Infrared Physics & Technology 85 (2017) 44–51

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

Eye pupil detection system using an ensemble of regression forest and fast radial symmetry transform with a near infrared camera

Mira Jeong, Jae-Yeal Nam, Byoung Chul Ko *

Department of Computer Engineering, Keimyung University, Sindang-dong, Dalseo-gu, Daegu 704-701, Republic of Korea

HIGHLIGHTS

- Using a near IR camera to detect eye pupil region.
- Using an ensemble of regression forest and fast radial symmetry transform.
- The pupil displacement is estimated to maintain the level of accuracy.
- Pupil detection accuracy is higher than those of related algorithms.

ARTICLE INFO

Article history:
Received 8 May 2017
Revised 27 May 2017
Accepted 27 May 2017
Available online 30 May 2017

Keywords:
Pupil center detection
Near infrared
Driver-state monitoring system
Fast radial symmetric transform
Weighted random forest regressor

ABSTRACT

In this paper, we focus on pupil center detection in various video sequences that include head poses and changes in illumination. To detect the pupil center, we first find four eye landmarks in each eye by using cascade local regression based on a regression forest. Based on the rough location of the pupil, a fast radial symmetric transform is applied using the previously found pupil location to rearrange the fine pupil center. As the final step, the pupil displacement is estimated between the previous frame and the current frame to maintain the level of accuracy against a false locating result occurring in a particular frame. We generated a new face dataset, called Keimyung University pupil detection (KMUPD), with infrared camera. The proposed method was successfully applied to the KMUPD dataset, and the results indicate that its pupil center detection capability is better than that of other methods and with a shorter processing time.

1. Introduction

Pupil detection for the human eye has an enormous number of applications in various industrial fields, including human-machine interaction (HCI) in games and entertainment biometric systems for use in iris recognition, and drowsiness estimation of drivers using a driver-state monitoring system (DSM). Among the several applications related to pupil detection, this research focuses on a DSM, because 2.4% of fatal crashes have been reported to have been caused by drowsy drivers [1] and pupil detection is the best solution for estimating a driver's inattention or drowsiness during a real driving situation.

Eye pupil detection can be classified into two categories based on the camera device: monocular- and IR-camera based detection. First monocular-camera based methods [2–6] have been studied particularly for HCI and biometric applications because a monocular camera is a nonintrusive device, unlike a head-mounted device, and it requires no additional illuminators. Sippl et al. [2] presented a real-time gaze-tracking algorithm for public display for estimating the focus of visual attention in pervasive advertising scenarios. Markus et al. [3] estimate a pupil location based on an ensemble of randomized regression. With this algorithm, a face-bounding box is first detected and eye regions are estimated using simple anthropometric relations. Each eye region is estimated using a chain of multi-scale tree ensembles. Zhang et al. [4] use a convolutional neural network (CNN) to learn the mapping, from a detected 2D face angle and eye image, to gaze angles in a normalized space for gaze direction estimation. Vrânceanu et al. [5] determine an eye-bounding box using an integral and edge image projection function. The eye components are retrieved, and the relative position of the iris is extracted, through a template matching mechanism. Kacete et al. [6] estimate a 2D pupil location in an image space from projected random forest outputs on the Hough space in a pyramidal manner through a merging process. Although these approaches show robust results for certain specific applications, some approaches [2–4] estimate the gaze direction based on both

* Corresponding author.
E-mail addresses: jeongmr@kmu.ac.kr (M. Jeong), jynam@kmu.ac.kr (J.-Y. Nam), niceko@kmu.ac.kr (B.C. Ko).

http://dx.doi.org/10.1016/j.infrared.2017.05.019
1350-4495/

the head pose and rough pupil location. In addition, these approaches are ineffective in environments with poor illumination, such as at nighttime or under uneven lighting.

Another approach adopts range infrared cameras with one or more illuminators. Several studies [7–10] have focused on estimating a pupil location by analyzing the difference in specular reflection between the corneal and pupil regions. Yan et al. [7] proposed a pupil center detection algorithm based on a radial symmetry transform and proved that, with an infrared image, the proposed algorithm can reduce the inference caused by asymmetric luminance, eyelashes, hair, or glasses. Gwon et al. [8] presented a gaze-tracking system using four illuminators and a near IR (NIR) camera. During the initial stage, their proposed system determines whether a user is wearing glasses by counting the number of pixels in an image captured using a low-exposure camera and by turning on or off the illuminators sequentially to reduce any noise occurring from reflections from the user's glasses. Vicente et al. [9] proposed a driver gaze-tracking system for use in a driving environment. This system computes the head pose and gaze direction using tracked landmarks and a 3-D face model. This system then detects when the driver's eyes are off the road using a 3-D geometric analysis. Jeong et al. [10] estimate a driver's head pose and gaze direction using a weighted random forest regressor (WRFR) consisting of five stages. With their WRFR, the initial position of each pupil is moved toward to the target position as the iteration steps are increased.

Although these algorithms use several illuminators and a low-exposure camera to overcome the limitations of an IR camera, reflections and noises from the surfaces of a user's glasses caused by ambient lighting or the illuminators themselves are large obstacles decreasing the accuracy of the gaze or pupil detection.

To overcome the limitations related to eye pupil detection under various changes in illumination, noises, and reflections from the user's glasses, we propose a combination of pupil regression based on a cascade regression forest and fast radial symmetry transform (FRST) [11]. For this study, we used an NIR camera instead of monocular camera to design a system invariant to changes in illumination, in which two illuminators are positioned, one at each end of the camera. The camera installation and dataset generation are described in tion 4. The major contributions of this paper are as follows:

- We first find four eye landmarks in each eye using a regression forest method, which is a modified version of that described in [12]. The initial positions of the two pupils are estimated as the centroid of the four eye landmarks.
- After the initial pupil centroids of both eyes are detected, a cascade local regression based on a random forest can be used to coarsely locate the pupil center.
- To detect the pupil center more accurately in noisy low-resolution images, FRST is applied based on the previously determined pupil center to all the real pupil center to be rearranged.
- To maintain the level of accuracy against a false locating result occurring in a single frame, pupil displacement between the previous and current frames is computed based on the individual eyes. If the two pupil displacements are different, only a random forest regressor is applied again based on the previous pupil center.
- The proposed method was successfully applied to the KMUPD dataset, and we confirmed that its tracking accuracy is higher than that of other related methods and with a shorter processing time.

The remainder of this paper is organized as follows. In Section 2, detection of eye landmarks and the initial pupil center are described. In Section 3, a stepwise eye pupil refinement algorithm using WRFR and FRST is presented. In Section 4, we introduce the KMUPD followed by a discussion of the experimental results of the proposed algorithms as compared with the results of previous related methods. Finally, in Section 5, we provide some concluding remarks as well as areas of future work.Fig. 1

2. Detection of eye landmarks and initial pupil center

In previous researches [3,6], face detection has been conducted in the first stage, followed by eye detection within the face region using a similar learning algorithm. However, an eye detector cannot detect the eye region exactly in the face area because it is sensitive to the head pose or changes in illumination. To detect the eye region more accurately, we adopt a facial landmark detection algorithm based on the local WRFR with a global face model (GFM) [12]. In this study, facial landmarks are classified as stable or unstable; seven stable points, whose positions are not changeable against occlusions or differing facial expressions, and 12 unstable landmarks, which are changeable based on each particular facial expression. After the initial landmark is determined based on the face detection and mean face matching, the WRFR can be used to locate the facial landmarks. To train the WRFR, the training dataset is divided into two subsets, in-of-bag (IOB) and out-of-bag (OOB) subsets. An IOB subset is selected using a bagging method to build each tree, whereas an OOB subset is used to evaluate the regression capability of the tree learning from the IOB subset, and then generates the weight of each tree based on its accuracy. After training the WRFR, each leaf node of a decision tree predicts and stores a 2D offset vector using training samples in the leaf node. After the regression is applied for individual landmarks, the detected landmarks are refined using the GFM that was constructed based on the spatial relations of the different landmarks accounting for the facial variations. From the final facial landmarks (Fig. 2(a)), we only extract eight landmarks around the eye regions (Fig. 2(b)). From the landmarks of each eye, the center location of the four eye corners is estimated as the initial pupil center, as shown in Fig. 2(c).

3. Stepwise pupil center refinement

After the initial pupil locations of the two eyes are estimated, the center of the pupil should be accurately detected. To detect the pupil center concisely, we combine two algorithms in a sequential manner; first, the WRFR [12] is applied to the initial pupil location for a coarse estimation of the pupil center, and based on the rough location determined through the first approach, FRST is then used to refine the location of the regressor.

3.1. WRFR for detecting coarse pupil center

As a similar approach to that in [10], a local patch centered on the initial pupil center is created. The scale of this local patch gradually decreases from stage 0 to stage 2, which is different from facial landmark detection, to reduce the computation time. At every stage, a patch is divided into 4×4 non-overlapping blocks. The scale of a patch at stage 0 is 90 pixels $\times$ 90 pixels, at stage 1 is 60 pixels $\times$ 60 pixels, and at stage 2 is 40 pixels $\times$ 40 pixels. We extract eight-dimensional oriented center symmetric-local binary patterns (OCS-LBP) [13] from each block in a patch, and these features of each stage are concatenated, generating the splitting functions of the decision tress of the WRFR. In addition, random sampling consensus (RANSAC) [14] is adopted during the middle of the tree generation to remove the outliers included in the sample data.

A summary of the WRFR training used for estimating the coarse pupil center is described in Algorithm 1: