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