Expert Systems with Applications 42 (2015) 510–518

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Review

Eye-gaze estimation under various head positions and iris states

Reza Jafari *, Djemel Ziou

Departement d'informatique, Universite de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada

ARTICLE INFO

Article history:
Available online 22 August 2014

Keywords:
Gaze estimation
Bayesian logistic regression
Kinect

ABSTRACT

This paper describes a method for eye-gaze estimation under normal head movement. In this method, head position and orientation are acquired by Kinect depth data and eye direction is obtained from high resolution images. We propose the Bayesian multinomial logistic regression based on a variational approximation to construct a gaze mapping function and to verify iris state. Our method eliminates limitation of head movements, eye closure and light source as common drawbacks in most conventional techniques. The efficiency of the proposed method is validated by performance evaluation for multiple people with different distances and poses to the camera under various eye states.

1. Introduction

Eye-gaze deals with the estimation of the line of sight of a person. The eyes are the windows to the soul (William Shakespeare), therefore one of the logical steps to understanding human behavior and motivation should involve the study of eye gaze tracking. Eye-gaze mapping is important for many applications such as the pedophilia, training, and marketing. Indeed, in the presence of children, it has been shown that the map of pedophile is different from that of a person who is not (Renaud et al., 2009). During the training of surgeons, the eye-gaze is recorded and analysed for more reliable assessment of surgical skill (Law, Atkins, Kirkpatrick, & Lomax, 2004). In marketing, that is suitable to determine what features of the product attracts buyer attention (Khushabaa et al., 2013).

In this paper, we present a method for eye-gaze estimation that robustly detects the location and orientation of a person's head, from the depth data obtained by Kinect. There are two cameras embedded in the Kinect, one operating in the visible spectrum (RGB) and the other in the IR (infrared). Unfortunately, Kinect RGB camera is in low resolution to obtain iris images. Hence, another camera can be used simultaneously with the IR Kinect camera to acquire high resolution eyes images. In order to estimate eye-gaze we use the iris center and the reference point provided by the IR camera. However, the iris center and the reference point vary significantly with head position and orientation. This prompts us to consider head orientation and location in a gaze mapping model. To do so, the head location and orientation in 3D space is calculated from Kinect. Then, the resulting measures are used for an eye gaze mapping function. Since a gaze mapping function cannot be assumed beforehand, the variational Bayesian multinomial logistic regression (VBMLR) is used as a model to estimate it. Most of gaze estimators work when eye is open. However, winking is a necessity for humans. Therefore, in order to estimate iris center we need to detect the state of the eyes (i.e. whether they are open or closed). For this work, we introduce a method that uses Histogram of Oriented Gradients (HOG) and VBMLR to detect whether the eyes are open or closed.

Preliminary results of gaze-estimation model are reported in Jafari and Ziou (2012a, 2012b). In early version of this work, we used a PTZ camera, we did not provide a detailed description of the proposed model such as the iris state verification and the comparison between different matching functions.

Our paper is organized as follows. We start by summarizing related work in Section 2. In Section 3, the suggested eye-gaze scheme is explained. Section 4 presents experiment results and Section 5 the conclusion.

2. Related work

Many traditional techniques for eye-gaze estimation require some existing equipments to be put in physical contact with the user such as contact lenses, electrodes, and head mounted devices. The resulting situation can be massively inconvenient and uncomfortable for the user. Due to recent advances in computers and video cameras technology, eye-gaze estimation have been widely investigated based on digital video analysis (Hansen & Ji, 2010). Since it does not need anything attached to the user, the video

* Corresponding author.

http://dx.doi.org/10.1016/j.eswa.2014.08.003
0957-4174/

technology opens the most encouraging direction to build a remote eye-gaze tracker.

The remote systems can be classified into two different techniques (Guestrin & Eizenman, 2006; Miyake, Haruta, & Horihata, 2005, 2009; Morimoto & Mimica, 2005; Villanueva et al., 2009): interpolation based gaze estimation and 3D model-based gaze estimation. Interpolation methods use general purpose equations, such as linear or quadratic polynomials, to map the image data to gaze coordinates. 3D model-based techniques, on the other hand, directly compute the gaze direction based on a geometric model of the eye. All methods require calculating some parameters such as calibration procedure in which the user is asked to look at certain points on the screen. Moreover most of these techniques use a reference point to estimate gaze direction. The reference point can be generated by using something such as marker paper sticks to the face (Miyake et al., 2005, Miyake, Asakawa, Yoshida, Imamura, & Zhang, 2009). Unfortunately, these methods are cumbersome for users. Corneal reflection or glint is another well-known method to create a reference point. Glint is generated by an active light source on the cornea surface. Thus the vector from the glint to the center of the iris describes the gaze direction (Guestrin & Eizenman, 2006; Morimoto & Mimica, 2005; Villanueva et al., 2009). These approaches have problems with changes in lighting conditions, the reflection of light sources on glasses, awkward calibration process and a limitation on distance between the user and camera. Moreover, the small head motions tolerate by such devices have considerable influence on their accuracy, therefore experiments are usually done using a chin rest to restrict head motion, which greatly reduces the user's comfort.

Usually, a person moves the head to a comfortable position before turning the eye. Therefore, a 3D head pose needs to be modeled and integrated within a gaze estimation algorithm. There has been research for 3D human pose detection and tracking in the past years by depth image via Time of Flight (ToF) camera (Diraco, Leone, & Siciliano, 2013; Almansa-Valverdea, Castilloa, & Fernndez-Caballero, 2012; Zhu, Dariush, & Fujimura, 2008). Unlike 2D intensity images, depth images are robust to color and illumination change. Unfortunately existing commercial ToF cameras such as the Swiss Ranger SR4000 (C.C.S. d'Electronique, 2009) and PMD Tech (Video Sensor, 2009) are quite expensive and low resolution. Fortunately, Microsoft has launched the Kinect, which is cheap and easy to use. Recently, EyeCharm uses Microsoft Kinect for eye tracking (Eye tracking gets kickstarted, 2013). Since Kinect RGB camera is in low resolution to obtain eye images and EyeCharm uses Kinect RGB for eye detection, therefore, this system has a limitation on distance between Kinect and user (about 80 cm). Unlike mentioned systems, Tobii (Be first to get, 2013) and GazePoint (Products, 2014) have some low cost eye trackers on the market. Unfortunately, these systems do not consider head movements in mapping function, therefore, they have a restriction on changing head position and orientation.

In this study we propose a novel method to overcome common drawbacks that most of the existing gaze tracking systems share. First, by using Kinect we avoid the use of physical marking and corneal reflection that are very sensitive to distance and lighting condition. Second, with web camera and 3D reference point, we overcome the head movements and distance limitation, therefore the user can move and rotate his/her head freely in front of the camera. Third, since blinking is a physiological necessity for humans, our gaze estimator verifies eye state in order to detect: iris is visible or not. Fourth, we propose a discriminative Bayesian formalism for the estimation of eye-gaze mapping which eliminate individual calibration procedure in classical methods. Finally, proposed method use low cost devices (web camera and Kinect) for eye-gaze estimation.

3. Proposed method

The architecture of the proposed eye-gaze system includes Kinect sensor and Logitech HD web camera for acquisition of the same scene. The Logitech web camera is low cost camera that pans and tilts to automatically track the user face in the visible spectrum with high resolution images. The Microsoft Kinect is also a low cost peripheral, used as a game controller with the Xbox 360 game system. The basic principle of Kinect's depth calculation is based on stereo matching. The stereo matching requires an image to be captured by an infrared camera, and another to operate in the visible spectrum of the scene. In addition to these cameras, we use the head tracker system of Kinect that continuously computes the head location and orientation from depth data. Note that, location and orientation are in term of Kinect reconstructed geometry in Cartesian space as shown in Fig. 1. This system uses random regression forests to estimate the 3D head pose in real time (Fanelli, Weise, Gall1, & Gool, 2011). It basically learns a mapping between simple depth features and real-valued parameters such as 3D head position and rotation angles. Fig. 2 shows the location and orientation of a head estimated using Kinect. Since the RGB image of Kinect is in low resolution, another camera is concurrently used to detect observer's irises.

Our eye-gaze estimation method is based on the relative displacements of the irises in terms of the reference point. The reference point can be extracted via corneal reflection, eye's corner or physical mark on the face. Unfortunately, these methods need specialized hardware device, individual calibration or are not user friendly (Hansen & Ji, 2010). In contrast, the reference point can be simply measured via depth information of the head which we explain in the next section. As shown in Fig. 3, the displacement of the irises is quantified by the *RM* vector, where *R* and *M* are the reference point and the midpoint between the right and left irises center respectively. The *RM* vector is expressed by two components, d_x and d_y, which are measured along the horizontal and vertical axes. Thus eye-gaze can be estimated by comparing d_x and d_y components with predetermined threshold values where a person is looking at different target points. Here, we assumed that a person is stationary in a specific location. In Section 3.1 we extend our method to deal with face movement and rotation. The reference point is calculated by the depth Kinect images that is described in Section 3.4. We verify iris state in order to detect whether the iris is visible or not, then the iris center is calculated which are described in Sections 3.2 and 3.3 respectively.

3.1. Gaze estimation

The user's gaze point can be accurately estimated based on the extracted *RM* vector when the user does not move his/her head significantly. However, if the head changes its position and orientation, the eye-gaze method will fail to estimate the target point because the vector *RM* (Fig. 3) changes. To overcome this problem,

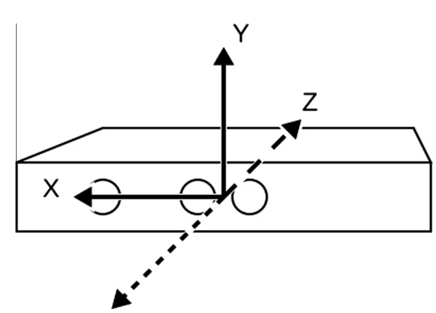

Fig. 1. Kinect coordinate system.