Signal Processing: *Image Communication* 36 (2015) 29–42

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

On visual gaze tracking based on a single low cost camera

Evangelos Skodras*, Vasileios G. Kanas, Nikolaos Fakotakis

Department of Electrical and Computer Engineering, University of Patras, Patras, Greece

ARTICLE INFO

Article history:
Received 26 November 2014
Received in revised form
9 April 2015
Accepted 27 May 2015
Available online 4 June 2015

Keywords:
Gaze estimation
Gaze tracker
Eye tracking
Human computer interaction
HCI

ABSTRACT

Gaze tracking technologies provide an unconventional way of human–computer interaction, envisaged to advance practical applications and industrial products in a multitude of fields. The success of such systems depends on selecting the best calibration setup and image features that correspond to a person's line of sight. The purpose of this study is to estimate eye gaze from a single, low cost web-cam, under natural lighting. Facial traits are extracted from the sensory data, from which distance vectors related to gaze are derived. Different experimental setups are studied to evaluate the robustness of the proposed method with respect to various calibration setups, camera position and head movements. The use of new additional features improves the modeling of the subtle eye movements in the vertical direction, while a new calibration setup is proposed that further enhances the performance. The results demonstrate that the proposed framework is able to track gaze with good accuracy, consolidating the use of inexpensive equipment and techniques towards an ever-expanding range of gaze tracking applications.

1. Introduction and motivation

The swift growth of functional sophistication in computing over the last decades has inevitably induced a growing interest in improving all aspects of interaction between humans and computers. This emerging field is gaining momentum for scientists across several different disciplines such as computer science, engineering, psychology and neuroscience. Along with speech, eye gaze comprises the most natural and comfortable means for human–computer interaction (HCI), giving rise to the ever-growing interest to develop systems that take advantage of gaze tracking technologies.

Eye gaze is defined as the direction of a person's line of sight, revealing a person's focus of attention. It comprises a significant source of information about the cognitive and affective state of human beings, providing implicit cues of intention and interest. As a control input, in conjunction with the standard input methods, gaze can greatly increase efficiency and usability. Gaze monitoring can be applied in a wide range of applications [1], including non-glasses type 3D technologies [2], monitoring of drivers attention and vigilance [3–5], visual attention analysis (*e.g.* for marketing purposes [6]), interactive gaze-based interfaces for disabled people [7], diagnostic purposes [8,9] and attentive HCI interfaces [10,11].

Despite active research in the field, ubiquitous gaze tracking is beyond the grasp of the current systems. The vast majority of research in academia and industry is directed towards gaze tracking using active light sources, *i.e.* infrared (IR) illumination, achieving high accuracy rates [12,13]. Hitherto, numerous commercial products making use of this well known approach are already on the market. However, active light approaches require dedicated hardware equipment which is usually high priced and the intrusiveness of which is controversial. Moreover, as they usually require a controlled environment to prevent undesired reflections in the eyes, their applicability during day time is precluded. Other common approaches employ 3D techniques [14] (using multiple cameras or depth sensors) and wearable devices such as helmets or glasses [15], being cumbersome for the users. Universal gaze tracking from completely unobtrusive, remotely located low-cost sensors (*e.g.* web-cams) still remains one of the most sought-after goals among researchers.

* Corresponding author. Tel.: +30 6972965909.
E-mail addresses: evskodras@upatras.gr (E. Skodras), vaskanas@upatras.gr (V.G. Kanas), fakotaki@upatras.gr (N. Fakotakis).

http://dx.doi.org/10.1016/j.image.2015.05.007
0923-5965/

Although web-cam based gaze trackers have so far demonstrated inferior performance compared to active light approaches, they are apt in applications where accuracy can be traded off for low cost, simplicity and flexibility. In this paper, a feature-based gaze tracking system is proposed, based on a single web-cam. Instead of the most common approach of detecting eye corners in every frame, in our framework, salient feature points serving as anchor points are utilized to extract discriminative features. Additional features corresponding to the vertical position of the eyelids are proposed in order to increase accuracy and robustness. Despite the number of gaze tracking approaches in the literature, a common basis is not yet established, since standard databases for evaluating gaze direction have only recently emerged. For this purpose, different calibration setups and the influence of head movements are extensively evaluated. Furthermore, a new arrangement of the calibration points is presented, contributing towards an even higher accuracy rate. In summary, the characteristics of the proposed system, which also correspond to the desired attributes of the entirety of gaze trackers, are the following:

- It estimates gaze with fairly high accuracy; this is suitable for most gaze tracking applications.
- It presents minimal intrusiveness and obstruction, given that only a single camera is required. Not requiring any special equipment, gaze tracking is not bound to a specific *ad hoc* computer specialized for this purpose but can run universally.
- The setup is straightforward and flexible. The procedure is autonomous (no expectation of manual initialization), any camera available can be used and is free from the need to zoom into the eyes or perform any other special actions.
- It works under different illumination conditions and is also independent of the special characteristics of each subject.
- It is demonstrated to function for lower resolution cameras, thus constituting a low-cost approach and offers the capability for real time processing.

The layout of the paper is organized as follows: in Section 2 a review of the literature in the area of natural lighting approaches (without IR illumination) and not using any wearable equipment is presented. The model used for gaze estimation is also defined in the same section, while in Section 3 an in-depth description of the proposed algorithm is given. Section 4 presents the experimental setup and Section 5 the obtained results, as well as a meticulous analysis of them. Finally, the discussion of the main points in Section 6 is followed by a recapitulation in Section 7 which concludes this work.

2. Related work and background

2.1. Gaze estimation methodologies

The problem of gaze estimation is an active research topic with several recent publications [16], the overwhelming majority of which are using hardware-based approaches (*i.e.* IR light sources, high-resolution cameras, multiple cameras and wearable equipment). The focus of the current overview is on approaches working under natural illumination, using a single, remotely located camera.

The current subset of gaze estimation methods can be subdivided into two broad categories, *i.e. feature-based* and *appearance-based* methods [16].

Feature-based methods use computer vision techniques to extract and track local eye features such as eye centers, corners and contours. The extracted features can be used to derive feature vectors which can be directly related to gaze. According to the approach used for relating the image features with gaze direction, *feature-based* methods can be further divided into *geometric* (or *model based*) and *interpolation based* (or *regression based*). *Geometric* methods [4,17–20] compute directly the gaze direction from the image features based on a geometric model of the eye, while *interpolation based* methods [21–26] built a mapping function between them, using parametric (*e.g.* polynomial) or non-parametric forms (*e.g.* neural networks).

Torricelli et al. [21] use image processing algorithms to extract and track the eye features and perform mapping with gaze direction using neural networks. In [22], Zhu et al. detect and track eye centers and corners using subpixel accuracy and a linear interpolation model for inferring gaze coordinates. The works in [17,18] use geometric models which rely upon facial feature tracking for estimating head pose and eye orientation. The feature vectors between eye centers and eye corners are used in [27,23,24] to derive gaze estimation through 2D interpolation mapping. A comparison between a common polynomial mapping function and a geometric model is performed in [28], giving a slight edge to the latter. Ishikawa et al. [4] use Active Appearance Models (AAM) to detect and track facial points and employ an edge-based algorithm for iris refinement. They subsequently employ geometric models to derive gaze direction combining head pose and eyeball orientation. Authors in [29] also use AAM for iris and eyelid tracking in order to derive gaze information. The gaze angle is geometrically defined combining head pose and eyes position information. In the work of Salam et al. [30] the head pose is derived using a 2.5D global AAM, while a multi-texture AAM is used for iris localization. The contribution of each eye to the final gaze direction is weighted depending on the detected face orientation. In general, global appearance-based methods are very robust in detecting the overall rough positions of the facial features. However, as they depend on the convergence of the full model (*i.e.* by satisfying a minimization function or reaching a maximum number of iterations), they do not ensure localization of each feature with high precision, thereby adversely affecting the gaze estimation accuracy. Chen and Ji in [19] also use a geometric model to localize facial points, manually extract pupil centers and build a 3D gaze estimation model, tailored with person-specific eye parameters. Heyman et al. [20] track the 3D pose of the head using Canonical Correlation Analysis and extract the positions of the irises using blob detection and 4-connected component labeling. Valenti et al. in [25] employ their eye detection approach based on isocenters, to also detect eye corners which are used as *anchor* points for gaze estimation. Given the vectors between the detected eye centers and eye corners they perform 2D linear mapping to screen coordinates. Their work is extended in [31] where pose