Image and Vision Computing 71 (2018) 17–24

Contents lists available at ScienceDirect

Image and Vision Computing

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

# Hybrid eye center localization using cascaded regression and hand-crafted model fitting☆

Alex Levinshtein[a,*], Edmund Phung[a], Parham Aarabi[a,b]

[a] ModiFace Inc., Toronto, Canada
[b] Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada

ARTICLE INFO

*Article history:*
Received 23 June 2017
Received in revised form 8 January 2018
Accepted 10 January 2018
Available online xxxx

*Keywords:*
Eye center detection
Eye tracking
Facial feature alignment
Cascaded regression

ABSTRACT

We propose a new cascaded regressor for eye center detection. Previous methods start from a face or an eye detector and use either advanced features or powerful regressors for eye center localization, but not both. Instead, we detect the eyes more accurately using an existing facial feature alignment method. We improve the robustness of localization by using both advanced features and powerful regression machinery. Unlike most other methods that do not refine the regression results, we make the localization more accurate by adding a robust circle fitting post-processing step. Finally, using a simple hand-crafted method for eye center localization, we show how to train the cascaded regressor without the need for manually annotated training data. We evaluate our new approach and show that it achieves state-of-the-art performance on the BioID, GI4E, and the TalkingFace datasets. At an average normalized error of $e < 0.05$, the regressor trained on manually annotated data yields an accuracy of 95.07% (BioID), 99.27% (GI4E), and 95.68% (TalkingFace). The automatically trained regressor is nearly as good, yielding an accuracy of 93.9% (BioID), 99.27% (GI4E), and 95.46% (TalkingFace).



## 1. Introduction

Eye center localization in the wild is important for a variety of applications, such as eye tracking, iris recognition, and more recently augmented reality applications for the beauty industry enabling virtual try out of contact lenses. While some techniques require the use of specialized head-mounts or active illumination, such machinery is expensive and is not applicable in many cases. In this paper, we focus on eye center localization using a standard camera. Approaches for eye center localization can be divided into two categories. The first, predominant, category consists of hand-crafted model fitting methods. These techniques employ the appearance, such as the darkness of the pupil, and/or the circular shape of the pupil and the iris for detection [1–10]. These methods are typically accurate but often lack robustness in more challenging settings, such as low resolution or noisy images and poor illumination. More recently, a second category emerged - machine learning based methods. While there are approaches that train sliding window eye center detectors, recent success of cascaded regression for facial feature alignment [11–14] has prompted the community to apply these methods for eye center localization [15–17]. These new methods have proven to be more robust, but they lack the accuracy of the model fitting approaches and require annotated training data which may be cumbersome to obtain.

In this paper, we propose a novel eye center detection method that combines the strengths of the aforementioned categories. In the literature on facial feature alignment, there are two types of cascaded regression methods, simple cascaded linear regressors using complex features such as SIFT or HoG [14,18] and more complex cascaded regression forests using simple pairwise pixel difference features [11,13]. Our first contribution is a new method for eye center localization that employs complex features and complex regressors. It outperforms simple regressors with complex features [17] and complex regressors with simple features [15,16]. Similar to [15,16] our method is based on cascaded regression trees, but unlike these authors, following [11–14], our features for each cascade are anchored to the current eye center estimates. Moreover, based on the pedestrian detection work of Dollar et al. [19] we employ more powerful gradient histogram features rather than simple pairwise pixel differences. Finally, while the aforementioned eye center regressors bootstrap the regression using face or eye detectors, given the success of facial feature alignment methods we make use of accurate eye contours to initialize the regressor and normalize feature

☆ This paper has been recommended for acceptance by Hugo Proenca.
* Corresponding author.
*E-mail address:* alex@modiface.com (A. Levinshtein).

https://doi.org/10.1016/j.imavis.2018.01.003
0262-8856/

locations. We show that the resulting method achieves state-of-the-art performance on BioID [20], GI4E [21], and TalkingFace [22] datasets.

The proposed cascaded regression approach is robust, but suffers from the same disadvantages of other discriminative regression-based methods. Namely, it is inaccurate and requires annotated training data. To make our approach more accurate, in our second contribution we refine the regressor estimate by adding a circle fitting post-processing step. Employing robust estimation and prior knowledge of iris size facilitates sub-pixel accuracy eye center detection. We show the benefit of this refinement step by evaluating our approach on GI4E [21] and TalkingFace [22] datasets, as well as performing qualitative evaluation.

Finally, for our third contribution, rather than training our regressor on manually generated annotations we employ a hand-crafted method to generate annotated data automatically. Combining recent advances of eye center and iris detection methods, we build a new hand-crafted eye center localization method. It performs well compared to its hand-crafted peers, but is inferior to regressor-based approaches. Despite the noisy annotations generated by the hand-crafted algorithm, we show that the resulting regressor trained on these annotation is nearly as good as the regressor trained on manually annotated data. What is even more unexpected is that the regressor performs much better than the hand-crafted method used for training data annotation.

In summary, this paper proposes a new state-of-the-art method for eye center localization and has three main contributions that are shown in Fig. 1. First, we present a novel cascaded regression framework for eye center localization, leading to increased robustness. Second, we add a circle fitting step, leading to eye center localization with sub-pixel accuracy. Finally, we show that by employing a hand-crafted eye center detector the regressor can be trained without the need for manually annotated training data. This paper builds on the work that was presented in a preliminary form in [23].

## 2. Related work

The majority of eye center localization methods are hand-crafted approaches and can be divided into shape and appearance based methods. In the iris recognition literature there are also many segmentation based approaches, such as methods that employ active contours. An extensive overview is given by Hansen and Li [24]. Shape-based techniques make use of the circular or elliptical nature of the iris and pupil. Early methods attempted to detect irises or pupils directly by fitting circles or ellipses. Many techniques have roots in the iris recognition and are based on the integrodifferential operator [25] . Others, such as Kawaguchi et al. [26], use blob detection to extract iris candidates and use Hough transform to fit circles to these blobs. Toennies et al. [27] also employ generalized Hough transform to detect irises, but assume that every pixel is a potential edge point and cast votes proportional to gradient strength. Li et al. [5] propose the Startburst algorithm, where rays are iteratively cast from the current pupil center estimate to detect pupil boundaries and RANSAC is used for robust ellipse fitting.

Recently, some authors focused on robust eye center localization without an explicit segmentation of the iris or the pupil. Typically, these are either voting or learning-based approaches. The method of Timm and Barth [8] is a popular voting based approach where pixels cast votes for the eye center based on agreement in the direction of their gradient with the direction of radial rays. A similar voting scheme is suggested by Valenti and Gevers [9], who also cast votes based on the aforementioned alignment but rely on isophote curvatures in the intensity image to cast votes at the right distance. Skodras and Fakotakis [6] propose a similar method but use color to better distinguish between the eye and the skin. Ahuja et al. [1] improve the voting using radius constraints, better weights, and contrast normalization.

The next set of methods are multistage approaches that first robustly detect the eye center and then refine the estimate using circle or ellipse fitting. Świrski et al. [7] propose to find the pupil using a cascade of weak classifiers based on Haar-like features combined with intensity-based segmentation. Subsequently, an ellipse is fit to the pupil using RANSAC. Wood and Bulling [10], as well as George and Routray [4], have a similar scheme but employ a voting-based approach to get an initial eye center estimate. Fuhl et al. propose the Excuse [2] and Else [3] algorithms. Both methods use a combination of ellipse fitting with appearance-based blob detection.

While the above methods are accurate, they still lack robustness in challenging in-the-wild scenarios. The success of discriminative cascaded regression for facial feature alignment prompted the use of such methods for eye center localization. Markuš et al. [15] and Tian et al. [16] start by detecting the face and initializing the eye center estimates using anthropometric relations. Subsequently, they use a cascade of regression forests with binary pixel difference features to estimate the eye centers. Inspired by the recent success of the SDM

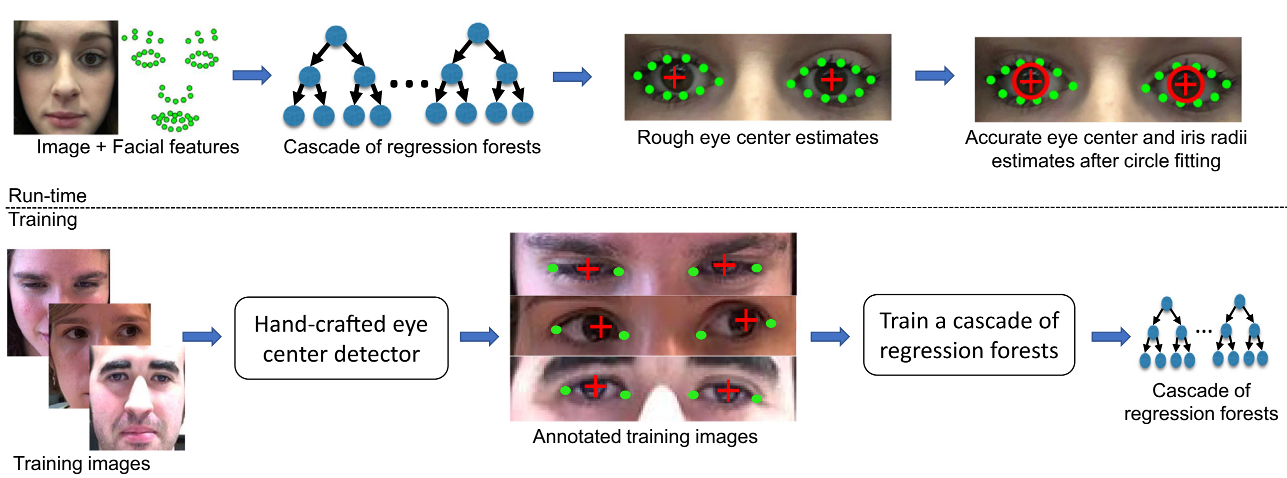


**Fig. 1.** Overview of our method. Top: At run-time an image and the detected facial features are used to regress the eye centers, which are then refined by fitting circles to irises. Bottom: Rather than training the regressor from manually annotated data, we can train the regressor on automatically generated annotations from a hand-crafted eye center detector.