Neurocomputing 129 (2014) 136–145

Contents lists available at ScienceDirect

Neurocomputing

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

Spectral embedding based facial expression recognition with multiple features

Kaimin Yu [a], Zhiyong Wang [a,*], Markus Hagenbuchner [b], David Dagan Feng [a]

[a] *School of Information Technologies, The University of Sydney, NSW 2006, Australia*
[b] *School of Computer Science and Software Engineering, University of Wollongong, Wollongong, NSW 2522, Australia*

ARTICLE INFO

Article history:
Received 28 March 2013
Received in revised form
17 July 2013
Accepted 16 September 2013
Communicated by Huiyu Zhou
Available online 7 November 2013

Keywords:
Facial expression recognition
Spectral embedding
Multi-modal
Multiple features
Data fusion

ABSTRACT

Many approaches to facial expression recognition utilize only one type of features at a time. It can be difficult for a single type of features to characterize in a best possible way the variations and complexity of realistic facial expressions. In this paper, we propose a spectral embedding based multi-view dimension reduction method to fuse multiple features for facial expression recognition. Facial expression features extracted from one type of expressions can be assumed to form a manifold embedded in a high dimensional feature space. We construct a neighborhood graph that encodes the structure of the manifold locally. A graph Laplacian matrix is constructed whose spectral decompositions reveal the low dimensional structure of the manifold. In order to obtain discriminative features for classification, we propose to build a neighborhood graph in a supervised manner by utilizing the label information of training data. As a result, multiple features are able to be transformed into a unified low dimensional feature space by combining the Laplacian matrix of each view with the multiview spectral embedding algorithm. A linearization method is utilized to map unseen data to the learned unified subspace. Experiments are conducted on a set of established real-world and benchmark datasets. The experimental results provide a strong support to the effectiveness of the proposed feature fusion framework on realistic facial expressions.

1. Introduction

Automatic facial expression recognition is an active research topic with a wide range of potential applications including human–computer interactions, augmented reality and affective computing [1]. Although recent years have witnessed significant progress in the field [1], accurate recognition of facial expressions remains a challenging problem, particularly for realistic facial expressions. One of the key reasons is that high variations exist in facial expression images of the same type, which are caused by human face appearance, age, gender and ethnic groups. They are commonly observed when different people execute the same expression. The problem is further hampered by high similarities among different facial expression types, which are found when the same person executes different expression without explicit exaggeration. If the intensity of an expressions is low, the differences among facial expressions can easily be shadowed by facial appearance, thus increasing the difficulties for recognition.

Due to the high intraclass variations and interclass similarities, effective feature extraction is vital to facial expression recognition. Existing feature expression features can be categorized into two groups: appearance features [2–5] and geometric features [6–10]. The appearance features model the appearance change of faces, such as wrinkles and furrows, by directly utilizing pixel values. It can be extracted from either an entire face or from local regions of a face image. Geometry based features utilize the shape and locations of facial components, such as the eyes and mouth, to represent the face geometry.

It is known that different features extracted from a same pattern can reflect different characteristics of the pattern [11]. Hence, it is anticipated that the performance of facial expression recognition methods can benefit from multiview representations, where a view is defined as a type of feature that describes a subset of facial expression characteristics. However, there is often no obvious way to select and combine different types of features. Moreover, the likelihood of incorporating noisy or redundant features increases with the number of different types of features, which may affect the recognition performance adversely. To make the matter worse, facial expression features are often of high dimension. A simple concatenation of different features can lead to the adverse effects of the *curse of dimensionality*, increase the computational cost, and negatively affect recognition and generalization results.

In this paper, we present a feature selection and fusion framework for facial expression recognition based on Multiview Spectral

---

* Corresponding author. Tel.: +61 2 9351 3766.
E-mail address: zhiyong.wang@sydney.edu.au (Z. Wang).

0925-2312/$ - see front matter Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.09.046

Embedding (MSE) [11]. Inspired by the recent success of multiview features in related domains [12], our proposed framework treats feature selection and fusion as a multiview dimension reduction problem and aims to find a unified low dimensional subspace that captures information from all sources (*e.g.* different features and labels) by preserving local geometric properties of the original features. Specifically, by assuming that facial expression features extracted from one type of expressions form a manifold embedded in high dimensional feature space, we construct a neighborhood graph that encodes the structure of the manifold locally. In order to maximize the discriminative power, we propose to build the neighborhood graph in a supervised manner by utilizing the label information that is available with the training data. We then combine the corresponding graph Laplacian matrix of each view with the multiview spectral embedding algorithm. A unified low dimensional feature space is obtained by performing spectral analysis of the combined matrix. Finally, a linearization method is utilized to map unseen data to the learned unified subspace for facial expression recognition.

The main contributions of our work are

1. A supervised multi-view spectral embedding algorithm is proposed to combine the appearance based and geometry based features for facial expression recognition. In order to solve the out-of-sample problem, we utilize a linearization method to map unseen data to the unified low dimensional subspace discovered by the MSE algorithm, where facial expression recognition can be performed.
2. We perform a comprehensive study of the widely used facial expression features, including Active Appearance Model (AAM) [13], Local Binary Pattern (LBP) [14], Multiscale Weber Local Descriptor (Multiscale-WLD) [15], Scale-Invariant Feature Transform (SIFT) descriptor [16] and Gabor filters [17]. Experimental results show that their abilities to describe facial expression are varied based on the type of facial expression and suggest that a combination of different features can lead to better recognition performance.
3. We compare the proposed Multiview Spectral Embedding (MSE) based multi-feature fusion method with other popular approaches to demonstrate its effectiveness.

The rest of this paper is organized as follows: Section 2 reviews popular facial expression features and feature level fusion techniques. Section 3 presents the theory and method of the MSE used for facial expression features fusion. A description of the facial expression features studied in the experiments are described in Section 4. Experimental results on three datasets are reported and analyzed in Section 5. Final conclusions are drawn in Section 6.

2. Related work

2.1. Appearance and geometry based face representation

Face representation has been studied intensively for automatic expression recognition over the past decades, and a large number of approaches have been presented. In general, these approaches can be divided into two groups: geometry based approaches [2–5,18] and appearance based approaches [6–10].

Geometry based facial expression feature extracts the shape and locations of facial components to represent the face geometry. In an early work by Zhang et al. [2], 34 fiducial points were utilized to represent a face image. The fiducial points were manually selected at facial landmarks and the image coordinates of these points were used as features, which results in a 68-dimension feature vector. Tian et al. [3] proposed a Multistate Face Component Model to detect and track changes of facial components in near frontal face images. This model represented facial movements by measuring the state transitions of corresponding facial components. In an image sequence, the facial movements could be modeled by measuring the geometrical displacement of facial feature points between the current frame and the initial frame. Valstar et al. [4] manually selected 20 facial points and recognized Facial Action Units (AUs) by classifying features calculated from tracked facial points. Their experiments demonstrated that the facial representation based on tracked facial points was well suited for facial expression analysis. This approach was further extended by adopting a fully automatic facial movement detection system that could automatically localize facial points in the initial frame and recognize the facial movements using the most representative features selected by AdaBoost [5]. However, extracting geometric features usually requires accurate and reliable facial feature detection and tracking. The automatic detection and tracking of facial features is still an open problem in many real world situations, and relies on manual labor which is very time expensive and error prone. Therefore, appearance based features for facial expression analysis have also been investigated.

Appearance based facial expression features model the appearance change of faces, such as wrinkles and furrows, by directly utilizing pixel values. Holistic spatial analysis including Principal Component Analysis (PCA) [19,20], Linear Discriminant Analysis (LDA) [21], and Independent Component Analysis (ICA) [6] has been applied to the whole face or specific face regions to extract the facial appearance changes. Typically these methods project face images onto a subspace, find a set of basis images, and represent faces as a linear combination of those basis images. The Active Appearance Models (AAM) [13] was also applied to facial expression recognition, which used PCA to model both shape and texture variations. The models can be fitted to new images by varying the shape and texture parameters within limits learned from a training set. Abboud et al. [22] applied LDA to the AAM parameters to obtain the most discriminative features and represent facial expression images. Sung et al. [23] combined the AAM with Active Shape Models (ASM) [24] to reduce the average model fitting errors. Ashraf et al. [25] utilized AAM derived representations for recognizing facial expression of pain.

In recent years, researchers have turned toward local descriptor based facial expression features as local descriptors have been shown to be more robust to occlusion, misalignment and moderate pose changes than traditional holistic methods [7–9,26]. Ojala et al. [27] proposed the Local Binary Pattern (LBP) as a computational effective texture descriptor. The original LBP operator labels the pixels of an image by thresholding a 3×3 neighborhood of each pixel with the center value and considering the results as a binary number, and use the derived binary number to represent texture primitives. In order to capture dominant features at a large scale, the original LBP operator was later extended to use neighborhood of different sizes. Using circular neighborhoods and bi-linearly interpolating the pixel values allow any radius and number of pixels in the neighborhood [14]. The LBP histogram contains information about the distribution of the local micro-patterns. Thus face images can be effectively represented by LBP histograms as shown in these works [10,28]. Shan et al. [10] performed a comprehensive study on facial expression recognition using LBP features. Different machine learning methods have been employed to classify expressions on several datasets, including SVM, Linear Discriminant Analysis (LDA) and linear programming. It is argued that LBP is more robust and efficient than Gabor wavelet features. By combining Gabor filtering with LBP, Local Gabor Phase Patterns (LGBP) [29] was proposed to extended LBP to multiple resolution and orientation. Xie et al. proposed Local Gabor XOR Pattern [30] to exploit the Gabor phase information. Liu and Wang [31] used a combination of multiple Gabor features