Knowledge-Based Systems 000 (2017) 1–9

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Learning a gaze estimator with neighbor selection from large-scale synthetic eye images

Yafei Wang [a,b], Tongtong Zhao [b], Xueyan Ding [b], Jinjia Peng [b], Jiming Bian [a], Xianping Fu [b,*]

[a] School of Physics and Optoelectronic Engineering, Dalian University of Technology, Dalian 116024, China
[b] Information Science and Technology College, Dalian Maritime University, Dalian 116026, China

ARTICLE INFO

Article history:
Received 11 July 2017
Revised 2 October 2017
Accepted 6 October 2017
Available online xxx

Keywords:
Gaze estimation
Neighbor selection
Learning-by-synthesis
Cross-subject

ABSTRACT

Appearance-based gaze estimation works well in inferring human gaze under real-world condition. But one of the significant limitations in appearance-based methods is the need for huge amounts of training data. Eye image synthesis addresses this problem by generating huge amounts of synthetic eye images with computer graphics. To fully use the large-scale synthetic eye images, a simple-but-effective appearance-based gaze estimation framework with neighbor selection is proposed in this paper. The proposed framework hierarchically fuses multiple *k*-NN queries (in head pose, pupil center and eye appearance spaces) to choose closest samples with more relevant features. Considering the structure characters of the closet samples, neighbor regression methods then can be applied to predict the gaze directions. Experimental results demonstrate that the representative neighbor regression methods under the proposed framework achieve better performance for within-subject and cross-subject gaze estimation.

1. Introduction

The eye gaze can represent feelings and desire, reveal human attention and play an important role in social communication. Therefore, gaze estimation has been extensively applied to better understand human activities in many low-resolution application scenarios such as human computer interaction and driver attention analysis [1].

Predicting gaze direction from low-resolution eye images in the wild is still a challenging problem. Traditional desktop-based gaze estimation methods are achieved by analyzing eye image in the laboratory condition which has video camera with one or more light sources and chinrest to fix user's head, but these methods can't be applied to people's daily use in natural environment.

Considering the conditions of natural light and large head rotation, appearance-based gaze estimation which is a data-driven method using image features and can work in real-world settings with single webcam. However, the significant limitation of many appearance-based methods is that person- and session- specific training are always required. Besides, to get satisfied accuracy of gaze estimation, large amount of training data are required, which means the process of data collection and annotation has to cost a lot of labor and time.

Eye image synthesis solves these problem by generating a large amount of training data with computer graphics automatically. Therefore gaze estimation can be done in a learning-by-synthesis way by taking advantage of synthetic eye images under different illumination changes.

This work mainly focusses on gaze estimation with neighbor selection from large scale synthetic eye image datasets. The contributions of this paper are three-fold:

- Different from previous nearest neighbor selection frameworks (in only eye appearance space, or in head pose and eye appearance space), a novel calibration-free appearance-based gaze estimation framework is proposed, with corresponding features neighbor selection (in head pose, pupil center, and eye appearance spaces). By combining head pose information and pupil center feature, the gaze estimation method is person- and head pose- independent, and has more relevant eye appearance features in neighbor space.
- To the best of our knowledge, both under the proposed gaze estimation framework and the previous frameworks, it is the first time that the representative neighbor regression methods (between eye appearance space and gaze angle space) are divided into two types and analysed on the public datasets. Experimental results show that the proposed framework improves their gaze estimation accuracy on within-subject and cross-subject test.
- Additionally, in order to detect pupil center accurately on low-resolution eye images, a pupil center localization method using im-

* Corresponding author.
E-mail addresses: wangyafei@dlmu.edu.cn (Y. Wang), fxp@dlmu.edu.cn (X. Fu).

https://doi.org/10.1016/j.knosys.2017.10.010
0950-7051/© 2017 Elsevier B.V. All rights reserved.

age thresholding and region pruning is presented, and it works well even on eye images with large head movement.

The rest of this paper is organized as follows. Section 2 introduces the recent appearance-based gaze estimation works. In Section 3, the neighbor selection framework is described with features extraction. Section 4 gives the neighbor regression methods. Several experiments are shown in Section 5. Finally, Section 6 gives a brief conclusion.

2. Related works

Model-based Gaze Estimation & Pupil Center Localization:

Generally, there are two main categories of eye gaze estimation methods: model-based and appearance-based [2]. Model-based methods aim to identify local distinctive features of the eyes. The pupil center and cornea reflections are commonly used for gaze estimation. Using high-resolution cameras and some other specialized hardware such as synchronized camera and light sources for calibration, more precise geometry properties can be extracted and more gaze estimation accuracy it will be.

However, these model-based methods are not stable to estimate the eye center in difficult scenarios. So, pupil center localization methods are often need to be invariant to changes in scale, pose, contrast and variations in illumination. In order to find the pupil center from eye images under natural light, Fridman et al. [3,4] introduced a method based on cumulative distribution function statistical intensity threshold. This method dynamically optimized for each eye image, and extracted the largest circular blob from the image to detect the pupil center. Relying on analyse the gradients of the eye region, Timm and Barth [5] searched for circular shapes and obtained the most probable pupil center. The pupil center is defined as the location with maximum value of objective function and intersecting most gradient vectors in the gradient filed. However, the performance of the algorithm degrades when edge detection fails in low-resolution images.

Appearance-based Gaze Estimation & Eye Image Synthesis:

Appearance-based methods are believed to work better in the wild. These methods map the entire eye image to a high-dimensional space (as a high-dimensional input feature). Then a regression model of gaze can be constructed to obtain the position of gaze or a two-dimensional rotation angle when using new eye images as input. Compared to model-based methods, appearance-based methods do not require any dedicated hardware, and show good robustness to outliers.

Appearance-based methods require large amounts of training data. Eye image synthesis approaches were proposed to solve this problem. There are four main categories of eye image synthesis methods: Optical Flow [6,7], 3D Eye Reconstruction [8,9], Eyeball Model method [10] and GANs (Generative Adversarial Networks) [11].

Synthetic images are complementary to geometrical models in some ways. The eye image synthesis process in [6] used 1D flows to simulate the appearance distortion caused by head pose moving. Wang et al. [7] introduced a 2D interpolation to synthesize the eye appearance variation caused by eyeball moving. These optical flow methods treat eye image synthesis as optical shift of original image and could not be utilized under large head rotation. Generating eye images by 3D eye reconstruction highly depends on the pre-trained face 3D model. Sugano et al. [8] recovered multi-view eye images from 3D shapes of eye region reconstructed from 8 cameras image capture system. While Wood et al. [9] relied on high-quality head scans to collect high-resolution eye images. In order to generate multi-part eye images, Wood et al. [10] also presented a morphable model of the facial eye region, as well as an anatomy-based eyeball model. Eyeball Model method tunes parameters to obtain high-resolution eye images, which coincide with the groundtruth situation. Shrivastava et al. [11] used GANs to generate synthetic eye images and learnt a refiner model that improved the realism of these synthetic images. However, GANs generates different synthetic images when the input image is same. Moreover, it is still not controllable to generate image with specific gaze angle.

Learning-by-synthesis:

With augmented synthetic eye images, learning-by-synthesis appearance-based gaze estimation can train a strong model more easily. Wood et al. [12] presented a novel method to synthesize large amounts of variable eye region images as training data, which addressed the limitation of learning-by-synthesis with respect to the appearance variability and the head pose and gaze angle distribution. Wang et al. [7] proposed an appearance-based gaze estimation method by supervised adaptive feature extraction and hierarchical mapping model, during which appearance synthesis method is proposed to increase the sample density. Zhang et al. [13] introduced a CNN-based gaze estimation method, which concatenated head pose vector in the hidden layer of neural network. This change improved the performance of CNN-based gaze estimation training by synthetic image dataset. Sugano et al. [8] presented a learning-by-synthesis approach for appearance-based gaze estimation and trained a 3D gaze estimator by Random Forests (RF). And in their experiments, k-Nearest Neighbors was selected as comparison, from which we can see that k-NN regression estimators can perform well with a large amount of dense training samples.

3. Proposed method with neighbor selection

Many existing appearance-based gaze estimation methods usually infer gaze via two steps: image appearance extraction from aligned eye image, and gaze prediction by non-linear regression. Among them, k-Nearest Neighbors, which is used to predict gaze with the mean of neighbor samples' gaze angles, has become a baseline method in appearance-based gaze estimation, and gets preferable accuracy in previous works [8,12]. Considering the structure characters of neighbor samples, this paper presents a gaze estimation framework which is a combination of neighbor selection (with pre-extracted features in different feature space) and neighbor regression (between appearance space and gaze angle space), as shown in Fig. 1.

Compared with the traditional gaze estimation methods, the proposed framework is practical and takes full use of head pose information, pupil center information and eye appearance. Moreover, the neighbor regression methods under this framework don't need any gaze calibration procedure, and are head pose- and person- independent.

The proposed gaze estimation framework consists of three steps:

Firstly, given a test eye image and pre-extracted corresponding features, like head pose information, pupil center feature and eye appearance feature. Head pose information [14] and pupil center feature [15] are commonly extracted and used with eye appearance features to build a gaze estimator.

Then, find the closest set of sample. To achieve this goal, the proposed method conducts joint k-NN searching in head pose, pupil center and eye appearance spaces to find closest samples in these spaces.

Finally, neighborhood-based gaze estimation can be done on the closet set of sample by building model between the appearance space and gaze angle space.

Generally, there are two types of gaze prediction, one is to learn the mapping between eye appearance and gaze coordinates using neighbor samples, another one is to calculate reconstruction coefficients corresponding to the neighbor samples in appearance space and to interpolate the test point in gaze angle space with the reconstruction coefficients.

Please cite this article as: Y. Wang et al., Learning a gaze estimator with neighbor selection from large-scale synthetic eye images, Knowledge-Based Systems (2017), https://doi.org/10.1016/j.knosys.2017.10.010