Neurocomputing 173 (2016) 418–424

Contents lists available at ScienceDirect

Neurocomputing

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

Learning multi-channel correlation filter bank for eye localization ☆

Shiming Ge [a,b], Rui Yang [a,b,c], Yuqing He [d], Kaixuan Xie [a,b,c], Hongsong Zhu [a,b,*], Shuixian Chen [a,b]

[a] State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[b] Beijing Key Laboratory of IOT Information Security Technology, Beijing 100093, China
[c] University of Chinese Academy of Sciences, Beijing 100049, China
[d] School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 30 June 2014
Received in revised form
20 March 2015
Accepted 25 March 2015
Available online 1 September 2015

Keywords:
Eye localization
Correlation filter
Filter bank
Multi-channel feature
Regression
Biometric security

ABSTRACT

Accurate eye localization plays a key role in many face analysis related applications. In this paper, we propose a novel statistic-based eye localization framework with a group of trained filter arrays called multi-channel correlation filter bank (MCCFB). Each filter array in the bank suits to a different face condition, thus combining these filter arrays can locate eyes more precisely in the conditions of variable poses, appearances and illuminations when comparing to single filter based or filter array based methods. To demonstrate the performance of our framework, we compare MCCFB with other statistic-based eye localization methods, experimental results show superiority of our method in detection ratio, localization accuracy and robustness.

1. Introduction

With the emergence of new media and social networking, face analysis in large-scale surveillance data and social media data is attracting more and more attention from both academia and industry due to the fact that facial photographs occupy a significant share in social network services (SNS) and huge surveillance data. Unlike conventional face recognition under a rather constrained environment, face related applications under SNS and realistic surveillance environment have to deal with large appearance variations and computational complexity for millions of users, which need to extract robust feature to represent face images. Accurate eye or landmark localization plays a preprocessing and key role in many face analysis related applications, such as face clustering [1] and face tagging [2] for social media, face detection for pose estimation [3], face verification [4], face recognition [5], drivers' fatigue monitoring [6], and low dimensional face representation [7], and thus has received much attention recent years.

Most of the convenient eye localization approaches process facial images in the spatial domain, and usually need to scan all the possible positions in the image to get an optimal result, which reduces the efficiency of the localizing progress. By contrast, correlation filter based methods process images in the frequency domain. Taking advantage of fast fourier transform (FFT), correlation could be efficiently calculated in frequency domain, which reduces the computational cost. Besides, eye localization needs to estimate eye coordinates in pixel or sub-pixel accuracy. By designing a proper output of the filtering operation, correlation filter could produce locating result in exact pixel. These advantages make correlation filter very suitable for eye localization task. Although many variants of correlation filter have been proposed, most of them are constituted by a single filter, which can hardly handle face variation in different environments.

In this paper, we propose a multi-channel correlation filter bank (MCCFB) based method to address eye localization task under diverse conditions. The MCCFB is composed with a group of filter arrays, each filter array suits to a certain environment adaptively, which increases the adaptability of the whole filter

☆This work is supported in part by the National Natural Science Foundation of China (No. 61402463), the Young Talent Program of Institute of Information Engineering Chinese Academy of Sciences (No. 1102008202), the "Strategic Priority Research Program" of the Chinese Academy of Sciences (No. XDA06040101), and the Xinjiang Uygur Autonomous Region Science and Technology Project (No. 201230122).

* Corresponding author at: Beijing Key Laboratory of IOT information security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China.

E-mail addresses: geshiming@iie.ac.cn (S. Ge), yangrui@iie.ac.cn (R. Yang), heyuqing@tju.edu.cn (Y. He), xiekaixuan@iie.ac.cn (K. Xie), zhuhongsong@iie.ac.cn (H. Zhu), chenshuixian@iie.ac.cn (S. Chen).

http://dx.doi.org/10.1016/j.neucom.2015.03.125
0925-2312/

bank comparing to a single filter. We formulate eye localization problem as an optimization problem with a well-defined cost function based on MCCFB. The optimal MCCFB is produced by an EM-like adaptive clustering approach, it contains several discriminative filters. We compare our method with several statistic-based eye localization methods, the experimental results show the advantages of our method in both localization performance and robustness.

2. Related works

According to the information used for building eye detector, the existing eye localization techniques can be classified into three categories: characteristics based methods, statistics based methods and hybrid methods. In the following, we firstly introduce characteristics-based methods and hybrid method, and then focus on statistics-based methods.

Characteristics based methods perform eye localization by measuring inherent eye features such as shape, contrast and context. In [8], Zhou and Geng used projection function to locate the *x*-coordinate of the eye corners and the *y*-coordinate of the eyelids. These techniques have limitations under the complex or uncontrolled conditions due to unreliable measuring of characteristics.

Hybrid methods integrate structural information into statistical appearance model to improve eye localization performance. These methods combine eye characteristics and appearance under a unified framework. The typical methods include enhanced pictorial structure model (EPS) [9], active shape model (ASM) [10], and active appearance model (AAM) [11]. Hybrid methods usually localize multiple features simultaneously, and provide a good mechanism to infer the position of an object by locating all of its parts. There are many methods for face alignment, which usually aim to detect tens of facial landmarks.

Statistics based methods learn statistical appearance model from a set of training images to extract useful visual features. These methods aim to find a function to discriminate eye and non-eye classes directly. In this way, the problem of eye localization is transformed to a binary classification issue, and then solved by training an eye classifier. A well-known classifier is proposed in [12], which uses Haar cascade classifier in face detection field [13] to locate the eyes. However, these classifiers are set to optimize classification accuracy rather than localization accuracy, thus the trained classifier may not give the maximal response at the right object location and reduce the localization accuracy.

One way to improve localization accuracy is to formulate localization task as a regression rather than a classification problem by incorporating the positions of the eyes. In this setting, given a set of input images with the corresponding eye positions, the training goal is to learn a regressor that maps from the input image to the predicted eye position. In [14], Sun et al. performed facial point detection with deep convolutional neural network (CNN). They carefully designed three-level convolutional networks which provide multi-level regression. The convolutional framework learns filters to optimize for minimizing reconstruction error of image patches instead of pattern localization. Other convolutional methods include separable filter learning [15] and many convolution based sparse coding methods [16]. Such methods learn those filters for feature representation rather than convolutional pattern detection.

Correlation filters are a class of discriminative template-based classifiers that can be used for pattern detection [17]. In [18], Xie et al. proposed a kernel correlation filter for redundant class-dependence feature analysis (CFA) to perform robust face recognition, which provides better dimensionality reduction and outperforms the baseline algorithm such as PCA and LDA. Correlation filters are specially suit to accurate pattern localization. In [19], Bolme et al. performed regression by constructing a correlation filter that exactly transforms each training image to its correlation image, then simply average all of these exact correlation filters to obtain the final learned filter called average synthetic exact filter (ASEF). As an extension, in [20], Bolme et al. proposed a minimum output sum of squared error (MOSSE) filter, which can get better outcome with fewer training samples. In [21], Heflin et al. adopted a similar method to perform eye localization task by first warping the face image through the candidate response pair and then selecting the one leading to face image with best quality. In [22], Boddeti et al. proposed a vector correlation filter (VCF), which train multi-channel descriptors for car landmark detection. In [23], Galoogahi et al. introduced a similar method called multi-channel correlation filters (MCCF), which saved the computational cost of VCF and was used into eye localization task. The crucial step in correlation filter based methods is filter construction. When the filter fits test samples, good outcome could be expected. However, one single correlation filter or filter array could hardly fit all the testing face images due to huge appearance variations of faces in lighting, expression, pose and so on.

Multi-channel features can be used to perform vision detection tasks which give robust results [24,25]. Typical discriminative multi-channel features such as HOG descriptor have gone on to be employed in vision tasks ranging from facial landmark localization/detection [26] to general image/video applications [27]. The HOG descriptor in conjunction with a linear SVM can learn a detector for pedestrian detection [28] or general object detection [29,30]. VCF [22] and MCCF [23] methods utilize a HOG descriptor to represent face image and give more accurate facial landmark localization than ASEF [19] and MOSSE [20] which both utilize gray-level feature for localization task. Motivated by this technique, our eye localization method also employs multi-channel feature for representing face images.

3. Proposed approach

In this section, we present the fundamental concepts of correlation filter and our proposed method for the task of eye localization. We only conduct localization on the left eye as the situation for the right eye is similar. We also note that the method suits for the localization of other facial points.

Notation: The training set is constituted by n face images $\{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_n\}$ and the corresponding eye coordinate $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. Matrices are presented in bold (e.g. $\mathbf{h}$). The symbol $\wedge$ applied on any matrix denotes the 2D Discrete Fourier Transform (DFT) of it. Complex conjugate and complex transpose are indicated by $*$ and $\dagger$ respectively. One of the most advantage of DFT lies in its ability to represent convolution/correlation in the spatial domain as a Hadamard product in the Fourier domain. The symbol $\odot$ denotes the correlation operation, Hadamard product is represented as the $\circ$ operator.

3.1. Correlation filter

Correlation filter is a spatial-frequency matrix trained for representing a particular pattern. In the task of localization, patterns of interest are searched by cross correlating the query image with the filter, then possible correlation peak is examined as the pattern location across the correlation output. Ideally, the correlation operation would generate the desired output which has a sharp peak at center of the target and (near) zero value for elsewhere. In order to approximate such filter, many approaches have been proposed.