Engineering Applications of Artificial Intelligence 33 (2014) 69–79

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Machine learning for multi-view eye-pair detection

Mahir Faik Karaaba *, Lambert Schomaker, Marco Wiering

Institute of Artificial Intelligence and Cognitive Engineering, University of Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 26 August 2013
Received in revised form
14 February 2014
Accepted 24 April 2014
Available online 14 May 2014

Keywords:
Eye detection
Eye-pair detection
Machine learning
Support vector machine
Restricted Boltzmann machine (RBM)

ABSTRACT

While face and eye detection is well known research topics in the field of object detection, *eye-pair* detection has not been much researched. Finding the location and size of an eye-pair in an image containing a face can enable a face recognition application to extract features from a face corresponding to different entities. Furthermore, it allows us to align different faces, so that more accurate recognition results can be obtained. To the best of our knowledge, currently there is only one eye-pair detector, which is a part of the Viola–Jones object detection framework. However, as we will show in this paper, this eye-pair detector is not very accurate for detecting eye-pairs from different face images. Therefore, in this paper we describe several novel eye-pair detection methods based on different feature extraction methods and a support vector machine (SVM) to classify image patches as containing an eye-pair or not. To find the location of an eye-pair on unseen test images, a sliding window approach is used, and the location and size of the window giving the highest output of the SVM classifier are returned. We have tested the different methods on three different datasets: the IMM, the Caltech and the Indian face dataset. The results show that the linear restricted Boltzmann machine feature extraction technique and principal component analysis result in the best performances. The SVM with these feature extraction methods is able to very accurately detect eye-pairs. Furthermore, the results show that our best eye-pair detection methods perform much better than the Viola–Jones eye-pair detector.

1. Introduction

Face alignment is an important requirement for a successful face recognition application. A human face in an image can be in a variety of scales, positions and poses. Without any alignment of the face entities in an image, recognition performance is very limited. An *eye-pair* is the image that contains a pair of eyes, and it is a significant part of a face. We believe that detection of it can be easier than other parts of a face. Still to the best of our knowledge, there is currently only one eye-pair detection method based on the Viola–Jones framework (Castrillón-Santana et al., 2008), but as we will show in this paper, this method is not very accurate. The aim of our work is to develop a system that can accurately detect eye-pairs. This will be useful to address in the future the problem of accurate face alignment and recognition.

Eye or eye-pair detection is a sub-field of object detection in images. The approaches can be classified into three fundamental methods: shape-based models, feature based models and appearance-based models. Shape-based models depend on a geometrical model of the eyes and use this model to decide whether an image patch contains an eye. It extracts contour properties of the image patch and compares these to the model using a similarity measure. In Kawaguchi et al. (2000), a separability filter is used for feature extraction and the Hough transform is used for model fitting. Some researchers focus on color images in order to exploit skin color of faces. So, a color conversion algorithm is applied to the image containing a face so that the separation of skin color from the background becomes easier. After the conversion the face is detected by means of a face mask calculation. In Kalbkhani et al. (2013) a non-linear RGB to YCBCr color conversion is adopted, and an eye mapping algorithm is applied using an already created face mask to find the eyes. In Huang et al. (2011) an algorithm which converts color pixels from the RGB color space to the HSL space is developed and used. Then, after some image enhancement operations specific to human skin, an object searching algorithm is used for finding eye candidates. Exploiting human-skin color as a discriminator can be very efficient, provided that the background is relatively simple and different from human skin color. In another eye detection and tracking system (Abdel-Kader et al., 2014), eyes are detected and tracked by a particle swarm optimization based multiple template matching algorithm. In another paper, the Hough transform algorithm is used in combination with directional image filters previously proposed for face detection

* Corresponding author.
E-mail addresses: m.f.karaaba@rug.nl (M.F. Karaaba), l.r.b.schomaker@rug.nl (L. Schomaker), m.a.wiering@rug.nl (M. Wiering).

http://dx.doi.org/10.1016/j.engappai.2014.04.008
0952-1976/

(Maio and Maltoni, 2000). In Ilbeygi and Shah-Hosseini (2012), luminance and chrominance values of colored image patches are extracted and given to a template matching algorithm to detect eyes. Shape-based eye detection models may be suitable for real-time eye-tracking applications if they require tracking only the iris and the pupil. However, they are sensitive to different rotation angles and image quality. Moreover, for obtaining more precise results, these models use more parameters to model the shape and this results in an extensive engineering effort and the application is computationally more demanding (Hansen and Ji, 2010).

Feature-based methods focus on finding local features related to the eye. For instance, the eyebrow, the pupil and the iris are basic parts of an eye and locating these features can be helpful for locating the eye. In Kim and Dahyot (2008), features of eyes and other facial parts (nose, mouth, etc.) of the face are extracted with the SURF algorithm (Bay et al., 2008). Then these features are given to a support vector machine (SVM) (Vapnik, 1998) to locate these facial parts. In Sirohey and Rosenfeld (2001) special linear and non-linear filters constructed from Gabor wavelets are used to detect the iris and corner features of the eyes. Then these features are further filtered to remove false features from the detected feature set. A voting mechanism is finally applied to compute the most accurate location of the iris. In Ando and Moshnyaga (2013), integral images are utilized for face tracking, face detection, and eye detection. There, instead of eyes themselves, the area between the eyes is exploited as a discriminator from other parts of a face. The face area is first obtained by subtracting the adjacent frames of video data and then a seven segmented rectangle template is used to slide through the image which contains the face. The output of the sliding window algorithm is given and processed by an algorithm according to their integral image output values. Since that application is designed for energy-constrained environments, the algorithm it uses is relatively simple which might give inaccurate results for some environments. While feature-based methods are robust to illumination and pose changes, they usually require high-quality images (Hansen and Ji, 2010).

Appearance-based models make a model from eye images by using the photometric appearance of the eyes. Since no specific a priori information related to eyes is used, a sufficient number of training data to learn the parameters for eye detection are needed. For the purpose of eliminating noise and reducing dimensionality, feature extraction and normalization operations to training data are usually applied. As for feature extraction techniques, principal component analysis (PCA), and edge detection methods are some of the techniques being used. After all these operations the output is given to a classifier for training. As classifiers, adaptive boosting (Freund and Schapire, 1995), neural networks and SVMs have been used. In Huang and Mariani (2000), patches of example eye images are processed by principal component analysis (PCA) to reduce the dimensionality and make a model eye for classifying unseen image patches if they contain an eye. In Vijayalaxmi and Rao (2012), a Gabor filter is used as a feature extractor and an SVM is used as a classifier. To make the final detector more robust to rotations, the face images are populated by rotation, translation and mirroring operations before giving them to a Gabor filter to be processed and then finally the output of it is fed into the SVM to train the classifier. In the face and eye detection method in Lin et al. (1997), some edge extraction techniques and a histogram equalization algorithm are applied to image patches before they are given to a probabilistic decision based neural network for detection. In Motwani et al. (2004), wavelet coefficients of image patches are given to a multilayer perceptron. In You-jia et al. (2010), the output of an orthogonal wavelet analysis on image patches is given to an SVM. The biggest advantage of the appearance-based methods is that they are applicable to all kinds of different objects, because they are based on machine learning algorithms to learn the model from training data. Therefore, they also often require almost no a priori knowledge and less engineering effort. A disadvantage is that they may need a lot of labeled data to learn a very good performing model.

In the Viola–Jones object detection framework (Viola and Jones, 2004), the eye-pair detector (Castrillón-Santana et al., 2008) adopted an appearance-based method as well. The framework exploits Haar wavelets as object features and these features are calculated using integral images, which makes the computation very efficient. Because of this efficiency fact, the face detector of Viola–Jones is known as a very time-efficient face detector and it is still a de facto standard for general platforms where speed can be preferred over accuracy. The method is based on using a cascaded classifier structure using weak Haar features to build a classifier. To train the cascaded structure an adaptive boosting algorithm is used. In this scheme, if a training example is misclassified by the detector, the weight of that example is increased so that the subsequent classifier is able to correct the errors made by the previous classifiers.

Primarily meant to be used for face detection, this framework has been extended for detecting facial parts such as eye, eye-pair, mouth, and nose. Nevertheless, this detector is not very accurate and may not be very suitable for platforms where source images are cluttered, noisy or have low-contrast. Since we aim to develop a very robust face recognition application useful for very different types of face images taken in challenging environments, we need high accuracy rather than high speed in order to minimize the recognition error caused by incorrectly aligned face images. Because of this reason, we will utilize a strong classifier and powerful feature extraction methods to increase the discrimination power of the system.

Contributions: In this paper a novel eye-pair detection method, addressing the problem of face alignment, is proposed. Our aim is to build a robust application, which can deal with many variances in different images, that can also be useful for robots. The system is constructed by using a feature vector extraction method that converts an image patch to the input of a support vector machine classifier. We have compared five different feature vector extraction methods. The first one is the linear restricted Boltzmann machine (RBM) (Smolensky, 1986) that extracts activities of latent variables which model the data. The second one directly uses pixel-intensity values. The third method uses principal component analysis to extract eigenvalues from an image patch, and the last two feature extraction methods use the difference-of-Gaussians edge detector and the Gabor wavelength filter before the image patch is given as an input to a linear RBM. These five feature extraction methods and the SVM classifier are implemented in a sliding window method to find the best matching eye-pair region in a face image. The detector is trained on images we collected from the Internet for which we manually cropped the eye-pair regions. We have compared our methods to the Viola–Jones eye-pair detector on three different test face image datasets (with 240, 450 and 566 face images). The results show that our eye-pair detection systems consistently perform better than the state-of-the-art Viola–Jones eye-pair detector. For almost all test images, the eye-pair regions are located very accurately with our system. Besides, we compare our eye-pair detector application with a single eye detector that we constructed in a similar fashion to show the superiority of using one single wider rectangle which contains two eyes instead of two smaller ones.

Paper outline: This paper is organized as follows: In Section 2, the classifier and feature extraction methods are described. In Section 3, the whole eye-pair detection algorithm is explained. After that, the experimental setup and results are described in Section 4. Section 5 discusses our findings and describes some directions for future work.