Computer Vision and Image Understanding 149 (2016) 157–170

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Model-based head pose-free gaze estimation for assistive communication

Stefania Cristina*, Kenneth P. Camilleri

Department of Systems and Control Engineering, University of Malta, Msida, MSD 2080, Malta

ARTICLE INFO

Article history:
Received 15 April 2015
Accepted 18 February 2016

Keywords:
Eye-gaze tracking
Videooculography
Head pose-free
Model-based
Augmentative and alternative communication
Cerebral palsy

ABSTRACT

The significance of employing video-based eye-gaze tracking as an assistive tool has long been recognised, especially in the domain of human–computer interaction to assist physically challenged individuals in operating a computer by the eye movements alone. Nonetheless, several operating conditions typically associated with existing eye-gaze tracking methods, relating to constraints on the head movement and prolonged user-calibration prior to gaze estimation, need to be alleviated in order to better assist individuals with motor disabilities. In this paper, we propose a method that is based on a cylindrical head and spherical eyeballs model to estimate the three-dimensional eye-gaze under free head movement from a single camera integrated into a notebook computer, alleviating any assumptions of stationary head movement without requiring prolonged user co-operation prior to gaze estimation. The validity of the proposed method has been investigated on a publicly available data set and real-life data captured through the voluntary collaboration of a group of normal subjects and a person suffering from cerebral palsy.

1. Introduction

Ever since the pioneering eye movement studies revealed an important link between the human eye movements and visual attention, considerable research work has been directed towards the development of different techniques that permit the measurement of these movements [1]. Methods for eye movement measurement have improved extensively over the years, eventually leading to the concept of videooculography (VOG) whereby the eye movements are captured in a stream of image frames by means of digital cameras without direct contact with the user [2]. Owing to its non-intrusive nature, the use of eye-gaze tracking by VOG has been applied to a host of domains, such as for human–computer interaction where eye-gaze tracking technology provides an alternative communication channel for individuals suffering from motor disabilities, permitting the user to operate a computer via the eye movements alone [3].

The field of eye-gaze tracking by VOG has been receiving increasing interest over the years, leading to the development of various methods that seek to estimate the eye-gaze reliably [2]. To achieve better estimation accuracy, state-of-the-art video-based eye-gaze tracking methods often impose specific conditions to operate, requiring the user to maintain a stationary frontal head pose during tracking [4–15] or the acquisition of close-up eye region images [8,9]. As will be further elaborated in Section 2, the inclusion of a personal calibration procedure for every individual user is generally required prior to the estimation of eye-gaze, permitting the computation of an image-to-screen mapping relationship to estimate a point-of-regard (POR) on a two-dimensional monitor screen [4,13–20] or the calculation of a set of user-dependent eyeball model parameters in order to estimate a gaze vector in three-dimensional space [5,9,12,21–26] from image information. Furthermore, the use of stereo-vision is also often proposed in order to estimate the eye-gaze from three-dimensional information of the eye region and face features [27–29].

Nonetheless, in the context of augmentative and alternative communication (AAC), several conditions typically associated with existing eye-gaze tracking methods need to be alleviated in order to assist individuals with motor disabilities. Indeed, it is desirable to compensate for the involuntary head movements associated with certain movement disorders, such as cerebral palsy [30], while the acquisition of close-up eye region images may not always be possible especially due to physical limitations in approaching the acquisition hardware emanating from the use of wheelchairs or buggies [31]. The use of multi-camera setups for stereo-vision [27–29] eliminate the possibility to estimate the eye-gaze on off-the-shelf portable devices with in-built imaging hardware, which may be mounted and used on a wheelchair.

* Corresponding author.
E-mail addresses: stefania.cristina@um.edu.mt (S. Cristina), kenneth.camilleri@um.edu.mt (K.P. Camilleri).

http://dx.doi.org/10.1016/j.cviu.2016.02.012
1077-3142/

Furthermore, the inclusion of a user-calibration procedure may hinder the estimation of eye-gaze entirely as reported by an assessment of eye-gaze tracking technology for individuals with profound and learning disabilities, which indicated difficulties in achieving correct calibration in cases where maintaining the correct head orientation of the participant for the duration of the calibration session was not accomplished [31]. This calls for methods that reduce the required calibration time and effort prior to gaze estimation.

In view of these challenges, we propose a method to estimate the eye-gaze from a single camera integrated into a notebook computer under free head movement, requiring minimal user co-operation prior to gaze estimation while the user sits at a distance from the camera. Under stationary head conditions the problem of gaze estimation is simplified because the appearance of the face region does not change during tracking, while the change in appearance of the iris and pupil regions corresponds with the direction of eye-gaze. In comparison, head motion changes the appearance of the face and eye regions significantly, resulting in many possible combinations of face appearance and eyeball orientation which correspond to the same eye-gaze direction. In order to model the change in appearance due to head rotation of a reference eye region inside the image space, we augment the cylindrical head model proposed by Xiao et al. [32] and later used by Valenti et al. [16] with spherical models of the human eyeball, hence proposing an extension to the approach in [16] that permits the estimation of three-dimensional gaze vectors in two consecutive stages; by first estimating the head pose, therefore compensating for distortion in the appearance of the eye region image resulting from head rotation, and subsequently the eyeball orientation angles. We investigate the validity of our method on images from the Columbia Gaze Data Set [33], due to the availability of ground truth head pose and eye-gaze information, and on real-life data captured through the voluntary collaboration of a group of normal subjects and a person suffering from cerebral palsy, comparing the achieved results with relevant state-of-the-art methods in the literature.

This paper is organised as follows. Section 2 reviews and discusses related state-of-the-art methods in the literature. An overview of the proposed approach together with a definition of the notation is provided in Section 3. Section 4 details the implementation of the proposed method. The experimental results are presented and evaluated in Section 5 and compared with the state-of-the-art in Section 6. Finally, Section 7 draws the concluding remarks and outlines suggestions for future work.

2. Related work

Eye-gaze tracking by VOG may be achieved by active or passive techniques [2]. Active VOG facilitates gaze estimation by illuminating the face and the eyes with infra-red (IR) illumination in order to brighten the pupil region and produce glints that reflect off the lens and cornea. This permits the estimation of gaze direction according to the relative positioning between the corneal glints and the pupil centre [2]. The use of specialised illumination sources and imaging hardware, however, reduces the suitability of active VOG in less constrained conditions where factors, such as the lighting conditions and head movement, may affect the visibility of the corneal glints, the image quality or the pupil size and brightness [34]. In the absence of IR illumination, passive VOG exploits the surrounding illumination alone to estimate the eye-gaze [2]. This allows for increased portability in different environments and the capability to estimate the eye-gaze on off-the-shelf devices comprising integrated hardware without requiring further hardware modification. In view of the advantages associated with passive VOG, our approach proposes to estimate the eye-gaze from images captured by a single webcam integrated inside a notebook computer, under ambient lighting.

Passive VOG methods may be further subdivided into two categories. The first category comprises methods that estimate a two-dimensional POR following the computation of an image-to-screen mapping relationship [4,13–20]. The POR estimates computed by these methods are inherently constrained to a two-dimensional plane, such as a monitor screen, which highly limits the applicability of these methods. The second category comprises methods that compute a three-dimensional gaze vector that coincides with the optical or visual axis of the eyeball by modelling the anatomical structure of the human eyeball [5,9,12,21–26]. The POR may be subsequently estimated by intersecting the computed gaze vector with the surface of surrounding objects, in the knowledge of their position in three-dimensional space, such as a monitor screen or objects of interest within assistive environments. Hence, given the higher versatility of the second category of methods, our approach aims for three-dimensional gaze estimation.

The three-dimensional gaze estimation methods generally model the anatomical structure of the human eyeball, characterising the model parameter values through a personal calibration procedure for every individual user [9,22–24,27,28]. Stereo-vision based methods are often proposed in order to determine the model feature coordinates, such as the eyeball and iris centres, in three-dimensional space [27–29]. This information permits the eyeball model to be positioned within a dense [29] or sparse [27,28] three-dimensional reconstruction of the face features computed by stereo-vision, following which the gaze direction may be determined from the vector that joins the three-dimensional positions of the eyeball and iris centre points [27–29]. The use of stereo-vision, nonetheless, necessitates intrinsic and extrinsic camera calibration in order to compute a projective transformation that relates the world coordinates to image coordinates. The use of a multi-camera setup, in turn, eliminates the possibility of estimating the eye-gaze on off-the-shelf devices with in-built imaging hardware. To alleviate this problem, single camera methods have been proposed that acquire a three-dimensional reconstruction of the face features via monocular techniques, such as shape-from-motion factorisation [23] and three-dimensional model-fitting [26], but which nonetheless require an estimation of the intrinsic camera parameters in order to position the spherical eyeball model within the three-dimensional reconstruction [23,26]. Hence, while many of these approaches permit three-dimensional gaze estimation under free head movement, the calibration effort required to calculate the camera and eye model parameters can significantly limit their usability and portability. Within an AAC scenario, personal calibration may prove difficult for users with motor disabilities affected by involuntary head and face movements, while the requirement for camera calibration hinders immediate use of off-the-shelf hardware, such as portable devices with integrated webcam that may be mounted and used on a wheelchair.

In light of these challenges, we propose a method to estimate the eye-gaze in three-dimensional space under free head movement from a single integrated camera inside a notebook computer. The eye-gaze is estimated without necessitating prior camera calibration and limiting user calibration to frontal eye and head pose detection, taken as reference according to which all subsequent gaze directions are measured. We build on the cylindrical head model proposed by Xiao et al. [32] and later used by Valenti et al. [16], by augmenting their head model to include spherical models of the human eyeballs. While spherical eye models have been considered by other methods in the literature [15,23,26–29], in the absence of camera calibration and three-dimensional data acquired by stereo-vision [27–29] or the use of corneal reflections [15] for gaze estimation, our cylindrical head and spherical eyeballs model allows us to extract the gaze information by projecting the

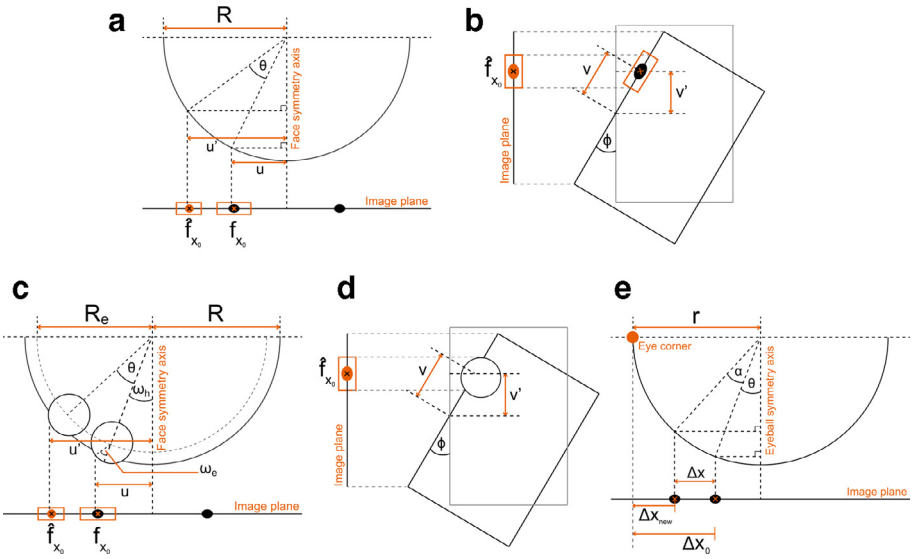

Fig. 1. We propose to augment the cylindrical model (a) and (b) in [16] with spherical models of the eyeballs (c)–(e) for three-dimensional gaze estimation. The reference eye region images, $f_{\mathbf{x}_o}(u', v')$, re-projected on the cylindrical model alone (a) and (b) and the proposed cylindrical head and spherical eyeballs model of radii R and r respectively (c)–(e), are rotated by the head yaw angle θ (a, c) and pitch angle ϕ (b, d). The displacement, $\Delta\mathbf{x}$, between the rotated reference and observed iris centres is projected on the spherical eyeballs model (e).

change in appearance due to head rotation of a reference eye region inside the image space. The introduction of the spherical eye models permits an extension of the approach proposed by Valenti et al. [16], which two-dimensional PORs are calculated following image-to-screen calibration, by estimating three-dimensional gaze vectors based on the image displacement between a rotated reference eye region image and a newly observed image. We characterise our model parameters with anthropometric measurements of the human eyeball and investigate the sensitivity of the gaze estimation error by considering the mean, and the lower and upper bound anthropometric values exhibited by the population. Furthermore, we investigate any improvement in the gaze estimation accuracy by the proposed cylindrical head and spherical eyeballs model over the cylindrical model alone [16], where the eye regions are flat on the cylindrical surface, by quantifying the error for both models over the same ranges of anthropometric values. The results achieved by the proposed method are compared and contrasted with relevant state-of-the-art approaches in the literature.

3. Overview of the algorithm

This section gives an overview of the proposed algorithm together with a definition of the notation that will be referred to in the subsequent sections. Without loss of generality, the following notation applies to both the left and right eye regions respectively.

We define cylindrical head and spherical eyeball models to estimate the respective rotation angles in three-dimensional space, as shown in Fig. 1(c)–(e) where the spherical eyeballs are embedded on a cylindrical head surface. The head model rotates a frontal reference image by the estimated head yaw, pitch and roll angles, resulting in a transformed image with known head pose and eyeball rotation. This transformation accounts for the changes in the appearance of the eye region under head rotation. Following image transformation by the head model, the reference image is then aligned to an observed eye region image to account for any head translation in the horizontal and vertical directions. The planar displacement between the iris centres inside the observed image and the reference image with known eyeball rotation is finally projected onto a spherical eyeball model to estimate the new eyeball orientation. Relevant to our method is the approach in [16], which estimates the head pose by employing a cylindrical head model as well, fit to the face region by sampling the underlying image in a grid-like structure. The head pose estimate is required to warp a rotated eye region back to a frontal pose, hence satisfying the constraints of an eye localisation algorithm based on the detection of isophotes which necessitates a frontal eye pose to identify circular patterns reliably. In comparison with the method in [16], we choose to rotate the reference frontal eye image to a known head pose because this image holds the most visible information, hence eliminating the possibility of holes in the image that would correspond to occluded patches in the rotated head pose image. The approach in [16] subsequently proceeds with a user-calibration procedure that is required to map the image displacement between a rotated pupil centre and a reference centre estimated from a frontal view onto screen coordinates. While this deviates from our work, we have chosen to consider the cylindrical model in [16] as well, as illustrated in Fig. 1(a) and 1(b) where the eye regions are flat on the cylindrical surface, in order to compare the gaze estimation results with those obtained through our proposed model as detailed further in Section 5.

Specifically, the proposed algorithm, therefore, initially searches for a frontal eye and head pose by checking the frontality of an upright face region inside the image space, as will be outlined in Section 4. Upon detection of an upright and frontal eye and head pose, the frontal eye regions $f_{\mathbf{x}_o}$, where $\mathbf{x}_o$ denotes the iris centre coordinates, are localised while the head yaw θ, pitch ϕ and roll γ angles, and the eyeball yaw α and pitch β angles are initialised to zero. The torsional component of the eyeball during eye movement is assumed not to contribute to the estimation of gaze direction [35] and hence our model reduces this degree-of-freedom for simplicity. Each frontal eye region image $f_{\mathbf{x}_o}$ around $\mathbf{x}_o$, is assumed to be a projection of the true eye region onto a plane from a sphere of known radius, r.

Let us consider the camera axes, (c_x, c_y, c_z), such that c_x is along the columns of the sensor, c_y is along the rows of the sensor, and c_z is orthogonal to both and directed into the sensor. The sensor is centred at the origin of this system of axes. With the foregoing initialisation of frontal eye and head pose, the axes of the head and its cylindrical model, and the axes of the eyes and their spherical

model are aligned to the camera axes. Therefore, following this initialisation the angles θ, ϕ and γ of the head are updated at every image frame according to the estimated head pose. By projecting pixels (x, y), in the reference image $f_{\mathbf{x}_o}$, orthographically onto the surface of the cylindrical head and spherical eyeballs model, the position on the model will be $\mathbf{u} = (u, v, z)$ where $u = x$ and $v = y$, while the depth z does not need to be taken into consideration because the projection is orthographic:

$$f_{\mathbf{x}_o}(\mathbf{u}) = P(f_{\mathbf{x}_o}(\mathbf{x})) \tag{1}$$

where P represents the orthographic projection function onto the model, $\mathbf{u} \in \Re^3$ are coordinates on the model surface, while u and v are along axes c_x and c_y respectively. This reference image projected onto the model is then transformed by a head rotation $\mathbf{R}_h$ according to the head pose $O_h = (\theta, \phi, \gamma)$,

$$f_{\mathbf{x}_o}(\mathbf{u}') = \mathbf{R}_h(f_{\mathbf{x}_o}(\mathbf{u}); O_h) \tag{2}$$

Fig. 1 illustrates the transformation according to rotation angles θ and ϕ.

This head rotated eye patch is then projected back onto the image plane according to (3) followed by the interpolation operation in (4) which gives the transformed reference image, $\hat{f}_{\mathbf{x}_o}(i, j)$,

$$f_{\mathbf{x}_o}(u', v') = P(f_{\mathbf{x}_o}(\mathbf{u}')) \tag{3}$$

$$\hat{f}_{\mathbf{x}_o}(i, j) \leftarrow f_{\mathbf{x}_o}(u', v') \tag{4}$$

Now let the iris centre in $\hat{f}_{\mathbf{x}_o}(\mathbf{i})$ be defined by $\mathbf{x}_o$. To align the rotated reference image with a newly observed eye region image, $g_{\mathbf{x}}$, having an iris centre at $\mathbf{x}_{\text{new}}$ and detected during tracking, displacements $\Delta\mathbf{x}_o$ and $\Delta\mathbf{x}_{\text{new}}$ are calculated with respect to a reference point inside each of the rotated reference and observed images respectively. The designated reference points are the inner eye corners, chosen due to their stability in the presence of facial movement and good visibility at different head rotation angles. The displacement,

$$\Delta\mathbf{x} = \Delta\mathbf{x}_o - \Delta\mathbf{x}_{\text{new}} \tag{5}$$

is finally projected onto the surface of a sphere with known radius r such that the eyeball rotation angles α and β may be estimated. Fig. 1(e) illustrates the eyeball rotation by the yaw angle α, and similarly for the pitch angle β.

4. Algorithm implementation

Following the overview of the proposed algorithm in Section 3, the next sections detail the implemented methods to extract the required information from the image frames and eventually estimate the eyeball rotation angles.

4.1. Face feature detection and tracking

There exist different methods for head pose estimation that have been proposed over the years, several of which necessitate specific requirements such as the availability of stereo images, the collection of training data that serves to train a classifier in estimating the head pose from face images, or an accurate initialisation for a complex face model to be fit to the face region [36]. We chose to opt for a geometric method that exploits the relative configuration of facial features, namely the eyes, nose and mouth, in order to estimate the head pose without necessitating prior user-calibration or a training stage. This geometric method will be detailed further in Section 4.2.1, however detection and tracking of face features is required in advance.

4.1.1. Face feature detection

Detection and tracking of the eyes, nose and mouth features through a stream of image frames follows the method in [37] due to its applicability to our real-time application. To reduce the computational cost and the occurrence of false positive detections in searching for the facial features over an entire image frame, the method in [37] detects the face region first such that the search for the features is then constrained to within the bounding box which encloses the face region. The face and facial features are rapidly detected by the Viola–Jones algorithm [38], which detects features of interest by sliding rectangular windows of Haar-like operators over an image frame, subtracting the underlying image pixels that fall within the shaded regions of the Haar-like operators from the image pixels that fall within the clear regions. In our work, detection of the face and facial features is carried out by freely available cascades of classifiers that are included with the OpenCV library [39]. These classifiers had been previously trained on a wide variety of training images in order to ensure that detection generalises well across different users.

At this stage, a check for the frontality of the face region is carried out by calculating the distance between the position of the nose region and the centroid of the triangle that joins the two eye regions with the mouth. If this distance does not exceed a predefined threshold value, the head pose is taken to be frontal hence permitting the frontal eye region images, $f_{\mathbf{x}_o}$, to be captured and the head and eyeball rotation angles to be initialised to zero.

4.1.2. Face feature tracking

Following the initial detection of the facial features, a tracking stage updates their initial positions over successive image frames to account for their displacement during head movement. While performing detection on a frame-by-frame basis would suffice to follow the displacement of the facial features over a sequence of image frames, such a solution would be sub-optimal in terms of computational cost for a real-time application. Therefore, the facial features are tracked by template matching, constraining the search area for each feature inside the next image frame based on the last known position inside the previous image frame.

A template image of every feature of interest is captured and stored following earlier detection by the Viola–Jones algorithm. Template matching is then carried out inside windows of fixed size centred around the last known positions of the features of interest, using the normalised sum of squared differences (NSSD) as a measure of similarity. The new positions of the features of interest are specified by locations inside the search image that give the minimum NSSD values after template matching.

4.2. Iris centre localisation

The feature detection and tracking stage described earlier provides a rough estimate of the movement of the eye regions through a stream of image frames. In order to refine on this estimate and hence extract a better representation of the trajectory followed by the eyes through an image sequence, the iris or pupil centre coordinates are typically localised following detection of the eye regions [2]. Due to the low resolution of the eye region images inside the image space as the user sits at a distance from the camera, we choose to localise the centre coordinates of the iris regions rather than the pupil as these can be detected more reliably since they occupy a larger area inside the eye regions.

There exist different methods that permit localisation of these centre coordinates inside an image frame. Nonetheless, not all methods are suitable to localise the iris region from low-resolution eye region images, especially if the fine boundaries that separate different components of the eye need to be clearly distinguishable

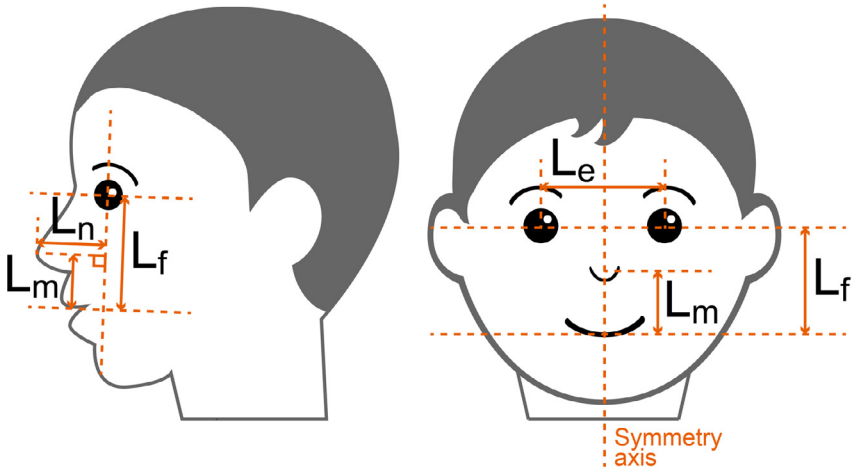

Fig. 2. The head pose is estimated from the relative distances between different facial features, where L_e denotes the distance between the eye regions, L_m and L_f denote the nose-to-mouth and eye-to-mouth distances respectively, while L_n defines the distance between the nose base and nose tip.

[9,40–42]. In view of these challenges, we follow the method proposed in our previous work [17] which localises the iris centre coordinates via a Bayes' classifier that segments the iris region based on its appearance rather than the shape. The Bayes' classifier is trained to classify between iris and non-iris pixels during an offline training stage, based on the red channel values of the pixels in the RGB colour space. During tracking, pixels residing inside the eye region are then classified as belonging to the iris region if the likelihood of their intensity value exceeds a pre-defined threshold value.

This method was found to be suitable in detecting the iris region inside low-resolution images at different angles of head and eyeball rotation and partial occlusion by the eyelids. The advantages associated with this method are related to its independency from geometrical information, performing localisation via statistical colour modelling instead. To alleviate the main susceptibility of this method to illumination variations, the Bayes' classifier in our work was trained on iris and non-iris pixels acquired under different illumination conditions. Localisation of the iris centre permits the initialisation of coordinates, $\mathbf{x}_o$, inside the frontal reference eye region image, $f_{\mathbf{x}_o}$, and localisation of coordinates, $\mathbf{x}$, in the observed image frames, $g_{\mathbf{x}}$.

4.2.1. Gaze estimation

Having determined the positions of the facial features and the iris centre coordinates, the final stage seeks to estimate the head pose and eyeball rotation. The head pose and the eyeball orientation are closely related to one another and collectively contribute towards the estimation of gaze: the head pose typically defines a coarse estimate of the gaze direction, while the eyeball orientation refines upon this estimate to define the gaze direction at a finer level.

The estimation of head pose follows the method in [37], which fits a generic face model to the facial features detected earlier and estimates the head pose in three-dimensions according to the relative distances between these features. Simplifying the face to a plane that passes through the eyes and the mouth, the symmetry axis of the face is first determined by joining the mouth to the midpoint of the line that connects the eye regions together. The nose base is taken to reside upon this symmetry axis and is subsequently located by calculating the ratio $R_m = \frac{L_m}{L_f}$, where the lengths L_m and L_f denote the nose-to-mouth and eye-to-mouth distances respectively, as illustrated by Fig. 2. Joining this nose base to the previously detected position of the nose tip allows the computation of the facial normal in three-dimensional space as detailed in [37], and hence the estimation of the head yaw θ, pitch ϕ and roll γ angles.

Following the estimation of the head rotation, we seek to estimate the rotation of the eyeball as the second component which contributes to the calculation of gaze direction. The projection of the eye inside the image space is affected by changes in head pose, such that even if the eyeball itself remains stationary with respect to the head, the overall appearance of the eye region inside an image frame changes according to the head rotation.

Therefore, following re-projection of the frontal reference image onto a cylindrical head and spherical eyeball model to produce the re-projected image, $f_{\mathbf{x}_o}(\mathbf{u})$, the cylinder and sphere having radii R and r respectively are rotated according to the estimated head rotation angles to produce the newly rotated image $f_{\mathbf{x}_o}(\mathbf{u}')$. The new image coordinates $\mathbf{u}'$ following rotation of the model by angles θ and ϕ are calculated as follows,

$$u' = \begin{cases} R\sin(\omega_h + \theta) & \text{if } u' \text{ is on the cylindrical part of the model} \\ r\sin(\omega_e + \theta) + R_e\sin(\omega_h + \theta) & \text{if } u' \text{is on the spherical part of the model delineated by the eye corners} \end{cases} \tag{6}$$

$$v' = \begin{cases} v\cos(\phi) & \text{if } v' \text{is on the cylindrical part of the model} \\ r\cos(\omega_e + \phi) + v\cos(\phi) & \text{if } v' \text{is on the spherical part of the model delineated by the upper and lower eyelids} \end{cases} \tag{7}$$

where ω_h is the angle which every pixel inside image $f_{\mathbf{x}_o}(\mathbf{u})$ makes with the symmetry axis of the face and ω_e is the angle which every pixel on the spherical parts of the model makes with the symmetry axis of the eyeballs, as illustrated in Fig. 1(c). Radius R_e, as illustrated in Fig. 1(c), determines the protrusion of the spherical eyeball inside the cylindrical head model and was set to $R_e = (R - 3.8\text{ mm})$ following anthropometric measurements [43]. The horizontal coordinates u and u' are defined with respect to the symmetry axis of the face, while the vertical coordinates v and v' are defined with respect to the nose region. To estimate a value

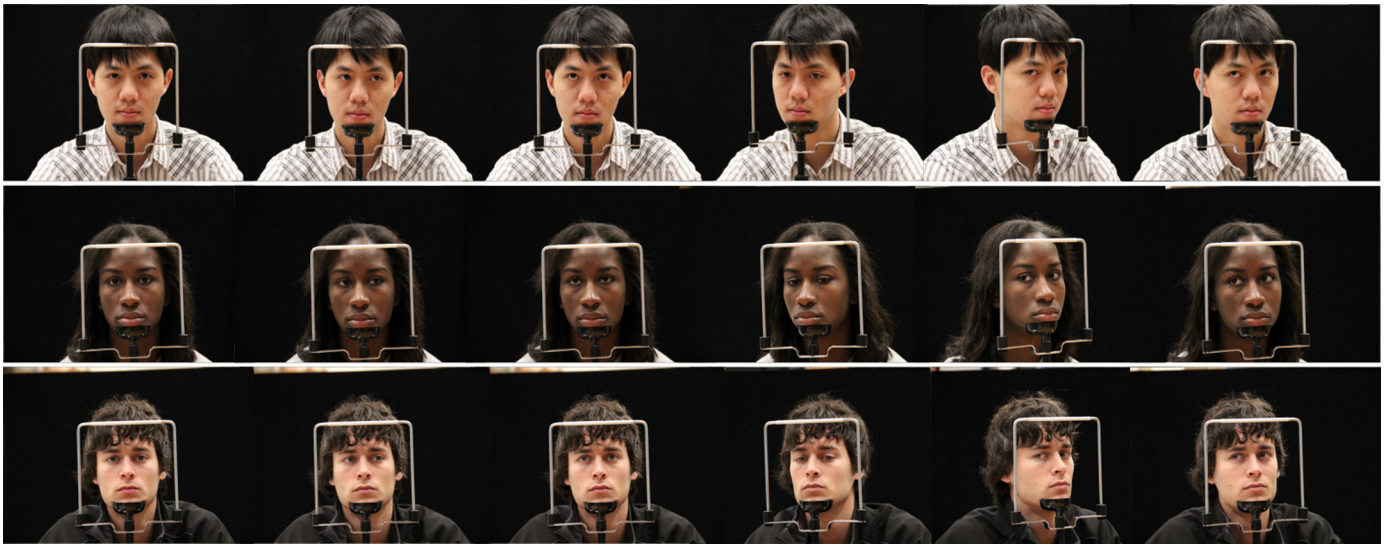

Fig. 3. Images of male and female subjects from the Columbia Gaze Data Set [33] having different gaze directions and head poses, and different ethnic backgrounds, were considered for evaluation purposes.

for R, a Bayes' classifier, which follows the theory outlined in [17], is trained during an offline training stage to identify skin pixels by their intensity value and is used to segment the skin region inside the initial face bounding box that is detected following the method in Section 4.1. The radius R is then estimated as half the width of the blob of pixels in the resulting binary image.

Following rotation by θ and ϕ, the re-projected image is further rotated by the head roll angle, γ:

$$u' = (u' - u'_c)\cos(\gamma) - (v' - v'_c)\sin(\gamma) + u'_c \tag{8}$$

$$v' = (u' - u'_c)\sin(\gamma) + (v' - v'_c)\cos(\gamma) + v'_c \tag{9}$$

where coordinates (u'_c, v'_c) denote the centre of image rotation.

Prior to the estimation of the gaze angles, the rotated reference image $f_{\mathbf{x}_o}(\mathbf{u}')$ and the observed eye region image $g_{\mathbf{x}}$ are aligned by corresponding the horizontal and vertical positions of the eye corners after these have been detected by the method in [44]. The displacement between the iris centres in these two images, defined by Eq. (5), is then projected to a spherical surface such that the eyeball yaw and pitch angles are estimated as follows,

$$\alpha = \sin^{-1}\left(\frac{x}{r}\right) - \theta \tag{10}$$

$$\beta = \sin^{-1}\left(\frac{y}{r}\right) - \phi \tag{11}$$

where x and y denote the x- and y-coordinates of the iris centre inside the observed image $g_{\mathbf{x}}$ and r refers to the eyeball radius in pixel coordinates. Fig. 1(e) illustrates the eyeball rotation by the yaw angle, α. Following the estimation of the iris radius r_i in pixel units inside the image space, estimated by fitting a circle to the frontal iris region in $f_{\mathbf{x}_o}$ via the Hough Transform algorithm [45], a value for r in pixels is estimated as follows,

$$r = r_i\left(\frac{12.2}{5.9}\right) \tag{12}$$

Hence, Eq. (12) estimates a value for r by relating the ratio between the iris and eyeball radii in millimetres, specified in the literature as 5.9 [46] and 12.2 mm [47] respectively, with the ratio between the corresponding iris and eyeball radii in pixel units.

The final eye-gaze direction is estimated as a function of the head rotation angles, θ and β, and eyeball rotation angles, α and ϕ, in the horizontal and vertical directions as follows,

$$\zeta = \theta + \alpha \tag{13}$$

$$\psi = \beta + \phi \tag{14}$$

where ζ and ψ denote the estimated gaze yaw and pitch angles respectively, under the simplifying assumption often made in the literature [27,28] that the estimated gaze vector coincides with the optical axis of the human eye, which joins together the eyeball and iris centre points, rather than the visual axis.

5. Experimental results and discussion

For evaluation purposes, the method proposed in this paper has been tested on a publicly available gaze dataset [33], having the availability of ground truth data, and on real-life video sequences captured through the voluntary collaboration of normal subjects and a person suffering from cerebral palsy in order to investigate the performance of our method in the context of AAC. The following sections describe and discuss the experimental results obtained through these evaluation procedures.

5.1. Evaluation on test data

To evaluate the gaze estimation accuracy of our method, we considered images of male and female subjects from the Columbia Gaze Data Set [33] captured at different gaze directions and head poses, and having different ethnic backgrounds as shown in Fig. 3. The subjects inside this dataset are seated in a fixed position at a distance of 2.5 m from a grid of visual stimuli that serve to induce horizontal eye rotations of 0°, ±5°, ±10° and ± 15°, and vertical eye rotations of 0° and ± 10°. In order to produce different head rotation angles of 0°, ±15° and ± 30°, the camera is rotated around a radius of 2 m while the subjects hold a frontal head pose aided by a chin-rest that is adjusted to position the eyes at 70 cm above the floor. The images inside the dataset have been captured at a resolution of 5184 × 3456 pixels, however owing to the real-time requirements of our eye-gaze tracking application we reduced the images to a resolution of 519 × 346 pixels. While this served to reduce the computational cost at every image frame, hence ensuring that a stream of image frames at this resolution could be

Table 1
Mean values and ranges of considered anthropometric measurements of the human eye, for the eyeball protrusion and the eyeball and iris radii as specified in the literature [43,46,47].

	Eyeball protrusion [43]	Eyeball radius [47]	Iris radius [46]
Mean/mm	3.80	11.99	5.9
Lower bound/mm	0.00	10.68	5.10
Upper bound/mm	9.40	13.30	6.50

Table 2
Mean absolute error (MAE) and standard deviation (SD) of the gaze estimates in visual angle calculated using ground truth head pose and eye corner positions, for the proposed model characterised by population mean, and lower and upper bound anthropometric measurements of the human eyeball.

	YAW			PITCH		
	Left eye (MAE(°), SD(°))	Right eye	Combined	Left eye (MAE(°), SD(°))	Right eye	Combined
Mean values	(7.40, 4.00)	(8.16, 3.56)	(7.09, 3.31)	(5.64, 3.98)	(4.42, 3.23)	(4.40, 3.15)
Lower protrusion	(5.17, 3.67)	(5.44, 3.48)	(3.47, 2.95)	(5.64, 3.98)	(4.42, 3.23)	(4.40, 3.15)
Upper protrusion	(15.03, 4.67)	(16.22, 4.03)	(15.73, 3.51)	(5.64, 3.98)	(4.42, 3.23)	(4.40, 3.15)
Lower eyeball radius	(9.94, 4.63)	(10.99, 4.16)	(10.18, 3.66)	(5.84, 4.03)	(4.54, 3.29)	(4.39, 3.13)
Upper eyeball radius	(5.86, 3.73)	(6.38, 3.36)	(5.02, 3.27)	(5.55, 3.94)	(4.42, 3.20)	(4.49, 3.16)
Lower iris radius	(5.77, 3.72)	(6.26, 3.36)	(4.96, 3.34)	(5.53, 3.93)	(4.44, 3.20)	(4.55, 3.17)
Upper iris radius	(8.97, 4.45)	(9.94, 3.99)	(9.03, 3.52)	(5.80, 4.02)	(4.51, 3.27)	(4.39, 3.14)
Optimal values	(5.90, 4.04)	(5.56, 3.64)	(5.11, 3.70)	(5.53, 3.93)	(4.54, 3.17)	(4.71, 3.21)

processed in real-time by the proposed method, we faced the additional challenge of estimating the head pose and eyeball orientation from lower resolution eye region images. For evaluation purposes, we selected an ethnically diverse subset of subjects in the dataset that do not wear prescription glasses or exhibit changes in facial expression, since these are not presently handled by the proposed method, or display evident changes in head pose which are not otherwise reflected in the ground truth information provided with the dataset.

We are interested in investigating the achievable gaze estimation accuracy following the estimation of model parameters from image information and anthropometric measurements of the human eye rather than through a user-calibration procedure, to find out whether calibration may be reduced to the detection and initialisation of a frontal head pose as proposed in our work. To this end, the eye-gaze is initially estimated using ground truth head pose information and hand-labelled eye corner positions together with population mean values of the human eye, in order to quantify the gaze estimation error that is mainly attributed to the calculation of the eye and head model parameters. The head pose and eye corner positions are subsequently estimated as described in Section 4 to investigate their contribution to the gaze estimation error. Furthermore, the sensitivity of the gaze estimation to different anthropometric measurements of the human eye was evaluated by considering the range of parameter values exhibited by the population, permitting quantification of the estimation error corresponding to the lower and upper bound human eye measurements. The same evaluation procedures employing ground truth information and ranges of anthropometric measurements were also repeated using a cylindrical model alone, where the eye regions are flat on the cylindrical surface as proposed in [16], in order to compare the resulting gaze estimation error with the results obtained by our cylindrical head and spherical eyeballs model. The ranges of considered anthropometric measurements according to the literature are provided in Table 1.

In order to quantify the error for each eyeball separately, the horizontal ground truth rotation angle for each eyeball was calculated to account for the effect of vergence in the eye movement. Hence, the rotation angle of each eyeball as the subject gazes at

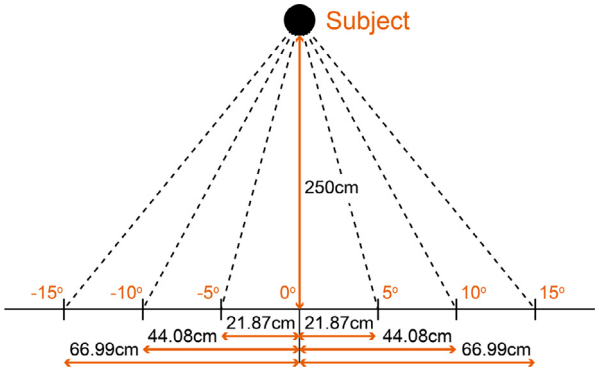

Fig. 4. Horizontal distance values for every stimulus on the grid of visual stimuli used in the collection of ground truth data for the Columbia Gaze Data Set [33].

different visual stimuli is defined by,

$$\tau = \tan^{-1}\left(\frac{d + d_{\text{mid}}}{250}\right) \tag{15}$$

where d_{mid} denotes the frontal distance between each eye region and the symmetry axis of the face inside the world space, manually extracted for each subject, while d denotes the distance between every visual stimulus and the centre of the grid calculated as follows,

$$d = 250\tan(\sigma) \tag{16}$$

where σ denotes the angle which each visual stimulus makes with the grid centre. The horizontal distance values for every visual stimulus are shown in Fig. 4.

The mean absolute error (MAE) and standard deviation (SD) values in visual angle calculated with and without using ground truth head pose and eye corner positions are given in Tables 2 and 3 calculated for different anthropometric measurements of the human eyeball, and similarly the results in Table 4 using a cylindrical model alone and ground truth information. The full yaw and pitch data for the chosen participants, computed using population mean values for the model parameters, are presented in Table 5. The

Table 3
Mean absolute error (MAE) and standard deviation (SD) of the gaze estimates in visual angle calculated using estimated head pose and tracked eye corner positions, for the proposed model characterised by population mean, and lower and upper bound anthropometric measurements of the human eyeball.

	YAW			PITCH		
	Left eye (MAE(°), SD(°))	Right eye	Combined	Left eye (MAE(°), SD(°))	Right eye	Combined
Mean values	(8.01, 4.18)	(8.55, 4.06)	(6.26, 3.59)	(6.19, 3.89)	(7.20, 3.98)	(6.23, 3.61)
Lower protrusion	(10.31, 4.77)	(8.69, 4.12)	(8.08, 4.20)	(6.18, 3.97)	(7.49, 4.04)	(6.29, 3.60)
Upper protrusion	(13.32, 4.42)	(16.59, 4.32)	(14.77, 3.95)	(6.08, 3.95)	(7.42, 4.02)	(6.35, 3.61)
Lower eyeball radius	(9.82, 4.57)	(10.65, 4.27)	(8.46, 3.98)	(6.07, 3.98)	(8.16, 4.53)	(6.47, 3.73)
Upper eyeball radius	(7.73, 4.18)	(7.22, 3.74)	(5.96, 3.49)	(5.95, 3.89)	(7.05, 3.96)	(6.10, 3.62)
Lower iris radius	(7.70, 4.17)	(6.98, 3.77)	(5.88, 3.46)	(5.90, 3.80)	(6.95, 3.88)	(6.09, 3.55)
Upper iris radius	(9.14, 4.35)	(9.86, 3.99)	(7.77, 3.58)	(6.40, 4.04)	(7.99, 4.49)	(6.56, 3.71)
Optimal values	(7.42, 4.37)	(6.56, 3.94)	(5.87, 3.73)	(5.57, 3.53)	(6.17, 3.76)	(5.49, 3.45)

Table 4
Mean absolute error (MAE) and standard deviation (SD) of the gaze estimates in visual angle calculated using ground truth head pose and eye corner positions, for the cylindrical model alone characterised by population mean, and lower and upper bound anthropometric measurements of the human eyeball.

	YAW			PITCH		
	Left eye (MAE(°), SD(°))	Right eye	Combined	Left eye (MAE(°), SD(°))	Right eye	Combined
Mean values	(9.84, 4.17)	(10.95, 3.89)	(10.12, 3.49)	(5.66, 3.98)	(4.43, 3.25)	(4.43, 3.16)
Lower eyeball radius	(9.25, 4.28)	(10.39, 4.14)	(9.26, 3.62)	(5.86, 4.04)	(4.54, 3.30)	(4.42, 3.15)
Upper eyeball radius	(10.56, 4.23)	(11.57, 3.91)	(10.92, 3.68)	(5.57, 3.95)	(4.43, 3.22)	(4.52, 3.17)
Lower iris radius	(10.86, 4.29)	(11.83, 3.97)	(11.24, 3.80)	(5.55, 3.94)	(4.45, 3.21)	(4.57, 3.18)
Upper iris radius	(9.33, 4.24)	(10.47, 4.08)	(9.40, 3.56)	(5.82, 4.03)	(4.51, 3.29)	(4.41, 3.15)
Optimal values	(9.02, 4.58)	(10.17, 4.49)	(8.66, 3.96)	(6.12, 4.13)	(4.75, 3.39)	(4.50, 3.16)

Table 5
Mean absolute error (MAE) and standard deviation (SD) of the gaze estimates in visual angle calculated using estimated head pose and tracked eye corner positions, for the proposed model characterised by mean anthropometric measurements of the human eyeball.

	YAW			PITCH		
Subject number	Left eye (MAE(°), SD(°))	Right eye	Combined	Left eye (MAE(°), SD(°))	Right eye	Combined
2	(6.79, 5.00)	(8.72, 4.43)	(7.09, 4.45)	(9.36, 4.06)	(14.43, 4.93)	(11.84, 4.09)
6	(7.49, 3.98)	(5.34, 3.10)	(6.04, 3.34)	(4.57, 3.67)	(5.58, 4.05)	(4.78, 3.82)
7	(14.49, 4.22)	(7.51, 3.07)	(5.72, 3.20)	(6.33, 6.18)	(8.86, 3.76)	(6.98, 3.76)
9	(6.13, 3.04)	(6.06, 2.68)	(5.45, 2.57)	(3.80, 2.67)	(3.76, 2.74)	(3.60, 2.40)
13	(6.42, 3.02)	(6.42, 3.04)	(2.74, 2.06)	(4.72, 3.56)	(5.00, 3.21)	(4.56, 3.26)
14	(4.88, 2.91)	(5.27, 3.06)	(4.29, 2.35)	(6.58, 4.33)	(10.80, 4.77)	(8.38, 4.51)
15	(6.01, 3.30)	(8.05, 3.21)	(3.59, 2.97)	(5.40, 3.61)	(8.26, 4.97)	(6.44, 4.36)
19	(7.28, 5.10)	(8.28, 5.31)	(6.86, 4.50)	(9.97, 4.67)	(10.98, 3.77)	(10.44, 3.89)
21	(7.13, 4.54)	(8.03, 3.87)	(6.25, 3.81)	(5.62, 3.17)	(7.45, 4.64)	(6.43, 3.60)
24	(9.85, 4.97)	(12.29, 8.97)	(7.13, 7.02)	(5.99, 3.18)	(3.32, 2.53)	(3.50, 2.21)
30	(5.93, 4.08)	(6.98, 4.55)	(5.39, 3.76)	(6.16, 4.41)	(6.45, 4.41)	(6.12, 4.46)
31	(5.91, 3.38)	(7.42, 3.33)	(4.95, 3.01)	(10.66, 5.32)	(9.85, 5.28)	(9.75, 4.59)
33	(8.91, 4.01)	(6.29, 2.92)	(5.42, 2.71)	(5.39, 3.20)	(6.91, 3.20)	(5.67, 2.89)
38	(6.23, 3.89)	(7.79, 4.52)	(6.45, 3.77)	(7.91, 4.24)	(7.18, 4.14)	(7.47, 4.12)
45	(8.87, 5.00)	(10.68, 3.46)	(5.33, 3.32)	(6.21, 3.49)	(6.47, 4.14)	(6.05, 3.75)
47	(15.09, 4.87)	(8.29, 5.18)	(11.36, 4.35)	(5.22, 3.76)	(5.88, 3.92)	(4.97, 3.86)
48	(9.24, 4.20)	(13.13, 3.98)	(10.24, 3.88)	(4.95, 4.15)	(5.42, 3.96)	(3.34, 3.00)
49	(9.60, 6.44)	(20.13, 5.25)	(10.72, 4.31)	(4.43, 3.24)	(5.78, 3.79)	(4.07, 2.99)
56	(6.03, 3.42)	(5.73, 3.23)	(3.84, 2.90)	(4.43, 3.08)	(4.37, 3.39)	(4.08, 2.99)
Mean	(8.01, 4.18)	(8.55, 4.06)	(6.26, 3.59)	(6.19, 3.89)	(7.20, 3.98)	(6.23, 3.61)

reported results quantify the gaze estimation error for the left and rights eyeballs separately, and the combined gaze estimation error. In the knowledge of the distance between the subjects and the grid of visual stimuli, the combined gaze was estimated by joining the two vectors projecting from each eyeball, into a single vector that connects the mid-point between the eyes together with the mid-point between the intersections of the two gaze vectors with the grid. The results in Table 5 for every participant, setting the model parameters values to population averages, indicate that the accuracy for several of the separate left and right gaze estimates improves if these are combined into a single gaze estimate. This improvement in the results is congruent with other methods which have reported that combining the two gaze vectors together tends to compensate for the effect of noise [27,28].

The results presented in Tables 2–4 indicate that the gaze estimation error is sensitive to changes in the model parameter values corresponding to different anthropometric measurements. Nonetheless, these results also reveal a significant difference in error between the lower and upper bound measurements, indicating that the proposed model consistently performed better across all

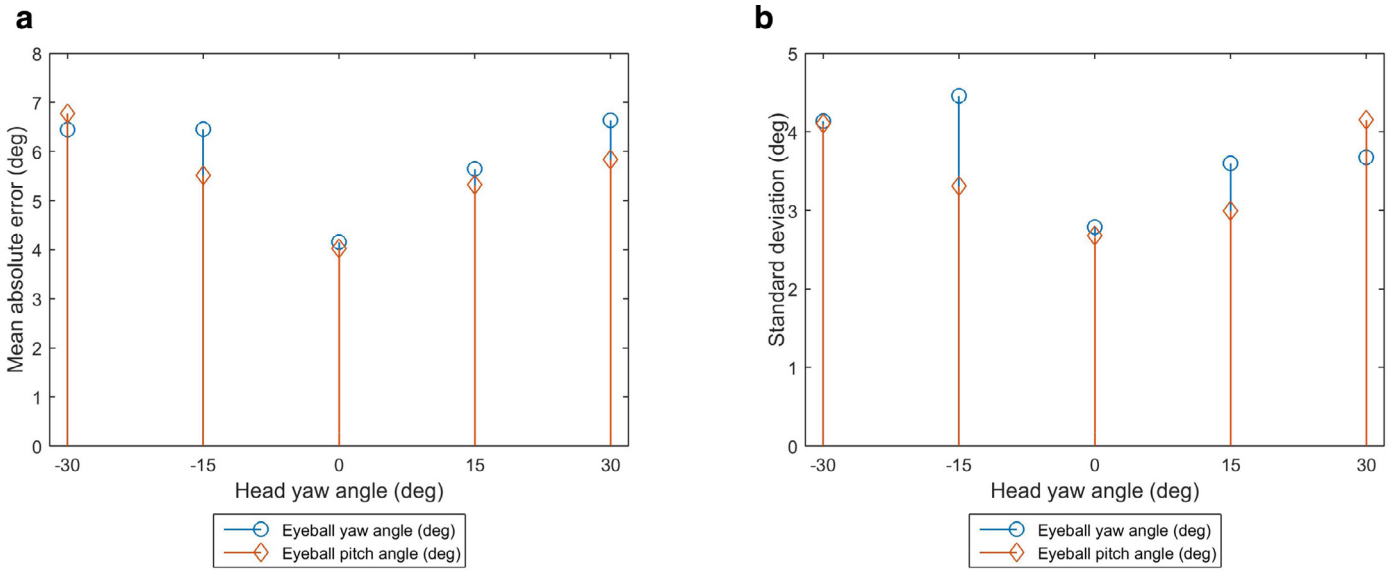

Fig. 5. The gaze estimation error increases in both yaw and pitch with larger head rotation angles due to increased distortion and self occlusions in the appearance of the regions surrounding the features of interest.

Table 6
Mean absolute error (MAE) and standard deviation (SD) of the head yaw and pitch estimates in degrees, calculated across all participants with respect to the ground truth head rotation angles provided with the Columbia Gaze Data Set [33].

Head rotation (Yaw/°, Pitch/°)	(0, 0)	(15, 0)	(30, 0)	(-15, 0)	(-30, 0)
Yaw error (MAE/°, SD/°)	(1.70, 1.50)	(10.30, 2.73)	(9.83, 1.50)	(7.04, 1.61)	(13.62, 2.19)
Pitch error (MAE/°, SD/°)	(1.94, 1.47)	(3.00, 1.63)	(2.54, 1.71)	(2.68, 1.72)	(2.86, 1.96)

participants when the measurements marked in bold were considered. It may also be noted in Tables 2 and 3 that, setting the remaining parameter values to population averages, the lowest gaze estimation errors obtained with and without the use of ground truth information correspond to the same lower or upper bound measurements marked in bold, indicating further consistency across all participants. In comparison, the results obtained for the cylindrical model alone in Table 4 exhibit the lowest error values at different lower and upper bound measurements. The eyeball protrusion measurements were not considered in Table 4, since in this case the spherical shape of the eyeballs was not modelled. The model parameter values were further set to the anthropometric measurements that achieved the lowest gaze estimation errors, marked in bold, permitting a further improvement in the gaze estimation accuracy for the optimal results reported in Tables 3 and 4.

A further comparison between the results presented in Tables 2 and 3 reveals a degradation of approximately 0.5°–2° in the gaze estimation accuracy without the use of ground truth information. This degradation in accuracy typically arises due to inaccurate tracking of the facial features, which in turn increases the head pose estimation error since the head rotation angles in this work are being estimated according to the relative positioning between the facial features. Indeed, the results in Table 6 reveal an increase in the estimation of the head yaw and pitch angles across all participants with increasing head rotation angles, due to increased distortion and self occlusions in the appearance of the regions surrounding the features of interest that arise during head rotation. It may also be observed in Fig. 5 that the gaze estimation error increases with larger head rotation angles as well, indicating that the gaze estimation error depends upon the accuracy of the head pose estimation. Nonetheless, it may also be noted that while the head pose estimation error increases to a maximum of 13.62° in yaw and 3.00° in pitch with respect to ground truth, the resulting degradation in the estimation of gaze remains within a significantly smaller range that approximately spans between 0.5° and 2° in both yaw and pitch.

A comparison with the results obtained for the cylindrical model employing ground truth head pose and eye corner positions in Table 4, reveals an inferior performance for the cylindrical model when compared with the results obtained for the cylinder head and spherical eyeballs model being proposed in this work. The proposed enhancement over the cylindrical model has been found to produce a significant 2.5-fold improvement in the accuracy of the yaw estimates when the lowest gaze estimation errors in Tables 2 and 4 were compared. While the proposed model improves over the pitch estimates as well, the error values in the vertical direction for the two models are comparable as expected since the Columbia Gaze Data Set does not comprise changes in the head pitch rotation angles. The improvement in yaw may also be corroborated qualitatively in Fig. 6, where it may be seen that the reference eye region images rotated by the proposed model in Fig. 6(g)–(j) are visually closer to the observed eye region images in Fig. 6(c)–(f), than the reference images rotated by the cylindrical model alone in Fig. 6(k)–(n).

5.2. Evaluation on real data

To evaluate our method in the context of an AAC application, a video sequence has been captured in a real-life scenario by having a person suffering from cerebral palsy gaze upon a set of visual targets appearing in succession on a 23" monitor screen. The video sequence has been captured at a resolution of 800 × 448 pixels and a frame rate of 30 fps, while the subject sat at a distance of 65 cm from the monitor screen and camera such that the average dimensions of the face and eye regions inside the image space were

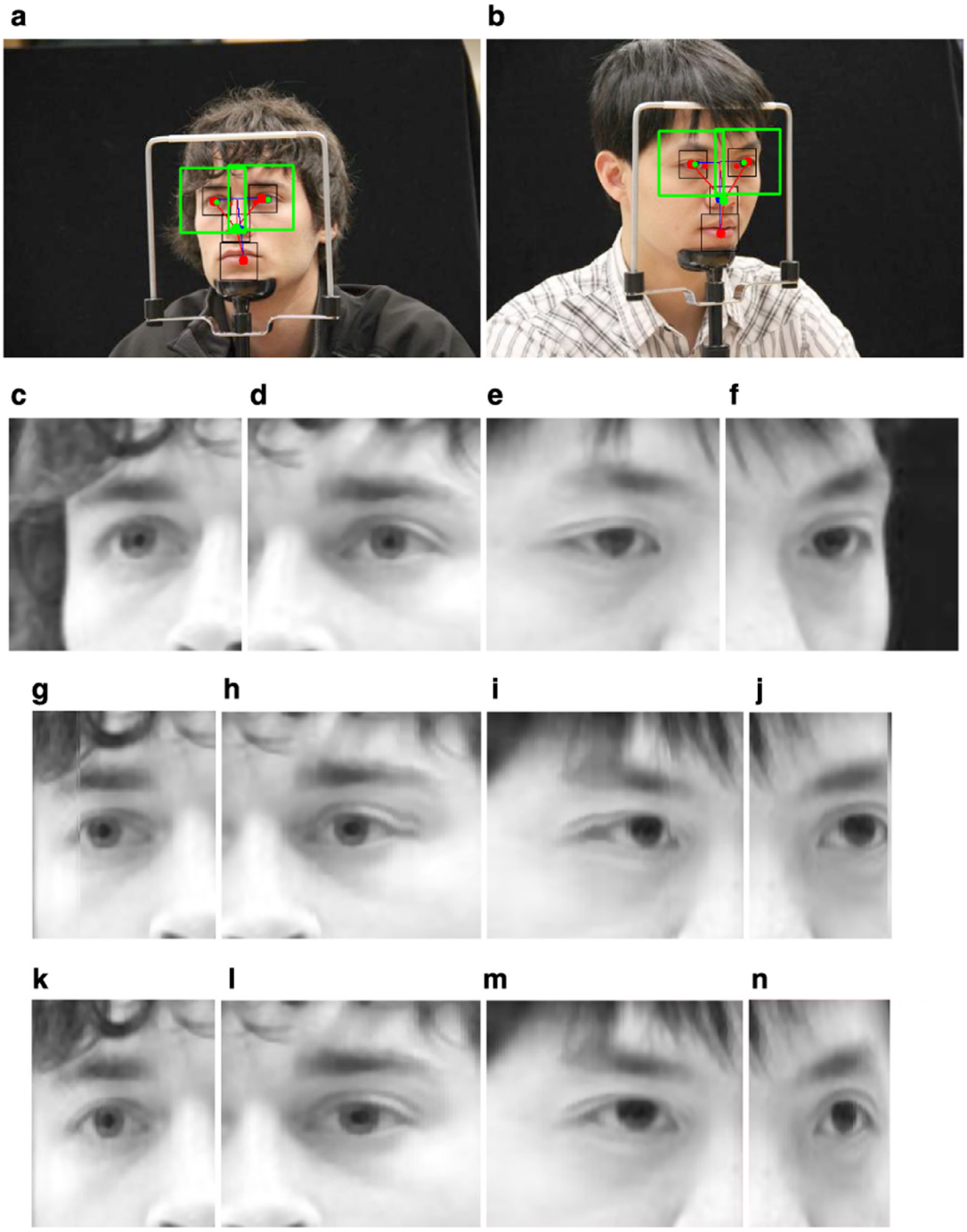

Fig. 6. The reference eye region images rotated by the proposed model in (g)–(j), corresponding to -15° and 30° head yaw rotation angles in (a) and (b) respectively, are visually closer to the observed eye region images in (c)–(f), than the reference images rotated by the cylindrical model alone in (k)–(n).

95 × 147 pixels and 29 × 13 pixels respectively. Each visual target was displayed for 5s, permitting the computation of 150 gaze estimates per visual target, and positioned according to the configuration shown in Fig. 7. Video sequences of four normal subjects have also been captured after these were requested to gaze upon the same set of visual stimuli, allowing qualitative and quantitative comparisons of the resulting gaze estimation errors, as affected by the involuntary head movements performed by the subject with cerebral palsy and the controlled head movements performed by the group of normal subjects, as indicated in Fig. 8. In order to demonstrate the proposed algorithm in operation, a video component has also been provided with the electronic version of this manuscript as described further in Appendix A.

Fig. 8 indicates a significant discrepancy between the head movements performed by the subjects throughout the experiment. While the normal subjects maintained an overall steady head pose, performing small head movements that enabled them to shift the gaze direction between different stimuli appearing on the monitor screen as shown in Fig. 9, the head movement data acquired from the subject with cerebral palsy exhibits continuous fluctuations which are especially pronounced in the yaw and roll directions. It is known that for the subject in question the medical condition hampers the vertical eyeball rotation, while head movement in the pitch direction was constrained by the use of a headrest attached to the wheelchair during the experiment. The subject, therefore, tended to compensate for the limitation in eyeball rotation by

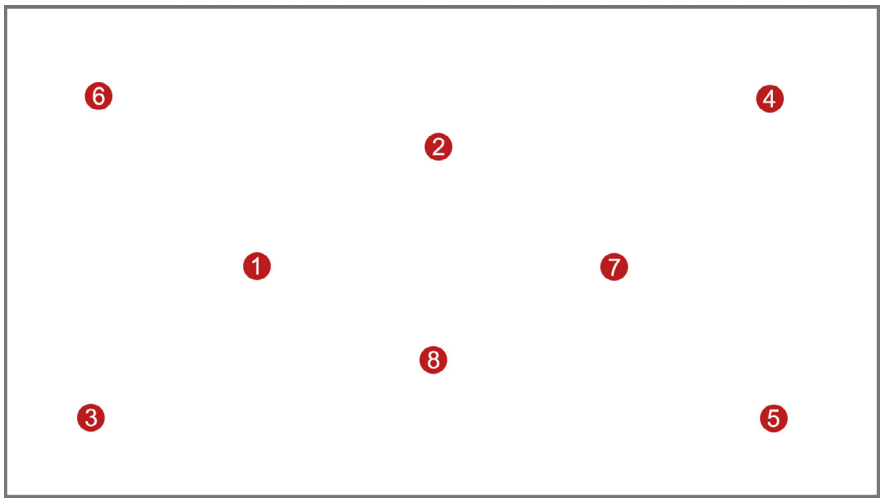

Fig. 7. Configuration of visual targets displayed in succession on a monitor screen.

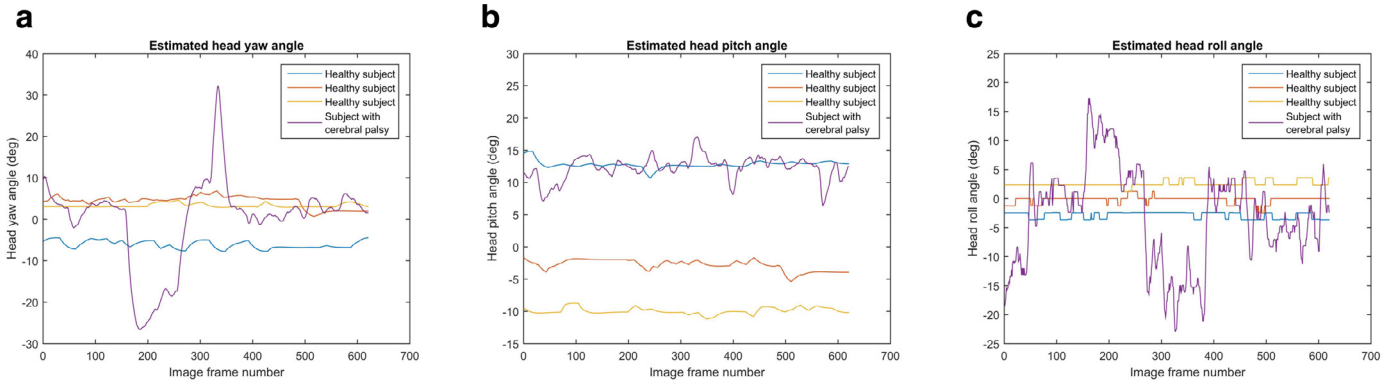

Fig. 8. Qualitative comparison between the involuntary head yaw (a), pitch (b) and roll (c) movements induced by cerebral palsy, and the corresponding head movements performed by a group of healthy subjects.

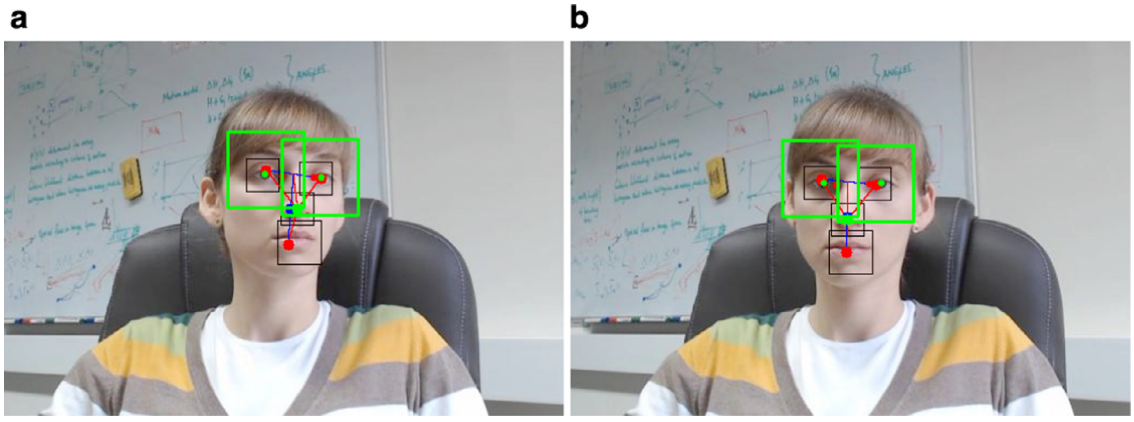

Fig. 9. The normal subjects maintained an overall steady head pose, performing small head movements that enabled them to shift the gaze direction between different stimuli appearing on the monitor screen.

the head yaw and roll movement in addition to other involuntary head rotations, as exhibited by the higher yaw and roll amplitudes in Fig. 8. This substantiates the need to lift any assumptions of stationary head movement for gaze estimation especially within an AAC context, which is in-line with one of the objectives of our research work.

In the knowledge of the monitor position in three-dimensional space and the distance between the subjects and the monitor screen, the separate gaze vectors projecting from each eyeball were combined into a single vector that joins the mid-point between the eyes together with the mid-point between the intersections of the two gaze vectors with the monitor screen. In the absence of ground truth information for the head and eye rotations performed by the subject with cerebral palsy, we recorded the three-dimensional positions of each eyeball with respect to the monitor screen by a Tobii X2-30 eye-gaze tracker. This enabled the computation of a single ground truth gaze vector that projects from the three-dimensional eyeball positions to the corresponding visual

Table 7
Mean absolute error (MAE) and standard deviation (SD) of the gaze estimates in visual angle, calculated between the intersection points of the estimated gaze vectors with a monitor screen and ground truth information recorded by a Tobii X2-30 eye-gaze tracker, for the normal subjects (1–4) and the subject with cerebral palsy (5).

	YAW			PITCH		
Subject number	Left eye (MAE(°), SD(°))	Right eye	Combined	Left eye (MAE(°), SD(°))	Right eye	Combined
1	(12.50, 4.11)	(13.15, 3.39)	(2.77, 1.86)	(2.44, 1.34)	(4.09, 2.06)	(2.41, 1.49)
2	(6.12, 4.15)	(8.58, 2.26)	(3.44, 1.83)	(3.95, 2.37)	(4.66, 3.17)	(2.69, 1.91)
3	(4.66, 3.43)	(10.13, 2.73)	(4.29, 2.52)	(2.76, 2.28)	(2.31, 1.41)	(2.15, 1.54)
4	(6.03, 3.38)	(4.19, 1.79)	(2.00, 1.50)	(2.19, 1.35)	(1.88, 1.08)	(1.62, 0.90)
Mean	(7.33, 3.77)	(9.01, 2.54)	(3.13, 1.93)	(2.84, 1.84)	(3.24, 1.93)	(2.22, 1.46)
5	(8.38, 3.01)	(8.21, 2.99)	(7.77, 2.98)	(7.72, 4.81)	(8.01, 4.92)	(7.40, 4.76)

Table 8
Comparison of the mean absolute error (MAE) and standard deviation (SD) results obtained by our method following evaluation on test data, with (a) and without (b) ground truth information, and evaluation on real data collected from a group of normal subjects (c) and a subject with cerebral palsy (d), with results reported by relevant state-of-the-art three-dimensional gaze estimation methods [9,22–24,27,28].

Method	Yaw (MAE/°, SD/°)	Pitch (MAE/°, SD/°)	Number of subjects	Range of head rotation (α/°, β/°, γ/°)	User-camera distance (cm)	User-target distance (cm)
(a)	(3.47, 2.95)	(4.40, 3.15)	19	(±30, 0, 0)	200	250
(b)	(5.87, 3.73)	(5.49, 3.45)	19	(±30, 0, 0)	200	250
(c)	(3.13, 1.93)	(2.22, 1.46)	4	($\leq$ 5, $\leq$ 5, $\leq$ 5)	60–70	60–70
(d)	(7.77, 2.98)	(7.40, 4.76)	1	(±30, 5-20, ± 20)	60–70	60–70
[22]	(1.87, 0.82)		5	(±34, ± 21, ± 32)	N/A	80
[22]	(2.19, 1.14)		5	(±36, ± 23, ± 30)	N/A	130
[23]	5.3	7.7	5	N/A	220	240
[24]	5.953		4	N/A	400 (with zoom)	400
[9]	2.36	2.11	5	None	200	400
[27]	$\leq$3		1	N/A	80	80
[28]	7		1	N/A	50	50

target positions on the monitor screen, hence permitting a comparison between the gaze estimates obtained through our method and those obtained via the Tobii tracker for every visual target. The calculated gaze estimation error is presented in Table 7, which has been divided into two sections; the upper section presents the results for the normal subjects, numbered 1–4, while the lower section tabulates the error results for the subject with cerebral palsy. The presented results reveal a significant difference between the mean yaw and pitch error values calculated for the normal subjects in comparison with the gaze estimation error obtained for the subject with cerebral palsy. Further to the discussion in Section 5.1, inaccuracies in face feature tracking during large head rotations performed by the subject with cerebral palsy are a major source of error that lead to higher gaze estimation errors. It may be observed from Fig. 8 that while the normal subjects generally perform head rotations that do not exceed 5° in yaw, pitch and roll, the subject with cerebral palsy performs significantly larger head rotations within the range, (±30°, 5°–20°, ± 20°). A comparison between the results in Tables 2 and 7 reveals an improvement in the gaze estimation accuracy for the normal subjects when compared with the results obtained from the Columbia Gaze Data Set, due to a smaller range of head movements performed by the normal subjects. On the other hand, the subject with cerebral palsy performs head rotations that exceed the head rotation angles considered in the Columbia Gaze Data Set, resulting in increased gaze estimation error when compared with the results in Table 2. Furthermore, it has been noted that the involuntary face movements performed by the subject with cerebral palsy tended to increase the head pose estimation error, since the head pose in this work is calculated according to the relative positions of the facial features under the assumption of a rigid face model. Hence, a possible avenue for future work that permits an improvement in the gaze estimation accuracy within the context of AAC, would be to preserve the assumption of rigidity by stabilising the movement of the facial features prior to head pose estimation.

6. Comparison with the state-of-the-art

In order to put the proposed method into context, we have compared the results presented under Section 5 with other state-of-the-art three-dimensional gaze estimation methods in the literature. Table 8 has been divided into two parts, the first of which tabulates the results obtained by our method following evaluation on test and real data, while the second part presents gaze estimation results in visual angle reported by relevant three-dimensional gaze estimation methods [9,22–24,27,28], specifying the number of subjects participating during data collection, the range of head rotation angles performed by the participants, and the distances between the participants and the camera and visual targets.

The range of head rotation angles performed by the participants during data collection is an important factor to consider when comparing between different methods, if this impacts on the gaze estimation accuracy as further discussed in Section 5. Tying with the observations in the previous section, a scenario in which the participants are sitting at close range to a monitor screen displaying the visual targets [28], where the range of head rotations is generally small enabling the participants to comfortably shift the gaze direction between different visual stimuli, is expected to be characterised by lower gaze estimation error in comparison to a scenario in which the participants perform larger head rotations.

Hence, a comparison between the results in Table 8 reveals that the MAE and SD values obtained by the proposed method, with and without ground truth head pose and eye corner positions (a, b), exhibit a gaze estimation performance that is comparable to the three-dimensional gaze estimation methods of [23,24] both of which operate at a significant distance from the user ($\geq$200),

or better than the gaze estimation performance achieved by the method of [28] that operates at a closer distance to the user ($\leq$80). A further comparison between the error values obtained by the group of normal subjects (c), using the proposed cylindrical head and spherical eyeballs model, and the state-of-the-art three-dimensional gaze estimation methods of [27,28] that operate at close range ($\leq$ 80) also reveals comparable [27] or better [28] performance. The methods which perform better than the proposed approach facilitate head pose estimation by fitting a plane to a set of three markers attached to the face [22], or do not consider free head movement [9]. In view of these results it is important to mention that, while other state-of-the-art methods include a personal calibration session to estimate a set of user-dependent eyeball model parameters, the proposed method achieves a comparable performance under free head movement merely by requiring detection of an initial frontal head pose.

7. Conclusions

In this paper, we have proposed a method for three-dimensional eye-gaze estimation under free head movement from a single integrated camera inside a notebook computer, as an AAC tool to assist individuals with motor disabilities, such as cerebral palsy, affected by involuntary head and face movements. Specifically, the proposed research work built upon the cylindrical head model proposed by Xiao et al. [32] and later employed by Valenti et al. [16], by augmenting their head model to include spherical models of the human eyeballs. This permitted an extension of the two-dimensional method proposed by Valenti et al. [16] that allowed us to project the change in appearance due to head rotation of a reference eye region inside the image space, permitting the estimation of three-dimensional gaze vectors based on the image displacement between a rotated reference eye region image and a newly observed image. The three-dimensional eye-gaze has been estimated without necessitating prior camera calibration and reducing user calibration to frontal eye and head pose detection.

The proposed method has been evaluated on publicly available test data from the Columbia Gaze Data Set [33], and real-life data captured by having a group of normal subjects and a person suffering from cerebral palsy gaze upon a set of visual targets appearing in succession on a monitor screen. The experimental results obtained on the test data reveal that the gaze estimation error is sensitive to different anthropometric measurements of the human eye exhibited by the population, nonetheless it has been observed that the proposed method consistently performed better when specific anthropometric measurements, as outlined in Section 5.1, were used across all participants. Furthermore, while a degradation in the gaze estimation accuracy was observed with increasing head rotation angles, the resulting degradation in the estimation of gaze remained within a significantly smaller range in comparison with increasing head pose estimation error. In addition, the proposed enhancement over the cylindrical head model, where the eye regions are flat on the cylindrical surface, by augmenting the cylindrical model with spherical eyeballs, yielded a 2.5-fold improvement in the gaze estimation accuracy hence demonstrating the benefit of this model augmentation.

Furthermore, the achieved gaze estimation errors for the Columbia Gaze Data Set, with and without the use of ground truth information, and the group of normal subjects reveal a comparable performance with relevant state-of-the-art methods that have been tested in similar conditions, relating to the range of head movements performed by the participants during data collection. Increased gaze estimation errors for the subject with cerebral palsy were attributed to inaccuracies in face feature tracking during large head rotations that exceed the head rotation angles considered in the Columbia Gaze Data Set, due to distortion and partial occlusion of the features of interest associated with large head pose angles. Involuntary face movements performed by the subject with cerebral palsy were also noted to increase the head pose estimation error, since the head pose has been calculated according to the relative positions of the facial features under the assumption of a rigid face model. It is important to mention that, while other state-of-the-art methods required a personal calibration session to estimate a set of user-dependent eyeball model parameters prior to gaze estimation, the proposed method achieved a comparable performance under free head movement merely by requiring detection of an initial frontal head pose.

Possible avenues for future work aim to improve upon the achieved three-dimensional gaze estimation accuracy by enhancing the robustness of face feature tracking under large head rotations, and to handle the involuntary face movements performed by persons with motor disabilities which may affect the gaze estimation accuracy within an AAC context.

Appendix A. Supplementary Video Data

A video component accompanies the electronic version of this manuscript in order to demonstrate the proposed algorithm in operation. In the supplementary video data, the tracked facial features, namely the eyes, eye corners, nose and mouth, are displayed in the upper left window where each feature is enclosed by a black bounding box. The cropped eye region images, marked by a green bounding box in the upper left window, are subsequently binarised via a Bayes' classifier to detect the iris regions and displayed in the lower left region of the video. The windows in the mid-right region display the cropped eye region images, whereas the upper right windows display the reference eye images transformed according to the estimated head pose. The estimated head and eyeball rotation angles are visualised by the pin model in the upper left corner.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.cviu.2016.02.012.

References

[1] R.J.K. Jacob, What you look at is what you get: eye movement-based interaction techniques, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People, ACM New York, NY, USA, 1990, pp. 11–18.
[2] D.W. Hansen, Q. Ji, In the eye of the beholder: A survey of models for eyes and gaze, IEEE Trans. Pattern Anal. Mach. Intell. 32 (3) (2010) 478–500.
[3] J.L. Levine, An Eye-controlled Computer, Res. Rep. RC-8857, IBM Research Division, T.J. Watson Research Center, 1981.
[4] E. Demjen, V. Abosi, Z. Tomori, Eye tracking using artificial neural networks for human computer interaction, Physiol. Res. 60 (2011) 841–844.
[5] R. Ohtera, T. Horiuchi, S. Tominaga, Eye-gaze detection from monocular camera image using parametric template matching, in: Proceedings of the 8th Asian Conference on Computer Vision, Springer Berlin Heidelberg, 2007, pp. 708–717.
[6] S. Kim, B. Hwang, M. Lee, Gaze tracking based on pupil estimation using multilayer perceptron, in: Proceedings of International Joint Conference on Neural Networks, Institute of Electrical and Electronics Engineers (IEEE), 2011, pp. 2683–2689.
[7] A. Al-Rahayfeh, M. Faezipour, Classifiers comparison for a new eye gaze direction classification system, in: IEEE Long Island Systems, Applications and Technology Conference (LISAT), Institute of Electrical and Electronics Engineers (IEEE), 2014, pp. 1–6.
[8] D. Li, D. Winfield, D.J. Parkhurst, Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model based approaches, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Workshops, vol. 3, Institute of Electrical and Electronics Engineers (IEEE), 2005, pp. 79–86.
[9] H. Wu, Y. Kitagawa, T. Wada, T. Kato, Q. Chen, Tracking iris contour with a 3d eye-model for gaze estimation, in: Proceedings of the 8th Asian Conference on Computer Vision, Springer Berlin Heidelberg, 2007, pp. 688–697.
[10] W. Sewell, O. Komogortsev, Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network, in: Proceedings of the 28th International Conference Extended Abstracts on Human Factors in Computing Systems, ACM New York, NY, USA, 2010, pp. 3739–3744.

[11] N.H. Cuong, H.T. Hoang, Eye-gaze detection with a single webcam based on geometry features extraction, in: 2010 11th International Conference on Control, Automation, Robotics and Vision, Institute of Electrical and Electronics Engineers (IEEE), 2010, pp. 2507–2512.
[12] J. Wang, E. Sung, R. Venkateswarlu, Estimating the eye gaze from one eye, Comput. Vis. Image Underst.—Spec. Issue Eye Detect. Track. 98 (1) (2005) 83–103.
[13] C. Holland, O. Komogortsev, Eye tracking on unmodified common tablets: Challenges and solutions, in: Proceedings of the Symposium on Eye Tracking Research and Applications, ACM New York, NY, USA, 2012, pp. 277–280.
[14] K. Kunze, S. Ishimaru, Y. Utsumi, K. Kise, My reading life—Towards utilizing eyetracking on unmodified tablets and phones, in: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing, ACM New York, NY, USA, 2013, pp. 283–286.
[15] K. Takemura, S. Kimura, S. Suda, Estimating point-of-regard using corneal surface image, in: Proceedings of the Symposium on Eye Tracking Research and Applications, ACM New York, NY, USA, 2014, pp. 251–254.
[16] R. Valenti, N. Sebe, T. Gevers, Combining head pose and eye location information for gaze estimation, IEEE Trans. Image Process. 21 (2) (2012) 802–815.
[17] S. Cristina, K.P. Camilleri, Cursor control by point-of-regard estimation for a computer with integrated webcam, in: The 8th International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP), The International Academy, Research and Industry Association (IARIA), 2014, pp. 126–131.
[18] H. Wang, C. Pan, C. Chaillou, Tracking eye gaze under coordinated head rotations with an ordinary camera, in: 9th Asian Conference on Computer Vision, vol. 5995, Springer Berlin Heidelberg, 2010, pp. 120–129.
[19] J. Mansanet, A. Albiol, R. Paredes, J.M. Mossi, A. Albiol, Estimating point of regard with a consumer camera at a distance, Pattern Recog. Image Anal. (2013) 881–888.
[20] F. Lu, T. Okabe, Y. Sugano, Y. Sato, Learning gaze biases with head motion for head pose-free gaze estimation, Image Vis. Comput. 32 (3) (2014) 169–179.
[21] M. Reale, T. Hung, L. Yin, Viewing direction estimation based on 3d eyeball construction for HRI, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Institute of Electrical and Electronics Engineers (IEEE), 2010, pp. 24–31.
[22] R. Ronsse, O. White, P. Lefevre, Computation of gaze orientation under unrestrained head movements, J. Neurosci. Methods 159 (1) (2007) 158–169.
[23] H. Yamazoe, A. Utsumi, T. Yonezawa, S. Abe, Remote gaze estimation with a single camera based on facial-feature tracking without special calibration actions, in: Proceedings of the 2008 Symposium on Eye Tracking Research and Applications, ACM New York, NY, USA, 2008, pp. 245–250.
[24] M.J. Reale, S. Canavan, L. Yin, K. Hu, T. Hung, A multi- gesture interaction system using a 3-d iris disk model for gaze estimation and an active appearance model for 3-d hand pointing, IEEE Trans. Multimedia 13 (3) (2011) 474–486.
[25] P. Smith, M. Shah, N. da Vitoria Lobo, Monitoring head/eye motion for driver alertness with one camera, in: Proceedings of the 15th International Conference on Pattern Recognition, vol. 4, Institute of Electrical and Electronics Engineers (IEEE), 2000, pp. 636–642.
[26] J. Chen, Q. Ji, 3d gaze estimation with a single camera without ir illumination, in: 19th International Conference on Pattern Recognition, Institute of Electrical and Electronics Engineers (IEEE), 2008, pp. 1–4.
[27] Y. Matsumoto, A. Zelinsky, An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement, in: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Institute of Electrical and Electronics Engineers (IEEE), 2000, pp. 499–504.
[28] E. Pogalin, A. Redert, I. Patras, E.A. Hendriks, Gaze tracking by using factorized likelihoods particle filtering and stereo vision, in: Third International Symposium on 3D Data Processing, Visualization, and Transmission, Institute of Electrical and Electronics Engineers (IEEE), 2006, pp. 57–64.
[29] W. Waizenegger, N. Atzpadin, O. Schreer, I. Feldmann, P. Eisert, Model based 3d gaze estimation for provision of virtual eye contact, in: 19th IEEE International Conference on Image Processing (ICIP), Institute of Electrical and Electronics Engineers (IEEE), 2012, pp. 1973–1976.
[30] K.K. W., Cerebral palsy: An overview, Am. Family Phys. 73 (1) (2006) 91–100.
[31] D. Moore, A. Gorra, M. Adams, J. Reaney, H. Smith, Disabled Students in Education: Technology, Transition, and Inclusivity, first, IGI Global, 2011.
[32] J. Xiao, T. Kanade, J. Cohn, Robust full motion recovery of head by dynamic templates and re-registration techniques, Int. J. Imag. Syst. Technol. 13 (2003) 85–94.
[33] B. Smith, Q. Yin, S. Feiner, S. Nayar, Gaze locking: Passive eye contact detection for human–object interaction, in: ACM Symposium on User Interface Software and Technology, ACM New York, NY, USA, 2013, pp. 271–280.
[34] K.M. Evans, R.A. Jacobs, J.A. Tarduno, J.B. Pelz, Collecting and analyzing eye-tracking data in outdoor environments, J. Eye Move. Res. 5 (2) (2012) 1–19.
[35] J. Edinger, D. Pai, M. Spering, Rolling motion makes the eyes roll: torsion during smooth pursuit eye movements, J. Vis. 13 (9) (2013) 383.
[36] E. Murphy-Chutorian, M. Trivedi, Head pose estimation in computer vision: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 31 (4) (2009) 607–626.
[37] M. Sapienza, K. Camilleri, Fasthpe: A Recipe for Quick Head Pose Estimation, Technical Report TR-SCE-2011-01, University of Malta, 2011. https://www.um.edu.mt/library/oar/handle/123456789/859.
[38] P. Viola, M. Jones, Robust real-time object detection, Int. J. Comput. Vis. 57 (2) (2001) 137–154.
[39] OpenCV, Open source computer vision, 2016, http://opencv.org/.
[40] R. Valenti, T. Gevers, Accurate eye center location and tracking using isophote curvature, in: Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, Institute of Electrical and Electronics Engineers (IEEE), 2008, pp. 1–8.
[41] R. Valenti, J. Staiano, N. Sebe, T. Gevers, Webcam-based visual gaze estimation, in: Proceedings of the 15th International Conference on Image Analysis and Processing, Springer Berlin Heidelberg, 2009, pp. 662–671.
[42] T. Moriyama, T. Kanade, J. Xiao, J.F. Cohn, Meticulously detailed eye region model and its application to analysis of facial images, IEEE Trans. Pattern Anal. Mach. Intell. 28 (5) (2006) 738–752.
[43] C. Wilkinson, S. Mautner, Measurement of eyeball protrusion and its application in facial reconstruction, J. Forensic Sci. 48 (1) (2003) 12–16.
[44] J. Shi, C. Tomasi, Good features to track, in: Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Institute of Electrical and Electronics Engineers (IEEE), 1994, pp. 593–600.
[45] R. Duda, P. Hart, Use of the hough transformation to detect lines and curves in pictures, Commun. ACM 15 (1) (1972) 11–15.
[46] P. Caroline, M. Andre, The effect of corneal diameter on soft lens fitting, Part 1, 2002, http://www.clspectrum.com/articleviewer.aspx?articleid=12130.
[47] I. Bekerman, P. Gottlieb, M. Vaiman, Variations in eyeball diameters of the healthy adults, J. Ophthalmol. 2014 (2014) 1–5.

Stefania Cristina graduated with the B.Eng.(Hons.) degree in electrical engineering from the University of Malta in 2008, and received the M.Sc.(Melit.) by Research degree from the same University in 2010. She is presently pursuing the Ph.D. degree from the University of Malta and also employed as a Research Officer with the Department of Systems and Control Engineering. Her research interests include computer vision, human–computer interaction and assistive technology.

Kenneth P. Camilleri graduated with the B.Elec.Eng.(Hons.) degree in electrical engineering from the University of Malta and received the M.Sc. degree in signal processing and machine intelligence and the Ph.D. degree in image processing and pattern recognition in 1994 and 1999, respectively, from the University of Surrey, Guildford, UK. He is currently the Head of the Department of Systems and Control Engineering and Director of the Centre for Biomedical Cybernetics at the University of Malta. His research interests include machine vision, thermal image analysis, and biomedical engineering, in particular brain signal analysis applied to the diagnosis of brain diseases and to brain–computer interfacing.