Image and Vision Computing 36 (2015) 51–60

Contents lists available at ScienceDirect

Image and Vision Computing

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

Precise localization of eye centers in low resolution color images[☆]

Evangelos Skodras [*], Nikos Fakotakis

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

ARTICLE INFO

Article history:
Received 30 March 2014
Received in revised form 18 December 2014
Accepted 6 January 2015
Available online 19 February 2015

Keywords:
Eye localization
Eye detection
Facial feature detection
Radial symmetry

ABSTRACT

The localization of eye centers and tracking of gaze constitutes an integral component of many human–computer interaction applications. A number of constraints including intrusiveness, mobility, robustness and high-price of eye tracking systems have hindered the way of the integration of eye trackers in everyday applications. Several 'passive' systems based on a single camera have been lately proposed in the literature, exhibiting though subordinate precision compared to the commercial, hardware-based eye tracking devices. In this paper we introduce an automatic, non-intrusive method for precise eye center localization in low resolution images, acquired from single low-cost cameras. To this end, the proposed system uses color information to derive a novel *eye map* that emphasizes the iris area and a radial symmetry transform which operates both on the original eye images and the *eye map*. The performance of the proposed method is extensively evaluated on four publicly available databases containing low resolution images and videos. Experimental results demonstrate great accuracy in challenging cases and resilience to pose and illumination variations, achieving significant improvement over existing methods.

1. Introduction

Eyes represent the most distinctive features of the human face, while their position and movements are a significant source of information about the cognitive and affective state of human beings; eyes hold a key role in expressing interest, intention and attention. Precise eye center localization constitutes the cornerstone for gaze monitoring and can be applied in an ever increasing range of applications [1]. These include face alignment and normalization, liveness detection (*e.g.* for security applications), non-glasses type 3D technologies, monitoring of drivers' attention and vigilance, visual attention analysis (*e.g.* for marketing purposes), attentive human–computer interaction (HCI) interfaces and interactive gaze-based interfaces for disabled people.

Although many commercial, off-the-shelf products for eye detection and tracking are available in the market, they all require dedicated, high-priced hardware. The most common approaches in research and commercial systems use active infrared (IR) illumination, to obtain accurate eye location through corneal reflection [2,3]. Although these approaches yield high-precision localization, their intrusiveness is controversial and they cannot be used outdoors, in daytime applications, due to ambient infrared illumination. Other hardware approaches require the use of special equipment such as contact lenses, special helmets and electrodes [4,5], causing discomfort to the users and introducing limitations, thus rendering them cumbersome for everyday applications. Algorithmic approaches constitute non-intrusive techniques which can be incorporated in many applications where the use of extra dedicated hardware is impracticable.

Despite active research in the field, eye center localization with high precision from completely unobtrusive and remotely located (*i.e.* not requiring special helmets, glasses or chin rests) image-based systems, remains a very challenging task. Eyes present great variability in shape and color depending on eye state (open/closed or anything in between), iris direction, facial expression, head pose and ethnicity. Occlusions caused by hair, glasses, reflections, shadows or pose (self-occlusions) make the localization process notably difficult. Furthermore, imaging conditions such as lighting, contrast, camera characteristics and further processing (*e.g.* compression) have a strong influence on how eyes appear in the image. The localization process becomes even more challenging when dealing with low resolution images derived from inexpensive imaging devices such as webcams, mobile devices or pinhole cameras.

In this paper, a fast, fully automatic method for accurate eye center localization in low resolution images and videos is presented. The goal of the proposed system is to locate eye centers accurately and robustly for HCI applications, reporting accuracies comparable to the commercial hardware-based eye trackers, with the use of a single low-cost camera.

The layout of the paper is organized as follows. In Section 2, a literature review in the area of eye localization is presented. In Section 3 a detailed description of the proposed algorithm is given. Section 4 presents the experimental setup and Section 5 the results obtained using the proposed algorithm, compared with the state-of-the-art systems for eye

---

[☆] This paper has been recommended for acceptance by Qiang Ji.

[*] Corresponding author at: Wire Communications Laboratory, Department of Electrical and Computer Engineering, University of Patras, GR-26 504, Rion, Patras, Greece. Tel.: +30 2610 969806.

E-mail addresses: evskodras@upatras.gr (E. Skodras), fakotaki@upatras.gr (N. Fakotakis).

http://dx.doi.org/10.1016/j.imavis.2015.01.006
0262-8856/

localization. The concluding Sections 6 and 7 report implications and inferences of the work presented in this paper.

2. Related work

Over the last decades, a great number of methods have been employed for the task of eye detection and tracking [6,7]. Eye localization methods, working under natural illumination and using a single camera, can be coarsely divided into two broad categories: *appearance-based* and *feature-based*.

Appearance-based methods, also known as holistic or image-based methods, incorporate eye knowledge implicitly by using the intensity distribution or filter responses of the eye area and its surroundings to train a system, using example datasets. They generally require a large amount of training data and powerful non-linear algorithms in order to learn the high variability of eyes. To this end, many machine learning algorithms have been employed, including neural networks [8,9], Bayesian models [10], hidden Markov models (HMMs) [11], support vector machines [12] and Adaptive Boosting (adaboost) [13,14]. Template matching has been used by Grauman et al. [15] to detect and track eyes by searching the image for the highest correlation with a moving template image. Though simple and straightforward, this method is prone to erroneous detections due to pose, facial expressions and other changes in the face appearance. Pentland et al. [16] were the first to extend the eigenface technique to describe facial features (*i.e.* eigeneyes, eigennoses, eigenmouths). Subspace methods report better results when compared to template matching; nevertheless their performance is largely dependent on the training set and fail to localize eyes precisely. A prevalent method for eye localization is to employ a cascade of boosted classifiers working with Haar features [14]. The main advantage of this approach is its high efficiency and computational speed. It suffers however from high false positive rates, and its discriminative capability may be limited in cases of challenging illumination conditions or pose variations. In a nutshell, although *appearance-based* methods can achieve remarkably high accuracy in detecting the eye area, they often fail to provide an accurate detection of the eye centers, a critical feature for gaze monitoring.

Feature-based methods make explicit use of *a priori* eye knowledge in order to derive features such as shape, geometry, color and symmetry. Geometrical information of edges has been widely used for eye detection [17,18], also combined with several other cues [19]. For detailed modeling of the eye shape, parametric models and complex shape-based methods have been employed. They achieve localization by constructing a generic eye model in which the eye is fitted through energy minimization; deformable-template models were proposed by Yuille et al. in [20] and widely used later on [21]. Despite the accuracy of these methods, they are computationally demanding, require high-resolution images, and a close to the eye initialization. A number of methods have been employed in order to model the circular shape of the iris using the Hough transform [22,23], however the circularity shape constraints render the method applicable only to frontal or near frontal faces in high resolution images. In order to overcome these issues, ellipse fitting algorithms have been also proposed [24,25]. Other popular techniques employed to localize the eye center by modeling the iris shape are the *Starburst* algorithm [26] and the *Integro Differential Operator* [27]. Hansen and Pece [28] also model the iris as an ellipse, fitting locally the ellipse to the image through an EM and RANSAC optimization scheme. The iris region is located in [29] using intensity differences between the center region and its neighboring regions using a contrast operator, and the detection accuracy is further enhanced using a Kalman tracker. Valenti et al. [30–33] propose a technique based on isophote curvatures and a voting process for real-time eye center detection with high accuracy, while a modified Mean Shift procedure is used for tracking. The desired features of the eyes can be enhanced using several filter responses [34]. To this end, Gabor filters have attracted much popularity [35,36]. However, filtering methods yield coarse estimates of the eye and usually additional techniques for finer localization are used. The idea of projection functions has been studied by Zhou and Geng [37] in order to locate exact iris centers. Experiments show that this method is sensitive to face orientation and lighting conditions. Symmetry operators have also been investigated, usually in combination with other techniques, for the purpose of automated eye detection [38–41]. Finally, color models of the eye have received very little attention [6] and color information has been solely used in order to distinguish the eye regions from the rest of the skin area [42–44], or for tracking the eye region [45].

The proposed method is based on a synergy of color and radial symmetry to precisely and robustly localize eye centers. The use of color is based on the finding that the color distribution of the eye and particularly the iris is consistently different to its surroundings. Symmetry uses as a basis the inherent radially symmetric brightness patterns of the iris and the pupil. The main contribution of the proposed method is the use of color information for accurately localizing eye centers rather than for defining the rough areas of the eye regions, which constitutes what was studied so far in the literature. The proposed method substantially differs to existing approaches as it uses chrominance information to build *eye maps* which distinguish and enhance the circular shape of the iris, leading to a highly radially symmetric pattern (Section 3.2). The effectiveness of using color information is supported by a statistical analysis of the color distribution of the skin, the eye areas and the irises. The novel *eye map* built contributes to a cumulative radial symmetry transform with the original eye region, demonstrating high accuracy even in lower resolution images.

3. Proposed system

The proposed eye center localization system is summarized as follows: once a face is detected in a given image, regions containing the eyes are defined. Color information is used to build an *eye map* which emphasizes the iris area. Subsequently, a radial symmetry transform is applied both to the *eye map* and the original eye image. The cumulative result of the transforms indicates the precise positions of the eye centers. The procedure is illustrated in Fig. 1; a detailed description of the different stages of the algorithm is given below.

3.1. Face detection and eye regions definition

The detection of faces in a given image is carried out using the real-time face detector proposed by Viola and Jones [46]. Based upon an ensemble of boosted cascade detectors working with Haar features, it represents the state-of-the-art method in face detection. Within each detected face a *Region of Interest* (*ROI*) containing each of the eyes is defined based on face geometry (Fig. 1). The dimensions of the *ROIs* are amply specified so as to contain the whole eye regions even when reaching the detection limits of the face detector, regarding the in-plane and out-of-plane rotations [46]. Therefore, the width and height of the eye regions are determined as $EyeRegionWidth = FaceWidth/3$ and $EyeRegionWidth = FaceHeight/4$. Subsequently, the proposed procedure is applied to each of the cropped eye *ROIs* in order to localize the exact positions of each eye center.

3.2. Eye map construction

3.2.1. Skin color model

The modeling of the human skin color requires an appropriate color space in which skin can be discriminated from any other area in the image, following a certain distribution in the color space. The *YCbCr* color space has been extensively used in the literature for modeling the skin regions [42,47], presenting significant advantages compared to other color spaces [42,48]. In *YCbCr* the luminance information (Y) is decoupled from the chrominance information (Cb, Cr), while the