Information Sciences 320 (2015) 346–360

Contents lists available at ScienceDirect

Information Sciences

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

Real time gaze estimation with a consumer depth camera

Li Sun [a,*], Zicheng Liu [b], Ming-Ting Sun [c]

[a] College of Computer Science, Zhejiang University, Hangzhou, China
[b] Microsoft Research, Redmond, WA, United States
[c] Department of Electrical Engineering, University of Washington, Seattle, WA, United States

ARTICLE INFO

Article history:
Received 13 June 2014
Received in revised form 3 January 2015
Accepted 1 February 2015
Available online 16 February 2015

Keywords:
Gaze estimation
Eye tracking
Gaze direction
System calibration
Human–computer interaction

ABSTRACT

Existing eye-gaze-tracking systems typically require multiple infrared (IR) lights and high-quality cameras to achieve good performance and robustness against head movement. This requirement limits the systems' potential for broader applications. In this paper, we present a low-cost, non-intrusive, simple-setup gaze estimation system that can estimate the gaze direction under free head movement. In particular, the proposed system only uses a consumer depth camera (Kinect sensor) positioned at a distance from the subject. We develop a simple procedure to calibrate the geometric relationship between the screen and the camera, and subject-specific parameters. A parameterized iris model is then used to locate the center of the iris for gaze feature extraction, which can handle low-quality eye images. Finally, the gaze direction is determined based on a 3D geometric eye model, where the head movement and deviation of the visual axis from the optical axis are taken into consideration. Experimental results indicate that the system can estimate gaze with an accuracy of 1.4–2.7° and is robust against large head movements. Two real-time human–computer interaction (HCI) applications are presented to demonstrate the potential of the proposed system for wide applications.

1. Introduction

Eye gaze estimation is used to determine the gaze direction of a person, either the line of sight or the point of regard (PoR). Researchers are vigorously investigating the eye-gaze-tracking technology with the ultimate goal of providing low-cost, non-intrusive, easy-calibration/calibration-free, highly accurate and robust systems that can be applied for general public uses. Gaze-tracking technology is highly valuable, with abundant interactive and diagnostic applications, such as HCI, virtual reality, video oculography, eye disease diagnosis and human behavior studies. For example, a real-time driver vigilance monitoring system is described in [12] for monitoring various visual bio-behaviors. Because gaze direction reflects a person's intention, gaze-tracking systems represent a unique and effective tool for disabled people [16] and also demonstrate great potential for video gaming and gaze-contingent interactive graphic displays [29]. In addition, gaze-tracking technology has been used for image annotation [9] and remote collaboration [19,24]. Working together with 3D object retrieval technique [7], gaze-tracking can provide comfortable interaction experience.

Despite numerous applications, existing gaze-tracking systems suffer from great system setup complexity, inflexible system configuration, the requirement of expensive devices (e.g., high quality cameras and lenses), cumbersome calibration

* Corresponding author.
E-mail addresses: lsun@zju.edu.cn (L. Sun), zliu@microsoft.com (Z. Liu), mts@uw.edu (M.-T. Sun).

http://dx.doi.org/10.1016/j.ins.2015.02.004
0020-0255/

procedures and low tolerance toward head movement, which hinders them from being widely used. In particular, gaze tracking is challenging due to the individuality of eyes, occlusion, and variation in scale, head poses, and light conditions.

In this paper, we present a low-cost, non-intrusive, simple-setup gaze estimation system allowing free head movement. The proposed gaze estimation system uses a single consumer depth camera (Kinect), which is easy to setup even for non-professional users. Based on a 3D geometric model, the gaze estimation system allows free head movement. With robust facial feature extraction and simple system calibration, the system works in realtime with good accuracy, which can be used for many applications. Our main contributions are as follows:

- A simple-setup real-time gaze estimation system using only a consumer depth camera.
- A 3D model-based gaze estimation method allowing free head movement.
- A simple camera-screen calibration method that is easy to carry out even for nonprofessional users.
- An iris center localization method that can handle relatively low-quality eye images.
- Experimental results demonstrating that the system can estimate gaze directions accurately (error 1.4–2.7°) under free head movement.
- Two real time HCI applications validating the effectiveness and efficiency of the proposed technique.

2. Related work

Most recent work on non-intrusive gaze tracking can be broadly classified into two categories: feature-based and appearance-based. In the following section, we briefly review each type of methods. For a detailed review on recent gaze-tracking techniques, we refer interested readers to [10].

2.1. Feature-based gaze estimation

Feature-based gaze estimation methods typically rely on extracting local features such as pupil/iris contours, eye corners and corneal reflections (glints generated by IR lights), which are related to gaze. Feature-based approaches can be further divided into two distinctive groups: 2D regression-based (mapping-based) [12,29,1,18,30,14,32] and 3D model-based [25,3,8,31,2,21]. 2D regression-based approaches focus on calibrating a gaze mapping function from the extracted 2D local eye movement features to the eye gaze (or PoR), whereas 3D model-based approaches calculate the 3D gaze direction by developing a 3D geometric model of the human eye.

For 2D regression-based gaze estimation methods, the most popular 2D local eye movement feature is the pupil center and corneal reflection (PCCR). Point of regard is estimated by tracking the relative position of PCCR, which are usually generated by dedicated IR lights. Many gaze-tracking systems [29,1,18,30] are based on the PCCR technique, most of which require the subjects to keep their heads still. Although these systems could achieve very high accuracy (error less than 1°), the calibrated regression function decays as the head moves away from the original calibration position [18]. The recently developed 2D regression-based methods [29,1,18,30,31] attempt to address the problem by compensating for the errors caused by the head movement. In [31], Zhu and Ji introduced a 2D regression-based method that incorporates 3D eye positions to compensate for the head movement. However, their method requires additional hardware. In [32], Zhu and Yang used the iris center and eye corner as regression variables. Its accuracy is in general lower than those of the PCCR systems, and the method is also sensitive to the head movement.

In contrast, 3D model-based gaze estimation methods enjoy the inherent advantage that the head movement is implicitly modeled by the 3D eye location. A 3D geometric eye model is used to determine the gaze direction for 3D model-based gaze-tracking systems, which is based on the anatomical structure of the eye [20] (see Fig. 1). Most 3D model-based methods require the calibration of the geometric relationships between the IR lights, the screen, and the camera. However, few work has described their calibration procedures. In [4], Francken et al. proposed a method to determine the screen's position and orientation using gray code reflections. The setup consists of an LCD screen, a digital camera, and a spherical mirror. The calibration procedure is conducted for two different sphere locations. In this paper, a simple screen-camera calibration method is described, which is easy to carry out for nonprofessional users.

In most 3D model-based gaze estimation methods, the gaze direction is estimated from the 3D cornea center (center of curvature of the cornea) and the pupil center. The glints generated by IR lights are usually used to derive the 3D cornea center, and the pupil/iris boundary is extracted from an image of the eye to find the 3D pupil/iris center. However, many 3D model-based gaze tracking systems are faced with the dilemma of trading off the head movement range for high-resolution eye images. Basically, a wide field of view is required to allow free head movement, but a narrow field of view is needed to obtain high-resolution eye images. There are generally two types of approaches to this problem, either by using multiple cameras or using stereo and active cameras (cameras mounted on motorized pan and tilt units or cameras together with mirrors mounted on motorized pan and tilt units), both of which increase the complexity and cost of a gaze-tracking system.

Recently, methods using a single camera without IR lights have been proposed [28,2]. In [28], Yamazoe et al. proposed to estimate gaze directions as 3D vectors connecting both the eyeball and the iris centers. However, their model did not consider the deviation of the visual axis from the optical axis, and the distance between the eye and the camera was fixed and assumed to be known. The method achieved an accuracy of 5° horizontally and 7° vertically. In [2], Chen and Ji used a 3D eye model to determine the gaze direction, where the deviation angle between the visual axis and the optical axis was modeled.