



Eye Centre Localisation: An Unsupervised Modular Approach

Journal:	<i>Sensor Review</i>
Manuscript ID	SR-06-2015-0098.R2
Manuscript Type:	Original Manuscript
Keywords:	Eye centre localisation, HCI, Pupil and iris analysis, Pattern recognition

SCHOLARONE™
Manuscripts

Review Only

Eye Centre Localisation: An Unsupervised Modular Approach

Abstract

Purpose – This paper introduces an unsupervised modular approach for eye centre localisation in images and videos following a coarse-to-fine, global-to-regional scheme. The design of the algorithm aims at excellent accuracy, robustness and real-time performance for use in real-world applications.

Design/methodology/approach – A modular approach has been designed that makes use of isophote and gradient features to estimate eye centre locations. This approach embraces two main modalities that progressively reduce global facial features to local levels for more precise inspections. A novel Selective Oriented Gradient (SOG) filter has been specifically designed to remove strong gradients from eyebrows, eye corners and self-shadows, which sabotage most eye centre localisation methods. The proposed algorithm, tested on the BioID database, has shown superior accuracy.

Findings – The eye centre localisation algorithm has been compared with 11 other methods on the BioID database and 6 other methods on the G14E database. The proposed algorithm has outperformed all the other algorithms in comparison in terms of localisation accuracy while exhibiting excellent real-time performance. This method is also inherently robust against head poses, partial eye occlusions and shadows.

Originality/value – The eye centre localisation method utilises two mutually complementary modalities as a novel, fast, accurate and robust approach. In addition, other than assisting eye centre localisation, the SOG filter is able to resolve general tasks regarding the detection of curved shapes. From an applied point of view, the proposed method has great potentials in benefiting a wide range of real-world HCI applications.

Keywords Eye centre localisation, HCI, Pupil and iris analysis, Pattern recognition

Paper type Research paper

1. Introduction

Eye centre localisation from images and videos has received a considerable amount of attention in the area of HCI through utilization of computer vision techniques. Its study is still rapidly expanding due to the increased availability of HCI devices and systems. The ability to accurately localise eye centres can promise to bring significant benefits to a HCI system that is designed to observe its users and to capture user attentiveness. With the knowledge of user attention and predicted user intentions, a HCI system can make informed decisions and therefore can respond to a user in a more intelligent and natural way. Compared to other facial cues of a particular user such as age and gender that remain unchanged during a HCI session, eye analysis provides a constant stream of information that can address the dynamic nature of HCI. Furthermore, eye analysis excels in remote and contactless HCI, which provides an ideal channel for elderly people and those with motor disabilities to access HCI systems.

Generally, eye centre localisation methods fall into two main categories: passive methods and active methods. The former type of method is based on inherent features from periocular appearance or geometry, while the latter is based on additive features created by active lighting. More specifically, an additive feature based method projects near-infrared illumination toward the eyes and results in reflections on the corneas, which are commonly referred to as 'glints' (Zhu and Ji, 2005). This type of method has been widely employed for commercial eye trackers (Holmqvist, 2011). Being highly reliant on dedicated devices, this type of method essentially alters the primary task of eye centre detection into corneal reflection detection as a simplified task. Therefore, a passive inherent feature based method is more generalizable since it employs characteristic features from the eye region itself and therefore becomes the method we explore in this paper.

Both types of methods have witnessed real-world applications in different areas. For example, smart solutions are available that monitor gaze directions of a driver in order to identify driver distraction/drowsiness and provide timely alerts (Tawari, Chen and Trivedi, 2014). These driver assistance systems, capable of detecting and acting on driver inattentiveness, are of great value to road safety. Eye/gaze tracking has also been employed for psychological and medical applications (Mele and Federici, 2012). Its ability to reflect human processing of visual information and to indicate levels of cognitive load can contribute to reading comprehension and presentation design. For example, statistics such as eye saccade, gaze direction and gaze duration are valuable user feedback and therefore can lead to adaptive e-learning systems (Rosch and Vogel-Walcutt, 2013) that can react and adapt to users' psychological responses.

Although research in eye centre localisation is becoming increasingly active and sophisticated, a few issues remain unresolved that restrict most works from being transformed into real-world applications that can benefit the daily lives of human beings. We summarise the three major general issues that undermine the practicability of these works as follows:

- 1) *Lack of accuracy in real-world scenarios.* Many research works are tested on controlled databases where ideal illumination, high-resolution images and desirable viewpoint are available. When tested under various types of real scenes with dynamic environmental factors, their performance drops severely.
- 2) *Undesirable real-time performance.* As reliable as they might be, sophisticated algorithms often incur large computational cost, rendering them unsuitable for real-time implementations.
- 3) *High dependence on expensive or inconvenient hardware configuration.* A high cost and complexity of algorithm implementation will limit the usability and applicability of any method. Inexpensive yet effective methods are in high demand in order to boost real-world applications.

Therefore, resolving these issues can bridge the gaps in the field of eye centre localisation under realistic scenes and will give rise to HCI applications that are more accessible and robust. To this end, our method for eye centre localisation aims to maintain high accuracy on low-resolution images by utilising two types of features, and to increase robustness to head pose and to decrease computational cost by following a global-to-regional scheme. We also design a Selective Oriented Gradient (SOG) filter specifically to strengthen robustness to self-shadowing and interfering facial edges. Moreover, our method is unsupervised and only requires a webcam to function. Therefore it is of great practical value due to its ease of use and implementation.

1.1 Related works

Eye centre localisation methods are commonly based on analysis geometrical/morphological features (Cuong and Hoang, 2010) that conform to a pre-designed eye model, or machine learning techniques that train classifiers with appearance features. Others combine these two types of methods to increase the accuracy and robustness (Hansen and Ji, 2010) of eye centre localisation in images containing a wide range of variations (e.g. illumination variation, different head poses, shadows and specularities, etc.)

Timm and Barth (2011) proposed to localise eye centres by means of gradients. In this approach, periocular geometry was expressed by an objective function that peaked at the centre of a circular object. Despite its capability in dealing with deformation of circular pupil/iris contours, its performance would decline in the presence of strong gradients from eyelids, eyebrows, shadows and occluded pupil/iris. This remains to be an unsolved problem shared by most eye centre localisation methods. Another unsupervised method using geometrical features investigated the self-similarity space, where image regions that can maintain particular characteristics under geometric transformations receive high self-similarity scores (Leo et al., 2014). This eye model is derived from the rotational invariance of a pupil/iris region. As a result, extraction of a pupil/iris region directly affects computation of self-similarity scores, because the inclusion of eyebrows and other interfering sharp edges can produce high self-similarity scores that surpass those from the pupil/iris. In addition, isophote patterns also proved to be effective as a type of geometrical feature (Valenti and Gevers, 2008). Characterising contours of equal pixel intensity, isophotes are invariant to linear lighting changes as well as in-plane rotations and thus can give excellent eye centre localisation results even under challenging experimental conditions. Furthermore, isophote features have been combined with the shape regression model by Wei, Pang and Chen (2014) for improved accuracy and robustness.

On the other hand, with regard to machine learning based methods (Alpaydin, 2014), training data pose a critical influence on the performance of the algorithms (Zhu and Ramanan, 2012). More specifically, variations posed by illumination and head rotation have a huge impact on the accuracy and robustness of most algorithms. Inspired by Fisher Linear Discriminant (FLD) (Duda, Hart and Stork, 2012), Kroon, Hanjalic and Maas (2008) designed a linear filter trained by image patches extracted from normalized face images. This method not only considers the high response from a filtered face image, but also examines a rectangular neighbourhood around the estimated eye centre positions. This is based on the observation that a pupil in an image is formed by a collection of dark pixels within a small region. Another machine learning based method (Niu et al., 2006) focuses on the design of a novel classifier rather than the extraction of representative features. This method introduces a 2D cascade AdaBoost classifier that combines bootstrapping positive samples and bootstrapping negative samples (Viola and Jones, 2001). The final localisation of an eye is achieved by fusion of multiple classifiers. More recently, invariant eye patterns were investigated by Ren et al. (2014) in order to facilitate localisation of eyes with arbitrary angles in face images. They constructed a codebook of Scale Invariant Feature Transform (SIFT) features and validated its relative scale and rotation invariance on near-frontal face images rotated by from 0 to 180 degrees. In general, unsupervised methods have advantages that they are independent of training data and are therefore less biased toward a certain type of environmental setting. For unsupervised methods, models characterising eye regions are vital in determining the accuracy and robustness of the resulting algorithms.

Many eye centre localisation methods are designed for only frontal faces and thus deteriorate with the presence of head rotations and/or eye movements. Asadifard and Shanbezaheh (2010) employed a cumulative distributed function (CDF) for adaptive centre of pupil detection on frontal face images. Their approach firstly extracts the top-left and top-right quarters of a face image as the regions of interest and then filters each region of interest with a CDF. An absolute threshold is defined for the filtering process given the fact that the pixels in the pupil region are darker than the rest of the eye region. Despite the limitation that this method only accounts for frontal faces, the eye model in this method only considers the intensity values of an eye region in a greyscale image. Only when a complete pupil can be extracted by means of erosion would this method give an accurate estimation of the eye centre. However, under realistic scenes, specularities on a pupil or those caused by a pair of spectacles will split the pupil into several disconnected regions while self-cast shadows would easily change the values calculated by a CDF. Another study on frontal faces (Türkan, Pardas and Cetin, 2007) explored edge projections for eye localisation. With a face image available, their method firstly defines a rough horizontal position for the eye region according to facial anthropometric relations. After the eye band is cropped, it gathers eye candidate points that are extracted by a high-pass filter of a wavelet transform. A Support Vector Machine (SVM) based classifier (Chang and Lin, 2011) is then used to estimate the probability value for every eye candidate. This type of method normally requires that all face images are

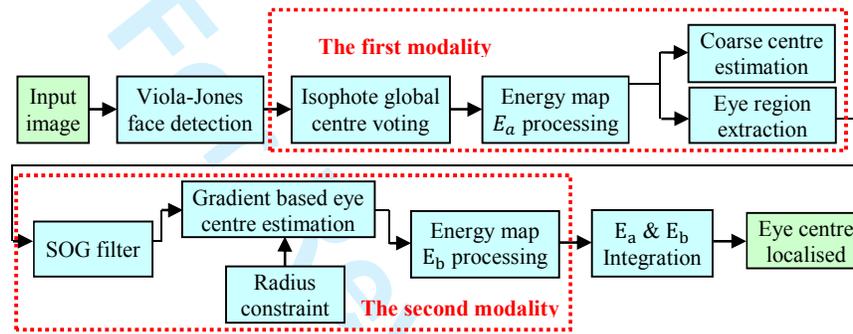
perfectly aligned so that the facial geometry agrees with facial anthropometric relations as the prior knowledge. Any misalignment will cause inconsistency to the features and will thus lead to poor results.

In summary, the attention that eye analysis receives has never faded despite the variety of observed facial features. A facility for eye centre localisation and exploitation of its practical application offers huge potential for HCI applications. The challenges of this research area largely arise from poor illumination conditions that create shadows and specularities around the eye region. Further complications arise from changes in head pose and eye movement, long distance scenarios and dependency on dedicated devices or complex system structures. Apart from the studies reviewed in this section, we provide a detailed summary of 11 state-of-the-art methods in section 3 and a comparison with our eye centre localisation method.

2. Eye centre localisation – an unsupervised modular approach

We propose a hybrid method that can perform accurate and efficient localisation of eye centres in low-resolution images and videos in real time. An overview of the proposed method can be found in Figure 1.

Figure 1 An overview of the eye centre localisation algorithm



The algorithm includes two main modalities, using global isophote features and regional gradient features, respectively. The first modality performs a global estimation of eye centres over a face image and extracts eye regions. Results from the first modality are then fed into the second modality as prior knowledge, which lead to a local and more precise estimation of eye centres. The two energy maps generated by the two main modalities are fused to enable the final estimation of eye centres.

2.1 Isophote based global centre voting and eye detection

A pupil and an iris in an image can be represented by contours of equal intensity values, i.e. isophotes (Lichtenauer, Hendriks and Reinders, 2005). Equation (1) is then formulated (Valenti and Gevers, 2008) to calculate displacement vectors which point from pixels to the centres of isophotes they belong to.

$$\{D_x, D_y\} = -\frac{\{I_x, I_y\}(I_x^2 + I_y^2)}{I_y^2 I_{xx} - 2I_x I_{xy} I_y + I_x^2 I_{yy}} \quad (1)$$

where I_x, I_y, I_{xx}, I_{xy} and I_{yy} are first-order and second-order derivatives of the luminance function $I(x, y)$. The importance of each vote is indicated by the curvedness of the isophote since the iris and pupil edges that are circular obtain high curvedness values as opposed to flat isophotes. The curvedness (Koenderink and Doorn, 1992) can be calculated as:

$$cd(x, y) = \sqrt{I_{xx}^2 + 2 \times I_{xy}^2 + I_{yy}^2} \quad (2)$$

We also consider the brightness of the isophote centre in the voting process based on the fact that the pupil is normally darker than the iris and the sclera. Therefore, an energy map $E_a(x, y)$ is constructed that collects all the votes to reflect the eye centre position following equation (3), where α is the maximum greyscale in the image ($\alpha = 255$ in the experiments).

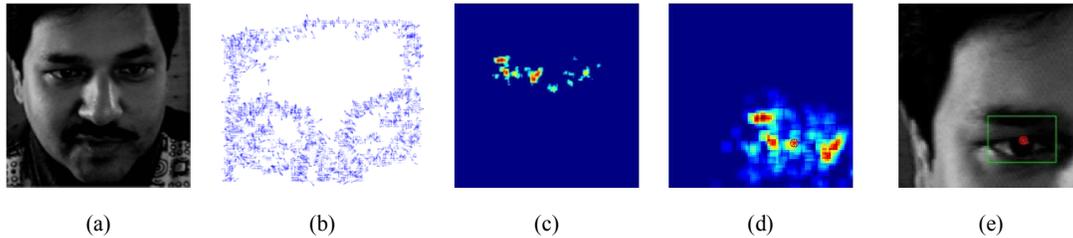
$$E_a(x + D_x, y + D_y) = [\alpha - I(x + D_x, y + D_y)] \times cd(x, y) \quad (3)$$

Our isophote based modality is *different* from other isophote feature based methods, in that it extracts isophote features for the *whole face* instead of periorcular regions. Pixels with intensities below 30% of the maximum value are removed. The lower half of $E_a(x, y)$ is simply removed since it is unlikely to concern any eye region information regardless of normal head rotations. For the remainder of the left and right half of the energy map $E_{aul}(x, y)$ and $E_{aur}(x, y)$, we further calculate the energy centre, i.e. the first moment divided by the total energy, which is selected instead as the optimal eye centres. Taking $E_{aul}(x, y)$ as an example, this can be formulated as equation (4)

$$\{cx_{aul}, cy_{aul}\} = \frac{\sum_{x=1}^m \sum_{y=1}^n \{x, y\} \cdot E_{aul}(x, y)}{\sum_{x=1}^m \sum_{y=1}^n E_{aul}(x, y)} \quad (4)$$

where $C_{aul} = \{cx_{aul}, cy_{aul}\}$ is the optimal estimation of the left eye centre, m and n are the maximum row and column number in E_{aul} . The eye region to be analysed by our second modality is then selected which centres at the optimal eye centre estimation (its width being 1/10 of the face size and its height being 1/15 of the face size). As a result, our method does not require an eye detector and is robust to head rotations since *global* isophotes are investigated. This process is shown in Figure 2.

Figure 2 An illustration of the first modality for eye centre estimation and eye detection



Note: (a) A face image (BioID Technology Research, 2001) detected by the Viola-Jones face detector (Viola and Jones, 2004). (b) Displacement vectors calculated by equation (1). (c) The energy map calculated by equation (3). (d) The energy centre found by equation (4), denoted by a circled asterisk sign. (e) The optimal eye centre position found and the eye region selected.

The isophote features employed by the first modality are extracted on the global level, i.e. from a whole face image rather than a pre-detected eye region. Despite its computational efficiency, it is deemed a coarse estimation due to global variations such as uneven lighting and strong edges from outside the eye region. Therefore, a second modality is needed to analyse gradient features iteratively at the local level, i.e. from within a refined eye region, for enhanced precision. The modality introduced in Section 2.2 is further boosted by an iris radius constraint (Section 2.3) and a SOG filter (Section 2.4).

2.2 Gradient based eye centre estimation

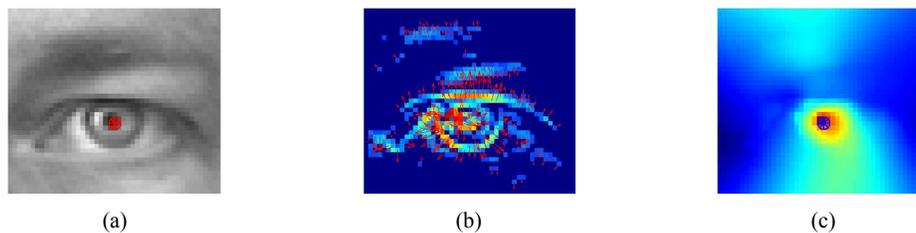
A pupil and an iris can also be characterised by radial vectors pointing inward to an eye centre. Specifically, prominent gradient vectors on a circular iris/pupil boundary should agree with the radial directions. Therefore the dot product of each gradient vector with its corresponding radial vector is maximised. This is formulated by Timm and Barth (2011) as an objective function:

$$c^* = \arg \max_{\mathbf{c}} \left\{ \frac{1}{N} \sum_{x=1}^m \sum_{y=1}^n I_c(x, y) \cdot (\mathbf{d}^T(x, y) \cdot \mathbf{g}(x, y))^2 \right\} \quad (5)$$

$$\mathbf{d}(x, y) = \frac{\mathbf{p}(x, y) - \mathbf{c}}{\|\mathbf{p}(x, y) - \mathbf{c}\|_2}, \quad \forall x \forall y: \|\mathbf{d}(x, y)\|_2 = 1, \quad \|\mathbf{g}(x, y)\|_2 = 1 \quad (6)$$

where \mathbf{c} is the centre candidates, c^* is the optimal centre, N is the number of pixels in the eye region to be analysed, $\mathbf{d}(x, y)$ is the displacement vector connecting a centre candidate \mathbf{c} and $\mathbf{p}(x, y)$ which is a pixel different from \mathbf{c} , $\mathbf{g}(x, y)$ is the gradient vector, I_c is the intensity value at an isophote centre and m and n have the same definition as in the preceding subsection. The displacement vectors and gradient vectors are both normalised to unit vectors. We only preserve the gradients whose directions are reverse to the corresponding displacement vectors based on the fact that the pupil should be darker than its neighbouring regions and thus generates inward gradients. A sample implementation of this approach is illustrated in Figure 3.

Figure 3 A sample implementation of the second modality using local gradient features



Note: Using gradient features for eye centre localisation. (a) An eye image where the localised eye centre is displayed. (b) The display of gradient magnitude where the gradient directions (inverted) are represented by arrows. The gradients with magnitude below 70% of the maximum are removed. (c) The resulting energy map for eye centre candidates.

To resolve the challenges posed by strong shadows inside an eye region and sharp edges outside the eye region, we introduce a radius constraint and design a SOG filter that can effectively deal with the circularity measure and automatically remove strong gradients from eyebrows and eyelids.

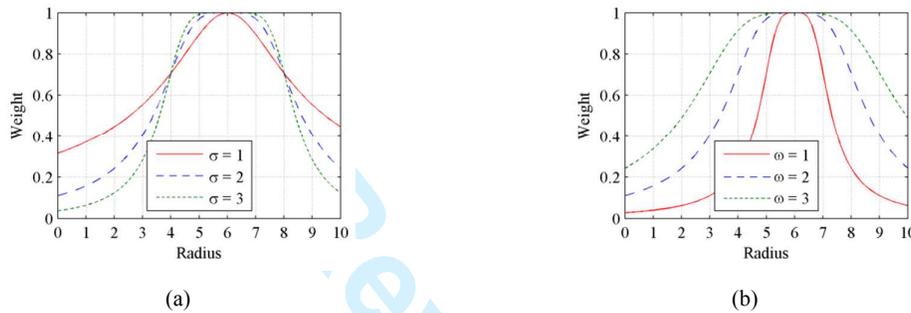
2.3 Iris radius constraint

We introduce an iris radius constraint based on the assumption that shadows and eyebrow segments have random radius values, while the iris radii are more constant relative to the size of a face. This provides a way to differentiate circular clusters of various radii and to determine their weights in energy map accumulation. The function for the significance measure emulates the frequency response of a Butterworth low pass filter:

$$rw(x, y) = \frac{1}{\sqrt{1 + \left(\frac{\|d(x, y)\|_2 - D}{\omega}\right)^{2\sigma}}} \quad (7)$$

where $\|d(x, y)\|_2$ is the L_2 norm of the displacement vector without being normalised to a unit vector. D is the estimated radius of the iris. σ and ω correspond to the order and the cutoff frequency of the filter. The curves corresponding to varying σ and ω following equation 7 are shown in Figure 4.

Figure 4 The radius weight function curves emulating the frequency responses of a Butterworth low pass filter



Notes: (a) Curves with varying σ ($\sigma = 1, 2, 3$) and constant ω ($\omega = 2$). (b) Curves with varying ω ($\omega = 1, 2, 3$) and constant σ ($\sigma = 2$).

The radius weight function is maximally flat around the estimated centre D and drops rapidly when the radius is out of the flatness band. Increasing ω while decreasing σ will enhance the rigidity of the constraint which could be assumed for circumstances where strong shadows are present. D is a value relative to the size of the face region. This objective function allows adjustable coefficients that effectively alleviate the problematic issues posed by edges around eyelids, eye corners and shadows.

2.4 SOG filter

We specifically design a novel SOG filter that discriminates gradients of rapid change in orientation from those of less change. It is perfectly tailored to reinforce the two main modalities despite its versatile applicability. The basic idea takes the form of a statistical analysis of gradient orientations within a window centred at a pixel position. For each $D_x \times D_y$ window centred at pixel i , the gradients in x and y directions are calculated whose orientations follow:

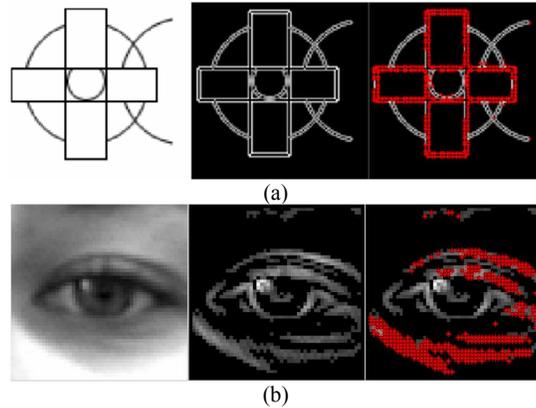
$$\text{orient}G = \tan^{-1}\left(\frac{I_y}{I_x}\right) \cdot \frac{180^\circ}{\pi} \quad (8)$$

We then stack gradient orientations into k ($k < 360$) orientation bins, where each bin contains the count of the orientations from $b \cdot \frac{360^\circ}{k}$ to $(b + 1) \cdot \frac{360^\circ}{k}$ ($0 \leq b \leq k - 1$) within the window. If the count recorded in a bin exceeds a threshold, the corresponding pixels that accumulate the bin will have their gradient vectors halved, i.e. their weights reduced. As a result, the objective function becomes:

$$c^* = \arg \max_c \left\{ \frac{1}{N} \sum_{x=1}^m \sum_{y=1}^n sw(x, y) \cdot rw(x, y) \cdot [\alpha - I(x, y)] \cdot (d^T(x, y) \cdot g(x, y))^2 \right\} \quad (9)$$

where $sw(x, y)$ is the weight of a gradient vector adjusted by the SOG filter.

The threshold for the counts is determined by an absolute value as well as a value relative to the number of pixels with non-zero gradients within the window. As a result, the pixels that maintain similar gradient orientations to their neighbours will have their weights reduced and they are referred to as 'impaired pixels' in the rest of the paper. When a curve has low local curvature, it comprises more 'impaired pixels'. Figure 5 demonstrates the effectiveness of a SOG filter applied to an image containing irregular curves and an image of an eye region.

Figure 5 Gradient filtering using a SOG filter

Note: (a) An example of curved shape detection using a SOG filter where the ‘impaired pixels’ are detected and marked in red. (b) An example where a SOG filter is applied to an eye image. The magnitudes of the gradients are computed where the edges on the eye pouch and eye lid are successfully detected and marked in red.

It is shown in Figure 5 that the SOG filter has successfully distinguished curves with low and high curvatures and that it is effective in dealing with intersected and occluded curves or curve segments. In the eye image, the eyelid and shadowed eye pouches are detected as ‘impaired pixels’ whose weights are to be reduced in the accumulation of the energy map while gradients around the iris and the pupil maintain their original weights. This approach resolves challenges brought by shadows, facial makeup and edges on eyelids, eyebrows and other facial parts outside the iris, which are extremely problematic for most geometrical feature based eye centre localisation approaches.

2.5 Energy map integration

In the final stage, the two energy maps E_a and E_b are integrated into $E_f(x, y)$ so that they both contribute to the election of the eye centre. It is critical, prior to the integration, to determine the confidence of each modality, to estimate the complexity of the eye image, and thus to determine their weights in the fusion mechanism.

The left eye region is taken as an example to illustrate the fusion mechanism. If the equivalent centroid C_{aul} calculated by equation (4) is close to the pixel position C_{maxl} that has the maximum value in the first energy map $E_{aul}(x, y)$, C_{aul} is considered confident since the isophote centre and the equivalent centroid coincide. In this case, more gradients from the pupil/iris are present, allowing the second modality to be more robust and precise. The two main modalities are then utilised and fused following equation (10):

$$E_f(x, y) = \frac{1}{\|C_{aul} - C_{Eamax}\|_2} \cdot E_a(x, y) + E_b(x, y) \quad (10)$$

where $0 < \|C_{aul} - C_{Eamax}\|_2 \leq \epsilon$, with ϵ taking a value relative to the width of the eye region ϵ_f . In our experiments, $\epsilon = 0.3\epsilon_f$ pixels. The coefficient of 0.3 is an empirical value found to be a suitable weight adjustment factor for the first modality. It corresponds to the average diameter of the iris in our experiments. When C_{aul} and C_{maxl} disagree and have a large Euclidean distance, the first energy map will have high energy clusters sparsely distributed, potentially caused by severe shadows and specularities. The second modality will be influenced by ‘impaired pixels’ and produce erroneous centre estimates. Therefore only the equivalent centroids C_{aul} and C_{aur} are selected to be the final eye centres.

The maximum response in the final energy map will represent the estimated eye centre. Estimation of the final right eye centre follows the same approach.

3. Eye centre localisation experiments and results

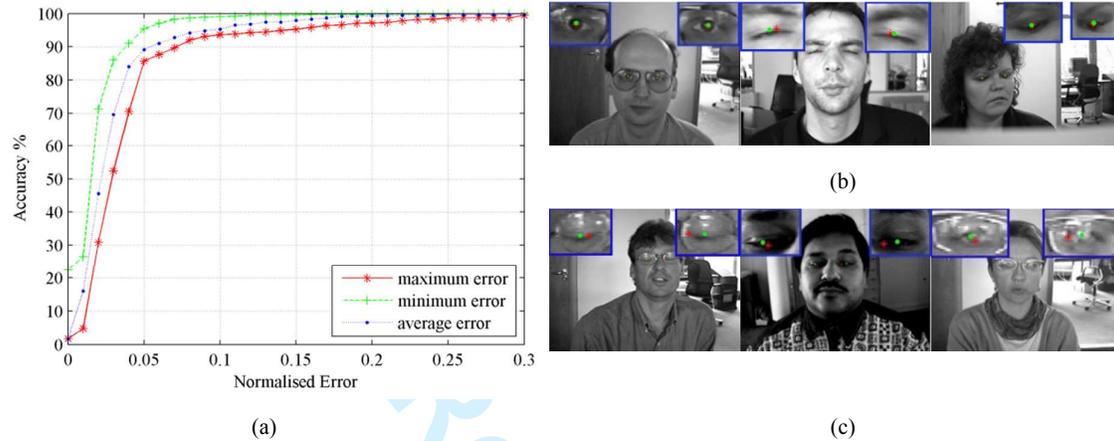
We employed two publicly available datasets for algorithm evaluation, i.e. the BioID dataset and the GI4E dataset. The BioID dataset (BioID Technology Research, 2001) consists of 1520 images with ground truth data of eye centre coordinates. It has been popular in the literature for the evaluation of other eye centre localisation algorithms. The variations in this dataset include illumination, face scale, head pose and the presence of glasses. The GI4E dataset (Villanueva et al., 2013) contains images of 103 subjects with 12 different gaze directions. Therefore, the impact of eye and head movement on the proposed algorithm can be evaluated. The relative error measure proposed by Jesorsky, Kirchberg and Frischholz (2001) was used to evaluate the accuracy of the algorithm. This approach firstly calculates the absolute error, i.e. the Euclidean distance between the centre estimates and the ground truth provided by the dataset, and then normalises the Euclidean distance relative to the pupillary distance. This is formulated by equation (12).

$$e = \frac{\max(d_{left}, d_{right})}{\omega} \quad (12)$$

where d_{left} and d_{right} are the absolute errors for the eye pair, and ω is the pupillary distance in pixels. The maximum of d_{left} and d_{right} after normalisation is defined as 'max normalised error' e_{max} . Additionally, the accuracy curve for the minimum normalised error e_{min} and the average normalised error e_{avg} are calculated. A relative distance of $e = 0.25$ corresponds to half the width of an eye.

Evaluation results on the BioID dataset are shown in Figure 6 where representative examples of accurately and inaccurately localised eye centres are demonstrated, as well as the overall accuracy curve on this dataset. The proposed algorithm is further compared with 11 other methods in the literature, summarised in Table 1.

Figure 6 Results of the proposed eye centre localisation method tested on the BioID dataset



Note: (a) Accuracy curve of the proposed method on the BioID dataset. (b) Examples of accurately localised eye centres. (c) Examples of inaccurately localised eye centres. The localised eye centres are marked in red, compared to the ground truth (provided by the dataset) marked in green.

Table 1 Comparison of the accuracy for eye centre localisation on the BioID dataset.

Method	Accuracy under minimum and maximum normalised error						Score
	$e_{max} \leq 0.05$	$e_{min} \leq 0.05$	$e_{max} \leq 0.10$	$e_{min} \leq 0.10$	$e_{max} \leq 0.25$	$e_{min} \leq 0.25$	
The proposed method	85.66%	95.46%	93.68%	99.06%	99.21%	99.93%	6
(Ren et al., 2014)	77.08%	\	92.25%	\	<u>98.99%</u>	\	1
(Leo et al., 2014)	80.67%	\	87.31%	\	93.86%*	\	0
(Valenti and Gevers, 2012)	86.09%	<u>96.07%</u>	91.67%	97.87%	97.87%	100%*	3
(Timm and Barth, 2011)	82.50%	93.50%*	93.40%	<u>98.50%*</u>	98.00%	100%*	1
(Asadifard and Shanbezadeh, 2010)	47.00%	\	86.00%	\	96.00%	\	0
(Kroon, Hanjalic and Maas, 2008)	65.00%	\	87.00%	\	98.80%	\	0
(Valenti and Gevers, 2008)	84.10%	96.28%	90.85%	97.94%	98.49%	100%*	2
(Campadelli, Lanzarotti and Lipori, 2006)	62.00%	\	85.20%	\	96.10%	\	0
(Niu et al., 2006)	75.10%*	\	93.00%	\	96.30%*	\	0
(Hamouz et al., 2005)	58.00%*	\	76.00%*	\	90.80%*	\	0
(Cristinacce, Cootes and Scott, 2004)	57.00%*	\	96.00%*	\	97.10%*	\	2

Note: Those with a '*' notation are not explicitly provided by the author(s) but are measured from the accuracy curves available. Those with a '\ ' notation are neither explicitly nor implicitly provided by the author(s). The numbers in bold are the highest accuracy in their corresponding ranges and those underlined the second highest. The accuracy measure for $e_{min} \leq 0.25$ does not contribute to scores since very similar results have been achieved by all the methods.

The proposed method gains the best results for the accuracy measure $e_{min} \leq 0.10$ as well as $e_{max} \leq 0.25$, and the second best for $e_{max} \leq 0.05$ and $e_{max} \leq 0.10$. Except for the accuracy measure for $e_{min} \leq 0.25$ where very similar results have been achieved by all the listed method, a score of 2 is assigned to every first rank and a score of 1 is assigned to every second rank. The proposed method gains a total score of 6, outperforming all the other methods in comparison.

Figure 7 The Accuracy curves of the proposed method on the GI4E dataset, in comparison with 6 other methods

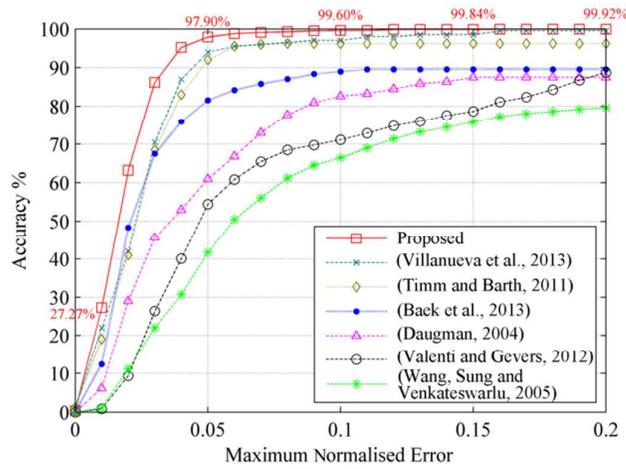
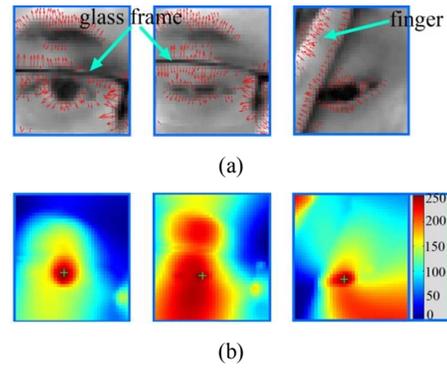


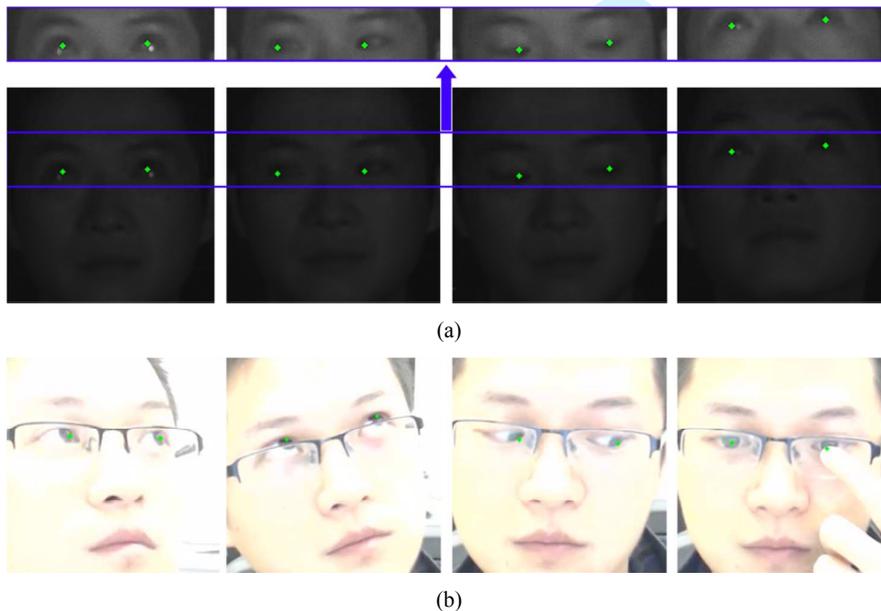
Figure 8 Results of eye centre localisation on representative image frames from a video



Note: (a) eye regions from the raw video frames, where the arrows represent strong gradients on edges. (b) the energy maps corresponding to all pixels in (a).

Similarly to the experiments by Villanueva et al. (2013) and Baek et al. (2013), we also evaluated the proposed method on the GI4E dataset and compared it to 6 other methods. As shown by Figure 7, the proposed method outperforms all the other methods in comparison and proves to be robust against eye/head movement by achieving 97.9% accuracy for $e_{max} \leq 0.05$. Tests on video data also validated the superior accuracy and efficiency of the proposed method. Results of a few representative image frames from a video recording can be found in Figure 8, which further reflect the accuracy and robustness of gradient features against facial accessory, dynamic eye morphology and occlusion. In addition, we further evaluated the robustness of the proposed method against head poses and eye occlusions in challenging illumination conditions. In this test, an under-illuminated environment was created by two near-infrared (NIR) lamps while a NIR bandpass filter was mounted to the camera lens. An over-exposed condition was simulated by adjusting camera settings (i.e. exposure and gain). Results can be found in Figure 9 which displays representative frames captured during this test.

Figure 9 Eye centre localisation results demonstrating the robustness of the proposed algorithm against challenging illuminations, head poses and eye occlusions



Note: (a) Representative results generated in an under-illuminated environment. For assist with visualisation, we increased the brightness of the eye regions in (a) and displayed them separately. In (a), from left to right, challenging conditions include specularities, eye movement, half-closed eyes and head pose. (b) Representative results generated from over-exposed image frames. In (b), from left to right, challenging conditions include head pose, facial accessory, extreme eye position and eye occlusion.

We further demonstrate the simplicity and efficiency of the proposed method by comparing it to Timm and Barth (2011) which claims to have achieved excellent real-time performance as one of its key features. Take the image containing a 41×47 eye region, i.e. 1927 pixels, as an example (Figure 3), the method in comparison performs per-pixel estimation of the eye centre, assuming that every pixel is an eye centre candidate. Therefore 1927 iterations are needed before the optimal candidate can be selected. The proposed method, on the other hand, resolves the problem by utilising the prior knowledge from the first modality which, through an initial estimation, avoids the per-pixel candidate assumption. The removal of low-energy pixels in the first modality largely reduces the number of candidates, i.e. number of iterations in the second modality. In the same sample image of the 41×47 eye region, the iterations decrease to only 67, making our algorithm 29 times faster. With a webcam running at 30 frames per second, the proposed algorithm is capable of localising eye centres in detected faces accurately in real time.

4. Discussion and conclusion

This paper introduces an unsupervised modular approach for eye centre localisation, suitable for HCI system implementations. The novelty of this approach can be summarised as follows:

1) The proposed modular scheme is different from the majority of existing approaches that follow the conventional three stages, i.e. pre-processing, feature extraction and classification. In fact, the two main modalities in our method are based on different eye models and employ different types of features. As a result, the proposed method can benefit from the relative rotational invariance brought by the isophote features, as well as the iterative circularity measurements by the gradient features.

2) Most existing methods require the eye regions to be located prior to eye centre localisation. This is commonly achieved by image cropping according to facial anthropometrics or pre-trained eye detectors. While the former type of approach cannot deal with different head poses, the latter type is complex and often yields inaccurate results in undesirable environments. In contrast, the proposed method first analyses features at the global level such that no eye detector is needed. Subsequently, global features are processed for regional analysis, meaning that the local features are naturally robust to head poses.

3) The designs of the iris radius constraint and the SOG filter are tailored to this modular scheme. They effectively deal with interfering edges and shadows, which severely sabotage most eye centre localisation algorithms.

4) The interaction of the two main modalities not only results in increased accuracy and robustness, but it reduces the computation load by filtering eye candidates before performing iterative analysis.

A number of experiments have been carried out to justify the effectiveness of the proposed method on different datasets, under different illuminations, and with the presence of head poses and eye occlusions. Tested on the BioID dataset, the proposed method has outperformed the 11 other methods in comparison with the highest localisation accuracy. Tested on the GI4E dataset, the proposed method has achieved the highest accuracy, compared to 6 other methods. The proposed method will afford great potentials in implementations of a variety of HCI systems and will benefit the development of assistive technologies that can assist the elderly people and those with motor disabilities.

References

- Alpaydin, E. (2014), *Introduction to machine learning*, MIT press.
- Asadifard, M. and Shanbezhadeh, J. (2010), "Automatic adaptive center of pupil detection using face detection and cdf analysis", in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, IAENG, Hong Kong.
- Baek, S.J., Choi, K.A., Ma, C., Kim, Y.H. and Ko, S.J. (2013), "Eyeball model-based iris center localization for visible image-based eye-gaze tracking systems", *IEEE Transactions on Consumer Electronics*, vol. 59, no. 2, pp. 415-421.
- BioID Technology Research (2001), "BioID Face Database", available at: <https://www.bioid.com/About/BioID-Face-Database/> (accessed 15 March 2013).
- Campadelli, P., Lanzarotti, R., and Lipori, G. (2006), "Precise eye localization through a general-to-specific model definition", in *Proceedings of British Machine Vision Conference*, BMVA, pp. 187-196.
- Chang, C.C. and Lin, C.J. (2011), "LIBSVM: A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3.
- Cristinacce, D., Cootes, T.F., and Scott, I.M. (2004), "A Multi-Stage Approach to Facial Feature Detection", in *Proceedings of British Machine Vision Conference*, BMVA, pp. 1-10.
- Cuong, N.H. and Hoang, H.T. (2010), "Eye-gaze detection with a single webcam based on geometry features extraction", in *Proceedings of 11th IEEE International Conference on Control Automation Robotics & Vision*, IEEE, pp. 2507-2512.
- Daugman, J. (2004), "How iris recognition works", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 21-30.
- Duda, R.O., Hart, P.E. and Stork, D.G. (2012), *Pattern classification*, 2nd ed., John Wiley & Sons.
- Hamouz, M., Kittler, J., Kamarainen, J.K., Paalanen, P., Kalviainen, H. and Matas, J. (2005), "Feature-based affine-invariant localization of faces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 9, pp. 1490-1495.

- 1
2 Hansen, D.W. and Ji, Q. (2010), "In the eye of the beholder: A survey of models for eyes and gaze", *IEEE Transactions on*
3 *Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478-500.
- 4 Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. and Van de Weijer, J. (2011), *Eye tracking: A*
5 *comprehensive guide to methods and measures*, Oxford University Press.
- 6 Jesorsky, O., Kirchberg, K.J. and Frischholz, R.W. (2001), "Robust face detection using the hausdorff distance", *Audio-and*
7 *video-based biometric person authentication*, Springer Berlin Heidelberg, pp. 90-95.
- 8 Kapoor, A., Burleson, W. and Picard, R. W. (2007), "Automatic prediction of frustration", *International journal of human-*
9 *computer studies*, vol. 65, no. 8, pp. 724-736.
- 10 Koenderink, J.J. and Doorn, A.J.v. (1992), "Surface shape and curvature scales", *Image and vision computing*, vol. 10, no. 8,
11 pp. 557-564.
- 12 Kroon, B., Hanjalic, A. and Maas, S.M. (2008), "Eye localization for face matching: is it always useful and under what
13 conditions?", in *Proceedings of the 2008 international conference on Content-based image and video retrieval*, ACM, New
14 York, NY, pp. 379-388.
- 15 Leo, M., Cazzato, D., De Marco, T. and Distanto, C. (2014), "Unsupervised Eye Pupil Localization through Differential
16 Geometry and Local Self-Similarity Matching", *PloS one*, vol. 9, no. 8.
- 17 Lichtenauer, J., Hendriks, E. and Reinders, M. (2005), "Isophote properties as features for object detection", in *Proceedings*
18 *of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 649-654.
- 19 Mele, M.L. and Federici, S. (2012), "Gaze and eye-tracking solutions for psychological research", *Cognitive processing*, vol.
20 13, no.1, pp.261-265.
- 21 Niu, Z., Shan, S., Yan, S., Chen, X. and Gao, W. (2006), "2D cascaded Adaboost for eye localization", in *Proceedings of*
22 *18th International Conference on Pattern Recognition*, IEEE, pp. 1216-1219.
- 23 Ren, Y., Wang, S., Hou, B. and Ma, J. (2014), "A novel eye localization method with rotation invariance", *IEEE*
24 *Transactions on Image Processing*, vol. 23, no. 1, pp. 226-239.
- 25 Rosch, J.L. and Vogel-Walcutt, J.J. (2013), "A review of eye-tracking applications as tools for training", *Cognition,*
26 *technology & work*, vol. 15, no. 3, pp.313-327.
- 27 Tawari, A., Chen, K. H. and Trivedi, M. M. (2014), "Where is the driver looking: Analysis of head, eye and iris for robust
28 gaze zone estimation", in *Proceedings of 17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*,
29 IEEE, pp. 988-994.
- 30 Timm, F. and Barth, E. (2011), "Accurate Eye Centre Localisation by Means of Gradients", in *Proceedings of International*
31 *Conference on Computer Vision Theory and Applications*, pp. 125-130.
- 32 Türkan, M., Pardas, M. and Cetin, A.E. (2007), "Human eye localization using edge projections", in *Proceedings of*
33 *International Conference on Computer Vision Theory and Applications*, pp. 410-415.
- 34 Valenti, R. and Gevers, T. (2008), "Accurate eye center location and tracking using isophote curvature", in *Proceedings of*
35 *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 1-8.
- 36 Valenti, R. and Gevers, T. (2012), "Accurate eye center location through invariant isocentric patterns", *IEEE Transactions on*
37 *Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1785-1798.
- 38 Villanueva, A., Ponz, V., Sesma-Sanchez, L., Ariz, M., Porta, S. and Cabeza, R. (2013), "Hybrid method based on
39 topography for robust detection of iris center and eye corners", *ACM Transactions on Multimedia Computing,*
40 *Communications, and Applications*, vol. 9, no. 4.
- 41 Viola, P. and Jones, M.J. (2004), "Robust real-time face detection", *International journal of computer vision*, vol. 57, no. 2,
42 pp. 137-154.
- 43 Viola, P. and Jones, M.J. (2001), "Rapid object detection using a boosted cascade of simple features", in *Proceedings of the*
44 *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, vol. 1, pp. I-511-518.
- 45 Wachs, J. P., Stern, H. I., Edan, Y., Gillam, M., Handler, J., Feied, C., and Smith, M. (2008), "A gesture-based tool for sterile
46 browsing of radiology images", *Journal of the American Medical Informatics Association*, vol. 15, no.9, pp. 321-323.
- 47 Wang, J.G., Sung, E. and Venkateswarlu, R. (2005), "Estimating the eye gaze from one eye", *Computer Vision and Image*
48 *Understanding*, vol. 98, no. 1, pp. 83-103.
- 49 Wei, C., Pang, Z. and Chen, D. (2014), Combining shape regression model and isophotes curvature information for eye
50 center localization, in *Proceedings of 7th IEEE International Conference on Biomedical Engineering and Informatics*, IEEE,
51 pp. 156-160.
- 52 Zhu, X. and Ramanan, D. (2012), "Face detection, pose estimation, and landmark localization in the wild", in *Proceedings of*
53 *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 2879-2886.
- 54 Zhu, Z. and Ji, Q. (2005), "Robust real-time eye detection and tracking under variable lighting conditions and various face
55 orientations", *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 124-154.
- 56
57
58
59
60