Image and Vision Computing 57 (2017) 147–164

Contents lists available at ScienceDirect

Image and Vision Computing

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

Eye detection in a facial image under pose variation based on multi-scale iris shape feature☆

Hyunjun Kim, Jaeik Jo, Kar-Ann Toh, Jaihie Kim *

School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, Republic of Korea

ARTICLE INFO

Article history:
Received 17 July 2015
Received in revised form 23 August 2016
Accepted 3 October 2016
Available online 16 November 2016

Keywords:
Eye detection
Eye localization
Facial feature detection
Eye verification

ABSTRACT

The accurate location of eyes in a facial image is important to many human facial recognition-related applications, and has attracted considerable research interest in computer vision. However, most prevalent methods are based on the frontal pose of the face, where applying them to non-frontal poses can yield erroneous results.
In this paper, we propose an eye detection method that can locate the eyes in facial images captured at various head poses. Our proposed method consists of two stages: eye candidate detection and eye candidate verification. In eye candidate detection, eye candidates are obtained by using multi-scale iris shape features and integral image. The size of the iris in face images varies as the head pose changes, and the proposed multi-scale iris shape feature method can detect the eyes in such cases. Since it utilizes the integral image, its computational cost is relatively low. The extracted eye candidates are then verified in the eye candidate verification stage using a support vector machine (SVM) based on the feature-level fusion of a histogram of oriented gradients (HOG) and cell mean intensity features.
We tested the performance of the proposed method using the Chinese Academy of Sciences' Pose, Expression, Accessories, and Lighting (CAS-PEAL) database and the Pointing'04 database. The results confirmed the superiority of our method over the conventional Haar-like detector and two hybrid eye detectors under relatively extreme head pose variations.

1. Introduction

Since the eyes are the salient features of a human face, the detection and localization of the eyes are necessary processes in various face- and eye-related applications. In such face-related applications as face recognition [1,2], age estimation [3], and three-dimensional (3D) facial reconstruction [4], the positioning of the eyes is a crucial step in the alignment of facial feature points and for face shape normalization. Other eye-related applications, such as driver behavior analysis [5,6], gaze estimation [7,8], and iris recognition [9], also require the detection of eye position for accurate decision making.

According to Refs. [10,11,12], existing methods for eye detection are based on (i) the measurement of eye characteristics, (ii) the learning of a statistical appearance model, and (iii) the exploitation of structural information.

Methods based on the *measurement of eye characteristics* use intuitive visual characteristics, such as shape, the difference in intensity between the iris and neighbor regions, or reflections of the eye in infrared images, as templates for detection. These methods have the advantage of being simple and fast in implementation because they use intuitive algorithms. Jing et al. [13] detected eye-like regions by using the image contrast between an eye and neighboring areas of the skin, and localized the position of the eyes by projecting the image in the horizontal and vertical directions. San et al. [14] proposed the detection of eyes based on color information and symmetric characteristics of the face. Skodras et al. [15] achieved eye center localization in a low-resolution image by using color and radial symmetry information. Jian and Lam [16] used a salient map to detect eyes with strong saliency in the face region. Takano et al. [17] used gradient information regarding the iris and the eyelid to detect eyes. Ito et al. [18] used Hough transformations to detect circular border of the iris of the eyes. Kim et al. [19] proposed eye detection using iris shape features. Horng et al. [6] developed an eye tracking method using template matching and edge gradients. Zhao and Grigat [20] proposed an eye detection method that uses the difference in reflection properties between the pupil and the iris using active infrared light.

However, most above methods were developed in the early stages of eye localization research, and their limitations have become more pronounced with the consideration of increasingly complex and uncontrolled conditions, where the characteristics on which these methods depend may not be reliably measurable [10].

☆ This paper has been recommended for acceptance by Matthew Turk.

* Corresponding author.
E-mail addresses: kimhj8574@gmail.com (H. Kim), jaeik@yonsei.ac.kr (J. Jo), katoh@yonsei.ac.kr (K.-A. Toh), jhkim@yonsei.ac.kr (J. Kim).

http://dx.doi.org/10.1016/j.imavis.2016.10.003
0262-8856/

Methods that using *learning a statistical appearance model* use a statistical model of a photometric appearance feature from eye patch images as template for detecting eyes. The statistical appearance model can use eye information, which may be ignored or may not be measurable by the intuitive approaches mentioned above [10]. With regard to representative methods, Viola and Jones [21] proposed an object detection framework that uses Haar-like features with an AdaBoost classifier. This algorithm yielded both a high detection rate and a high frame rate by combining Haar-like features and an integral image. Mu [22] further improved the eye detection performance of Viola and Jones' method by selecting appropriate Haar-like features for eye detection. Chen and Liu [23] developed discriminating Haar-like feature-based eye detection, and Kroon et al. [24] used local binary patterns (LBPs) to localize the position of the eyes. Wang et al. [25] proposed an eye detection method using the Log-Gabor transform. Monzo et al. [26] proposed an eye localization method using the histogram of oriented gradients (HOG). The eye region was detected in this method by using the Viola–Jones detector, and eye position was localized using the HOG. Savakis et al. [27] reduced the number of dimensions in the HOG by using principal component analysis (PCA) for eye detection.

In general, statistical model-based methods struggle to detect the position of the eyes in cases involving variation in head pose because these methods use 2D features, and do not consider 3D information. Efraty et al. [28,29] overcame this problem by using a number of detectors trained for various poses. However, the training of detectors for each pose is inconvenient, and incurs high computational cost.

Methods involving the *exploitation of structural information* use structural information, such as spatial information of facial feature points, for eye detection. These methods are typically used in combination with statistical modeling techniques to detect initial feature points. The Active Shape Model (ASM) [30], the Active Appearance Model (AAM) [31], the Constrained Local Model (CLM) [32], and the Pictorial Structure Model (PS) [33] are representative techniques developed under this approach. Liang et al. [34] proposed a component-based discriminative search method that first detects initial feature points, and fits an ASM face model to localize accurate facial feature points.

In general, structural information-based methods can accurately detect facial points even if there is an error in the detection of the initial points. However, if the error in initial points' detection is large, model fitting may fall in a local minimum because ASM model fitting uses iterative search. Moreover, these methods have relatively high computational costs because the number of feature points is large and iterative methods are used. Table 1 shows a comparison of the methods proposed according to each of the three approaches.

In order to overcome the difficulties described above, we propose in this paper an eye detection method tolerant to head pose variation. Our method first detects eye candidates from face images in various head poses based on multi-scale iris shape features. The size of an iris in a face image varies with head pose, but our proposed multi-scale features detect eyes better than those of existing methods under changes of head pose.

Following this, the detected eye candidates are verified by using feature-level fusion of the histogram of oriented gradients (HOG) and cell mean intensity (CMI) features. The HOG uses shape information from edge gradient images, and the CMI features use the mean values of regional pixel intensity. Since these two sets of features represent independent characteristics of the eyes in the facial image, eye detection performance can be improved when these features are combined in feature-level fusion. The fused feature is used to train a support vector machine (SVM) classifier, and the eye candidates are classified into eye and non-eye candidates. To show the effectiveness of feature-level fusion, eye detection performance following fusion is compared with that when only the HOG or CMI features are used. To exhibit the robustness of our method against head pose variation, we performed an experiment on the Chinese Academy of Sciences' Pose, Expression, Accessories, and Lighting (CAS-PEAL) database and the Pointing'04 database.

The contributions of this paper are twofold. First, our proposed method can detect eye candidates for various head pose images using multi-scale iris shape features. Second, by using feature-level fusion of the HOG and CMI features, the eye candidate verification process exhibits better performance than methods that use a single feature without additional computational cost.

The rest of this paper is organized as follows: in Section 2, we describe our proposed method for eye candidate detection and eye candidate verification. Section 3 is devoted to a description of our experimental setup, such as the profile of the databases used in our experiments and the evaluation method employed. In Section 4, we report the experimental results of multi-scale fusion, eye verification, and the overall performance of our method given head pose variation. Some concluding remarks are provided in Section 5.

2. Proposed method

2.1. Flow chart of the proposed method

Fig. 1 shows the flow of the proposed method. The face region is first detected from an input image using the Viola–Jones face detector [21]. Both frontal and profile pose face detector are used to detect the face under various head pose changes. Eye candidates are then extracted from the detected face region. In order to attain eye detection tolerant to head pose variation, we propose a multi-scale iris shape feature. Here, an integral image is used to accelerate feature calculation. At this stage, our goal is to not miss any real eye (high recall rate) rather than accurately detecting eye (precision), even at the expense of a high false detection rate. Subsequently, a verification process eliminates incorrectly detected eye candidates. The verification process adopts an SVM using the HOG and CMI features as input.

Table 1
Comparisons of past eye detection methods.

Category	Measurement of eye characteristics	Learning of a statistical appearance model	Exploitation of structural information
Representative methods	– Skin color [13,14,15], projection [13], symmetry [14], salient map [16], iris shape feature [19], edge gradient [17], infrared [20]	– Haar-like [21,22,23], LBPs [24], Gabor [25], HOG [26,27], multi-classifier [28,29]	– ASM [30,34], AAM [31], CLM [32], PS [33]
Strengths	– Simple and fast implementation – Low computational complexity	– Can obtain eye feature information that may not be measurable by intuitive methods	– Accurate localization based on facial structure information
Weaknesses	– Feasible only under designed environments – SOME algorithms require additional hardware, such as IR devices.	– Vulnerable to head pose variations – Needs a complicated combination of detectors trained for each pose in order to resolve pose variation issues – Requires training data	– Requires an initial points' detection process – Accuracy can degrade when face model fitting process fall into a local minimum – High computational cost for large numbers of feature points due to an iterative search process – requires training data