Neurocomputing 149 (2015) 354–363

Contents lists available at ScienceDirect

Neurocomputing

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

Efficient smile detection by Extreme Learning Machine

Le An*, Songfan Yang, Bir Bhanu

Center for Research in Intelligent Systems, University of California, Riverside, CA 92521, USA

ARTICLE INFO

Article history:
Received 3 August 2013
Received in revised form
1 April 2014
Accepted 7 April 2014
Available online 16 September 2014

Keywords:
Facial expression analysis
Smile detection
Extreme Learning Machine
Classification
Feature extraction

ABSTRACT

Smile detection is a specialized task in facial expression analysis with applications such as photo selection, user experience analysis, and patient monitoring. As one of the most important and informative expressions, smile conveys the underlying emotion status such as joy, happiness, and satisfaction. In this paper, an efficient smile detection approach is proposed based on Extreme Learning Machine (ELM). The faces are first detected and a holistic flow-based face registration is applied which does not need any manual labeling or key point detection. Then ELM is used to train the classifier. The proposed smile detector is tested with different feature descriptors on publicly available databases including real-world face images. The comparisons against benchmark classifiers including Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) suggest that the proposed ELM based smile detector in general performs better and is very efficient. Compared to state-of-the-art smile detector, the proposed method achieves competitive results without preprocessing and manual registration.

1. Introduction

Facial expression analysis plays a significant role in understanding human emotions and behaviors. The varying facial expressions are considered to be the most important cues in the psychology of emotion [1]. Analyzing facial expressions accurately has broad applications in areas such as human behavior analysis, human–human interaction, and human–computer interaction. Automatic recognition of emotion from images or videos of human facial expression has been a challenging and actively studied problem for the past few decades [2].

Six basic facial expressions and emotions that are commonly referred include anger, surprise, disgust, sadness, happiness and fear. To quantitatively study various facial expressions, facial action coding system (FACS) has been developed [3]. In total 44 Action Units (AUs) are defined to describe all possible and visually detectable facial changes.

One of the most common expressions in a person's daily life, smile, conveys the emotion indicating joy, happiness, satisfaction, etc. Smile expression involves two parts of facial muscle movements, namely Cheek Raiser (AU 6) and Lip Corner Puller (AU 12), as shown in Fig. 1. Detecting smile finds its applications mainly in the Human–Computer Interaction systems, such as interactive gaming, customer satisfaction rating, and patient monitoring. As a consequence, some research has been dedicated to smile detection [4–6]. Although accurate smile detection can be achieved in a laboratory controlled database, smile detection for real-world face images is still challenging due to the variations in pose, illumination and image quality. Fig. 2 shows some smile and non-smile real-world face images.

In this paper, we focus on detecting smiles from face images that contain either a smile or a non-smile facial expression. The discriminative classifier is trained in an efficient manner using Extreme Learning Machine (ELM) [7]. ELM is a recently proposed learning framework which has very low computational cost without the hassle of parameter tuning in contrast to the other learning methods. ELM has been previously applied to different tasks such as image super-resolution [8], genome analysis [9], human action recognition [10], face recognition [11], and face emotion recognition [12]. Before training the classifier, we use a standard face detector [13] to detect faces from the images. The faces are registered using a flow-based affine transformation automatically without any manual labeling or key-point detection. The trained model by ELM is then used to predict the smile status of a given face. Experiments, on a collection of laboratory controlled database and a real-world database, using various feature representations show that the proposed approach achieves high accuracy, efficiency, and better generalization performance compared to the benchmark classifiers. Compared to the state-of-the-art method, our method is competitive in detection accuracy, yet it is more efficient and does not require manual image registration.

* Corresponding author.
E-mail address: lan004@ucr.edu (L. An).

http://dx.doi.org/10.1016/j.neucom.2014.04.072
0925-2312/

Fig. 1. Two AUs (AU 6 and AU 12) that account for smile.

Fig. 2. Sample faces with smiles (top row) and neutral faces (bottom row).

The remainder of this paper is organized as follows. In Section 2 related work is reviewed. The proposed method for smile detection is presented in Section 3. Section 4 is dedicated to the experimental results. Finally, we conclude this paper and suggest future work in Section 5.

2. Related work

For facial expression recognition, mainly two directions have been explored: geometric-based approaches and appearance-based approaches.

Geometric-based approaches track the facial geometry information over time and infer expressions based on the facial geometry deformation. In [14] a set of points as the facial contour feature is defined, and an Active Shape Model (ASM) is learned in a low dimensional space. In this way the mismatch due to non-linear image variations is avoided. Lucey et al. [15] use Active Appearance Model (AAM)-derived representation for expression recognition together with normalization techniques to improve the recognition performance. A particle filtering scheme is used to track the fiducial facial points in a video for facial expression recognition in [16]. In [17] the geometrical displacement of certain grid nodes on face landmarks placed manually by users is used as input to the multi-class Support Vector Machine (SVM) for facial expression recognition in image sequences.

Appearance-based approaches emphasize on describing the appearance of facial features and their dynamics. Bartlett et al. [18] use a bank of Gabor wavelet filter to decompose the facial texture and the AdaBoost and SVM are used for subsequent classification. In [19] the dynamic texture, which combines the motion and the appearance along the time axis, is used by extracting the volume of local binary patterns (VLBP). Wu et al. [20] explore Gabor Motion Energy Filters (GME) as a biologically inspired facial expression representation. Most recently, Yang and Bhanu [21] propose a new image-based face representation and an associated reference image called Emotion Avatar Image (EAI), which aggregates the dynamics of the facial expression change into a single image. The experiments show that the EAI is a very strong yet compact cue for expression inference.

Apart from abundant literature for facial expression recognition, a few papers are dedicated to smile detection. In [22], smile detection is achieved by using Fisher weight map (FWM) and higher-order local auto-correlation (HLAC). In [6] a real-world image collection of thousands of subjects is introduced. Comprehensive studies are conducted by examining different aspects such as size of database, effect of image registration, image representation and classifier. Results show that human-level expression recognition accuracy can be achieved using current machine learning methods. Recently, Shan [5] proposes the usage of pixel intensity difference as features for smile detection. AdaBoost is used to choose the weak classifiers and a strong classifier is formed by combining the chosen weak classifiers. This approach achieves state-of-the-art results on a real-world smile detection database.

Besides image-based approach, some methods integrate multi-modal information for smile detection. Ito et al. [23] detect smile using an image-based facial expression recognition combined with an audio-based laughter sound recognition method. In [4] audio features from the spectogram and the video features extracted by estimating the mouth movement are used in stacked sequential learning for smile detection.

For commercial applications, some software tools have been developed for smile detection, such as VDFaceSDK [24] and Omron's smile measurement and analysis software [25].

3. Technical approach

The system pipeline of the proposed approach is shown in Fig. 3. Before face registration, the faces are extracted from the original images using a standard face detector. The detected faces are registered using a fully automated flow-based registration method without the manual labeling of key points. The facial features are then extracted from the registered faces. The ELM is used to train the smile/non-smile binary classifier and the learned model is used to predict the smile status of a given face for testing. In the following each step will be explained in detail.

3.1. Face detection

The faces from original images are extracted using Viola-Jones face detector [13] implemented in OpenCV which is suitable for real-time processing. The Viola-Jones face detector performs well and the detection accuracy is very high at low false alarm rate. The detected faces are normalized to a certain image size using the bicubic interpolation.

3.2. Face registration

Face registration/alignment normally requires the knowledge of the facial key points such as the locations of the eyes, nose, and mouth. Facial points' localization is either performed through manual labeling or by an automated detector (e.g., [26]). When