an ensemble of randomised trees. They used a standard face detector to localise face and eye regions. Ensemble of randomised trees model was trained using the eye regions and ground truth locations. Their method obtained good accuracy in BioID database. However, the accuracy of gaze estimation is not discussed in their work. Zhang *et al.* [20] proposed an appearance-based gaze estimation framework based on convolutional neural network (CNN). They have trained a CNN model with a large amount of data collected in real-world conditions. Normalised face images and the head poses obtained from a face detector were used as the input to the CNN to estimate the gaze direction. They obtained good accuracy in person and pose independent scenarios. However, the accuracy for person-dependent case is lower than current geometric model-based methods. The accuracy may increase with larger amount of training data, but the time taken for online data collection and training becomes prohibitive. Schneider *et al.* [21] proposed a manifold alignment-based method for appearance-based person independent gaze estimation. From the registered eye images, a wide variety of feature extraction methods such as local binary pattern (LBP) histogram, histogram of oriented gradients (HoGs), mHoG and discrete cosine transform (DCT) coefficients were extracted. A combination of LBP- and mHOG-based features obtained the best performance. Several regression methods were used for appearance-based gaze estimation. Sub-manifolds for each individual were obtained using the ground truth gaze locations. Synchronised Delaunay sub-manifold embedding method was used to align the manifolds of different persons. Even though their method achieved better performance compared with other appearance-based regression methods, the effect of head pose variations on the accuracy was not discussed.

Sugano *et al.* [22] proposed a person and head pose independent method for appearance-based gaze estimation. They captured the images of different persons using a calibrated camera, and images corresponding to various head poses were synthesised. An extension of random forest algorithm was used for training. The appearance of eye region and the head pose is used as the input to the algorithm which learns a mapping to the three-dimensional (3D) gaze direction.

Most of the methods proposed in literature fail when iris moves toward the corners. Another problem is regarding eye blinks, most of the algorithms return false positives when the eyes are closed. A stable reference point is required along with the IC location for PoG estimation. Learning-based methods require large amounts of labelled data for satisfactory performance. The performance of such methods deteriorates when imaging conditions are different. Training for person-dependent models requires large amounts of data and often require a considerable amount of time. This limits the deployment of such methods in mobiles, tablets etc.

In the proposed method, IC can be accurately localised even in extreme corner locations using the ellipse approximation. The high computational complexity is avoided using the two-stage scheme. An eye closure detection stage is added to prevent false positives. The localisation error can be minimised by tracking the IC in the subsequent frames. The estimated IC is used in a regression framework to estimate the PoG.

3 Proposed algorithm

Different stages of the proposed framework are described here.

3.1 Face detection and eye region localisation

Knowledge of the position and pose of the face is an essential factor in determining the PoG. Detection and tracking of the face help in obtaining candidate regions for eye detection. This reduces the false positive rate as well as computation time. Haar-like feature-based method [23] is used for face detection because of its higher accuracy and faster execution. An improved implementation of face detection and tracking has been proposed in our earlier work [24]. The modified algorithm can detect in-plane rotated images with an affine transform-based algorithm. Processing is carried out in the down sampled images to make the detection faster. The search space of detector algorithm is dynamically constrained based on the temporal information, which further increases face detection speed. Kalman filter (KF)-based tracking is used to remove the false detections and to predict the location of the face when it is not detected. The de-rotated eye region obtained is used in subsequent stages, which makes the performance of the algorithm invariant to in-plane rotations. The purpose of the de-rotation stage is only to provide a de-rotated ROI for the further processing stages. The accuracy of face rotation estimation in the pre-processing stage is only up to ±15°. More accurate in-plane face rotation is obtained in the later stage using the angle of the line connecting the inner ECs. With the improved face-tracking scheme, the frame rates of processing increase greatly (up to 200 fps). The analysis and trade-offs of the algorithm are presented in our earlier work [24].

3.2 IC localisation

The proposed method uses a coarse-to-fine approach for detecting the accurate centre of the iris. The two-stage approach reduces the computational complexity of the algorithm as well as false detection. The stages in IC localisation are shown in Fig. 1.

3.2.1 Coarse IC detection: Iris detection is formulated as circular disc detection in this stage. An average ratio between the width of face and iris radius was obtained empirically. For a particular image, the radius range is computed using this ratio and width of the detected face. The image gradient of iris boundary points will always be pointing outwards. The gradient directions and intensity information is used for the detection of eyes. The gradients of the image are invariant to uniform illumination changes.

A novel convolution operator is proposed to detect peak location corresponding to the centre of the circle. A class of convolution kernels known as Hough transform filters [8] are used for this purpose. In CHT filter, the 3D accumulator is collapsed to a 2D surface by selecting a range for the radii.

The 2D accumulator can be calculated efficiently using a convolution operator. A CHT filter is derived, which acts directly on the image without any requirement of edge detection. A vector convolution kernel is designed for correlating with the gradient image, which gives a peak at the centre of the iris.

The convolution operator is designed as a complex operator with magnitudes as unity. The operator detects a range of circles by taking dot products with orientation inside the radius range. The equation is similar to orientation annulus proposed by Atherton and Kerbyson [8]. The equation of convolution kernel is given as

$$O_{\mathrm{COA}}(m,n) = \begin{cases} \dfrac{1}{\sqrt{m^2+n^2}}(\cos\theta_{mn} + i\sin\theta_{mn}), \text{ iff, } R_{\min}^2 < m^2+n^2 < R_{\max}^2 \\ 0, \text{ otherwise} \end{cases} \tag{1}$$

where

$$\theta_{mn} = \tan^{-1}\left(\frac{n}{m}\right) \tag{2}$$

where m and n denote the coordinates of the kernel matrix with respect to the origin. The operator is scaled for equal contributions of circles in the radius range. A weighting matrix kernel ($\boldsymbol{W}_A$) is also used for finding regions with maximum dark values

$$\boldsymbol{W}_A(m,n) = \begin{cases} \dfrac{1}{\sqrt{m^2+n^2}}, \text{ iff, } m^2+n^2 < R_{\max}^2 \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

the gradient complex orientation annulus can written as

$$C_{\mathrm{GCOA}} = \mathrm{Re}(O_{\mathrm{COA}}) \otimes S_x + i\,\mathrm{Im}(O_{\mathrm{COA}}) \otimes S_y \tag{4}$$

where '$\otimes$' denotes the convolution operator; S_x and S_y denote the 3 × 3 Schaar kernels in x and y directions, respectively. Schaar differential kernel is used owing to its mathematical properties in gradient estimation. In most of the cases, the upper portion of the iris is occluded by eyelids. An additional weighing factor (β) is included to increase the contribution of horizontal gradients. Equation for convolution kernel can be made a real-valued kernel as

$$C_{\mathrm{RCC}} = \beta\,\mathrm{Re}(O_{\mathrm{COA}}) \otimes S_x + \frac{1}{\beta}\mathrm{Im}(O_{\mathrm{COA}}) \otimes S_y \tag{5}$$

where β denotes the weighting factor.

The average intensity of each point in image can be obtained by convolving the weighting kernel with the negated version of the image as

$$W = (255 - I) \otimes \boldsymbol{W}_A \tag{6}$$

where I and $\boldsymbol{W}_A$ denote the image and the kernel for computing the intensity component, respectively. The final correlation output (CO) can be obtained by combining the convolution results for both gradient and intensity kernels as

$$\mathrm{CO} = \lambda(I \otimes C_{\mathrm{RCC}}) + (1-\lambda)W \tag{7}$$

where $\lambda \in [0, 1]$ is a scalar, which is used to obtain the weighted combination of gradient information and image intensity to reduce spurious detections. IC corresponds to the maximum of correlation surface CO. Furthermore, it is possible to represent all these operations with a single real convolution kernel, which can be applied on the image without any pre-processing, making the IC localisation procedure even faster. For bigger circles, convolution can be carried out in Fourier domain for increasing the speed of the computation.

The peak of CO alone may lead to false detections in partially occluded images. Here, peak-to-side lobe ratios (PSRs) of the points are used to find the iris location. The PSR values calculated in each of the local maxima and the point with maximum PSR is considered as the IC. The PSR is computed as

$$\mathrm{PSR} = \left(\frac{\mathrm{CO}_{\max} - \mu}{\sigma}\right) \tag{8}$$

where $\mathrm{CO}_{\max}$ is the local maxima in the CO, μ and σ are the mean and standard deviation in the window around the local maxima. We have used a window size of 11 × 11 in this paper. The point with the maximum PSR is selected as the IC.

3.2.3 Sub-pixel edge refining and ellipse fitting: In this stage, the rough centre points obtained in the previous stage are used to refine the IC location. The objective is to fit the iris boundary with an ellipse. The constraints on the major and minor axes can be obtained empirically ($R_{\min}$ and $R_{\max}$). The algorithm presented searches in the radial direction similar to Starburst algorithm [25]. However, the search process finds only the strongest edges with similar gradients. Dominant edges with agreeing directions are selected with sub-pixel accuracy. An angle versus distance plot is obtained and the outlier points are filtered using median filter. An ellipse can be fitted to five points by the least-square method using Fitzgibbon's algorithm [26]. However, we used this algorithm in a random sample consensus (RANSAC) framework for minimising the effect of outliers. RANSAC algorithm is employed [27] for ellipse fitting, using the gradient agreement [28] of the detected boundary points and the fitted ellipse as the support function. Additionally, a modified goodness of fit (GoF) is evaluated as the integral of dot products of outward gradients over the detected boundary (only agreeing gradients). The parameters obtained are considered as false positives if the GoF is less than a threshold. The detailed algorithm for ellipse fitting is given in Table 1

$$\mathrm{GoF} = \sum_{x,\,y \in f(\lambda)} \left(\min\left(\frac{\nabla f(x,y)}{|\nabla f(x,y)|} \cdot \nabla I(x,y),\ 0 \right) \right) \tag{9}$$

3.3 Iris tracking

KF [29] is used to track the IC in a video sequence. The search region for iris detection can be limited with the tracking approach. Once the IC is detected with sufficient confidence, the point can be tracked in subsequent frames easily. Face detection stage can be avoided in this case. KF [30] estimates can be used as the corrected estimates for iris position.

In the current tracking application, constant velocity model is selected as the transition model. Coordinates of the centre of iris along with their velocities are used as states

$$X_{k+1} = F_k X_k + W_k \tag{10}$$

where X_k is the state containing x, y, v_x, v_y (coordinates and velocities in x and y directions, respectively) at the kth instant. The measurement noise covariance matrix is computed from the measurements obtained during the gaze calibration stage. The process covariance matrix is computed empirically. Measurements obtained from the IC detector are used to correct the estimated states.

Estimated positions from KF can be used as iris positions during temporary occlusions. KF continues to predict based on the information obtained from past observations. These predicted locations could be used in gaze tracking when eyes are closed reducing the jitter in PoG estimates.

3.4 Eye closure detection

The IC localisation algorithm may return false positives when the eyes are closed. Thresholds on the peak magnitude were used to reject false positives. However, the quality of peak may degrade in conditions such as low contrast, image noise and motion blur. The accuracy of the algorithm may fall in these conditions, and hence machine learning-based approach is used to classify the eye state as open or close. The HoG [31] features of eye regions are calculated and a support vector machine (SVM)-based classifier is used to predict the state of the eye. The HoG features are computed in the detected ROI for left and right eyes separately. The SVM classifier was trained offline from the database. If the eye state is classified as closed, then the predicted value from KF is used as the tentative position of the eye. If eyes are detected as open, then the result from the two-stage method is used to update the KF.

3.5 EC detection and tracking

The appearance of inner EC exhibits insignificant variations with eye movements and blinks. Therefore, this paper proposes to use inner ECs as reference points for gaze tracking. The ECs can be located easily in the eye ROI. The vectors connecting ECs and ICs can be used to calculate gaze position. Several methods have been proposed in the literature for the localisation of facial landmarks [32]. In the proposed method, Gabor jets [33] are used to find ECs in the eye ROI owing to its high accuracy. The detected ECs are tracked in the subsequent frames using optical flow and normalised cross-correlation-based method [34, 35]. The tracker is automatically reinitialised if the correlation score is less than a pre-set threshold value.

Table 1 Algorithm for iris boundary refinement

```
1. Input: The grey-scale eye region $I$ and the estimated centres $O = (c_x,\ c_y)$
2. Output: Fitted ellipse parameters $\lambda = (c_x,\ c_y,\ a,\ b,\ \psi)$
3. Initialise: Candidate_Points = NULL.
   temp_best = [0, 0]
   For $\theta = 0{:}2\pi$
   For $r = R_{\min}{:}R_{\max}$
   $Pt = (c_x + r\cos\theta,\ c_y + r\sin\theta)$
   Calculate the gradients and magnitude at the points
   $\boldsymbol{g} = \frac{G_x\hat{x} + G_y\hat{y}}{\| G \|}$
   If $\| g \| <$ threshold
   Break;
   Else
   $\boldsymbol{r} = r\cos\theta\hat{x} + r\sin\theta\hat{y}$
   Calculate the dot product of normalised gradient vector
   If $\cos\theta = \boldsymbol{r}\cdot\boldsymbol{g} >$ threshold
   If tempbestmag < Pt · mag, Pt · angle
   temp_best = Pt, tempbestmag = [cos $\theta$, $||g||$]
   Else
   Continue:
   End For
   Candidate_Points.append(temp_best)
   End For
4. Filter the detected points with angular median filter
5. Fit ellipse with RANSAC algorithm
6. Return the parameters of ellipse $\lambda = (c_x,\ c_y,\ a,\ b,\ \psi)$
```

3.6 Gaze estimation

Gaze point can be computed from the IC location and a reference point. Earlier works [36, 37] have used the eye centre and corneal reflections as reference points. False detection in any of the corners will result in performance degradation of the algorithm. Hence, the inner ECs are used as reference points in this paper. Detection of the EC in every frame might increase the error rates and computational complexity. We avoid this issue by tracking of ECs in the frames which ensure stable reference points. If $(x_1,\ y_1)$ and $(x_2,\ y_2)$ denote the coordinates of EC and IC, respectively, the EC–IC vector (with reference to the corner) can be obtained as: $(x,\ y) = (x_2 - x_1,\ y_2 - y_1)$. The EC–IC vector is calculated separately for the left and right eyes.

3.6.1 Calibration: In the calibration stage, subjects were asked to look at uniformly distributed positions on the screen. The EC–IC vectors along with gaze points are recorded. The mapping between EC–IC vector and screen coordinates is non-linear because of the angular movement of the iris. We used two different models for the mapping between EC–IC vector and PoG: (i) polynomial regression and (ii) a radial basis function (RBF) kernel-based method. In polynomial regression, a second-order regression model is used for determining the PoG since it offers the best trade-off between model complexity and accuracy

$$\text{screen}\, X_i = a_0x_i + a_1y_i + a_2x_iy_i + a_3x_i^2 + a_4y_i^2 + a_5 \tag{11}$$

$$\text{screen}\, Y_i = b_0x_i + b_1y_i + b_2x_iy_i + b_3x_i^2 + b_4y_i^2 + b_5 \tag{12}$$

where $(x_i,\ y_i)$ are the components of EC–IC vector and (screen X_i, screen Y_i) the corresponding screen positions. The data obtained from calibration stage is used in the least-square regression framework to calculate unknown parameters.

In the RBF kernel-based method, we used non-parametric regression [38] for estimating PoG. The components of EC–IC vector are transformed into kernel space using the following expression:

$$k(p_i,\ p_l) = \exp\left(-\frac{\| p_i - p_l \|^2}{2\sigma_k^2}\right) \tag{13}$$

where p_i and p_l denote of EC–IC vector and the landmark points, respectively. σ_k denotes the standard deviation of the RBF function. We have tested the algorithm in both 3 × 3 and 4 × 4 calibration grids. Instead of using all the samples as landmark points, we have used only one landmark per calibration point. The number of landmarks used was 9 and 16 for 3 × 3 and 4 × 4 grid, respectively. For each calibration point on the grid, the landmark vector is calculated as the median of the components of EC–IC vector at the particular point. The dimension of the design matrix is reduced by the use of landmark points (since the data points are clustered around the calibration points). Regression is carried out after transforming all the points to kernel space [39] which improved the accuracy of PoG estimation. The training procedure is carried out for left and right eyes.

3.6.2 Estimation of PoG: The parameters obtained from calibration procedure are used to determine the gaze position. The regression function obtained is used to map the EC–IC vectors to screen coordinates. The gaze position is computed as the average position returned by the left and right eye models. The head position is assumed to be stable during the calibration stage. After calibration, the estimated gaze point will be on the calibration plane [i.e. with respect to (w.r.t.) the position of the face during the calibration stage]. Deviation from this face position would cause errors in the estimated gaze locations. The effect of 2D translation is minimal for moderate head movements since the reference points for EC–IC vector are ECs, which also move along with face (thereby providing a stable reference invariant to 2D translational motion). Even though the method is invariant to moderate amount

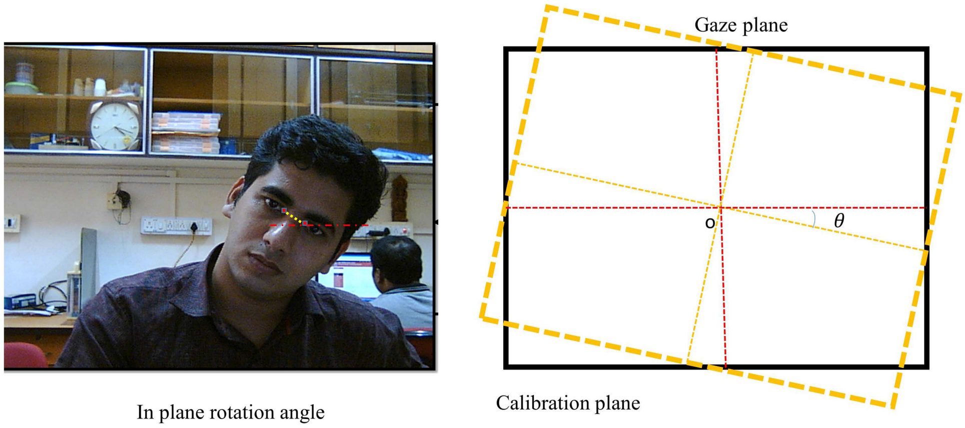
In plane rotation angle

Fig. 2 *Transformation of estimated gaze point to screen coordinates for in-plane rotation*

of translation, the accuracy falls when there is a rotation. This error can be corrected using the face pose information. The in-plane rotation of face can be calculated from the angle of the line connecting the inner corners of the left and right eyes as shown in Fig. 2. The rotation matrix can be computed as

$$\boldsymbol{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{14}$$

where θ is the difference in angle from the calibration stage. The corrected PoG can be found from coordinate transformation with the screen centre as the origin. The exact 3D pose variations can be corrected using more computationally intensive models such as active appearance models [40], constrained local model [41] etc.

4 Experiments

We have conducted several experiments to evaluate the performance of the proposed algorithm. The algorithm has been evaluated using standard databases and a custom database. The IC localisation accuracy is evaluated in standard databases and compared with the state-of-the-art methods. The accuracy in PoG estimation and eye closure detection is assessed in the custom dataset.

4.1 Experiments on IC localisation

4.1.1 Evaluation method: Face detection is carried out using Viola–Jones method [42]. The eye regions are localised based on anthropometric ratios.

The normalised error is used as the metric for comparison with other algorithms. The normalised measure for worst eye characteristics (WECs) [43] is defined as

$$e_{\text{WEC}} = \frac{\max(d_{\text{l}}, d_{\text{r}})}{w} \tag{15}$$

where d_{l} and d_{r} are the Euclidean distances between ground truth and detected ICs (in pixels) of left and right eyes, respectively, and w is the true distance between the eyes in pixels. The average [average eye characteristic (AEC)] and best of eye detection [best eye characteristic (BEC)] errors are also calculated for comparison. They are defined as

$$e_{\text{AEC}} = \frac{(d_{\text{l}} + d_{\text{r}})}{2w}, \quad e_{\text{BEC}} = \frac{\min(d_{\text{l}}, d_{\text{r}})}{w} \tag{16}$$

where e_{BEC} is the minimum error in both the eyes and e_{AEC} is the average error of both the eyes.

4.1.2 Experiments in BioID and Gi4E databases: A comparison of the proposed method with the state-of-the-art methods is carried out for BioID [44] and Gi4E [45] databases. The BioID database consists of images of 23 individuals taken at different times of the day. The size, position and pose of the faces change in the image sequences. The contrast is very low in some images. In some images, eyes are closed. There are images where subject wear glasses and glints are present due to illumination variations. The database contains a total of 1521 images with a resolution of 384 × 288 pixels. The ground truth files for left and right ICs are also available.

Gi4E dataset consists of 1380 colour images of 103 subjects with a resolution of 800 × 600. It contains sequences where the subjects are asked to look at 12 different points on the screen. All the images are captured at indoor conditions at varying illumination levels and different backgrounds. The database represents realistic conditions during gaze tracking, head movements, illumination changes and movement of eyes toward corners and occlusions with eyelids. The ground truth of left and right eye positions is also available with the database.

Fig. 3 shows some of the correct detections and failures of the algorithm in BioID database. The face detection accuracy obtained was 94.74%. In most of the cases, errors are due to partial closure of eyes and eyeglasses. The algorithm performs well when eyes are visible even with low contrast and varying illumination levels. Fig. 4 shows the performance of the proposed algorithm in BioID and Gi4E databases. The value of λ and β used were 0.95 and 2, respectively. The proposed algorithm gives an accuracy (WEC) of 85.08 for $e \leq 0.05$.

In Gi4E database, the worst-case accuracy (WEC) is 89.28 for $e \leq 0.05$. Fig. 5 shows results from the algorithm. The face detection accuracy obtained was 96.95%. The main advantage is that the algorithm performs well in different eye gaze positions which is essential in gaze-tracking applications.

The performance of the algorithm may vary depending on the distance of the user from the monitor. This effect is emulated using images with different spatial resolutions. The performance of the proposed algorithm in different spatial resolutions in BioID and Gi4E databases is shown in Fig. 6. The accuracy of iris localisation falls as the image resolution decreases. However, the detection accuracy (WEC-0.5) is more than 80% for scaling up to 0.8 (307 × 230 resolution) and 0.6 (480 × 360 resolution) in BioID (82.72%) and Gi4E (82.24%) databases, respectively.

4.1.3 Comparison with state-of-the-art methods: We have compared the algorithm with many state-of-the-art algorithms in BioID and Gi4E databases. The algorithms maximum isocenter (MIC) [10], Timm and Barth [12] and the proposed methods are tested in BioID database. The evaluation is carried out with normalised WEC. The results are shown in Fig. 7. The WEC data is taken from receiver operating characteristic (ROC) curves given in author's papers. The algorithm proposed is the second best in BioID database as shown in Table 2. Isophote method (MIC) performs well in this database. The proposed algorithm fails to detect accurate positions when eyes are partially or fully closed (eye closure detection stage was not used here). Presence of glints is another major problem. The failure of face detection stage and reflections from the glasses causes false detections in some cases.

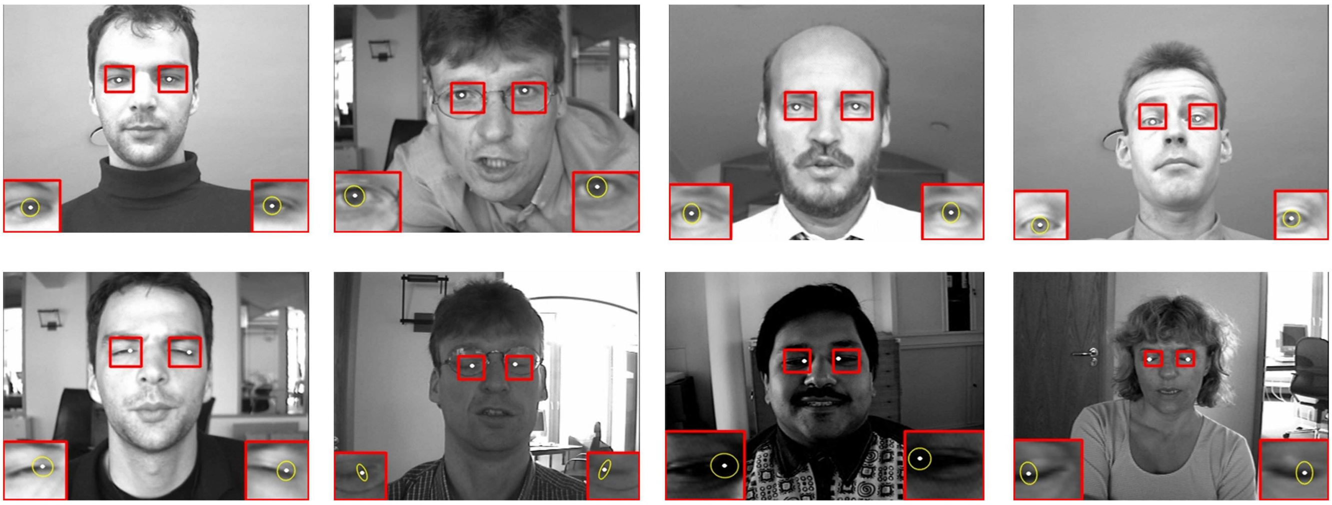

Fig. 3 *Few samples showing successful detections (first row) and failures (second row) in BioID database*

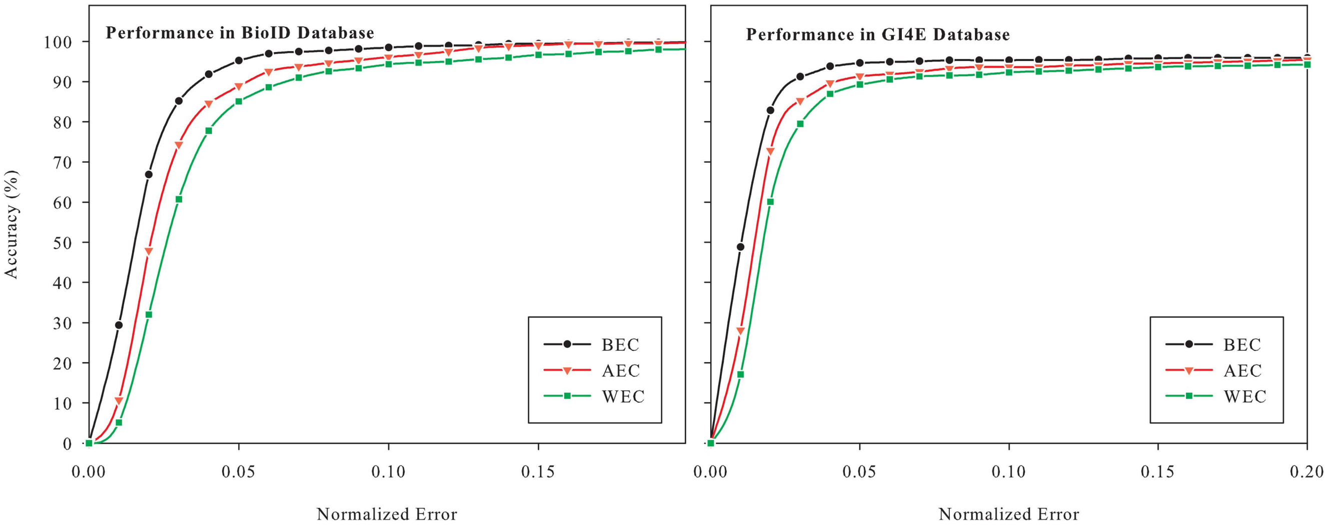

Fig. 4 *Performance of the proposed algorithm in BioID and Gi4E databases. The graph shows three normalised measures corresponding to WECs – worst eye characteristics, AECs – average eye characteristics and BECs – best Eye characteristics*

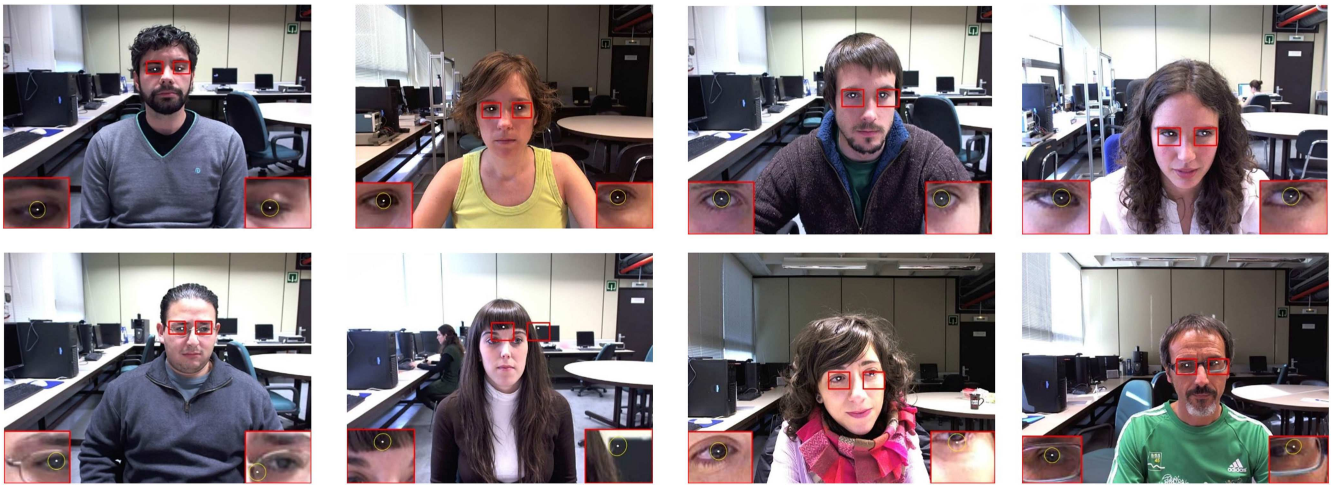

Fig. 5 *Some samples showing successful detections (first row) and failures (second row) in Gi4E database*

The addition of a machine learning-based classification of local maxima may improve the results of the proposed algorithm.

Gi4E is a more realistic database for eye tracking purposes. It contains images with head and eye movements. The algorithms for comparison are chosen as vertical edge (VE) [19], IDO [14], MIC [10] and elliptical shape iris center (ESIC) [15]. The results are compared with WEC values obtained from ROC curves reported in Baek *et al.* [15]. It is seen (Table 3) that the proposed method outperforms all. The accuracy of MIC method is very low when the eyes move to the corners. The circle approximation of most of the algorithms fails when eyes move toward the corners, making them inapt for eye gaze-tracking applications. The performance evaluation is carried out on each frame separately. Addition of temporal information described in Section 3.3 might increase the accuracy of the algorithm greatly.

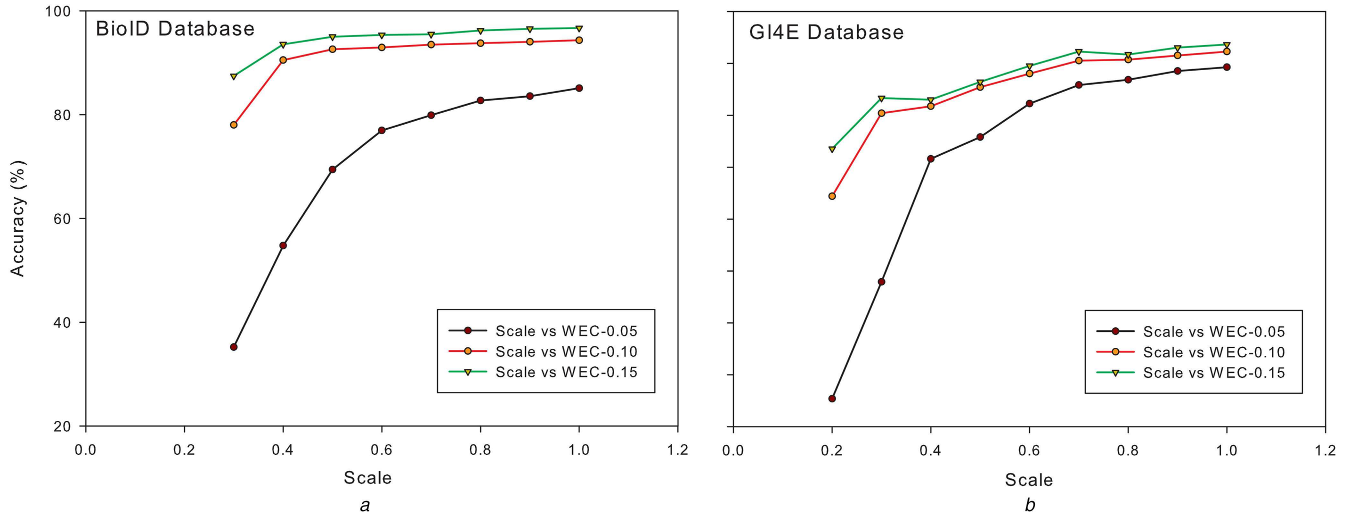

Fig. 6 *WEC performance of the proposed algorithm in#a BioID*
b Gi4E databases with different resolutions. Scaling parameter is w.r.t. the original image resolutions in the corresponding databases

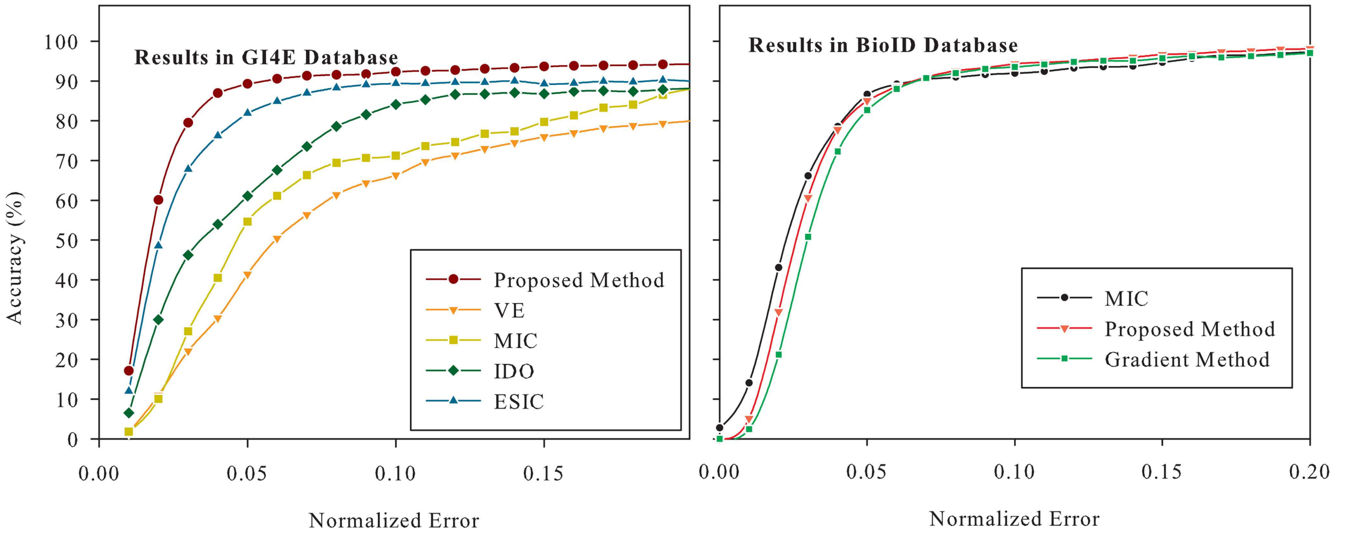

Fig. 7 *WEC performance comparison of the proposed method with state-of-the-art methods in Gi4E and BioID databases*

We have performed additional experiments on the Gi4E dataset to evaluate the performance when the iris moves to the corner. A subset of 299 images was selected according to the position of IC about the EC. We have compared the results with the gradient-based method for evaluating the accuracy with circle model and ellipse model. The WEC characteristics comparison is shown in Fig. 8. The ellipse approximation improves the accuracy significantly compared with the circle approximation.

Table 2 Comparison of proposed method with state-of-the-art algorithms in BioID database

Method	$e \le 0.05$	$e \le 0.10$	$e \le 0.15$	$e \le 0.20$
MIC [12]+sift kNN[b]	**86.09**	91.67	94.5[a]	96.9[a]
Proposed	85.08	**94.3**	**96.67**	**98.13**
Timm and Barth [12]	82.5	93.4	95.2	96.4

[a]Approximated from graph [10].
[b]K-nearest neighbour (kNN)

Table 3 Comparison of the proposed method with state-of-the-art algorithms in Gi4E database

Method	$e \le 0.05$	$e \le 0.10$	$e \le 0.15$	$e \le 0.20$
proposed	**89.28**	**92.3**	**93.64**	**94.22**
VE	41.4	66.3	75.9	80[a]
MIC	54.5	71.2	79.7	88.1[a]
IDO	61.1	84.1	86.7	88.15[a]
ESIC	81.4	89.3	89.2	89.9[a]

[a]Approximated from graph [15].

4.2 Experiment with our database

4.2.1 Experiment for gaze estimation accuracy evaluation: An experiment was performed on ten subjects using a standard webcam and a 15.6 in monitor with a resolution of 1366 × 768. The subjects were seated 60 cm from the screen and asked to follow the red dot on the monitor. The videos of the eye movements were recorded at 30 fps with a resolution of 640 × 480. The subjects were asked to look at the calibration patterns two times. We used both 9 point and 16 point calibration and compared the results. Fig. 9 shows some of the images from the dataset.

We have evaluated the IC localisation accuracy on a subset of images in the in-house dataset. A subset of 1000 images was selected, and the IC localisation accuracy was evaluated. The proposed approach obtained WEC accuracies of 90.2 and 92.9% for $e \le 0.05$ and 0.10, respectively.

For 3 × 3 and 4 × 4 calibration grids, we have tested with polynomial regression and kernel space-based methods. The samples from the first session were used in the training stage. The parameters for regression were found from the training data. In the testing stage, the samples in the second session were used to estimate the PoG. The mean position computed from the left and right eyes PoG is used as the final gaze point. The error in the

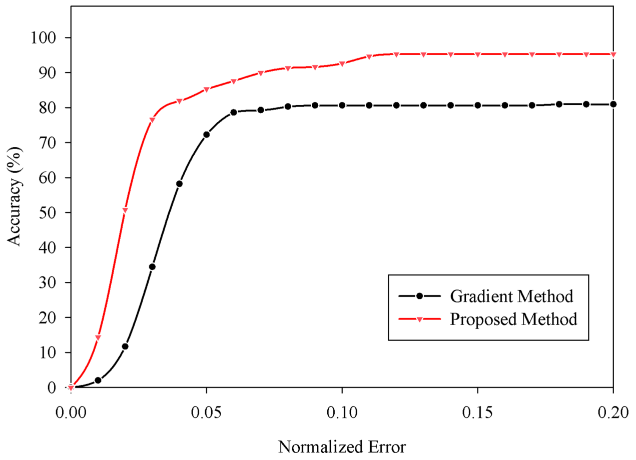

Fig. 8 *WEC performance comparison of the proposed method with gradient-based method in extreme corner cases*

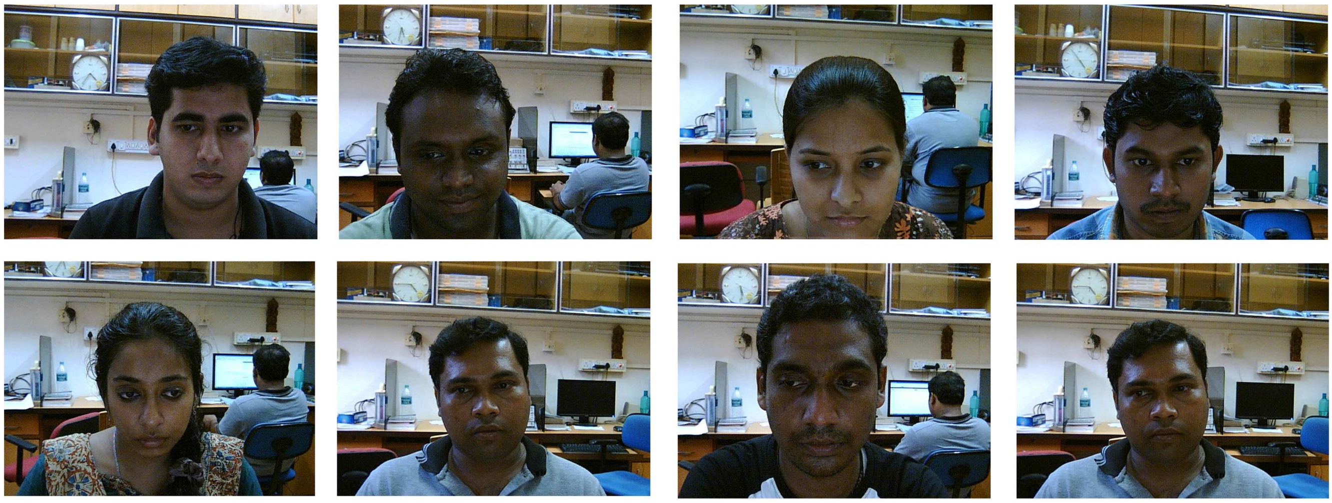

Fig. 9 *Sample images of subjects participated in the experiment*

estimation is computed using the ground truth. The MAE in visual angles in horizontal, vertical and overall accuracy is computed using the head distance from the screen

$$\text{accuracy} = \tan^{-1}\left(\frac{\text{error}}{\text{head distance}}\right) \tag{17}$$

The average errors are high when the EC–IC vectors are computed on a frame-by-frame basis. We further computed the PoG using KF estimates which reduced the jitter significantly. The results with and without KF on 3 × 3 and 4 × 4 calibration grids are tabulated in Table 4. The qualitative results of gaze estimation stage are shown in Fig. 10.

4.2.2 Experiment for eye closure detection: The eye regions obtained from the face detection stage are histogram equalised and resized to the size of 30 × 30. A dataset of 4000 images containing 2000 samples for open and 2000 samples for closed eyes were formed from our dataset. HoG features were extracted from various pixel per cell windows and eight orientations. The extracted HoG features were used to train the SVM classifier. Ten times ten-fold cross-validation was used to examine the accuracy of the trained classifiers. The results show that the proposed method achieves an average accuracy of 98.6% with linear SVM. The results obtained are shown in Table 5.

4.3 Discussion

The proposed method contains cascaded stages of many algorithms. The gaze estimation accuracy is a good proxy for the combined accuracy of all the cascaded stages. Face and IC are tracked using KFs independently due to their distinct dynamics. The tracking-based framework increases the robustness by reducing the effect of per-frame localisation errors.

For successful eye tracking using webcams, the normalised error should be <0.05. The proposed algorithm performs better in realistic conditions for webcam-based gaze tracking. The accuracy of gaze estimation was evaluated with the proposed approach in both 3 × 3 and 4 × 4 calibration grids. RBF kernel-based non-parametric regression method was found to perform better than second-order polynomial models. The average error rate obtained with the per-frame-based detection was 2.71°. The accuracy of the gaze tracking improved significantly by the use of KF, which uses the temporal information effectively to reduce the error rate to 1.33°.

One of the advantages of the proposed algorithm is low computational complexity. The eye detection, being a convolution-based method can be implemented in Fourier domain [46] for faster computation. Multi-resolution convolution can be used to reduce the search space even further. The algorithm was implemented in a 2.5 GHz core 2 duo personal computer (PC) with 2 GB random access memory. C++ implementation using OpenCV library [47] (without multi-threading) was used for the evaluation experiments in Ubuntu 14.04 OS (32 bit) environment. It detects the face and ECs in the first frame and tracks the ECs over time. The temporal information is used to reduce the search space for face detection using a KF. The images were acquired using a 60 fps 640 × 480 webcam. The online processing speed was limited only by the lower frame rate of the camera. The offline processing speed of the entire algorithm is well over 100 fps on the recorded video. This is suitable for normal PC-based implementation with 30 fps webcams. The proposed method can also be implemented in smart devices such as mobile phones and tablets due to its low computational overhead. The low computational complexity makes it possible to extend the pose tracking with more complex 3D models, which could make the PoG estimation invariant to out-of-plane rotations as well.

5 Conclusion

This paper describes an algorithm for a fast and accurate localisation of IC position in low-resolution grey-scale images. A framework for tracking faces in video at very high frame rates, well above 100 fps is also presented. A two-stage iris localisation is carried out, and the filtered candidate iris boundary points are used to fit an ellipse using a gradient aware RANSAC algorithm. The proposed algorithm is compared with the state-of-the-art methods and found to outperform edge-based methods in low-resolution images. The computational complexity of the algorithm is very less since it uses a convolution operator for IC localisation. This paper also proposes and implements a gaze-tracking framework. Inner ECs are used as the reference for calculating

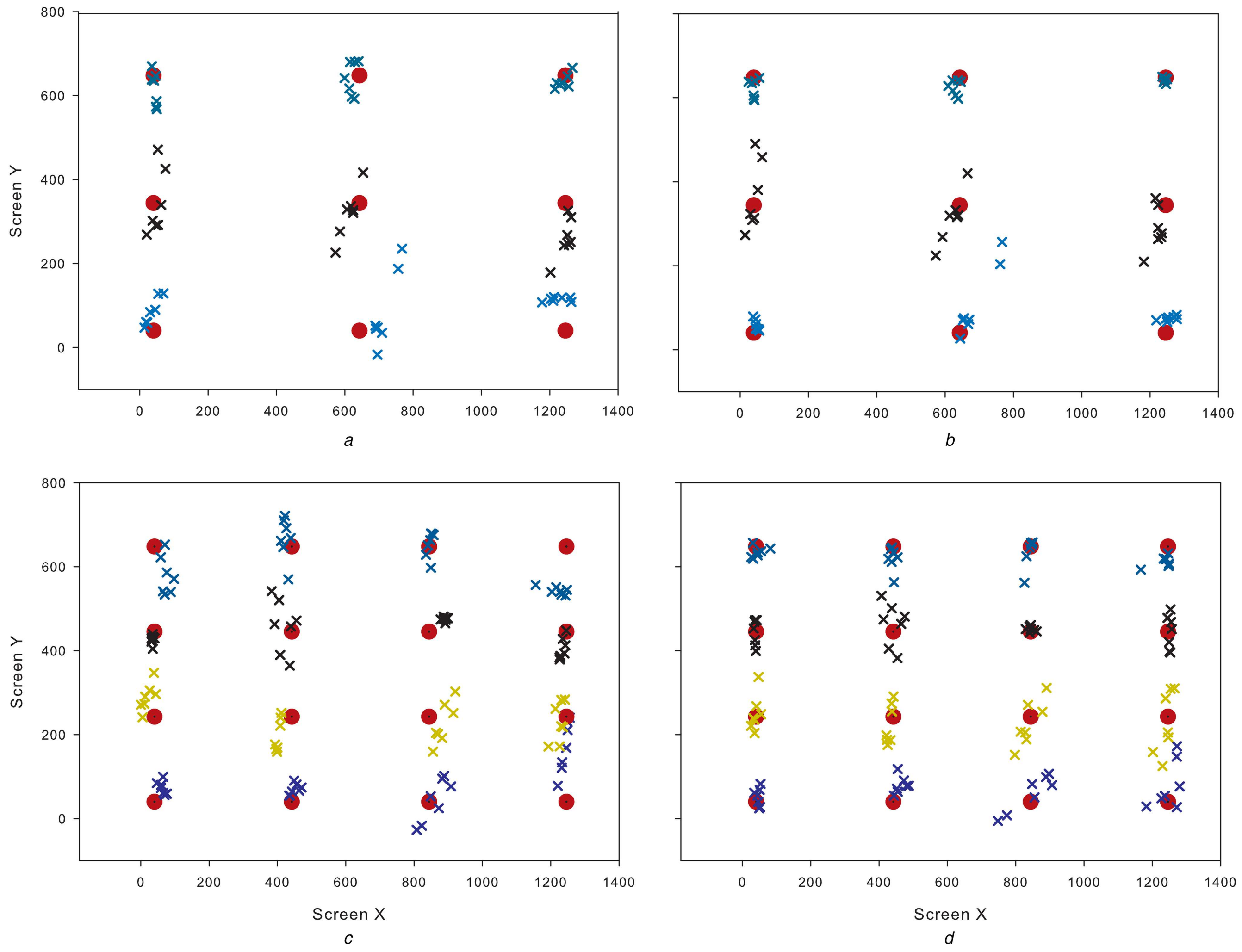

Fig. 10 *PoG estimates with 16 and 9 point calibration grids#a, c Polynomial regression*

b, d RBF kernel

Dots and crosses denote the target points and estimated gaze positions, respectively

Table 4 Gaze estimation accuracy

		Raw gaze position			With KF		
Method	Calibration points	MAE, deg	MHE, deg	MVE, deg	MAE, deg	MHE, deg	MVE, deg
polynomial	9	3.46	1.05	2.36	2.03	0.67	1.36
	16	2.97	0.98	2.01	1.95	0.62	1.32
RBF Kernel $\sigma_k = 5$	9	2.81	0.93	1.91	1.53	0.47	1.05
	16	2.71	0.87	1.83	1.33	0.40	0.91

MAE – mean absolute error; MHE – mean horizontal error; and MVE – mean vertical error.

Table 5 Accuracy of eye closure detection

Pixel per cell in HoG	RBF kernel SVM	Linear SVM
2	97.5% (SD = 3.1%)	98.3% (SD = 3.2%)
4	97.2% (SD = 2.7%)	98.6% (SD = 2.4%)

gaze vector. KF-based tracking is used to estimate the gaze accurately in video. Furthermore, ellipse parameters obtained from the algorithm can be combined with geometrical models for higher accuracy in gaze tracking. We have considered only in-plane rotations in this paper. However, pose invariant models can be developed by using more computationally complex 3D models.

6 Acknowledgment

The authors thank the subjects for participation in the experiment.

7 References

[1] Hansen, D.W., Ji, Q.: ‘In the eye of the beholder: a survey of models for eyes and gaze’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, **32**, (3), pp. 478–500

[2] Paperno, E., Semyonov, D.: ‘A new method for eye location tracking’, *IEEE Trans. Biomed. Eng.*, 2003, **50**, (10), pp. 1174–1179

[3] Zhu, Z., Ji, Q., Fujimura, K., *et al.*: ‘Combining Kalman filtering and mean shift for real time eye tracking under active IR illumination’. 16th Int. Conf. on Pattern Recognition, 2002. Proc., 2002, vol. **4**, pp. 318–321

[4] Markuš, N., Frljak, M., Pandžić, I.S., *et al.*: ‘Eye pupil localization with an ensemble of randomized trees’, *Pattern Recognit.*, 2014, **47**, (2), pp. 578–587

[5] Illingworth, J., Kittler, J.: ‘A survey of the Hough transform’, *Comput. Vis. Graph. Image Process.*, 1988, **44**, (1), pp. 87–116

[6] Young, D., Tunley, H., Samuels, R.: ‘*Specialised Hough transform and active contour methods for real-time eye tracking*’ (University of Sussex, Brighton, Cognitive & Computing Science, 1995)

[7] Smereka, M., Dulęba, I.: ‘Circular object detection using a modified Hough transform’, *Int. J. Appl. Math. Comput. Sci.*, 2008, **18**, (1), pp. 85–91

[8] Atherton, T.J., Kerbyson, D.J.: ‘Size invariant circle detection’, *Image Vis. Comput.*, 1999, **17**, (11), pp. 795–803

[9] Yang, P., Du, B., Shan, S., *et al.*: 'A novel pupil localization method based on Gabor eye model and radial symmetry operator'. , Int. Conf. on Image Processing, October 2004, (1), pp. 67–70

[10] Valenti, R., Gevers, T.: 'Accurate eye center location through invariant isocentric patterns', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (9), pp. 1785–1798

[11] Valenti, R., Sebe, N., Gevers, T.: 'Combining head pose and eye location information for gaze estimation', *IEEE Trans. Image Process.*, 2012, **21**, (2), pp. 802–815

[12] Timm, F., Barth, E.: 'Accurate eye centre localisation by means of gradients'. VISAPP, March 2011, pp. 125–130

[13] D'Orazio, T., Ancona, N., Cicirelli, C., *et al.*: 'A ball detection algorithm for real soccer image sequences'. 16th Int. Conf. on Pattern Recognition, 2002. Proc., 2002, (1), pp. 210–213

[14] Daugman, J.: 'How iris recognition works', *IEEE Trans. Circuits Syst. Video Technol.*, 2004, **14**, (1), pp. 21–30

[15] Baek, S.J., Choi, K.A., Ma, C., *et al.*: 'Eyeball model-based iris center localization for visible image-based eye-gaze tracking systems', *IEEE Trans. Consum. Electron.*, 2013, **59**, (2), pp. 415–421

[16] Sewell, W., Komogortsev, O.: 'Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network'. CHI'10 Extended Abstracts on Human Factors in Computing Systems, 2012, pp. 3739–3744

[17] Zhou, Z.H., Geng, X.: 'Projection functions for eye detection', *Pattern Recognit.*, 2004, **37**, (5), pp. 1049–1056

[18] Bhaskar, T.N., Keat, F.T., Ranganath, S., *et al.*: 'Blink detection and eye tracking for eye localization'. Conf. on Convergent Technologies for the Asia-Pacific Region, 2003, (2), pp. 821–824

[19] Wang, J., Sung, E., Venkateswarlu, R.: 'Eye gaze estimation from a single image of one eye'. Ninth IEEE Int. Conf. on Computer Vision, 2003. Proc., October 2003, pp. 136–143

[20] Zhang, X., Sugano, Y., Fritz, M., *et al.*: 'Appearance-based gaze estimation in the wild'. Computer Vision and Pattern Recognition, 2015, vol. **1**

[21] Schneider, T., Schauerte, B., Stiefelhagen, R.: 'Manifold alignment for person independent appearance-based gaze estimation'. IEEE Int. Conf. on Pattern Recognition, 2014, pp. 1167–1172

[22] Sugano, Y., Matsushita, Y., Sato, Y.: 'Learning-by-synthesis for appearance-based 3D gaze estimation'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2014, pp. 1821–1828

[23] Viola, P., Jones, M.: 'Rapid object detection using a boosted cascade of simple features'. Proc. of the 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2001, 2001, (1), pp. I–511

[24] Dasgupta, A., George, A., Happy, S.L., *et al.*: 'A vision-based system for monitoring the loss of attention in automotive drivers', *IEEE Trans. Intell. Transp. Syst.*, 2013, **14**, (4), pp. 1825–1838

[25] Li, D., Winfield, D., Parkhurst, D.J.: 'Starburst: a hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches'. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition-Workshops, June 2005, p. 79

[26] Fitzgibbon, A., Pilu, M., Fisher, R.B.: 'Direct least square fitting of ellipses', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1999, **21**, (5), pp. 476–480

[27] Fischler, M.A., Bolles, R.C.: 'Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography', *Commun. ACM*, 1981, **24**, (6), pp. 381–395

[28] Świrski, L., Bulling, A., Dodgson, N.: 'Robust real-time pupil tracking in highly off-axis images'. Proc. of the Symp. on Eye Tracking Research and Applications, March 2012, pp. 173–176

[29] Yoon, Y., Kosaka, A., Kak, A.C.: 'A new Kalman-filter-based framework for fast and accurate visual tracking of rigid objects', *IEEE Trans. Robot.*, 2008, **24**, (5), pp. 1238–1251

[30] Kiruluta, A., Eizenman, M., Pasupathy, S.: 'Predictive head movement tracking using a Kalman filter', *IEEE Trans. Syst. Man Cybern. B, Cybern.* , 1997, **27**, (2), pp. 326–331

[31] Dalal, N., Triggs, B.: 'Histograms of oriented gradients for human detection'. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2005, June 2005, (1), pp. 886–893

[32] Cristinacce, D., Cootes, T.F.: 'Facial feature detection and tracking with automatic template selection'. Seventh Int. Conf. on Automatic Face and Gesture Recognition, 2006, April 2006, pp. 429–434

[33] Vukadinovic, D., Pantic, M.: 'Fully automatic facial feature point detection using Gabor feature based boosted classifiers'. 2005 IEEE Int. Conf. on Systems, Man and Cybernetics, October 2005, (2), pp. 1692–1698

[34] Lewis, J.P.: 'Fast normalized cross-correlation', *Vis. Interface*, 1995, **1**, (10), pp. 120–123

[35] Tomasi, C., Kanade, T.: '*Detection and tracking of point features*' (School of Computer Science, Carnegie Mellon University, Pittsburgh, 1991)

[36] Pires, B.R., Hwangbo, M., Devyver, M., *et al.*: 'Visible-spectrum gaze tracking for sports'. 2013 IEEE Conf. on Computer Vision and Pattern Recognition Workshops, June 2013, pp. 1005–1010

[37] Sigut, J., Sidha, S.A.: 'Iris center corneal reflection method for gaze tracking using visible light', *IEEE Trans. Biomed. Eng.*, 2011, **58**, (2), pp. 411–419

[38] Nadaraya, E.A.: 'On estimating regression', *Theory Probab. Appl.*, 1964, **9**, (1), pp. 141–142

[39] Kohn, R., Smith, M., Chan, D.: 'Nonparametric regression using linear combinations of basis functions', *Stat. Comput.*, 2001, **11**, (4), pp. 313–322

[40] Cootes, T.F., Edwards, G.J., Taylor, C.J.: 'Active appearance models', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (6), pp. 681–685

[41] Cristinacce, D., Cootes, T.F.: 'Feature detection and tracking with constrained local models'. BMVC, 2006, vol. **1**, (2), p. 3

[42] Viola, P., Jones, M.J.: 'Robust real-time face detection', *Int. J. Comput. Vis.*, 2004, **57**, (2), pp. 137–154

[43] Jesorsky, O., Kirchberg, K.J., Frischholz, R.W.: 'Robust face detection using the Hausdorff distance', *Audio Video-based Biometric Person Authentication*, 2001, pp. 90–95

[44] 'BioID Database'. Available at https://www.bioid.com/About/BioID-Face-Database, accessed April 2014

[45] Ponz, V., Villanueva, A., Cabeza, R.: 'Dataset for the evaluation of eye detector for gaze estimation'. Proc. of the 2012 ACM Conf. on Ubiquitous Computing, September 2012, pp. 681–684

[46] Burrus, C.S.S., Parks, T.W.: '*DFT/FFT and convolution algorithms: theory and implementation*' (John Wiley & Sons, Inc., 1991)

[47] Bradski, G.: 'The OpenCV library', *Doct. Dobbs J.*, 2000, **25**, (11), pp. 120–126