Machine Vision and Applications manuscript No.
(will be inserted by the editor)

xSDL - Stroboscopic Differential Lighting Eye Tracker with Extended Temporal Support

Frank H. Borsato · Antonio Diaz Tula · Carlos H. Morimoto

Received: date / Accepted: date

Abstract Eye tracking (ET) for gaze interaction in wearable computing imposes harder constraints on computational efficiency and illumination conditions than remote ET. In this paper we present xSDL, an extended temporal support computer vision algorithm for accurate, robust, and efficient pupil detection and gaze estimation. The robustness and efficiency of xSDL partly come from the use of stroboscopic differential lighting (SDL), an extension of the differential lighting pupil detection technique developed in the 90's. Due to the erratic behavior of eye movements, traditional computer vision tracking techniques (such as Kalman Filters) do not perform well, so most ET techniques simply detect some eye feature (such as the pupil center) at every frame. Extended temporal support uses keyframes selected during eye fixations and a simple translation model of the pupil to further improve the computational performance of SDL. A prototype composed of two independent acquisition systems was developed to evaluate the performance of xSDL and other four state-of-the-art eye tracking techniques under similar conditions. Our results show that xSDL outperforms those 4 algorithms, both in speed (up to 500 Hz using 240 line frames) and accuracy, using a modest platform compatible with wearable computers.

Frank H. Borsato
Universidade Tecnológica Federal do Paraná, Via Rosalina Maria dos Santos 1233, Campo Mourão, 87301-899, Brazil
Tel.: +55-44-3518-1508
Fax: +55-44-3518-1400
E-mail: frankhelbert@utfpr.edu.br

Antonio Diaz Tula and Carlos H. Morimoto
Universidade de São Paulo, Rua do Matão 1010, São Paulo, 05508090, Brazil

1 Introduction

Eye trackers are devices that can estimate the point-of-gaze on a computer screen [27] or in the scene in front of the user [19]. Such devices are commonly used in usability studies [16], marketing research [22,29], medical diagnosis [20,32], communication for people with disabilities [4], psychological and psychophysical studies [5], and virtual reality [31,33].

Current eye tracking systems are mostly feature-based [13], i.e., they use one or more video cameras to detect and track the eye features and estimate the point of gaze on a planar surface (typically the computer screen). Most methods detect and track the iris or the pupil center and use active near infrared (NIR) illumination to improve the tracking performance. The use of NIR light is also desirable because it creates a corneal reflection that can be used as a reference point for gaze estimation [27].

The point of gaze can be estimated by a function that maps eye features (such as the pupil center) onto the observed surface. For example, a second order polynomial can be used where the coefficients can be computed using regression techniques with corresponding eye features and gaze positions obtained from a calibration procedure. Detecting eye features, such as the pupil or iris, in a video frame can be challenging due to noise, low image resolution, and motion blur.

When the eye is illuminated using a NIR light source placed away from the camera optical axis (off-axis), the pupil appears dark in the camera image. This facilitates the segmentation and contour detection of pupils within light-color irises, but the contrast between the pupil and iris is low for people with darker eyes. Segmentation and tracking of the iris or iris contour (also known as

limbus tracking) is also possible but it is more likely to be affected by occlusions of the eyelids and lashes.

Differential lighting (DL) [8,25] was introduced to improve the robustness of pupil detection methods. DL relies on two NIR illuminators. One is placed very close to the camera optical axis (on-axis), and one off-axis. The on-axis illuminator generates bright-pupil images because the camera is able to capture the light reflected from the back of the eye. DL alternates the off and on-axis illumination, producing a sequence of dark and bright-pupil images. By subtracting two consecutive frames (one dark and one bright), the overlap between the dark and bright pupils can be easily segmented as regions of high contrast.

Despite its advantages, DL requires camera synchronization with the NIR light sources to produce the sequence of dark and bright-pupil images. Figure 1 depicts such arrangement. Unfortunately most low-cost consumer cameras today do not provide a synchronization output. This is particularly true for digital web cameras used with computers. This might be one of the reasons why current low-cost eye trackers built with off-the-shelf web cameras use a single off-axis NIR illumination to detect and track the pupil [21,10,11,35].

In this paper, we present a high-performance, low-cost, stroboscopic differential lighting eye tracking technique with extended temporal support (xSDL). Our technique can be used with virtually any digital camera, and was particularly designed to be used with cameras without external synchronization. The dual NIR illuminators are synchronized by software.

The remaining of this paper is organized as follows. Section 2 presents the basic differential lighting technique developed for analog cameras. Section 3 describes how the use of stroboscopic lighting allows DL to be used with digital cameras without external synchronization output. In Section 4 we introduce the extended temporal support algorithm to improve the overall performance of SDL. Section 5 presents the evaluation of xSDL and its comparison with state-of-the-art algorithms in a typical gaze estimation experiment. Section 6 presents the results of the xSDL evaluation. In Section 7 we present a discussion about xSDL, and finally Section 8 concludes the paper.

2 Pupil Detection Using Differential Lighting

The differential lighting (DL) technique introduced by Ebisawa and Satoh [8] was developed as a robust pupil detection method to improve the performance of non-verbal communication tools for people with disabilities. In [7], Ebisawa shows several refinements to the basic DL such as noise removal using morphological operations and pupil brightness control.

Morimoto et al. [24] describe a pupil-corneal-reflection (PCR) gaze estimation technique using DL and a second order polynomial for mapping the PCR vector to target coordinates. Their system used an analog 30 Hz NTSC camera with external synchronization output. An external electronic circuitry was used to synchronize the even and odd frames of one interlaced camera image to the on and off-axis lights. Because each interlaced image contained a dark and bright-pupil image, the pupil could be detected at 60 fields per second (where the field has half the resolution of an image frame after de-interlacing). In [24] the pupil was computed as the center of mass of the blob detected from the differential image. Hennessey et al. [15] proposed, as a further DL refinement, the computation of the actual pupil contour in the bright or dark-pupil images, since the differential image only provides the overlap region when the eye is moving.

Morimoto and Flickner [23] also use DL to detect the eyes in a wider area to detect and track multiple faces. Ji and Yang [17] describe a gaze and face pose tracking system for monitoring driver's vigilance using DL. Due to the simplicity and good overall performance of the method, several other refinements and applications have been suggested in the literature [14,18,36].

DL was developed in the 90's to be used with analog cameras and *low performance* computers – low performance when compared to regular desktop computers today. As computers got more powerful and cameras more affordable and easier to setup and use (digital plug-and-play cameras), other pupil detection and tracking methods that do not require custom external hardware and use only visible light, such as [12], are preferred in practice despite DL's good performance. Nonetheless, as computational power continues to increase and hardware continues to reduce in size, the rise of technologies such as mobile, ubiquitous, and wearable computing demands more restrictive requirements for energy consumption and computational efficiency. Because an eye tracker can be used as a wearable input device that is always on, DL might become a stronger alternative because the active lighting allows the method to work under different lighting conditions, it is computationally very efficient, and its simplicity allows the method to be implemented in hardware [1].

Despite its advantages, DL using analog cameras requires the external lighting to be synchronized with the camera's odd and even fields [7,24]. A modern alternative would be the use of syncing-capable digital cameras with global-shutter, where all pixels are exposed within the same time window and a signal is generated to allow

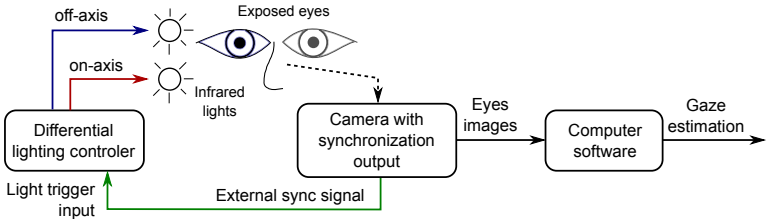


Fig. 1 Block diagram depicting the main components of the DL technique. The camera has a synchronization output which is used to trigger the light sources.

synchronization. While this option is attractive, such high-end cameras are still expensive. Most modern digital cameras (such as external webcams and cameras used in notebooks, tablets, and mobile phones) employ rolling shutters, i.e., each line of the frame is exposed to light at slightly different times. This sliding window creates image artifacts when fast moving objects, such as the eye, are present in the scene being captured. Also, because DL uses two light sources, it is possible that during an image scan part of the image is illuminated by one light and the rest of the image is illuminated by the other light source, as illustrated in Figure 2.

To reduce these artifacts, we have proposed in [2] the use of stroboscopic differential lighting (SDL) controlled by software. This technique is described next.

3 Stroboscopic Differential Lighting (SDL)

In [2] we have described how stroboscopic lighting can be synchronized with rolling-shutter digital cameras. The idea is to fire one very short light pulse for every frame. The use of short light pulses allows low-end cameras to capture very sharp images (reduces motion blur) and reduces artifacts due to the rolling-shutter. Nonetheless, when the lighting is not correctly synchronized with the camera frames other artifacts such as those shown in Figure 2 are created. The dark stripes correspond to sensor lines that were not lit by the light pulses.

Our method exploits the dark stripe artifacts to synchronize the lighting. In [2] we presented the computer vision algorithms to detect the stripe and to compute its spatial and temporal properties that are used to adjust the lighting parameters to conceal the stripes within hidden image lines. The basic idea is to first compute the position of the stripe by computing a vertical integral image, i.e., a column vector where each element corresponds to the integral of an image line. Once the stripe position is detected, the stroboscopic pulses are modulated to shift the stripe towards the hidden lines of the camera sensor.

One limitation of the method described in [2] was that it relied on the knowledge of sensor parameters.

a) b)

Fig. 2 Dark (a) and bright (b) pupil images with dark stripes created when the stroboscopic lights are not correctly synchronized with the camera frames. The stripes correspond to sensor lines that were not lit by the light pulses.

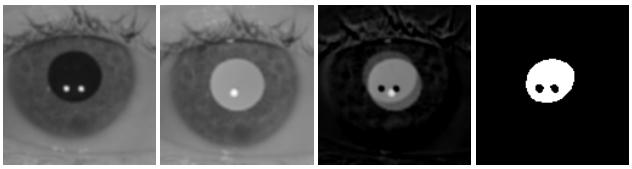
Because this information is not always available, the parameters had to be manually adjusted for some camera configurations used in the experiments. In [3] we presented a new solution that dynamically estimates the camera sensor exposure and number of lines from the dark stripe artifacts. Starting with a coarse estimation of the sensor parameters, the position and height of the stripe is computed using the vertical integral image. The stripe parameters are then used to refine the estimation of the sensor parameters before the adjustment of the firing of the illuminators, until a clear picture (without artifacts) is obtained.

While in previous papers we have focused on the SDL hardware and synchronization issues, the focus of this paper is on the computer vision software for accurate detection of eye features for gaze estimation. In the following subsections we describe how the pupil and corneal-reflections are detected using SDL.

3.1 Segmentation of pupil candidates

With the camera and light sources synchronized, the difference between two consecutive frames containing bright and dark-pupil images, can be used to detect pupil candidates by thresholding [26] as seen in Figure 3.

The resulting dominant blob most likely corresponds to the overlap pupil region. Instead of using a fixed threshold value to compute the blob, we initially use



a) b) c) d)

Fig. 3 Pupil candidate segmentation. a) dark pupil image; b) bright pupil image; c) Difference between bright and dark pupil images; d) Difference image thresholded.

an adaptive threshold technique similar to the procedure described in [21] to detect corneal reflections. The threshold is computed using (1) where $\mathbb{H}$ is the inverted cumulative histogram of the difference image and A is a geometric constraint used to eliminate noise, denoting the minimum expected pupil area in pixels.

$$\text{threshold} = arg\min_{i} |\mathbb{H}_i - A| \tag{1}$$

$\mathbb{H}$ is computed by (2), where i and j are bin numbers, h_j is the number of pixels that falls into the intensity interval defined by bin j, and k is the total number of bins.

$$\mathbb{H}_i = \sum_{j=i}^{k-1} h_j, \quad i = 0..k-1 \tag{2}$$

The algorithm varies the threshold from high to low values in large steps, speeding up the convergence. A new threshold is computed as the intensity which provides enough pixels to fill up an ellipse that encloses the contour of the largest area traced from the binarized image in the current iteration. We calculate the new threshold using (1) with A as the ellipse area.

Assuming that the pupil region corresponds to the largest elliptical blob, its area increases as the threshold is lowered. The search stops when the ratio between the largest blob and the remaining smaller blobs starts to decrease [21]. The threshold that maximizes the area A of the overlap region is selected. Additional geometric constraints regarding the expected size, shape, and position of the pupil are used to filter false candidates. Once this threshold is computed its value is used in future detections. The threshold is recomputed again only after long periods of miss or false pupil detections.

3.2 Segmenation of the corneal reflections

A similar adaptive threshold procedure based on the inverted cumulative histogram method is used to segment the corneal reflections (CRs) generated by the IR light sources. The reflections appear as bright small spots in the pupil image and they are commonly used to improve gaze estimation results [27]. Quite often they are within the pupil region and create artifacts in the difference images as seen in Figure 3.

Geometric constraints such as the expected size and position of the CRs are used to filter some of the false positive candidates. Unlike the pupil though, other candidates might remain, particularly due to the tear layer near the eyelids, and near eyelashes.

To improve the robustness of the CR detection, each CR candidate g is modeled as a 5-tuple $(l_g, c_g, r_g, d_g, \hat{r}_g)$, where l_g is the number of iterations in which g is present; c_g is the coordinate of the reflection center, r_g is the radius of the enclosing circle; d_g is the distance to the pupil center; and $\hat{r}_g$ is the ratio between the number of segmented pixels within the circumference with radius r_g and the number of pixels within the circumference with radius r_g+1, both centered at c_g. At each iteration of the adaptive threshold method, an ordered list of the corneal reflection candidates is stored. The candidates are sorted according to a quality function Q defined as

$$Q(g) = ((l_g + 1) \cdot r_g \cdot (1.0/(d_g + 0.1)) \cdot \hat{r}_g), \tag{3}$$

Therefore, the best quality CRs are those that are brighter $(l_g + 1)$, larger (r_g, within an expected range), closer to the pupil center $(1.0/(d_g + 0.1))$, and shaped like a circumference $(\hat{r}_g)$. The computation of the threshold stops when a number of appropriate CRs are detected and are stable between two iterations.

The position of the center (x_c, y_c) of one CR is estimated as the normalized center of mass that weights the pixels according to their intensities as follows:

$$(x_c, y_c) = \frac{1}{N} \sum_{i=1}^{N} (\mathrm{e}^{\varrho \cdot (I(x_i, y_i) - 1)} \cdot (x_i, y_i)) \tag{4}$$

where N is the total number of segmented pixels that belong to the CR, the function $I(.)$ returns the pixel intensity normalized to $[0, 1]$ and ϱ is a weighting constant. Higher values of ϱ are used to select the center closer to the brightest pixel while with small values all pixels are considered. Two weighting constant values are used, one for the bright pupil images and one for the dark pupil images. The weighting was introduced to reduce biasing on the center estimation when the CR surrounding pixels are bright. (4) assumes that the CR intensity profile follows a symmetric bivariate Gaussian distribution such as the one described in [21].

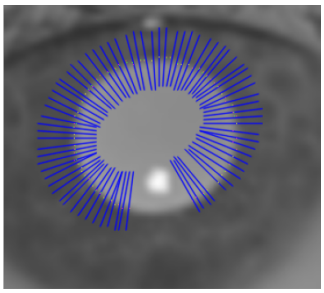


Fig. 4 Rays used to refine the edges of a bright pupil.

3.3 Pupil refinement with sub-pixel accuracy

Tough the pupil overlap region can be used for gaze estimation [24], more accurate results can be achieved using the true pupil contour. Starting from the pupil candidate obtained from thresholding, its contour can be described by an ellipse (e_o), defined by the 4-tuple $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \hat{\theta})$, where $\mathbf{a}$ and $\mathbf{b}$ are the major and minor axes, $\mathbf{c} = (c_x, c_y)$ is the ellipse center, and $\hat{\theta}$ is the rotation angle. The pupil refinement consists of projecting a number of rays R_k, $k = 1...m$ outwards from the pupil center to detect the actual pupil edges in the current frame, similar to [28], as seen in Figure 4. The length of each ray is proportional to the length of the pupil principal axis. The edge pixels with subpixel accuracy are then used to estimate the ellipse parameters.

Observe in Figure 4 that rays intersecting the CRs are discarded when they are detected within the pupil region to avoid further interference in the computation of the pupil contour.

Each ray profile is transformed to a column vector as seen in Figure 5 using bilinear interpolation. On each column (ray profile), a one dimensional convolutional Gaussian derivative edge detector is applied. Though the peak response in general corresponds to the pupil-iris boundary, other strong responses might be caused by eyelashes and scene reflections. To lower the effect of such noisy rays, a normalized weight is used to sort the rays as follows:

$$W_k = \underset{j=1...|R|}{max} \left(\frac{R_k^*(j)}{\sum_{i=1}^{|R|} R_k^*(i)} \right) \tag{5}$$

where $R_k^*(j)$ denotes the edge response at position j of ray R_k, and W_k the weight attributed to ray R_k. Typically, 10% of rays with the lowest weight are discarded.

For the remaining rays, subpixel accuracy is obtained by interpolating the values around the pixel with strongest edge response similar to [6]. Figure 5 shows the intensity profiles of the rays computed from a bright pupil image and the corresponding strongest responses along with the results of the Gaussian interpolation.

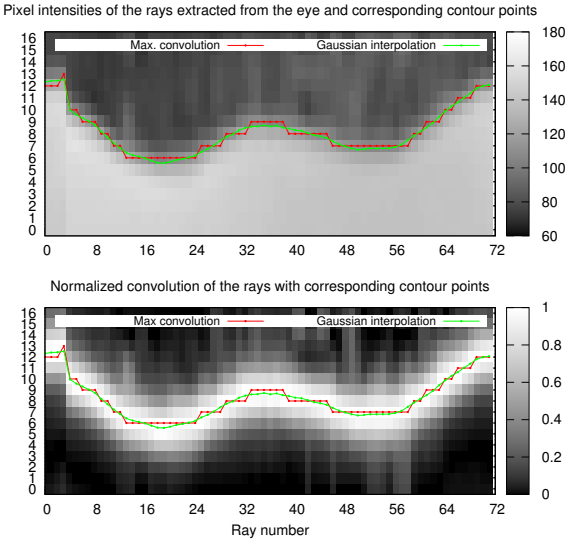


Fig. 5 The intensity profiles of rays computed from a bright pupil image. (Top) Rays resulting from bilinear interpolation of the pupil image along the ray trajectory. (Bottom) Results of the edge detector convolution along each ray. The red line shows the strongest filter response location, while the green represents the Gaussian estimator result.

The efficacy of removing outliers by discarding pupil-iris boundary candidates associated to the lowest weighted ray profiles depend on the chosen cut-off value. Therefore, before assuming subpixel locations on the remaining rays correspond to the actual pupil-iris boundary, a second filtering is performed, this time based on local statistics. This filter is intended to remove remaining outliers based on the spatial distribution of candidate positions. Consider the array C containing the subpixel locations of the peak edge responses of each ray, sorted by the associated angle of projection. Each position in this array is a candidate to be part of the pupil contour and, therefore, we expect the values in C to vary smoothly. To remove outliers, we compute an array (σ_L) with the local standard deviation (LSD). Each value in σ_L contains the standard deviation (SD) of a neighborhood around the corresponding value in C (the neighborhood is typically of size 7). Outliers are expected to have high LSD, which are detected by thresholding as the candidates with values higher than the mean plus the SD over all LSD, i.e. $threshold = \mu_{\sigma_L} + \sigma_{\sigma_L}$. An ellipse is fitted to the remaining candidates based on the direct least squares method [9].

Figure 6 shows a pupil partially covered by eyelashes and the candidates filtered by the threshold described.

When the ellipse is larger or smaller than the expected size and shape of a pupil or the ellipse is not completely contained in the image, the refinement process stops and reports that no pupil was detected in the frame.

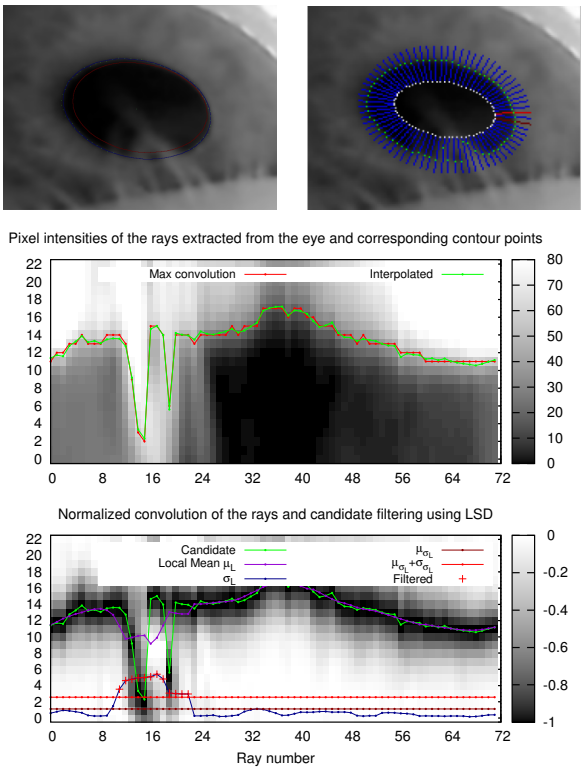


Fig. 6 Top: intensity profile of each ray extracted from a dark pupil image and corresponding convolutions. Bottom: the result of filtering using local statistics.

4 Extended Temporal Support

Basic eye movements can be classified into fixations, saccades, and smooth pursuits [30]. Because our vision is foveated, during a fixation the target of interest must be projected onto the fovea to be perceived at the highest resolution. The fovea only covers a small region of the retina, so the eye must stay somewhat stable during fixations. To perceive all the details of large objects, the eye must fixate on several locations. A saccade is a fast ballistic eye movement that moves the eye to different locations. Pursuits are eye movements used to foveate on moving objects, such as a flying bird.

Differential lighting works better when the overlap between the bright and dark pupil images is large. During saccades the overlap region between two consecutive frames may become quite small for slow frame-rate cameras. Fast frame-rate cameras are used when saccades are required to be tracked. Nonetheless, slow frame-rate cameras are still adequate when tracking saccades are not important, e.g., for most gaze interaction applications.

Independently of the camera frame-rate, an eye tracker based on DL typically detects the pupil at every frame, considering the overlap from the previous frame. Though

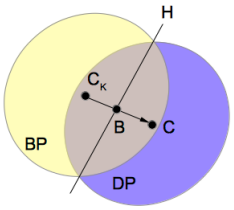


Fig. 7 xTS pupil position estimation process using a bright pupil (BP) as keyframe. A dark pupil (DP) center C is computed as a function of the keyframe pupil center C_k and the blob ellipse center B.

the pupil position could be predicted during pursuits by regular computer vision tracking algorithms, the erratic behavior of the eye during saccades makes pupil positions hard to predict.

Instead of using consecutive frames to estimate the pupil positions, our extended temporal support (xTS) technique computes most pupil displacements from selected keyframes, speeding up the computation of the new pupil position.

The basic idea is shown in Figure 7. Consider as keyframe a bright-pupil (BP) image. When any dark-pupil (DP) image with sufficient overlap is grabbed, the new pupil position C can be estimated as a function of the position of the pupil C_K in the keyframe and the center B of the overlap blob, computed as the center of mass of the blob. Assuming that BP and DP are about the same size, C can be computed as

$$C = B + (B - C_K) \tag{6}$$

The xTS algorithm maintains both a dark and a bright pupil keyframes. The algorithm for pupil tracking using extended temporal support with SDL (xSDL) uses those keyframes, whenever they are available, to compute the overlap region from the difference image of the most current frame and its appropriate keyframe (dark or bright). When the overlap is significant (typically, the minimum overlap width is set to be at least half the length of the keyframe's pupil minor axis), the new pupil center is computed using (6). Otherwise, xSDL tries to detect the pupil in the current frame using the previous frame. If successful, the new frame is considered as a new keyframe (dark or bright). In case no pupil is detected and no keyframe is available, the algorithm continues to detect pupils using consecutive frames until keyframes are once again available. Long periods without any pupil being detected is possible during blinks for example.

5 xSDL Empirical Evaluation

We compare the performance of xSDL with the following 4 eye tracking algorithms: Starburst [21], ExCuSe [10], ElSe [11], and the method proposed by Świrski et al. [35].

Li and Parkhurst [21] introduced Starburst, a hybrid method that integrates feature-based and model-based approaches to detect the pupil. Starting from a point within the pupil, rays are projected radially to detect pupil-iris boundary points, i.e. where the derivative is larger than a threshold. Those feature points are used to fit an ellipse using RANSAC. The process is repeated starting at each feature point and projecting rays on the opposite direction. To reduce possible bias from the selected initial point, the process is iterated replacing the start point by the average location of all feature points until convergence, i.e. the averaged center differs less than a given threshold from the starting point location. An ellipse is fitted to the feature points using RANSAC, followed by an image-aware model-based optimization used to improve the fitting. The algorithm also finds the CR using adaptive thresholding and remove it using interpolation.

ExCuSe (Exclusive Curve Selector), developed by Fuhl et al. [10], is based on edge filtering and oriented histograms calculated via the Angular Integral Projection (AIP) function. The method follows different processing flows depending on the normalized image histogram. When a bright peak is detected, the pupil is estimated using an edge-filtering approach. The edge-filtering comprehends several steps applied to the Canny-edge image. The first step discards single pixels, small rectangles, and straight lines. The next step selects the curve that most likely encloses the pupil (the darkest area). Finally, an ellipse is fitted using all points in the selected curve with a direct least-squares method. In case a peak is not detected in the normalized image histogram, a coarse pupil position is estimated using AIP functions on a thresholded image followed by a refinement step. Four AIP are computed 45°apart, and the pupil position is assumed to lie in the intersection of the strongest function responses. The pupil center is then refined by ellipse estimation similar to the one employed by the Starburst method [21].

ElSe (Ellipse Selector), from Fuhl et al. [11], is also based on edge filtering. After Canny-edge filtering, edges are filtered out (using edge thinning followed by edge straightening) and separated so that every edge has elliptical shape. The next step selects the best ellipse fitted to the edges, according to their area, shape, and intensity ratio between the inside and outside area. In case no ellipse is found, (e.g. due to motion blur), a coarse estimation of the pupil position is computed using convolution in the downscaled image, followed by a refinement step. As this second part of the method always computes an ellipse, a validation is performed as a last step to avoid returning a pupil position for a closed eye.

The last method used in the evaluation is the one proposed by Świrski et al [35]. The method uses a Haar-like feature detector to initially find a rough estimation of the pupil location. The intensity histogram of a region around the coarse position is clustered using k-means to refine the pupil center. Finally, the algorithm estimates the pupil contour with an image-augmented RANSAC ellipse fitting. Robustness and accuracy are improved by employing a support function that weights inliers according to their gradient direction and magnitude, preferring sets of points which agree with the ellipse gradient.

All these algorithms have a C or C++ source code available for download. The Starburst implementation used is available online [21] in both C and Matlab versions. The C version lacks the model-based optimization step which we have added for completeness. The experiments with Starburst, ElSe, and ExCuSe, used their default parameters, as provided by their source codes. The experiments with Świrski et al. algorithm also used default parameters, except for the pupil minimum and maximum radii, for which the technique showed to be quite sensitive. This parameter was adjusted once per user.

The experiment was designed to compare the performance of each method in a typical gaze estimation task. Ten people (4 female) from 30 to 59 years old (mean 36.2) volunteered for the experiment. All participants had normal vision without the need of any correction lenses.

During the experiment, each participant had to fixate their gaze on 35 small circular targets (7 pixels in diameter) arranged in a 5×7 grid. Targets were displayed one at a time in random order for about 3 seconds, and the 800 ms interval 500 ms before the next target presentation were considered for analysis. A 1680 × 1050 resolution, 60 Hz, 22" monitor was used, with a 20 pixel border at each side of the monitor. The distance between the monitor and the participants' eyes was about 70 cm.

5.1 Apparatus

The xSDL eye tracker prototype was built using a low-cost, off-the-shelf Play Station 3 video camera [34]. Because all algorithms except xSDL work with the dark-

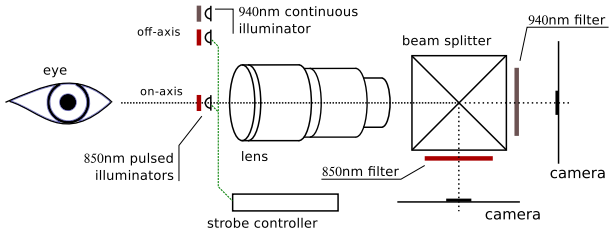


Fig. 8 System prototype used to compare eye tracking methods based on xSDL and dark-pupil-only methods.

pupil image, our prototype actually included 2 identical cameras and two independent lighting systems: one to capture the sequence of dark and bright-pupil images using structured illumination needed for xSDL, and the other to capture only the dark-pupil images for the other methods. Figure 8 depicts this setup.

The structured illumination system (used by the xSDL method) was composed by two light sources: on and off-axis, controlled by an Arduino board. The light was provided by 850 nm LEDs.

The continuous illumination system responsible for producing dark-pupil images only was built using 940 nm LEDs. A flat-convex lens of 20 mm focal length was used to improve the LED efficiency. This source was also filtered using a narrow bandpass filter centered on the emitter wavelength.

After passing through the objective lens, the light was separated into two paths by a custom made beam splitter. Each path projected light to a different video camera: one for xSDL algorithm and the other for dark-pupil-only algorithms. To avoid light interference between the different illumination systems, bandpass filters were placed in front of the camera sensors: a 850 nm pass filter for the xSDL camera and a 940 nm filter for the other camera. Thus, the eye image was captured from the same perspective using both illumination systems (the stroboscopic technique and constant illumination), as shown in Figure 9.

5.2 Data analysis

We used 9 (out of the 35) points to calibrate a second order polynomial function. This function was used to estimate the gaze position at all of the 35 points to compute the gaze estimation error.

We use three metrics to evaluate the methods: pupil detection robustness, gaze estimation accuracy and precision, and processing time.

Pupil detection robustness refers to the proportion of frames that a method is able to detect the pupil. All images collected were manually inspected and only those containing the pupil image were used for evaluation. For xSDL, the sequence of frames contained bright and dark pupil images. For all other methods, the sequence of frames contained only dark pupil images. Because Świrski's method and Starburst always return a pupil boundary even though these methods did not accurately detect the pupil, we decided to consider frames with a gaze estimation error above 5 degrees as no pupil detection.

The accuracy of each method is defined as the average gaze estimation error over all 35 target locations. Error in gaze estimation is the difference between the actual target position and the estimated gaze position in degrees of visual angle. The precision is given by the standard deviation of each participant's error.

The processing time is computed as the average time a method took to process each frame. All methods were timed using the same computer platform, a notebook equipped with an AMD Turion(tm) II P560 Dual-Core Processor with 2.5 GHz and 6 GB of RAM.

6 Results

This section presents the experimental results of pupil detection robustness, gaze estimation accuracy and precision, and processing time.

6.1 Pupil detection robustness

For each participant $\times$ method in all captured frames, we computed the proportion of frames where the pupil was detected (i.e. with a gaze estimation error below 5 degrees). Figure 10 shows the boxplot for all methods.

Because data is not normally distributed, as can be observed in Figure 10, we ran the non-parametric Friedman test. Results showed a significant effect of algorithm ($\chi^2(4) = 28.93$, $p < 0.01$). A post-hoc Wilcoxon signed rank test with Holm correction showed that xSDL (grand median 99.92%) was significantly different than ElSe (grand median 64.51%), ExCuSe (grand median 81.28%), and Starburst (grand median 97.48%), $p < 0.05$ in all cases. The method of Świrski (grand median 99.6%) had no significant difference with any other method. The difference between Starburst and ElSE was also significant ($p < 0.05$).

6.2 Gaze estimation accuracy and precision

For each method, the gaze estimation error was computed for all frames with a successfully detected pupil. For xSDL and Świrski, the error distribution was right-skewed, as shown in Figure 11. Starburst's distribution

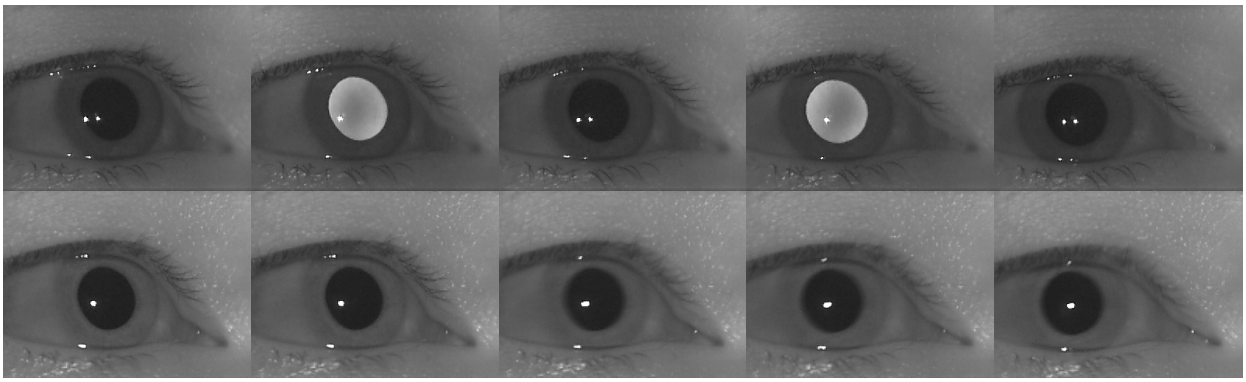


Fig. 9 Example capture of a participant with an ongoing saccade. On top, xSDL camera images and on bottom, continuous 940 nm illuminated images.

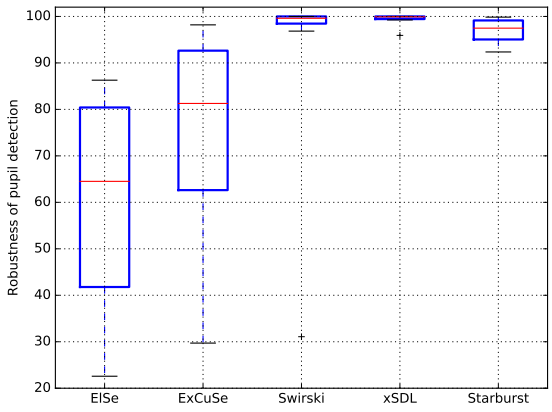


Fig. 10 Boxplot of pupil detection robustness for each method computed with data from the 10 participants.

was also right-skewed, though with a larger error. On the other hand, ElSe and ExCuSe had a slightly right-skewed distribution with a larger error and dispersion. Because of the different distributions, we computed the median error of each participant for each method. Figure 12 shows the boxplot of gaze estimation error for the 10 participants for each method.

As can be observed in Figure 12, xSDL was the most accurate method (grand median 0.57^o) and also the most precise (it had the smallest inter-quartile range (IQR) of 0.51^o). We run a Friedman test and found a significant effect of algorithm on accuracy ($\chi^2(4) = 30.88$, $p < 0.01$). A post-hoc Wilcoxon signed rank test with Holm correction showed that xSDL was significantly more accurate compared to ElSe (grand median 2.37^o), ExCuSe (grand median 2.2^o), and Starburst (grand median 0.95^o), $p < 0.05$ in all cases. There was also a marginal significant difference compared to Świrski (grand median 0.72^o), $p = 0.059$. We also found a significant difference between Starburst and ElSe ($p < 0.05$).

6.3 Processing time

We measured the time (in milliseconds) to process each frame. The histograms in Figure 13 show that the processing time distribution varied largely among methods. For Świrski's the distribution was more scattered compared to all other methods. ElSe and ExCuSe had a bimodal distribution, that might indicate different execution paths in their algorithms. Processing time of xSDL and Starburst had a similar distribution.

We computed the median processing time per frame for each participant and method. Figure 14 shows the boxplot of processing time for the 10 participants. Because of the large difference in processing time among methods, we used a logarithmic scale in Figure 14 to facilitate visualization. As can be observed, xSDL had the smaller processing time (grand median of 2.28 ms) and also the smaller variation (IQR 0.33 ms). Starburst had the second lowest average running time (grand median 6.31 ms, IQR 1.77 ms). On the other hand, Świrski had the larger processing time (grand median 79.89 ms) and also the larger variation (IQR 41.07 ms), followed by ElSe (grand median 42.65 ms, IQR 0.48 ms) and ExCuSe (grand median 18.25 ms, IQR 8.39 ms). A Friedman test showed a significant effect of algorithm ($\chi^2(4) = 39.28$, $p < 0.01$). A post-hoc Wilcoxon signed rank test with Holm correction showed that processing time per frame differed significantly between all methods ($p < 0.05$ in all cases).

7 Discussion

Experimental results support that xSDL is more robust, accurate, precise, and requires less processing power

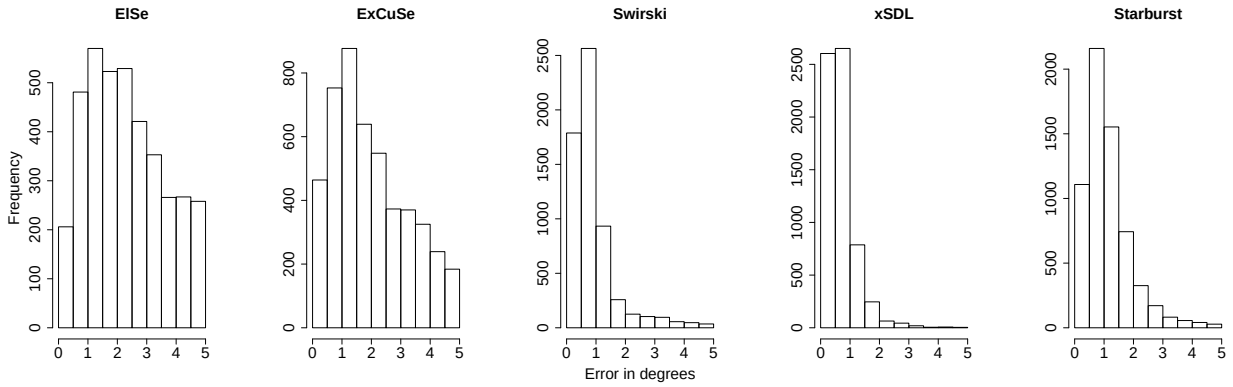


Fig. 11 Gaze estimation error histograms for all participants and methods.

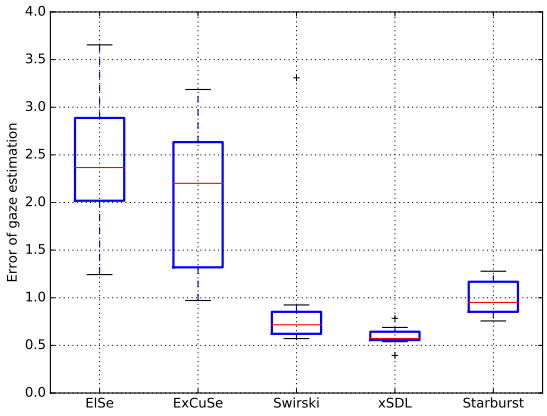


Fig. 12 Boxplot of accuracy for each method computed with data from the 10 participants.

compared to the other state-of-the-art methods that were evaluated.

The improved pupil detection robustness of xSDL could be attributed to the differential lightning technique, since it takes advantage of the sequence of dark and bright pupil images generated by the alternating light sources. The overlapping pupil area computed by the difference of consecutive (dark-bright or bright-dark) images serves as a robust estimator of the pupil location in the image, thus discarding false positives. Dark pupil image based methods rely on histogram analysis (such as ExCuSe and Świrski) or high contrast areas (borders detected by Canny-edge filter or derivative) as a rough estimation of the pupil location. These feature-based approaches could result in areas not corresponding to the pupil to be detected as such, e.g. a black blob in the captured image. Among these methods, Świrski was the only one that had a robustness close to xSDL, but at the cost of a much larger running time and hence computational power.

xSDL was also the most accurate and precise method, meaning it had the smallest error in gaze estimation. Accurate gaze estimation is very important for many applications, such as to point at small visual targets in a gaze-controlled interface such as a virtual keyboard. More accuracy allows to include more objects in an interface, e.g. punctuation, accents, and control commands in a virtual keyboard. Accuracy also reduces the number of selection errors. Precision is also very important, since gaze estimation is more consistent, improving the quality of gaze data. Better quality implies a better usability in gaze-controlled interfaces, and also more accurate results in medical applications that use gaze information.

Computational efficiency is another advantage of xSDL, as shown by the average processing time of each method. Not only xSDL is the fastest method, but it also holds the smallest dispersion among all methods. Our prototype was tested at 60 frames per second to keep infrared radiation within safe limits, as two independent illumination sources were in use. By reducing the infrared radiation emission (i.e. by using our technique alone), it is possible to run the prototype at 187 frames per second with the same low-cost PS3 video camera. By using faster cameras it would be possible to run xSDL at about 450 Hz with a similar computer processor, given the short average processing time of 2.28 ms per frame. It is noteworthy that ElSe and ExCuSe had a bimodal distribution processing time. This could be explained because those two algorithms have 2 possible execution paths. If the main execution path fails, then those methods use a simpler approach in an effort to detect the pupil. Though we did not verified whether ElSe and ExCuSe indeed ran different execution flows in our experiment, this hypothesis explains the bimodal distribution in their processing time histograms. Świrski et al. method possess the largest processing time, of about 80 ms per frame, and also the

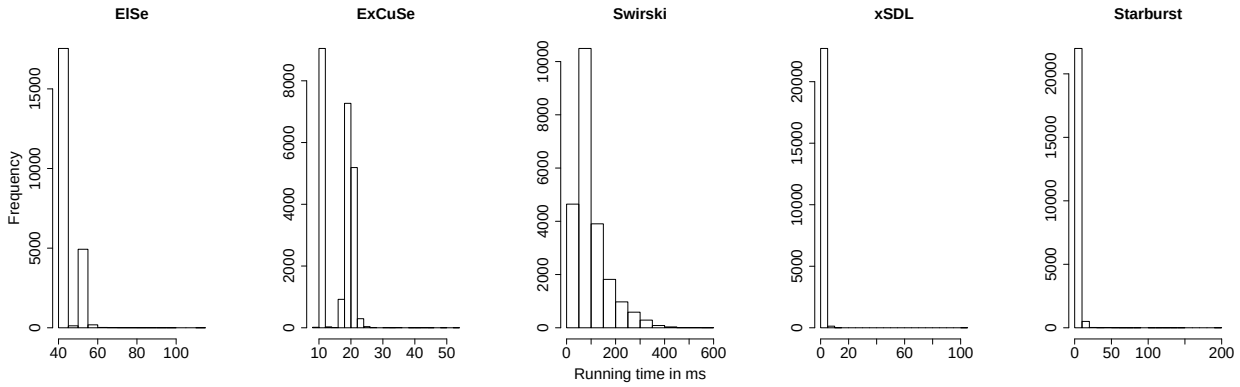


Fig. 13 Processing time per frame histogram for all participants and methods.

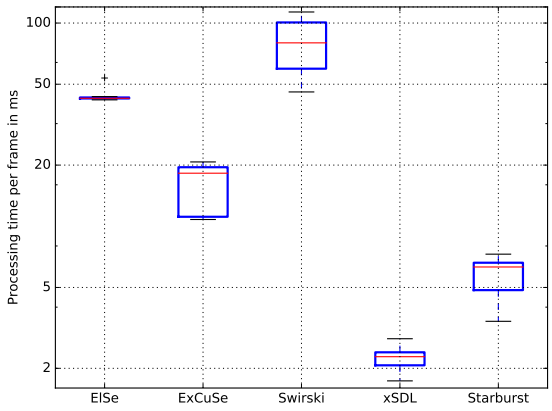


Fig. 14 Boxplot of processing time per frame for each method computed with data from the 10 participants. Note that the y axis has a logarithmic scale to facilitate visualization.

largest variation, making it most computation expensive method we have evaluated.

xSDL can be used with any digital, off-the-shelf camera. The only additional hardware requirement is a micro-controller to control the firing of the stroboscopic lighting. Therefore, xSDL arises as a low-cost, high-performance, and robust eye tracker alternative for gaze-based applications.

8 Conclusion

Differential lighting (DL) is a robust and computationally efficient pupil detection technique developed in the 90's for analog cameras. We have shown, in previous papers, that the use of stroboscopic differential lighting (SDL) allows DL to be used with any modern and low-cost rolling-shutter digital camera. We believe that its improved performance will compensate the trouble of building the extra synchronization hardware (in case it is not already available) in gaze enhanced wearable computers.

This paper introduced the extended temporal support computer vision algorithm that further improves the performance of SDL, we called xSDL. Traditional computer vision tracking techniques are not used for eye tracking in general because they are either computationally expensive (such as correlation based methods) or do not perform well due to the unpredictable eye movements during fixations. SDL detects the pupil at every frame by simple image differencing, and works better when the overlap between bright and dark pupil images is large, i.e., it works well during fixations. xSDL selects keyframes (dark and bright pupil images) to serve as reference images. The position of the pupil in keyframes are computed with high accuracy. This position is then refined by xSDL using the translation of the overlap region computed from the difference image.

To compare xSDL performance against other state-of-the-art eye tracking algorithms available in the literature, we have created a two-camera apparatus that simultaneously collects videos of the eye that are appropriate for xSDL (that requires bright and dark pupil images), and the other methods, that only use dark pupil images. Therefore, the methods were compared using videos showing exactly the same eye behaviors and with the same frame rate (and other video properties).

Our results show that xSDL outperforms those 4 algorithms. Using 240 line images, xSDL is able to process up to 500 fps using a modest platform compatible with a wearable computer, with an accuracy of about 0.6^o and detecting the pupil practically all the time. The second most accurate and robust method (Świrski et al. [35]) achieved an accuracy of about 0.7^o and similar robustness of over 99% of detections. This alternative though is much more computationally expensive, being

able to achieve about 13 fps using the same computing platform.

Acknowledgment

This work has been supported by the Fundação Araucária (DINTER project UTFPR/IME-USP) and FAPESP grant number 2011/00267-1, 2012/04426-0, and 2016/0446-2.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Amir, A., Zimet, L., Sangiovanni-Vincentelli, A., Kao, S.: An embedded system for an eye-detection sensor. Comput. Vis. Image Underst. **98**(1), 104–123 (2005). DOI 10.1016/j.cviu.2004.07.009. URL http://dx.doi.org/10.1016/j.cviu.2004.07.009
2. Borsato, F., Aluani, F., Morimoto, C.: A Fast and Accurate Eye Tracker Using Stroboscopic Differential Lighting. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 110–118 (2015)
3. Borsato, F.H., Morimoto, C.H.: Building structured lighting applications using low-cost cameras. In: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 15–22 (2017). DOI 10.1109/SIBGRAPI.2017.9
4. Diaz-Tula, A., Morimoto, C.H.: AugKey: Increasing Foveal Throughput in Eye Typing with Augmented Keys. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pp. 3533–3544. ACM, New York, NY, USA (2016). DOI 10.1145/2858036.2858517. URL http://doi.acm.org/10.1145/2858036.2858517
5. Diaz-Tula, A., Morimoto, C.H., Ranvaud, R.D.: A mathematical model of saccadic reaction time as a function of the fixation point brightness gain. Attention, Perception, & Psychophysics **77**(6), 2153–2165 (2015). DOI 10.3758/s13414-015-0902-9. URL https://doi.org/10.3758/s13414-015-0902-9
6. Dvornychenko, V.: Bounds on (deterministic) correlation functions with application to registration. Pattern Analysis and Machine Intelligence, IEEE Transactions on **5**(2), 206–213 (1983)
7. Ebisawa, Y.: Improved video-based eye-gaze detection method. IEEE Transactions on Instrumentation and Measurement **47**(4), 948–955 (1998). DOI 10.1109/19.744648
8. Ebisawa, Y., i. Satoh, S.: Effectiveness of pupil area detection technique using two light sources and image difference method. In: Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Societ, pp. 1268–1269 (1993). DOI 10.1109/IEMBS.1993.979129
9. Fitzgibbon, A., Pilu, M., Fisher, R.B.: Direct least square fitting of ellipses. Pattern Analysis and Machine Intelligence, IEEE Transactions on **21**(5), 476–480 (1999)
10. Fuhl, W., Kübler, T., Sippel, K., Rosenstiel, W., Kasneci, E.: Excuse: Robust pupil detection in real-world scenarios. In: G. Azzopardi, N. Petkov (eds.) Computer Analysis of Images and Patterns: 16th International Conference, CAIP 2015, pp. 39–51. Springer International Publishing, Valletta, Malta (2015). DOI 10.1007/978-3-319-23192-1_4. URL http://dx.doi.org/10.1007/978-3-319-23192-1_4
11. Fuhl, W., Santini, T.C., Kübler, T., Kasneci, E.: ElSe: Ellipse Selection for Robust Pupil Detection in Real-world Environments. In: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, ETRA '16, pp. 123–130. ACM, New York, NY, USA (2016). DOI 10.1145/2857491.2857505. URL http://doi.acm.org/10.1145/2857491.2857505
12. George, A., Routray, A.: Fast and accurate algorithm for eye localisation for gaze tracking in low-resolution images. IET Computer Vision **10**(7), 660–669 (2016). DOI 10.1049/iet-cvi.2015.0316
13. Hansen, D.W., Ji, Q.: In the eye of the beholder: A survey of models for eyes and gaze. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(3), 478–500 (2010). DOI 10.1109/TPAMI.2009.30
14. Haro, A., Flickner, M., Essa, I.: Detecting and tracking eyes by using their physiological properties, dynamics, and appearance. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662), vol. 1, pp. 163–168 vol.1 (2000). DOI 10.1109/CVPR.2000.855815
15. Hennessey, C., Noureddin, B., Lawrence, P.: A Single Camera Eye-gaze Tracking System with Free Head Motion. In: Proceedings of the 2006 Symposium on Eye Tracking Research & Applications, pp. 87–94 (2006)
16. Jacob, R.J.K., Karn, K.S.: Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research pp. 573–603 (2003). URL http://www.sciencedirect.com.lt.ltag.bibl.liu.se/science
17. Ji, Q., Yang, X.: Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance. Real-Time Imaging **8**(5), 357–377 (2002). DOI 10.1006/rtim.2002.0279. URL http://www.sciencedirect.com/science/article/pii/S1077201402902792
18. Kapoor, A., Picard, R.W.: A real-time head nod and shake detector. In: Proceedings of the 2001 Workshop on Perceptive User Interfaces, PUI '01, pp. 1–5. ACM, New York, NY, USA (2001). DOI 10.1145/971478.971509. URL http://doi.acm.org.ez67.periodicos.capes.gov.br/10.1145/971478.971509
19. Kassner, M., Patera, W., Bulling, A.: Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp. 1151–1160. ACM (2014)
20. Kempinski, Y.: System and method of diagnosis using gaze and eye tracking (2016). URL https://www.google.com/patents/US20160106315. US Patent App. 14/723,590
21. Li, D., Parkhurst, D.J.: Starburst: A robust algorithm for video-based eye tracking. Elselvier Science p. 6 (2005)
22. Menon, R.V., Sigurdsson, V., Larsen, N.M., Fagerstrøm, A., Foxall, G.R.: Consumer attention to price in social commerce: Eye tracking patterns in retail clothing. Journal of Business Research **69**(11), 5008–5013 (2016). DOI 10.1016/j.jbusres.2016.04.

072. URL http://www.sciencedirect.com/science/article/pii/S0148296316302351
23. Morimoto, C., Flickner, M.: Real-time multiple face detection using active illumination. In: Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), pp. 8–13. Grenoble, France (2000). DOI 10.1109/AFGR.2000.840605
24. Morimoto, C., Koons, D., Amit, A., Flickner, M., Zhai, S.: Keeping an eye for hci. In: Computer Graphics and Image Processing, 1999. Proceedings. XII Brazilian Symposium on, pp. 171–176 (1999). DOI 10.1109/SIBGRA.1999.805722
25. Morimoto, C.H., Koons, D., Amir, A., Flickner, M.: Pupil detection and tracking using multiple light sources. Image and vision computing **18**(4), 331–335 (2000)
26. Morimoto, C.H., Koons, D., Amir, A., Flickner, M.D.: Pupil detection and tracking using multiple light sources. Image and Vision Computing **18**(4), 331–335 (2000)
27. Morimoto, C.H., Mimica, M.R.: Eye gaze tracking techniques for interactive applications. Computer Vision and Image Understanding **98**(1), 4–24 (2005). DOI 10.1016/j.cviu.2004.07.010. URL http://www.sciencedirect.com/science/article/pii/S1077314204001109. Special Issue on Eye Detection and Tracking
28. Ohno, T., Mukawa, N., Yoshikawa, A.: FreeGaze: A Gaze Tracking System for Everyday Gaze Interaction. In: Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, pp. 125–132 (2002)
29. Oliveira, D., Machín, L., Deliza, R., Rosenthal, A., Walter, E.H., Giménez, A., Ares, G.: Consumers' attention to functional food labels: Insights from eye-tracking and change detection in a case study with probiotic milk. LWT - Food Science and Technology **68**, 160–167 (2016). DOI 10.1016/j.lwt.2015.11.066. URL http://www.sciencedirect.com/science/article/pii/S0023643815303571
30. Oyster, C.: The Human Eye: Structure and Function. Sinauer Associates (1999). URL https://books.google.com.br/books?id=GLPBYgEACAAJ
31. Piumsomboon, T., Lee, G., Lindeman, R.W., Billinghurst, M.: Exploring natural eye-gaze-based interaction for immersive virtual reality. In: 2017 IEEE Symposium on 3D User Interfaces (3DUI), pp. 36–39 (2017). DOI 10.1109/3DUI.2017.7893315
32. Samadani, U., Zahid, A.B., Lockyer, J., Dammavalam, V., Grady, M., Nance, M., Scheiman, M., Master, C.L.: Eye tracking a biomarker for concussion in the paediatricpediatric population. British Journal of Sports Medicine **51**(11), A5–A5 (2017). URL http://bjsm.bmj.com/content/51/11/A5.2
33. Soler-Dominguez, J.L., Camba, J.D., Contero, M., Alcañiz, M.: A proposal for the selection of eye-tracking metrics for the implementation of adaptive gameplay in virtual reality based games. In: S. Lackey, J. Chen (eds.) Virtual, Augmented and Mixed Reality: 9th International Conference, VAMR 2017, pp. 369–380. Springer International Publishing, Vancouver, CA (2017). DOI 10.1007/978-3-319-57987-0_30. URL https://doi.org/10.1007/978-3-319-57987-0_30
34. Sony, C.: PlayStation Eye Camera. [Online, 14/jan/2014] (2014). URL http://us.playstation.com/ps3/accessories/playstation-eye-camera-ps3.html
35. Świrski, L., Bulling, A., Dodgson, N.: Robust real-time pupil tracking in highly off-axis images. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 173–176. ACM (2012)
36. Zhu, Z., Ji, Q.: Robust real-time eye detection and tracking under variable lighting conditions and various face orientations. Computer Vision and Image Understanding **98**(1), 124–154 (2005). DOI 10.1016/j.cviu.2004.07.012. URL http://www.sciencedirect.com/science/article/pii/S1077314204001158