frontiers in ICT

ORIGINAL RESEARCH
published: 14 October 2019
doi: 10.3389/fict.2019.00018

Eyelid and Pupil Landmark Detection and Blink Estimation Based on Deformable Shape Models for Near-Field Infrared Video

Siyuan Chen[1]* *and Julien Epps*[1,2]*

[1] School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW, Australia,
[2] Data61, CSIRO, Sydney, NSW, Australia

The eyelid contour, pupil contour, and blink event are important features of eye activity, and their estimation is a crucial research area for emerging wearable camera-based eyewear in a wide range of applications e.g., mental state estimation. Current approaches often estimate a single eye activity, such as blink or pupil center, from far-field and non-infrared (IR) eye images, and often depend on the knowledge of other eye components. This paper presents a unified approach to simultaneously estimate the landmarks for the eyelids, the iris and the pupil, and detect blink from near-field IR eye images based on a statistically learned deformable shape model and local appearance. Unlike the facial landmark estimation problem, by comparison, different shape models are applied to all eye states—closed eye, open eye with iris visible, and open eye with iris and pupil visible—to deal with the self-occluding interactions among the eye components. The most likely eye state is determined based on the learned local appearance. Evaluation on three different realistic datasets demonstrates that the proposed three-state deformable shape model achieves state-of-the-art performance for the open eye with iris and pupil state, where the normalized error was lower than 0.04. Blink detection can be as high as 90% in recall performance, without direct use of pupil detection. Cross-corpus evaluation results show that the proposed method improves on the state-of-the-art eyelid detection algorithm. This unified approach greatly facilitates eye activity analysis for research and practice when different types of eye activity are required rather than employ different techniques for each type. Our work is the first study proposing a unified approach for eye activity estimation from near-field IR eye images and achieved the state-of-the-art eyelid estimation and blink detection performance.

Keywords: landmark detection, deformable shape model, eyelid estimation, pupil estimation, blink detection

OPEN ACCESS

Edited by:
Anton Nijholt,
University of Twente, Netherlands

Reviewed by:
Federica Marcolin,
Politecnico di Torino, Italy
Md. Atiqur Rahman Ahad,
University of Dhaka, Bangladesh

***Correspondence:**
Siyuan Chen
siyuan.chen@unsw.edu.au
Julien Epps
j.epps@unsw.edu.au

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in ICT

Received: *27 October 2018*
Accepted: *26 September 2019*
Published: *14 October 2019*

Citation:
Chen S and Epps J (2019) Eyelid and Pupil Landmark Detection and Blink Estimation Based on Deformable Shape Models for Near-Field Infrared Video. Front. ICT 6:18. doi: 10.3389/fict.2019.00018

INTRODUCTION

Eye activity has been of great interest since observations of human attention and intention began (Duchowski, 2007). The types and applications of eye activity that have been investigated up to now include gaze direction as a pointing device for paralyzed people (Duchowski, 2007); pupil size, blink, and eye movement (fixation and saccade) for cognitive load measurement (Chen et al., 2011), emotion recognition (Lu et al., 2015), visual behavior change (Chen et al., 2013), human activity recognition (Bulling et al., 2011), and mental illness diagnosis (Vidal et al., 2012); eyelid

closure for emotion recognition (Orozco et al., 2009), and fatigue detection (Yang et al., 2012; Daniluk et al., 2014). As opposed to these dynamic changes in eye components (eyelid opening, pupil size and location, blink length and depth) which form eye activities, static eye images are also of interest especially in biometrics. For example, eye shape is part of the face for face verification (Vezzetti et al., 2016, 2017), and iris texture is extracted from eye images for identity verification (Bowyer and Burge, 2016). Nevertheless, the basis for capturing these eye activities or biometrical information from an eye image are the eyelid contour and pupil contour. With specific algorithms, eyelid closure, pupil size, blink event, eye movement, and gaze direction can also be measured or developed. Therefore, robust and accurate estimation of eyelid and pupil contours is essential for these applications.

Historically, eyelid contour estimation is often conducted on eye images under normal light conditions and in the far field. In this scenario, only the iris boundary is visible instead of the pupil; hence the pupil contour estimation is unavailable, and the eyelid contour is only able to estimate blink but no other types of eye activity. To obtain a reliable pupil contour, infrared (IR) illumination is required. With IR cameras, the bright or dark pupil effect can help the pupil to be distinguished from the background (Duchowski, 2007; Hansen and Ji, 2010) while for precise pupil size change, wearable IR cameras are desired which give close-up IR eye images, however, this makes all other structures, such as eyelids and iris, in IR eye images inferior for eye activity detection. Recent applications to mental state and behavioral studies demand as many types of eye activity as possible (Chen et al., 2011). Therefore, a wearable eye-directed IR camera is ideal to capture all types of eye activity in real life, especially for pupil size measurement.

Although methods for pupil size estimation in near-field IR eye images have often been studied, few investigations (Fuhl et al., 2017) have been undertaken into eyelid contour estimation. Apart from being able to estimate eye closure and blink, robust eyelid contour estimation can certainly improve pupil contour estimation as it reduces background noise by limiting the search of the pupil inside eyelid contour.

However, the challenges are: (i) unlike eye images under normal light conditions, IR illuminance eliminates color information, and renders some robust features such as eye corners and iris edges weak and indistinctive, while pupil and eyelash features are stronger but can change dramatically with eye movement as opposed to a constant iris size, making distinctive eye features more complex in general; (ii) near-field high resolution eye images introduce more eyelash details than far-field, which are unwanted noise for eyelid contour estimation. In addition, fine-grained eyelid trajectories can result in blink detection not being as simple as detecting only two states of closed and open eye. Furthermore, motion blur due to eyelid movement, which seldom occurs in far-field eye images, can impair some distinctive eye features, further compounding the problem. Some examples can be found in **Figure 1**.

Facial landmark detection from the images in the wild has now attained strong performance (Xiong and De la Torre, 2013; Feng et al., 2015), and by analogy these computing techniques can be applied to eye images. However, the eye has its own distinct characteristics which are different from the face. The greatest distinction is that eye appearance can be completely deformed when the eye is closed, because the major components—the pupil and iris—disappear, as well as the eye corners, fundamentally changing the geometrical structure of the eye image. The problem due to this fundamental structure change is arguably more severe than occlusion of one facial component during facial landmark detection because the modeled components sometimes do not exist when applying model-based approaches.

In this paper, we aim to estimate eyelid and pupil landmarks, and blink events from near-field IR eye images simultaneously which is important to describe eye activity in a full spectrum. Specifically, we want to answer two questions: (i) whether eyelid and pupil landmarks can be reliably and robustly detected using deformable shape models for near-field IR eye images, where variability due to blink, detailed eyelashes and low contrast around eyelid during tasks is magnified in a close-up view, which has not been investigated; and (ii) how to reliably separate blinks from eye activities using a deformable shape model for eye landmarks that do not employ any knowledge of other eye components, as opposed to existing algorithms for blink detection or pupil size estimation which often use thresholds to indicate eyelids being close enough or pupil size being small enough for blink and to discard pupil size when it is small enough during pupil size estimation. These existing algorithms requires accurate thresholds, but they are hardly generalized for everyone and every context. The proposed model-based approach can overcome this limitation. Meanwhile it is novel to be integrated into deformable shape model as facial landmark detection does not have the problem of significant geometrical structure change.

RELATED WORK

The majority of studies on eyelid and blink detection are for frontal face images taken under normal light conditions and in the far field. They usually begin with face recognition and then estimate eye regions, cropped according to topography rules (Moriyama et al., 2006; Bacivarov et al., 2008; Yang et al., 2012; Mohanakrishnan et al., 2013; Daniluk et al., 2014; Yahyavi et al., 2016) or utilizing facial landmarks (Fridman et al., 2018). The eye images thus are in low resolution with clear views of only the iris and the eyelid contour. Few studies (Alabort-i-Medina et al., 2014) investigated high resolution or close-up eye images where the eyelashes were clearly visible.

Eyelid Detection

The general principal of eyelid detection has been to utilize the difference of intensity or edges between the skin, iris, and sclera. Methods to date include finding the points which have maximum response to an upper eyelid and lower eyelid filter (Tan and Zhang, 2006; Daniluk et al., 2014); heuristically examining the polynomials fitted to eye corner points with each, pairs and triples of edge segment above the iris center for topography criteria (Sirohey et al., 2002); and searching among parabolas which pass through the eye corner points to maximize an objective function involving edges, intensity and area (Kuo and Hannah,

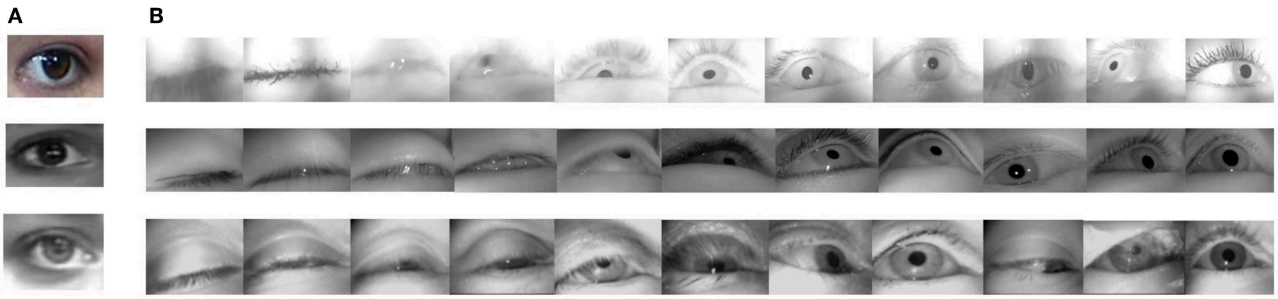

FIGURE 1 | Why near-field IR eye activity estimation is difficult: unlike far-field non-IR eye images in high resolution [first row of **(A)**] (Alabort-i-Medina et al., 2014) and low resolution [second and third rows of **(A)**] (Mohanakrishnan et al., 2013), where the eyelash is barely visible and the intensity between skin, the sclera and iris is often distinctive, IR near-field eye images, **(B)** enhance the contrast between the pupil and its surroundings but weaken the contrast between skin, the sclera and iris. The close-up view highlights greater eye appearance difference due to ethnicity, blinking, motion blur, eyelash density, illumination changes, and camera positions than the far-field view. The first two rows of **(B)** are from the Task Analysis 1 dataset (Epps and Chen, 2018) and Task Analysis 2 dataset (Chen et al., 2013), respectively, which are natural eye movements recorded during tasks. The third row is from the publicly available (Fuhl et al., 2017) dataset.

2005). Similarly, Fuhl et al. (2017) detected the upper and lower eyelid from close-up IR eye images based on the intensity change around the edges of eyelid. The lower eyelid was firstly detected, where the initial point for searching along a line was among the locations corresponding to the peaks of the intensity histogram. The locations near the line which satisfied the criteria of maximum area difference, gradient threshold, and concave polynomials were preserved. The upper eyelid search area was limited according to topography rules. Within this area, the polynomials fitted to three points were evaluated on the intensity change values and the maximum change was selected.

These techniques require reliable and distinctive eye corner or/and iris features and may be unreliable for IR eye images where the intensity of skin, iris, and sclera are indistinct. Moreover, parameters or design choices in these methods were empirical and did not consider different degrees of occlusion to eye components (iris, pupil, and lower eyelid), limiting its generalizability to eyelid detection across a diverse range of eye appearance in unconstrained scenarios.

For more sophisticated model-based approaches, pre-defined models have been initialized to novel eye images before being aligned to the optimal locations. Moriyama et al. (2006) hand crafted a 2D generative eye shape model based on the anatomical eye structure, which included the upper and lower eyelid, sclera, eye corners, and iris. The upper and lower eyelid were further composed of three or four sub-parts. Eleven parameters were deployed to control the eye structure, the motions of eyelid and 2D iris movement. Fitting the parameterized eye model to a novel image was posed as a problem of tracking the motion parameters with Lucas-Kanade gradient descent. Variations due to individual eye appearances and illuminance reflections were mitigated by the manual initialization of the size and texture of each eye part. Orozco et al. (2009) also used a parametrized shape model but derived from a hand-crafted standard design from the computer animation industry for the appearance-based tracker. This also required careful manual initialization, as the texture was learned online from near-frontal images. Yang et al. (2012) employed a four-landmark deformable template and fitted it to a likelihood map of the eye region by maximizing the total likelihood. The likelihood map was constructed from the Mahalanobis distance to the skin color distribution, which was obtained by clustering the color and texture descriptors.

All these methods were demonstrated for low-resolution eye images and employed a single shape model for both eye open and eye closed images, so that even when the eye is closed, the iris was still a valid and visible part of the shape model (Orozco et al., 2009), which makes little sense.

Tan and Zhang (2006) firstly determined the eye state of open or close by examining the existence of the iris using intensity and edge information. Then the eye feature was modeled as either a straight line or a deformable template before manual initialization and application of a standard Lucas Kanade framework for tracking.

Recently, Alabort-i-Medina et al. (2014) compared three standard deformable model fitting techniques, Active Appearance Models (AAM), Constrained Local Models (CLM), and Supervised Descent Method (SDM) to track the deformation and motion of eyelid, the iris and pupil. In order to represent the open and fully closed eye correctly, two sets of shape models were used for the two eye states. In contrast to manual initialization of the shape or texture as in previous work, the initial shape model was statistically learned from training data, making fewer assumptions about the shape and texture of an individual's eye. Learning and evaluation were conducted for open and closed eye images individually. The results demonstrated that for open eye images, AAM performed best when the Cumulative Error Distribution (CED) normalized to eye size was <0.05, otherwise SDM was the best choice. For closed eye images, SDM performed best.

This work demonstrated the potential for applying facial landmark detection approaches, especially the SDM technique, to far-field high resolution eye images for open eye motion tracking, however, their performance for fully closed eye images was poor. The feasibility for near-field IR eye images is unknown, neither of how to determine the open and closed eye states and how the

trained shape model affects individual eye alignment whose state is unknown.

Blink Detection

There are a variety of blink detection techniques in the literature, which can be categorized into four main groups. The first depends on iris or pupil existence (Tan and Zhang, 2006; Chen and Epps, 2014). For example, for near-field IR eye images, apart from declaring a blink when there was no pupil blob, Chen and Epps (2014) employed two ellipse fittings to the detected pupil blob. One was fitted to the whole pupil contour, and the other was fitted to the bottom half contour, so that blink was determined by the degree of pupil occlusion by the eyelid. However, these methods require prior knowledge of pupil or iris blob and are not suitable for cases when the bottom of the blob is occluded.

The second is based on the eyelid distance with a pre-defined threshold, e.g., the ratio between the eye width and its aperture (Bacivarov et al., 2008). However, the decision for the threshold may be especially difficult for the fine-grained eyelid trajectory.

The third involves eye motion. Mohanakrishnan et al. (2013) proposed motion vector difference in the face region and eye region to detect blink, since the motion vector in the eye region can be "random" during a blink, while it is typically similar to the face region when the eyelid is still relative to the face. Instead of determining a similarity threshold, Appel et al. (2016) extracted features from the intensity difference of two adjacent frames specifically for near-field IR images. However, these methods are probably not able to detect long blinks since during these, the eyelid also does not move. In the latest work, Fogelton and Benesova (2018) used motion vectors in the eye region and learned a sequence model of blink using Recurrent Neural Network to detect blink completeness and achieved similar or slightly better performance than other methods in four datasets.

The last detects blink directly from the eye appearance in images. Yahyavi et al. (2016) employed PCA and artificial neural networks for open and closed eye images. Mohammadi et al. (2015) detected the open and closed eye states by finding an appropriate threshold for the intensity change. Bacivarov et al. (2008) used the AAM parameters' difference in open and closed eye images to identify blink. Sun et al. (2013b) used SVM to classify each frame as the onset, apex and offset of blink and construct a temporal model of HMM-SVM to determine blink event. They found that this temporal model and the intensity feature were significantly better than using multi-class SVM and HOG, LBP, Gabor and optic flow features to detect blink.

The reported blink detection performance among previous studies is usually high, above 90% in accuracy. On one hand, most of these results were for low resolution non-IR eye images, and it is questionable whether close-up IR images containing a variety of detailed eye appearance changes can be classified simply into open and closed eye states. On the other hand, most datasets for evaluation were collected during leisure scenarios where voluntary blink often occurred, evidenced by completely closed eyes. However, during tasks, most blinks are partial and most full blinks have the lids approaching each other but not necessarily touching (Brosch et al., 2017). These involuntary blinks may have different duration, amplitude and speed ratio of eye closing and opening to voluntary blinks (Abe et al., 2014).

Pupil Contour Estimation

Since the pupil is only distinct under IR illumination and its size can be accurately measured only when the resolution of pupil is good enough, pupil detection is always conducted on IR eye images. Estimated pupil contour can be fitted by an ellipse model to obtain the pupil center, pupil size, and other features of interest. One straightforward approach is to segment the pupil from the background through binarization, however, it is challenging to find an adaptive threshold for a variety of eye images with large variations. Chen and Epps (2014) proposed a self-tuning threshold method which requires minimum parameter to tune to handle these variations for near-filed IR images. Other algorithms operate on remote IR images, with the aim of pupil center detection. These algorithms (Fuhl et al., 2016) involve a combination of multi-thresholding, edge filtering, morphological operation, intensity gradient, and iterative search of appropriate points satisfying a series of criteria. Fuhl et al. (2016) compared six such start-of-the-art algorithms for pupil center detection with head-mounted IR eye images. They found that their dual algorithm outperformed the others. In this algorithm, a first approach using edge image, morphological operations, and heuristic selection of the best edges is trialed, and if this fails, advanced blob detection is used.

Nevertheless, determining the best parameter settings is not always easy for an unknown dataset, and the sensitivity to different parameter settings, which can be examined through cross-corpus evaluation, is unknown. Moreover, sophisticated model-based approaches have not been seen in pupil detection.

Proposed Three-State Deformable Eye Model

As the pupil, iris and eyelid always interact with each other in normal eye behaviors, some eye activities cannot be detected alone. For example, if blink is detected based on pupil size, robust pupil detection is required. However, pupil size estimation is often unreliable due to blink or partial eyelid occlusion (e.g., eyelash down) and blink is not the only factor changing pupil size, therefore, reliable pupil size estimation requires prior knowledge of blink occurrence. In this study, we propose a unified approach by employing deformable shape models to detect eyelid contour, pupil contour, and blink simultaneously.

However, different to the problem of partial occlusion in facial landmark detection, where the majority of face components are still in position and recognizable, the majority of eye components can completely disappear when the eye is closing or closed or suffering from motion blur, as shown in **Figure 1B**. For shape models, each component's shape is required to be pre-defined; but these pre-defined shapes cannot be guaranteed to appear in each eye image, which may affect the parameter optimization process. This is resolved by distinguishing different eye states. The proposed framework is shown in **Figure 2**, where under each eye state (e) and given initial landmark locations ($\boldsymbol{s}$), the increment of each landmark ($\boldsymbol{\Phi}$) is trained given the annotated landmarks. Meanwhile, a model of local

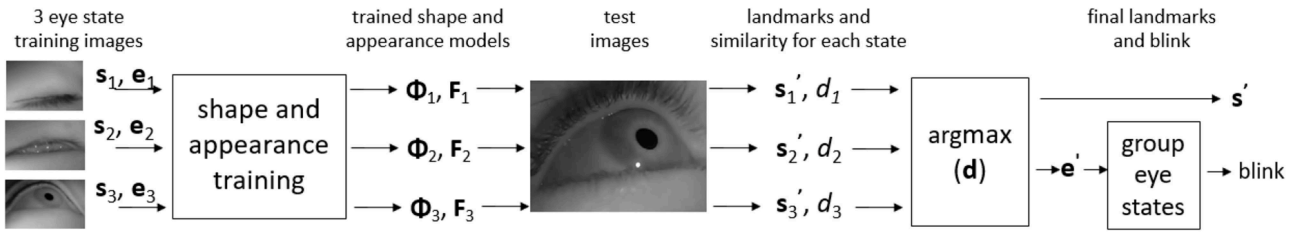

FIGURE 2 | An overview of the proposed unified approach for eyelid contour, pupil contour, and blink estimation, where $\boldsymbol{s}$ are the landmark locations, e are the eye states, $\boldsymbol{\Phi}$ is the trained shape model, $\mathbf{F}$ is the trained local appearance model, d is the similarity between the local appearance at the final landmarks and the trained appearance model.

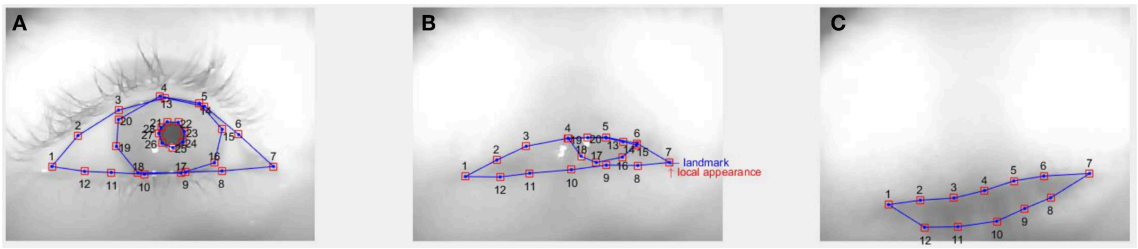

FIGURE 3 | (A) L = 28 landmarks (denoted by blue dots) for the state of open eye with the pupil and the iris ($e = 3$), **(B)** L = 20 landmarks for open eye with the iris ($e = 2$), and **(C)** L = 12 landmarks for closed eye ($e = 1$). Centering each landmark is a small patch (denoted by a red square) for local appearance. These landmarks comprise the eye shape model for each eye state. All three eye images shown are from the same subject, yet dramatic differences between them can be observed.

appearance represented by the feature of the patch at each annotated landmark ($\mathbf{F}$) was also trained. During the testing phase, the initial landmarks on a test eye image move incrementally according to the trained shape models under each eye state, and the local appearance at final landmarks ($\boldsymbol{s}'$) is collected. Only the eye state where the similarity (d) between the local appearance and trained appearance model is maximum is selected and the landmarks in the selected eye state is used.

Three-State Deformable Eye Model

We propose three distinct eye states for eyelid landmark detection before recognizing blink. They are closed eye (e_1), open eye with iris only (e_2), and open eye with iris and pupil (e_3). Each eye state applies a pre-defined different number of landmarks ($\mathbf{s}_i^{'} = [x_1, y_1, \ldots, x_{Li}, y_{Li}]^T$, where L_i is the number of landmarks, $i = 1, 2, 3$), as shown in **Figure 3**. These landmarks do not contain the furrow and bulge texture of the eye as in Moriyama et al. (2006) and Alabort-i-Medina et al. (2014) since they are not distinguishable in IR eye images. The states of closed eye and open eye with iris are associated with blink since sight cannot occur without the pupil. The reason for including an open eye with iris state is that it is common during tasks to observe that the eyelid moves fast during a blink but is not always fully closed. In this state, the pupil is fully occluded, but the lower half of the iris is still visible. This eye state is not rare in tasks (Brosch et al., 2017) as we often observe this partially-open-eye blink in our datasets shown in **Figure 1B**.

Learning Shape and Appearance

For training data with eye landmarks for each eye state (e_i, $i = 1$, 2, 3), we firstly warp the image $\mathbf{I}'$, $\mathbf{I} = W(\mathbf{I}')$, and its landmarks, $\mathbf{s} = W(\mathbf{s}')$, to the average size of all training images, which includes scale and translation. $W(\bullet)$ is the warp operator. We then not only train a shape model $\boldsymbol{\Phi}_i$ but also an appearance model for each eye state, $\boldsymbol{F}_i$, $i = 1, 2, 3$.

The recently proposed SDM (Xiong and De la Torre, 2013) was chosen to train these shape modes. SDM is a supervised discriminative model for solving general non-linear optimization problems, which has attracted great attention in facial landmark detection. This technique defines a local appearance model around pre-defined landmarks in an image, and builds a non-linear mapping function $\boldsymbol{\Phi}$ between the local appearance features $\{\boldsymbol{f}(\mathbf{I}, \mathbf{s})\}$ and shape update through the process of training a cascaded regressor.

The discriminative model, $\Phi : \boldsymbol{f}(\mathbf{I}, \mathbf{s_0}) \rightarrow \delta\mathbf{s}$, is trained by minimizing the cost function with ground truth annotation (Alabort-i-Medina et al., 2014; Feng et al., 2015),

$$\frac{1}{2N}\sum_{n=1}^{N}\left\|\mathbf{s}_0(n) + \delta\mathbf{s}(n) - \mathbf{s}^*(n)\right\|_2^2 \tag{1}$$

where $\mathbf{s}^*(n)$ is the ground truth shape of the nth training image, $\delta\mathbf{s}(n) = \boldsymbol{\Phi}(\boldsymbol{f}(\mathbf{I}, \mathbf{s}_0))$ is the corresponding shape update, and $\mathbf{s}_0(n)$ is the initial shape estimate.

The non-linear mapping function is obtained by cascading a sequence of M linear regressors $\mathbf{R}$:

$$\Phi = \mathbf{R_1}\mathbf{R_2}\ldots\mathbf{R}_M \tag{2}$$

where $\mathbf{R}_m = \{\mathbf{A}_m, \mathbf{b}_m\}$. The regression matrix $\mathbf{A}_m$ and bias term $\mathbf{b}_m$ of each regressor $\mathbf{R}_m$ can be obtained in closed form by solving a linear least squares problem with training data.

Meanwhile, the local appearance features around each annotated landmark are extracted. The trained appearance model for the ith eye state is the average of the local appearance features:

$$\boldsymbol{F}_i = \frac{1}{N_i}\sum_{n=1}^{N_i}\boldsymbol{f}(\mathbf{I}_n, \mathbf{s}_n{}^*) \tag{3}$$

The shape models $\boldsymbol{\Phi}_i$ are used to predict the landmarks while the appearance models $\boldsymbol{F}_i$ are used to distinguish the eye states.

Inference of Shape and Blink

Landmarks are predicted by updating the shape according to the mapping function using their local appearance features. Given a novel eye image $\mathbf{I}'$, the eye landmarks are recursively updated for each eye state from its warped image $\mathbf{I}$, using the same warping parameters for training data:

$$\delta\mathbf{s} = \mathbf{A}_m\boldsymbol{f}(\mathbf{I}, \mathbf{s}_{m-1}) + \mathbf{b}_m \tag{4}$$

$$\mathbf{s}_m = \mathbf{s}_{m-1} + \delta\mathbf{s} \tag{5}$$

Usually, after $M = 5$ cascade levels, Equation (1) converges (Xiong and De la Torre, 2013; Feng et al., 2015). Therefore, we can obtain estimated eye landmarks $\mathbf{s}_i$ and the local appearance features $\{\boldsymbol{f}_i(\mathbf{I}, \mathbf{s}_{i,\boldsymbol{m}})\}$ in the final update. The eye state e is determined by finding the maximum cosine similarity (d_i) to the trained local appearance model:i

$$e = arg\max_i d_i = arg\max_i \frac{\boldsymbol{f}_i \cdot F_i'}{\|\boldsymbol{f}_i\|\,\|F_i'\|} \tag{6}$$

The final set of eye landmarks is the landmarks of the detected eye state. The shape is then warped back to the original images using the reciprocal scale and translation:

$$\mathbf{s}' = W^{-1}(\mathbf{s}_i\,(i = e)) \tag{7}$$

Meanwhile, blink is determined by

$$b\left(\mathbf{I}'\right) = \begin{cases} blink & otherwise \\ not\ blink & if\ e = 3 \end{cases} \tag{8}$$

Hence, the eyelid contour can be found from the detected eye landmarks when $e = 3$ or 2, while the upper eyelid contour can be found when $e = 1$. The pupil contour can only be found from the deformable model in non-blink images when $e = 3$.

EXPERIMENTS

We employed three datasets to evaluate and cross-corpus evaluate the performance of eye landmarks and blink detection. Furthermore, eyelid estimation performance was compared with that of the only algorithm for eyelid detection in near-field IR eye images, to the best of our knowledge.

Datasets

a) Task Analysis 1 (TA1) dataset (Epps and Chen, 2018): The IR near-field eye videos were recorded by a head-mounted off-the-shelf webcam (30 fps) from 24 subjects of different ethnicity while they were completing mental and physical tasks (ethics approval was obtained). For each subject, there was over 1 h of video recording. We selected 200 images within the first few minutes from each subject for annotation, so there are 4,800 images in total. These selected images contain noticeable eye movement change from the preceding video frame. This dataset suffers from strong lighting conditions, causing weak contrasts, and an invisible iris in the images, some of which are shown in the first row of **Figure 1B**. For model training and test, we manually placed 12, 20, and 28 landmarks for each eye state, fully closed eye (6%), open eye with iris (3%), and open eye with iris and pupil (91%), respectively, as shown in **Figure 3**. Half of the eye images from all subjects were used for training and the other half were used for test. During cross-corpus-validation, all eye images were used for test.

b) Task Analysis 2 (TA2) dataset (Chen et al., 2013): Similar to the TA1 dataset, IR near-field eye movements were recorded also during tasks but using a different webcam, a different 22 subjects, and in an indoor light condition (ethics approval was obtained), some images of which are shown in the second row of **Figure 1B**. We selected 4,400 eye images (200 from each subject) and annotated them with the same scheme as in TA1 dataset for training and test. They contain 8.6% closed eye, 4.1% open eye with iris, and 87.3% open eye with iris and pupil. The ratio of training and test images was the same as that in TA1 dataset.

c) Fuhl et al. dataset (2017): This public dataset contains 5101 IR near-field eye images from 11 subjects. They were recorded in realistic scenarios and may contain deliberate eye movements. Some examples are shown in the third row of **Figure 1B**. This dataset comes with an annotation of 10 eyelid landmarks (Fuhl et al., 2017) regardless of eye state. Therefore, we only can obtain the eyelid alignments. Half of the data were used for training and the other half were used for test with our method. The performance of predicted eyelid landmarks was compared with the recently proposed Fuhl et al. algorithm (2017), which only evaluated eyelid detection on this dataset without pupil landmark detection and without model-based approach.

Cross-corpus validation was also conducted in two ways. One was to use the Fuhl et al. algorithm (2017) to test all eye images from TA1 and TA2 without altering any parameter. The other was to use the models trained on TA1 or TA2 to test all eye images from

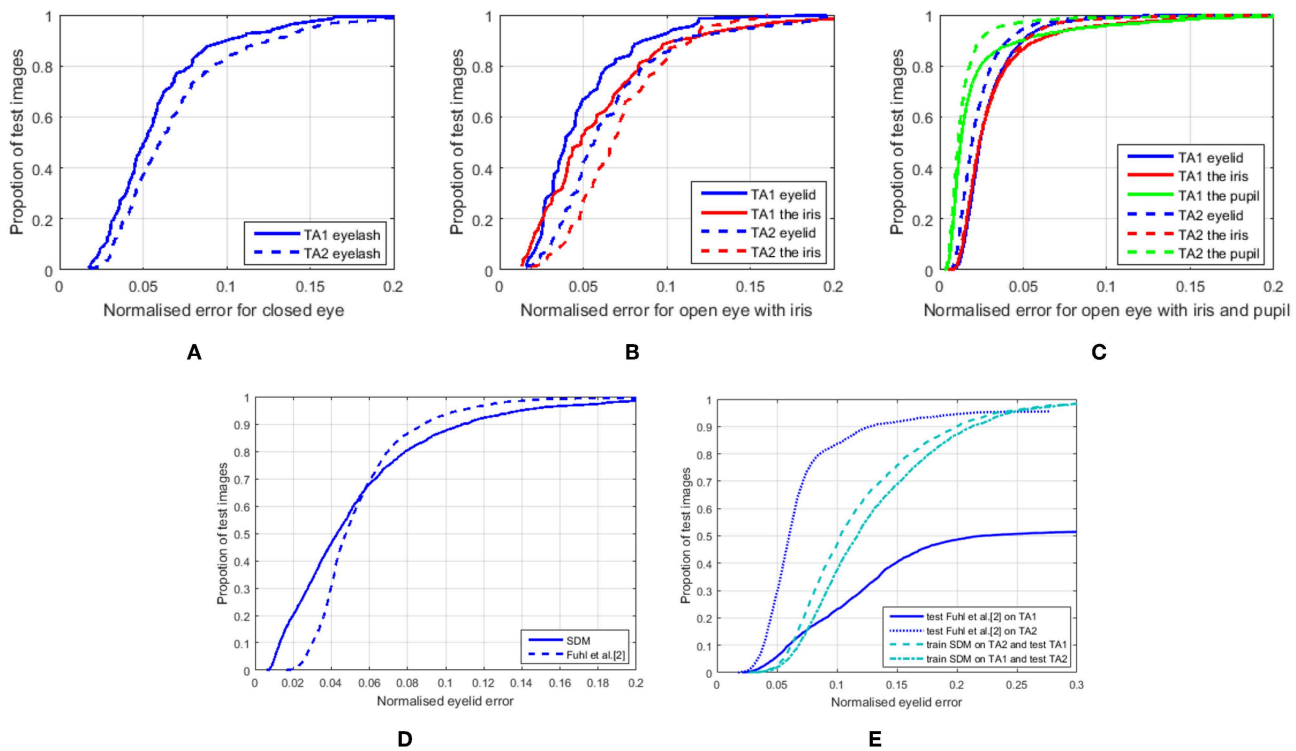

FIGURE 4 | (A–C) CED curves for three eye states on TA1 and TA2 datasets using SDM deformable eye models. **(D)** CED curves for eyelid landmarks only using Fuhl et al. algorithm (2017) and SDM deformable eye models on Fuhl et al. dataset (2017) and **(E)** the cross-corpus evaluation on TA1 and TA2 datasets (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article). It is worth noting that Fuhl et al. algorithm failed eyelid detection for 2,295 out of 4,800 test images in the TA1 dataset and 196 out of 4,400 test images in the TA2 dataset. These errors were not counted in **(E)**.

this dataset. It should be noted that only 10 eyelid landmarks in the states of open eye with iris and open eye with iris and pupil can be tested because our trained models in the closed eye state were eyelash contours instead of eyelid.

Experimental Settings

In all experiments, we set the radius of the local patch around a landmark to be 16 pixels, which was fixed for all dataset evaluations. A Histogram of Oriented Gradients (HOG) feature was used to represent the local appearance feature, $\boldsymbol{f}(\mathbf{I}, \mathbf{s})$, since it was reported effective (Feng et al., 2015). Six regressors were used in cascade to train the shape model, as it was found that performance did not change significantly when $M > 5$, similarly to Alabort-i-Medina et al. (2014) and Feng et al. (2015). The initial shape for each test image was scaled by the ratio of mean shape model from the training data and the test image size.

RESULTS AND DISCUSSION

Results of Landmark Detection Given True Eye States

Figures 4A–C show the CED curves for three eye states for the TA1 and TA2 datasets, given the true eye states. The errors were normalized by the distance between the two eye corners in each image, and averaged across all landmarks and images. The results show that the predicted iris landmarks (red curves) are generally not as accurate as eyelid landmarks (blue curves) for IR near-field images, probably because of weak contrast between the iris and sclera. Meanwhile, the open eye with iris and pupil state (**Figure 4C**) achieves the best performance among the three eye states. This state is also the most studied eye state for eyelid detection in non-IR far-field images. With the proposed three-state deformable eye model, eye landmark estimation can achieve similar normalized errors to state-of-the-art facial landmark alignment (Alabort-i-Medina et al., 2014; Feng et al., 2015). However, the performance of closed eye (**Figure 4A**) and open eye with iris (**Figure 4B**) were generally worse than the wide-open eye state with all eye components (**Figure 4C**). It is very likely that more variations, such as motion blur and eyelash changes mostly occurred during eyelid movement. Also, there is less training data for the former two states. Interestingly, the TA2 dataset has better image quality, which improves eye landmark estimation only for the state of open eye with iris and pupil while deteriorating the performance of the other two states. The tentative reason could be that clear views of the closing eye generate unwanted details of eyelash compared with blurred eye images.

Figures 4D,E are the CED curves for only 10 eyelid landmarks estimation using the Fuhl et al. algorithm and/or on their public dataset. As their algorithm outputs polynomial curves of eyelid

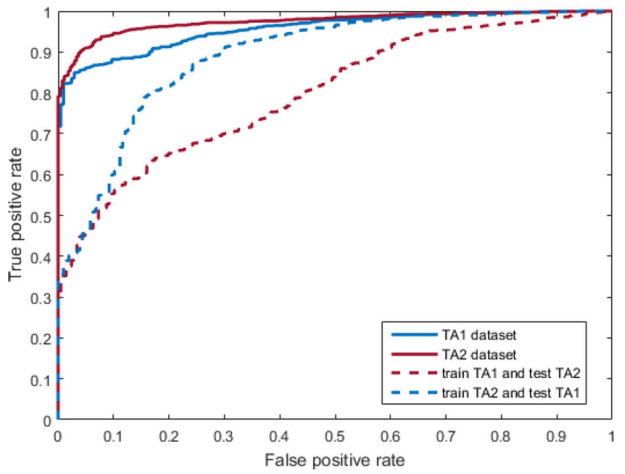

FIGURE 5 | ROC curves for blink detection on TA1 and TA2 datasets (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

locations, to compare with the annotated ground truth, 10 eyelid landmarks on the predicted curves were selected by firstly locating the two eye corners corresponding to the annotated ones, and then searching the nearest point to each of the remaining annotated landmarks.

From **Figure 4D**, we can see that the SDM-based deformable eye model achieved comparable performance to the Fuhl et al. algorithm on their public dataset. However, the Fuhl et al. algorithm failed on 6 out of 2,434 test images (output null), and these errors were not taken account into the normalized errors because the distance between the ground truth and the predicted null landmarks is impossible to calculate.

Figure 4E shows the cross-corpus validation performance of eyelid-only landmark estimation. Although all performances dropped compared with evaluations on the same datasets, the three-state deformable eye model was relatively more reliable than the Fuhl et al. algorithm, which output null for 2,295 out of 4,800 test images in the TA1 dataset and 196 out of 4,400 test images in the TA2 dataset. The normalized eyelid errors in **Figure 4E** do not include these null outputs. However, our proposed method never has null outputs although the landmarks could be very far away from the ground truth.

All these results suggest that for IR eye images, the deformable eye model can perform well for pupil landmark detection. This is not surprising because the infrared light enhances the distinction between the pupil and its surroundings. Meanwhile, when the eye is completely open and the eyelash is up, the variations are mainly due to pupil size change, pupil/iris position change, and eyelid shape change between individuals (**Figures 1B**, **6**). For the closed eye state and open eye with iris state, large variations are from pupil/iris shape (only in open eye with iris state), eyelid motion, eyelid shape and eyelash shape. Even within an individual, the variation can be substantial due to eyelid movement (**Figures 1B**, **6**). The eyelid landmark detection in these states becomes difficult, and the degradation

TABLE 1 | Confusion matrix for three eye states classification.

Actual \ Predict	Closed eye	Open eye with iris	Open eye with iris and pupil
Closed eye	0.77 (TA1) 0.76 (TA2)	0.21 (TA1) 0.18 (TA2)	0.02 (TA1) 0.06 (TA2)
Open eye with iris	0.17 (TA1) 0.29 (TA2)	0.75 (TA1) 0.69 (TA2)	0.07 (TA1) 0.01 (TA1)
Open eye with iris and pupil	0.06 (TA1) 0.04 (TA2)	0.05 (TA1) 0.05 (TA1)	0.89 (TA1) 0.91 (TA2)

of SDM-based deformable eye model performance and the cross-dataset performance confirms this challenge. Overall, SDM-based deformable eye model performance is better than the Fuhl et al. (2017) baseline algorithm because the eyelid landmark detection performance even in the worst cases (**Figure 4A** closed eye and **Figure 4B** open eye with iris) is comparable to that using Fuhl et al. algorithm for the best case (fully open eye) and one worst case (open eye with iris). Another advantage of the proposed three-state SDM-based deformable eye model is that we distinguished three eye states, and each fits the most appropriate eye model given all three models. Possible errors due to inappropriate models are reduced and the three-state model will not encounter eyelid detection failure unlike Fuhl et al. algorithm where the pre-condition of finding lower eyelid must meet for next processing.

Results of Blink Detection

Table 1 shows the confusion matrix of the three eye states detection on the TA1 and TA2 datasets. These results demonstrate that using the proposed approach to distinguish three eye states can achieve strong performance, above 69% in recall. Among the results, the open eye with iris and pupil state obtained higher recall performance than the other two states. This is most likely due to the higher accuracy of eye landmark estimation for this state.

Figure 5 shows the Receiver Operating Characteristic (ROC) curve for blink detection, which was obtained by grouping the states highlighted in blue in **Table 1** into a single blink class. Specifically, the recall of blink detection was 89% on TA1 dataset and 91% on TA2 dataset (solid curves). Note that blink detection using the proposed three-state deformable model approach based on pupil visibility is novel, rather than directly examining pupil or iris existence like existing methods (Bacivarov et al., 2008; Chen and Epps, 2014). The results are comparable to the general performance of blink methods for non-IR far-field eye images, which is often over 90% in recall. The dotted ROC curves are the blink detection performance of cross-evaluation, where the training and test data were from different datasets. The blink detection recall performance degraded significantly, while the precision dropped slightly. This means that a number of open eyes with iris and pupil states were misclassified as the other two eye states. Inaccurately predicted eye landmark is probably the main reason, since the local appearance extracted from the wrong landmarks could not be similar to the trained appearance model.

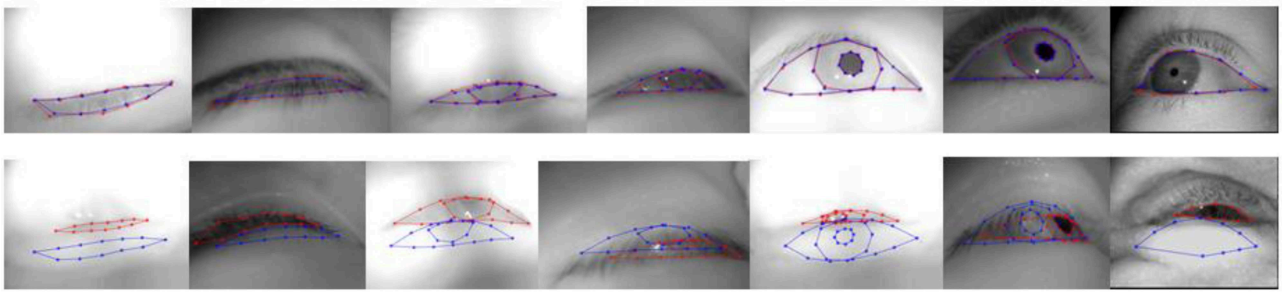

FIGURE 6 | Successful (1st row) and failed (2nd row) eye landmark detections selected from the TA1, TA2, and Fuhl et al. dataset. The red dots are the annotated landmarks and the blue ones are the predicted landmarks (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Meanwhile, the local appearance from different datasets might also vary due to the environments.

All these blink detection results suggest that using the proposed three-state SDM-based deformable eye model to detect blink is feasible for IR eye images along with eye landmark detection. It utilizes three distinct eye states and the best fitted eye model to find eye blink, which is different from current techniques mentioned in section Blink Detection. However, the proposed blink detection method relies on accurate eye landmark detection. Correctness in fitted eye models leads to high blink detection performance as indicated by the higher eye landmark detection performance in the open eye with pupil and iris state and in the cross-dataset blink detection performance (**Figure 5**). Although there is no parameter required in this method, the dependency of eye landmark detection plays a similar role as those parameters requiring setting in blink detection in existing techniques, but does not require manual tuning in our method.

Results of Landmark Estimation Given Detected Eye States

As mentioned earlier, eye components often interact with each other during eyelid movement, so it is difficult to determine which component should be detected first. Investigations of the proposed approach based on true vs. detected eye states also show this challenge. **Table 1** shows that the better the eyelid landmark estimation, the higher the recall performance of eye state, while **Table 2** shows how eyelid landmark estimation performance drops due to eye state estimation errors. This is because, as **Figure 2** shows, given a novel image, the final predicted eye landmarks are the ones from the detected eye state. From **Table 2**, we can see that on average, the landmark error increased by around 16.7% on the TA1 dataset and around 18.6% on the TA2 dataset compared with the performance given the true eye states. When given a trained deformable eye model from another dataset, the final eyelid landmark estimation performance was significantly degraded. The landmark error increased by around 160 and 180% compared with the performance given the true eye states and given the models trained from the same dataset.

These results give us the overview of what factors affect the performance significantly. In our study, the cross-dataset testing performance was unacceptable, which indicates that large variations due to different experimental settings, devices, and participants, were not grasped by the SDM-based deformable eye model. Therefore, more techniques need to be studied to further improve eye landmark estimation in order to obtain more accurate eye activity for wearable eye computing.

Some Typical Eye Landmark Detection Results

Figure 6 presents some good and bad examples from the three datasets used. In general, if the shape model performs well, the alignment can have very small errors, better than the Fuhl et al. algorithm as shown in **Figure 4D**, while the errors can be large if the shape model does not have the directions to deform which should be learned from SDM. These failures are mostly rare cases during tasks, therefore, there are few training examples for them.

Future work will focus on methods of improving the performance for these rare eye movements and cross-dataset performance. One direction is to collect and annotate more data for the closed eye state and open eye with pupil state from different people. More initial eye landmark locations need to be added to train the models and an algorithm for selecting useful initial locations for model training can be developed. These are to let the models learn better by exposing more variations. Another direction is using convolutional neural networks, which have been developed well in recent years and found to outperform other methods in facial landmark localization and to be performed for multi-task learning (Sun et al., 2013a; Jackson et al., 2016; Ranjan et al., 2017). The large amounts of parameters trained for neural networks are expected to grasp most variations presented in three eye states, however, the method may be very expensive due to requiring large amounts of annotation and computing resources.

CONCLUSIONS

Eye activity, including eyelid movement, pupil size, fixation, saccade, and blink, is attracting more and more attention in human mental state analysis with wearable cameras. However, most studies have focused on non-IR far-field eye images to compute individual eye activity such as blink or eye center.

TABLE 2 | Average error normalized by two eye corners' distance.

	Training and test from the same dataset				Train TA2	
	Given true three eye states		Given detected three eye states		Given true states	Given detected states
Test dataset	**TA1**	**TA2**	**TA1**	**TA2**	**TA1**	**TA1**
All three eye states	0.031	0.029	0.039	0.037	0.098	0.115
Closed eye	0.059	0.071	0.067	0.089	0.120	0.118
Open eye with iris	0.054	0.069	0.054	0.079	0.094	0.080
Open eye with iris and pupil	0.029	0.023	0.037	0.030	0.097	0.116

This work has presented a unified approach to obtain the eyelid contour, pupil contour, and blink event, from which most eye activity can be further extracted. In this approach, the eyelid contour and pupil contour are obtained from the update of the deformable shape models which are statistically learned for the most likely eye state. At the same time, blink is detected based on whether the most likely eye state contains the pupil. Results on three different datasets containing large variations of eye appearance in realistic situations demonstrate comparable performance with facial landmark detection and similar blink estimation performance to that in non-IR far-field eye images, therefore, a deformable shape model is suitable for eyelid and pupil landmark detection from near-field IR eye images. The proposed method also led to better and reliable results than the eyelid detection algorithm, Fuhl et al. (2017), based on intensity change for near-field IR images, which does not model the eye shape or extract pupil information and fails to find eyelid in a number of near-field IR eye images. However, due to the specific attributes of different eye activities in near-field IR eye images—not only eyelash and eyelid topological change due to the viewpoint, but also complete geometrical and texture change due to the interaction of eyelash, eyelid and pupil, the eyelash and eyelid landmark estimation in the cases of fully closed eye and slightly opened eye, as well as eye landmark estimation and blink detection in cross-corpus evaluation, are still challenging. More methods need to be studied to further improve eye landmark and blink detection for near-field infrared images. This is the first study focused on near-field IR eye images to obtain eyelid and pupil landmarks and blink based on deformable shape models, and to use different eye states to deal with the significant geometrical structure change in eye images. Future work will involve in using the wearable system to obtain all eye activities to conduct and improve human mental state analysis.

DATA AVAILABILITY STATEMENT

The datasets for this manuscript are not publicly available because there is no consent from participants for public availability. Requests to access the datasets should be directed to j.epps@unsw.edu.au.

ETHICS STATEMENT

At its meeting of 9th September 2014 the UNSW Human Research Ethics Advisory Panel was satisfied that this project is of minimal ethical impact and meets the requirements as set out in the National Statement on Ethical Conduct in Human Research. Having taken into account the advice of the Panel, the Deputy Vice-Chancellor (Research) has approved the project to proceed. This approval is valid for 5 years from the date of the meeting.

AUTHOR CONTRIBUTIONS

SC and JE contributed the ideas. SC did the literature survey, conducted experiments, and engaged in paper writing. JE provided insightful discussions and feedback on writing.

FUNDING

This work was supported in part by US Army ITC-PAC, through contract FA5209-17-P-0154. Opinions expressed are the authors', and may not reflect those of the US Army.

REFERENCES

Abe, K., Sato, H., Ohi, S., and Ohyama, M. (2014). "Feature parameters of eye blinks when the sampling rate is changed," in *Proceedings of the IEEE Region 10 Conference (TENCON)* (Bangkok).

Alabort-i-Medina, J., Qu, B., and Zafeiriou, S. (2014). "Statistically learned deformable eye models," in *Winter Conference on Applications of Computer Vision (ECCV) Workshops* (Zürich), 285–295.

Appel, T., Santini, T., and Kasneci, E. (2016). "Brightness-and motion-based blink detection for head-mounted eye trackers," in *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (Heidelberg), 1726–1735.

Bacivarov, I., Ionita, M., and Corcoran, P. (2008). Statistical models of appearance for eye tracking and eye-blink detection and measurement. *IEEE Trans. Consum. Electr.* 54, 1312–1320. doi: 10.1109/TCE.2008.4637622

Bowyer, K. W., and Burge, M. J. (eds.). (2016). *Handbook of Iris Recognition*. London: Springer.

Brosch, J. K., Wu, Z., Begley, C. G., Driscoll, T. A., and Braun, R. J. (2017). Blink characterization using curve fitting and clustering algorithms. *J. Model. Ophthal.* 1, 60–81.

Bulling, A., Ward, J. A., Gellersen, H., and Tröster, G. (2011). Eye movement analysis for activity recognition using electrooculography. *IEEE Trans. Pattern Anal. Machine Intel.* 33, 741–753. doi: 10.1109/TPAMI.2010.86

Chen, S., and Epps, J. (2014). Efficient and robust pupil size and blink estimation from near-field video sequences for human-machine interaction. *IEEE Trans. Cybernet.*44, 2356–2367. doi: 10.1109/TCYB.2014.2306916

Chen, S., Epps, J., and Chen, F. (2013). "Automatic and continuous user task analysis via eye activity," in *Proceedings of Intelligent User Interface (IUI)* (Santa Monica, CA), 57–66.

Chen, S., Epps, J., Ruiz, N., and Chen, F. (2011). "Eye activity as a measure of human mental effort in HCI," in *Proceedings of Intelligent User Interface (IUI)* (Palo Alto, CA).

Daniluk, M., Rezaei, M., Nicolescu, R., and Klette, R. (2014). "Eye status based on eyelid detection: a driver assistance system," in *International Conference on Computer Vision and Graphics* (Cham: Springer), 171–178.

Duchowski, A. (2007). *Eye Tracking Methodology Theory and Practice, 2 Edn*. London: Springer.

Epps, J., and Chen, S. (2018). Automatic task analysis: towards wearable behaviometrics. *IEEE Syst. Man Cybernet. Magazine* 4, 15–20. doi: 10.1109/MSMC.2018.2822846

Feng, Z. H., Hu, G., Kittler, J., Christmas, W., and Wu. X-J. (2015). Cascaded collaborative regression for robust facial landmark detection trained using a mixture of synthetic and real images with dynamic weighting. *IEEE Trans. Image Process.* 24, 3425–3440. doi: 10.1109/TIP.2015.2446944

Fogelton, A., and Benesova, W. (2018). Eye blink completeness detection. *Comput. Vision Image Understand*. 176–177, 78–85. doi: 10.1016/j.cviu.2018.09.006

Fridman, L., Reimer, B., Mehler, B., and Freeman, W. T. (2018). "Cognitive load estimation in the wild," in *The ACM CHI Conference on Human Factors in Computing Systems* (Montreal, QC). doi: 10.1145/3173574.3174226

Fuhl, W., Santini, T., and Kasneci, E. (2017). "Fast and robust eyelid outline and aperture detection in real-world scenarios," in *Winter Conference on Applications of Computer Vision (WACV)* (Santa Rosa, CA). doi: 10.1109/WACV.2017.126

Fuhl, W., Tonsen, M., Bulling, A., and Kasneci, E. (2016). Pupil detection for head-mounted eye tracking in the wild: an evaluation of the state of the art. *Machine Vision Appl.* 27, 1275–1288. doi: 10.1007/s00138-016-0776-4

Hansen, D. W., and Ji, Q. (2010). In the eye of the beholder: a survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Machine Intel.* 32, 478–500. doi: 10.1109/TPAMI.2009.30

Jackson, A. S., Valstar, M., and Tzimiropoulos, G. (2016). "A CNN cascade for landmark guided semantic part segmentation," in *European Conference on Computer Vision (ECCV)* (Amsterdam), 143–155.

Kuo, P., and Hannah, J. (2005). "An improved eye feature extraction algorithm based on deformable templates," in *IEEE International Conference on Image Processing (ICIP)* (Genova). doi: 10.1109/ICIP.2005.1530278

Lu, Y., Zheng, W., Li, B., and Lu, B. (2015). "Combining eye movements and EEG to enhance emotion recognition," in *Proceeding of International Joint Conference on Artificial Intelligence (IJCAI)* (Buenos Aires).

Mohammadi, G., Shanbehzadeh, J., and Sarrafzadeh, H. (2015). A fast and adaptive video-based method for eye blink rate estimation. *Int. J. Adv. Comput. Res.* 5, 105–114.

Mohanakrishnan, J., Nakashima, S., Odagiri, J., and Yu, S. (2013). "A novel blink detection system for user monitoring," in *IEEE/ACM International Conference on Utility and Cloud Computing (UCCV) Workshop* (Tampa, FL). doi: 10.1109/UCCV.2013.6530806

Moriyama, T., Kanade, T., Xiao, J., and Cohn, J. F. (2006). Meticulously detailed eye region model and its application to analysis of facial images. *IEEE Trans. Pattern Anal. Machine Intel.* 28, 738–752. doi: 10.1109/TPAMI.2006.98

Orozco, J., Roca, X., and Gonzlez, J. (2009). Real-time gaze tracking with appearance-based models. *Machine Vision Appl.* 20, 353–364. doi: 10.1007/s00138-008-0130-6

Ranjan, R., Patel, V. M., and Chellappa, R. (2017). Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Machine Intel.* 41, 121–135. doi: 10.1109/TPAMI.2017.2781233

Sirohey, S., Rosenfeld, A., and Duric, Z. (2002). A method of detecting and tracking irises and eyelids in video. *Pattern Recog.* 35, 1389–1401. doi: 10.1016/S0031-3203(01)00116-9

Sun, Y., Wang, X., and Tang, X. (2013a). "Deep convolutional network cascade for facial point detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Portland, OR), 3476–3483. doi: 10.1109/CVPR.2013.446

Sun, Y., Zafeiriou, S., and Pantic, M. (2013b). "A hybrid system for on-line blink detection," in *International Conference on System Sciences* (Hawaii, HI).

Tan, H., and Zhang, Y. J. (2006). Detecting eye blink states by tracking iris and eyelids. *Pattern Recog. Lett.* 27, 667–675. doi: 10.1016/j.patrec.2005.10.005

Vezzetti, E., Marcolin, F., Tornincasa, S., and Maroso P. (2016) Application of geometry to RGB images for facial landmark localization - a preliminary approach. *Int. J. Biometrics* 8, 216–236. doi: 10.1504/IJBM.2016.082597

Vezzetti, E., Marcolin, F., Tornincasa, S., Ulrich, L., and Dagnes, N. (2017). 3D geometry-based automatic landmark localization in presence of facial occlusions. *Multimedia Tools Appl.* 77, 14177–14205. doi: 10.1007/s11042-017-5025-y

Vidal, M., Turner, J., Bulling, A., and Gellersen, H. (2012). Wearable eye tracking for mental health monitoring. *Comput. Commun.* 35, 1306–1311. doi: 10.1016/j.comcom.2011.11.002

Xiong, X., and De la Torre, F. (2013). "Supervised descent method and its applications to face alignment," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*(Portland, OR), doi: 10.1109/CVPR.2013.75

Yahyavi, S. N., Mazinan, A. H., and Khademi, M. (2016). Real-time high-resolution detection approach considering eyes and its states in video frames through intelligence-based representation. *Complex Intel. Syst.* 2, 75–81. doi: 10.1007/s40747-016-0016-6

Yang, F., Yu, X., Huang, J., Yang, P., and Metaxas, D. (2012). "Robust eyelid tracking for fatigue detection," in *IEEE International Conference on Image Processing (ICIP)* (Lake Buena Vista, FL). doi: 10.1109/ICIP.2012.6467238

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.