Signal Processing 117 (2015) 1–10

Contents lists available at ScienceDirect

Signal Processing

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

Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection

Wei-Lun Chao [a], Jian-Jiun Ding [b,*], Jun-Zuo Liu [b]

[a] Department of Computer Science, University of Southern California, 3601 South Flower Street, Tyler 1, Los Angeles, CA 90089, USA
[b] Graduate Institute of Communication Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan, ROC

ARTICLE INFO

Article history:
Received 31 October 2014
Received in revised form
9 April 2015
Accepted 10 April 2015
Available online 30 April 2015

Keywords:
Facial expression
Expression-specific local binary pattern
Class-regularized locality preserving projection
Dimensionality reduction
Feature extraction

ABSTRACT

This paper provides a novel method for facial expression recognition, which distinguishes itself with the following two main contributions. First, an improved facial feature, called the *expression-specific local binary pattern* (es-LBP), is presented by emphasizing the partial information of human faces on particular fiducial points. Second, to enhance the connection between facial features and expression classes, *class-regularized locality preserving projection* (cr-LPP) is proposed, which aims at maximizing the class independence and simultaneously preserving the local feature similarity via dimensionality reduction. Simulation results show that the proposed approach is very effective for facial expression recognition.

1. Introduction

Because of its important role in human–computer interfaces (HCI), surveillance systems, and human entertainment, face recognition has attracted significant attention in pattern recognition and computer vision. Many algorithms about face verification [1,2], facial age estimation [3], gender identification [4], and facial expression recognition [5–8] have been developed in recent years. In this paper, we focus on the problem of image-based facial expression recognition.

In general, algorithms of facial expression recognition can be simply divided into two steps: *feature extraction* and *expression classification*. In the first step, features that are related to the facial appearance or geometry [9–11] are extracted from the input face image for compact representation. Then, in the second step, according to the extracted features, an expression classifier [12,13] is applied to assign the input face an expression label (e.g. seven expression labels—angry, disgust, fear, joy, sadness, surprise, and neutral, which are usually considered in the literature). Besides these two steps, some algorithms [14,15] further include *dimensionality reduction* [16–18] as an intermediate step for avoiding the overfitting problem and filtering out some irrelevant features on facial expression.

However, even with so much work as mentioned above, there is still a salient gap between human and machines' ability on facial expression recognition. More accurately, some technical challenges have not been well solved. For example, different people, or even the same person, can have different expression patterns at different time or on different conditions. Therefore, how to design a robust feature extraction method that can handle this variety is still a critical problem in facial expression recognition. In addition, there intrinsically exist correlations among the seven expression classes, making some pairs of classes easy to be recognized; some others, hard to be classified. Nevertheless, the classifiers applied to facial expression recognition usually assume the independency among these classes: The widely-used support vector machine (SVM) [12] and K-nearest neighbor classifier (KNN) [19] consider no correlation among classes. This mismatch unavoidably leads to the bottleneck of expression recognition

* Corresponding author. Tel.: +886 2 33669652; fax: +886 2 33663662.
E-mail addresses: weilunchao@gmail.com (W.-L. Chao), jjding@ntu.edu.tw (J.-J. Ding), jinzuo2011@gmail.com (J.-Z. Liu).

http://dx.doi.org/10.1016/j.sigpro.2015.04.007
0165-1684/

where some pairs of classes are always with low recognition rates.

Considering these two challenges, in this paper, we present an improved method for facial expression recognition, which is based on the conventional three-step framework and with the following two main contributions:

- Inspired by the human ability on expression recognition—with only partial information on faces, humans can still recognize facial expression with high accuracy—we propose the *expression-specific local binary pattern* (es-LBP) by computing the local LBP histograms [10] around some particular fiducial points of human faces. Besides, to further include the spatial information in each local LBP window, a symmetric extension for the es-LBP is also presented.
- To alleviate the mismatch between the properties of expression classes and classifiers, we propose the *class-regularized locality preserving projection* (cr-LPP), which aims to push the samples of each class towards some pre-defined locations during the process of LPP [16], a popular dimensionality reduction algorithm in recent pattern recognition researches. Through specifically setting the pre-defined locations, the independence among classes can be enhanced, therefore effectively reducing the degree of mismatch.

From the simulations on the widely-used Japanese Female Facial Expression (JAFFE) database [20], the proposed method does outperform other existing approaches on both the leave-one-person-out (LOPO) and the cross validation settings, demonstrating the effectiveness of our ideas.

This paper is organized as follows: In Section 2, an overview on the related work is presented. In Section 3, we introduce the framework of our method, and describe the proposed feature extraction approach in Section 4; the detail of our cr-LPP algorithm is then presented in Section 5. In Section 6, the experimental results are provided and discussed. In Section 7 we conclude our paper.

2. Previous and related work

Based on the existing literature, there are generally two tasks of facial expression recognition: action unit (AU) recognition based on the facial action coding system (FACS) [6], and discrete emotion recognition [5]. In this paper, we focus on the second task because of its higher semantic meanings, and decide to recognize seven basic expressions—angry, disgust, fear, joy, sadness, surprise, and neutral—which are widely considered in previous work.

As mentioned in Section 1, a facial expression recognition system can be simply divided into *feature extraction* and *expression classification*, and some work further includes an intermediate *dimensionality reduction* step. In the rest of this section, a broad overview of the previous and related work in each step is given, respectively.

2.1. Feature extraction

As summarized from the previous work, feature extraction algorithms for facial expression recognition can be categorized into three types: *geometric features*, *appearance features*, and *hybrid features*. The first type, geometric features, focuses on extracting the shapes and locations of facial components, and computes several measures (e.g., the corresponding locations between pairs of facial components) to represent the face. Representative algorithms for this type can be referred to [6,21–23]; some work further used the optical flow analysis to model muscular activities or estimate the displacements of feature points [24,25]. However, as mentioned in [8], the geometric features commonly require accurate and reliable facial component detection and tracking, which are difficult to accommodate in many situations.

The second type of features are appearance (texture) features, which include Gabor features [26–31], radon discrete cosine transform (DCT) features [32], curvelet transform features [33], contourlet transform features [34], and local binary pattern (LBP) [13–15,35,36] are representative features. Gabor features, which could extract the spatial-frequency information at specific locations, orientations, and scales, were used for texture representation in the early age [9], and later brought into expression recognition [26–31]. Gabor features do result in plausible recognition performance, but are both time and memory intensive to compute.

The local binary pattern (LBP) is another popular feature for expression recognition. It is similar to Gabor wavelets. The LBP was first proposed for texture recognition by Ojala et al. [37] and later became widely-used for extracting local information in face recognition and related topics. The LBP essentially converts the neighboring pixels into the binary form using some thresholds and then uses these binary values to construct a label. Later, through summarizing the labels of all the pixels by histograms, a LBP feature vector could be obtained. In regular settings [10], faces images are first separated into several local patches; the LBP features (histogram) are computed from each patch and then concatenated into a long vector. The most important property of LBP features are their tolerance against illuminant changes and their computational simplicity. To better selecting the locations of local patches and filtering out some irrelevant histogram bins, boosting [19] has been considered in the LBP for feature selection, resulting in the so-called boosted-LBP algorithm [5,14].

There are still other features that have been created or introduced into expression recognition. For instance, the local phase quantization (LPQ) [38], a powerful texture descriptor robust to image blurring, was recently used in expression recognition [39] and resulted in higher accuracy than the LBP features in many experiments. Berretti et al. [40] and Hu et al. [41] exploited the scaled invariant feature transform (SIFT) descriptor, which is widely-used in object recognition, extracted on a set of fiducial points for face representation.

2.2. Expression classification

Facial expression recognition is a multi-class classification problem. Hence, any such classification algorithm, such as the K nearest neighbors (KNN) [14], support vector machine (SVM) [5,15,26,32,42], neural networks [43,44], Bayesian networks [45], and rule-based classifiers [22,23,46], have all