Image and Vision Computing 32 (2014) 979–993

Contents lists available at ScienceDirect

Image and Vision Computing

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

Discrete area filters in accurate detection of faces and facial features ☆

Jacek Naruniec

Institute of Radioelectronics, Warsaw University of Technology, Faculty of Electronics and Information Technology, ul. Nowowiejska 15/19, 00-665 Warsaw, Poland

ARTICLE INFO

Article history:
Received 5 March 2013
Received in revised form 11 September 2014
Accepted 12 September 2014
Available online 5 October 2014

Keywords:
Face detection
Facial features detection
Discrete area filers
mLDA cascade

ABSTRACT

This paper introduces a new method for detection of faces and facial features. Proposed algorithm denies the thesis that bottom-up solutions can't work at reasonable speed. It introduces fast detection – about 9 frames per second for a 384×256 image – while preserving accurate details of the detection. Main experiments focus on the detection of the eye centers — crucial in many computer vision systems such as face recognition, eye movement detection or iris recognition, however algorithm is tuned to detect 15 fiducial face points. Models were trained on nearly frontal faces. Bottom-up approach allows to detect objects under partial occlusion — particularly two out of four face parts (left eye, right eye, nose, mouth) must be localized. Precision of the trained model is verified on the Feret dataset. Robustness of the face detection is evaluated on the BioID, LFPW, Feret, GT, Valid and Helen databases in comparison to the state of the art detectors.

1. Introduction

Accurate face and facial feature detection plays a major role in image processing techniques. In order to perform reliable face recognition, significant parts of the faces must be precisely localized, for example eye centers. Lack of accuracy highly decreases efficiency of the identification. Simple holistic face detection algorithms, for example the one used in popular Viola and Jones method, are trained in such fashion, that the eye centers are located in the specific locations of the bounding box. In practice the position of the detected faces isn't very precise and it gives only regional information. In order to perform reliable face recognition more sophisticated algorithms should be used.

Many methods cope with this problem, but they usually suffer from a subset of several disadvantages — model isn't flexible enough to cover all the variability of the human face (AAM, ASM), different treatment is needed for each part of the face [1], or only localisation of one face is possible.

There is also a possibility of refining face/facial feature position already given by the bounding box detector. Such algorithms often use content versus context approach [1] or detailed knowledge-based rules. Such top-down approach (from general information to details) often ensures high speed of the localization, but on the other hand, holistic methods tend to reject partially occluded faces. Some of the algorithms are trained using also such difficult examples, but the final detector tends to have rather lower accuracy on the non-occluded faces.

This article presents new bottom-up (from details to the general information) algorithm for detecting facial features (eyes, nose, mouth) and faces. It introduces new extraction and classifications methods for detecting characteristic face points (such as eye corner and mouth corner) — discrete area filters (DAF) and modified positive-to-negative linear discriminant analysis (mLDA). Points belonging to the specific categories are merged into face parts and implicitly to faces by the voting process combined with agglomerative clustering and the reference graph. Article proves that such an approach yields accurate detection comparing to manual marking, high detection rates and allows the detection of partially occluded faces.

Paper is divided as follows:

- Section 2 presents problem formulation and current state-of-the-art in the face and facial feature detection and localization.
- Section 3 provides basic information about each step of the algorithm, including discrete area filters and modified linear discriminant analysis.
- Section 4 describes performed experiments along with their interpretation.
- Section 5 concludes the paper.

2. Problem formulation and related work

In this paper *fiducial point detection* stands for defining number and positions of all the points of specific category, for example inner eye corners or nose tips. *Facial feature detection* or *face part detection* stands for defining number and positions of all complex parts of the face such as nose, mouth or eyes. *Face detection* stands for defining the number and position of all the faces present in the digital image. This can be achieved in several ways, for example by defining a bounding box, contour or its fiducial points. In the presented algorithms face detection is defined as a set of connected facial features. By the term *face localization* we can understand finding a position of the face present in the image.

First attempts of automatic facial features localization was made in the 1970s, for example in [2] or [3]. In the latter object was modeled

☆ This paper has been recommended for acceptance by Stefanos Zafeiriou.

http://dx.doi.org/10.1016/j.imavis.2014.09.004
0262-8856/

as a set of nodes connected with springs. Deviation from the predefined values caused higher cost function (spring bending). To decrease computations time caused by analyzing all of the possible object positions, the problem was solved using dynamic programming. Fishlers experiments based on a 45×50 pixel images, what gives over 100 times less data, than acquired with a modern simple web camera with resolution 640×480 pixels. In the 1990s increasing computational power stimulated development of the semantic image analysis, including face detection and recognition.

One of the most common pattern analysis techniques is Principal Components Analysis (PCA). It is usually used to reduce the dimensionality of the data (for example in work of Antonini [4] or Celuktitan [5]). Matas [6] used PCA to detect 10 fiducial face points by analyzing pixels in the neighborhood of the Harris corner detector responses. Work was continued by Hamouz [7] by replacing PCA with the Gabor filter responses. PCA is also used in the Active Appearance Models (AAM) to create flexible model of the face [8]. Because PCA doesn't consider information about the separation of resulting classes, Local Discriminant Analysis (LDA) was proposed. It is used mainly in face recognition [9], but there are also some publications proving, that it can also be used in face localization. Kim [10] proposed LDA for creating descriptor of the face components. The face was divided into 14 overlapping parts, each associated with the class separated by the discriminant analysis.

One of the particular LDA aspect is *searched point/non-searched point* detection (positive/negative). Hotta [11] noted, that the separation between face and background classes is more accurate, when treating each of the non-face example as a separate class. This fact was also confirmed by the author of this paper.

Another effective classification method used in face analysis is Support Vector Machines (SVM). Jee [12] proposed using SVM with radial basis function to detect eyes. Hamouz [7] used SVM to verify face coordinates localized using face fiducial points. SVM has been also used to verify area of eye light reflections caused by the IR illumination, for example in the work of Zhu [13]. Interesting notes can be found in the work of Ngyuen [14], who used SVM to depict points in window analysis, that are important from the view of the face detection.

There is a high number of publications on face detection using neural networks (NN) (for example [15]) and also in fiducial point detection. Reinders [16] used NN to localize 4 points of the eye in the video sequence. Geometric model was used to verify obtained results. Analysis was performed on gradient and edge orientation images. Duffner [17] proposed 6-layer architecture for detection of the characteristic face points.

One of the most popular face detector is Viola and Jones detector [18, 19]. It was based on the three main elements: Haar filters along with integral image, AdaBoost classification and cascade of the classifiers. These elements resulted in a fast and reliable algorithm. Extended set of Haar features and modified classification scheme was presented in the work of Lienhart [20]. Since 2001 it was used in many applications and object recognition problems, including detection of the face parts, gait, whole body and others.

One of the most accurate fiducial point descriptors are responses of the Gabor filters. It was proven [21], that they are more efficient in extraction process than PCA or LDA methods. It was used in face localization for example in work of Lades [22], who defined model of face responses of the Gabor filters in a rectangular grid. One of the most important articles, being motivation for many others, was the work of Wiskott [23]. For each fiducial point he proposed creating a model consisting of a set of Gabor responses (*Gabor jets*) for the representative examples — the opened eye, closed etc. The obtained results were better,

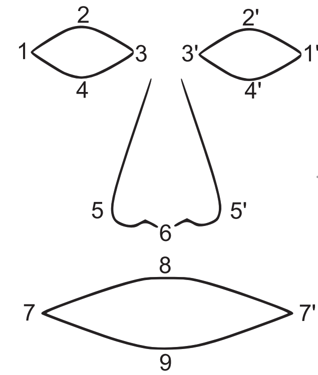

Fig. 1. Classified face fiducial points.

than in the work of Lades because points weren't sampled uniformly in a rectangular grid. These methods were also used for the depth images [24]. Vukadinovic [25] had proven accuracy of the responses of Gabor filters combined using GentleBoost algorithms. Interesting conclusions have been made by Fasel [26], who had shown, that the commonly used Gabor filter set is not optimal for different fiducial points.

Many of the algorithms proposed in the literature base on the geometric models. Most commonly used are Active Shape Models (ASM) [27] and their extension — Active Appearance Models (AAM) [8]. It was later extended by using 2D model instead of 1D in edge matchinhog, adding noise to the training data and using two active shapes models — on for the rough position estimation and the second for accurate localization [28]. Al Haj [29] proposed using color face detection and connected components to define initial position of the model. Lately [30] proposed an extension to the Active Shape Models by using Viterbi optimization along image contours to improve the localization accuracy. Belhumeur [31] used Bayesian model to combine outputs of local detector to face find fiducial points. Valstar et al. [32] used Support Vector Regression, Markov Random Fields and local appearance to achieve accurate frontal face detector, robust to the appearance variation.

One of the best results in the face and facial features detection have been achieved by using mixture of trees with a shared pool of parts [33]. Zhu et al. presented algorithm comparable to the commercial products in terms of precision, but their solution is still to slow for on-line processing.

Other algorithms base on our knowledge about human faces. Most commonly used information is: skin color [34,29,35], symmetry of the face [36], face appearance experts [37], and edges [38].

Very good results in precise localization of the facial features were presented by Ding [1]. It provides content-vs-context approach to discriminate facial features from their neighborhood. Latest feature localization methods [39] use regressors to map roughly localized points descriptors to the desired, correct position. Such an approach yields very precise results, in many cases comparable to the manual markings. Also new algorithms of regression analysis of the Constrained Local Models [40] provide very accurate face fitting techniques.

Algorithms presented here are an extension of the previous work [41]. Considerable effort has been made to improve the proposed method by:

- introducing extended set of extractors with removed frequency information,
- introducing voting method for detecting separate face parts,
- introducing cascade of the mLDA classifiers,
- allowing occlusions of the face parts,

Fig. 2. Detection scheme. a) represents the input image, b) represents Canny edge detection result (points of interest). Image c) illustrates pixel classification results. Each color corresponds to the different categories of points of the particular face part. Images correspond to the different face parts — from the left: left eye, right eye, nose and mouth. d) illustrates voting result for individual face parts. Images e) present the results of agglomerative clustering with constraints to maximal distance between nodes and minimal number of categories in cluster. f) illustrates centers of the points taking a part in a voting for clusters from image e). Final detection result g) is obtained by matching found face parts to the face reference graph. Final face part positions are calculated as the centers of the corners of each feature.