Pattern Recognition Letters 53 (2015) 62–68

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Facial feature points detecting based on Gaussian Mixture Models ☆

Junnan Wang, Rong Xiong*, Jian Chu

State Key Laboratory of Industrial Control Technology, Zhejiang University, Zheda Road 38, Hangzhou 310027, China

A R T I C L E I N F O

Article history:
Received 16 January 2014
Available online 22 November 2014

Keywords:
Facial feature points detecting
Model based Hough Voting
Gaussian Mixture Models
High-order Geometric constraint

A B S T R A C T

Detecting predefined facial feature points in a human face image is a well studied problem. Despite the impressive achievements that have been made, it is still open under unconstrained environments, with variations of illumination, expression, head pose, as well as partial occlusions. This paper proposed a novel method to locate facial feature points under the variations mentioned above. Support vector machines with probability outputs are trained to provide the observation probability of each facial feature point. The observation probabilities, as well as the distribution of face shape which serves as the prior to constrain the relative position of the facial feature points, are both approximated with Gaussian Mixture Models. The problem is solved by maximizing the posterior which combines the prior and observation probability within the framework of Bayesian Inference. An optimization algorithm is developed to maximize the posterior by iteratively maximizing the lower bound of it. The proposed method preserves the high-order geometric constraint within facial feature points. With a simple initialization method of Model based Hough Voting, the method shows competitive detecting rate and locating accuracy on the LFPW and LFW datasets, compared to the methods of state-of-the-art.

1. Introduction

Detecting facial feature points means locating a series of predefined landmarks in the face image, like the corners of the eyes, the tip of the nose, the corners of the mouth. The problem has been wildly studied due to its key importance to many applications, such as face recognition, expression recognition etc. Numerous of detecting methods have been proposed, and fairy high detecting rate and locating accuracy have been achieved on the near frontal face images taken in the controlled scenarios. Despite the impressive achievements have been made in the past few decades, it is still open to locate facial feature points on images taken in the uncontrolled scenarios, with great variations of imaging conditions, such as illumination, expression, head pose and partial occlusion due to hair, makeup or decorations.

In this paper, we present a Gaussian Mixture Models (GMMs) based detecting method to handle the variations of imaging conditions. For each facial feature point, an SVM with probability output is trained based on the local textures, and then applied to scan a predefined region of the input face image to generate a confidence map which indicates the probability of the occurrence of the corresponding facial feature point. The confidence maps generated are noisy and full of local extremes, GMMs are used to give an analytical approximation of each confidence map based on re-sampling to reduce the noise. GMMs are also utilized to model the distribution of face shape which is highly non-linear due to the variations of identity, head pose and expression. The GMMs of the confidence maps and face shape, which serve as the observation and prior respectively, are combined to calculate the posterior. The posterior is maximized subject to the face shape by iteratively maximizing a lower bound of it. Furthermore, a Model based Hough Voting Method is proposed to find an initial face shape for initialization. The framework of our method is shown in Fig. 1.

There are three main contributions of our method. First, GMMs are adopted to approximate the confidence maps to reduce the noise and simplify the optimization procedure, and a method based on re-sampling is proposed to estimate the parameters of the GMMs. Second, the face shape is directly estimated within the Bayesian Inference framework, with the consideration of high-order geometric constraint. Relative to the methods considering each point independently, our method improves the locating accuracy, especially for the unstable facial feature points, like the nose tip, points on the middle of the eyebrows. Third, a Model based Hough Voting method is proposed to estimate an initial face shape. Comparing to the methods using anchor points for initialization [13,19], our method combines the information of all local detectors, and is more robust to occlusions.

Our method was evaluated on two challenging datasets: LFPW and LFW, which are both constructed with images collected from the websites. The experiments shown comparative accuracy to the results of state-of-the-art on the stable facial feature points and much higher

☆ This paper has been recommended for acceptance by M. Tistarelli.

* Corresponding author. Tel.: +86 571 87952750/86 138 67417609.
E-mail address: rxiong@iipc.zju.edu.cn, rxiong_zju@hotmail.com (R. Xiong).

http://dx.doi.org/10.1016/j.patrec.2014.11.004
0167-8655/

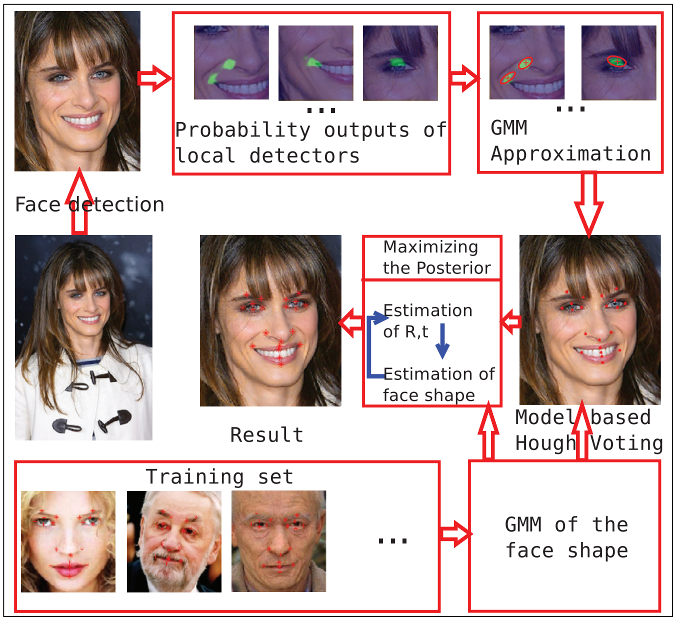

Fig. 1. Overview of the proposed method.

accuracy on the unstable ones. Our method also achieved slightly higher accuracy compared to human labeling.

2. Related works

Detecting facial feature points independently may suffer from the ambiguity of local appearances. The consideration of geometric constraint is of essential importance to rule out the non-plausible face shape configuration. According to how the geometric constraint been introduced, the methods can be divided into two classes: the global and the local geometric constraint based methods.

Global geometric constraint based methods take the whole face shape into consideration. Active Shape Model (ASM) [4] adopts Point Distribution Model (PDM) as the geometric constraint, and the PDM parameters are optimized to fit texture models of all points best. Many works have followed the ASM framework, but different in some details. Saragih et al. [18] model the local textures with Gaussian Mixture Models. Cristinacce and Cootes [7] use boosting to predict the target position based on the current texture. Gu and Kanade [11] use a 3D PDM as the global shape prior to improve the robustness to head pose variations. Multi-view ASM [15] trains several PDMs and texture models conditioned on head pose to accommodate the head pose variations. Amberg and Vetter [2] apply branch and bound to a set of potential facial feature point candidates to select a set of features which fit the PDM best.

Active Appearance Models (AAM) [5,14] is one of the most important extensions of the ASM class methods. Unlike ASM, AAM learns the model of the whole face texture based on shape free face texture. Multi-view AAM [17] method is proposed to adapt to head pose variations. The performance of AAM methods degrade greatly when applied to unseen face images. Cristinacce and Cootes [6] combine the advantages of AAM and ASM. It performs an exhaustive local search for each local detector and optimizes the PDM parameters to maximize the response of all local detectors jointly.

PDM is a linear model, and cannot properly model the nonlinearity which normally comes from great variations of head pose and expression. Belhumeur et al. [3] combine the outputs of local detectors with a set of non-parametric global models based on face shape samples to locate the facial feature points. Cao et al. [8] encode the geometric constraint into a set of regressors in a cascaded learning framework.

Relative to the global geometric constraint based methods, local geometric constraint was considered to simplify the procedure of optimization while still maintaining the geometric constraint in a certain extent. Arashloo et al. [1] model the face shape configuration as Markov random fields. Valstar et al. [19] use SVMs to predict the position of facial feature points and uses conditional Markov random fields to keep the prediction globally consistent. Roig et al. [16] model the face shape configuration with a pictorial structure.

3. Problem statement

Suppose the face has been detected, and the detection of feature points is limited in the face area. Let I denote the face image, and $\boldsymbol{x} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N]^T$ denote the locations of N facial feature points, where $\boldsymbol{x}_i = [x_i, y_i]^T$ is the coordinates of the ith feature point. The facial feature points are located by maximizing (1).

$$p(\boldsymbol{x}|I) \propto p(I|\boldsymbol{x})p(\boldsymbol{x}) \tag{1}$$

where $p(I|\boldsymbol{x})$ is the joint observation probability given $\boldsymbol{x}$. $p(\boldsymbol{x})$ is the shape prior. Suppose the joint distribution is independent,

$$p(I|\boldsymbol{x}) = \prod_{i=0}^{N} p(I(\boldsymbol{x}_i)) \tag{2}$$

where $I(\boldsymbol{x}_i)$ is the texture around $\boldsymbol{x}_i$, and $p(I(\boldsymbol{x}_i))$, an abbreviated form of $p(I|\boldsymbol{x}_i)$, is the observation probability of the ith facial feature point.

4. Local detectors and approximation

The local detectors are applied to estimate the observation probability. For each facial feature point, an SVM with probability output is trained on the Histogram of Gradient (HoG) descriptors extracted from the local texture around the target points. The HoG descriptor is similar to the one used in [3], which is consisted of two SIFT descriptors, with scales of 1/2 and 1/4 of the inter-ocular distance. The face samples are resized to maintain an inter-ocular distance of 48 pixels for scale consistency. The positive training descriptors are extracted right at the facial feature points, while the negative ones are sampled randomly around them, at least 8 pixels away. The SVM with probability output is obtained by combining a traditional SVM classifier with a logistic regression model.

SVMs are trained respectively and applied to predefined regions of the face image to get confidence maps. The item *confidence map* denotes the normalized output of SVMs, i.e. the blue areas in Fig. 2, while

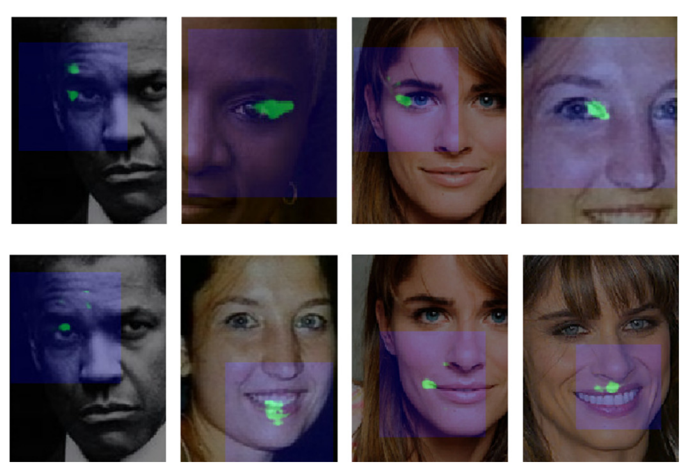

Fig. 2. Results of local detectors. The light blue square indicates the searching scale, the brightness of the green color indicates the value of probability. The true position of each detector is located in the center of the each light blue square. The blue area is also denoted as a confidence map. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)