Signal Processing: *Image Communication* 29 (2014) 177–188

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Linear subspaces for facial expression recognition

Niki Aifanti *, Anastasios Delopoulos

Multimedia Understanding Group, Information Processing Laboratory, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece

ARTICLE INFO

Article history:
Received 25 March 2012
Received in revised form
26 May 2013
Accepted 22 October 2013
Available online 5 November 2013

Keywords:
Face analysis
Expression recognition
Subspaces

ABSTRACT

This paper presents a method for the recognition of the six basic facial expressions in images or in image sequences using landmark points. The proposed technique relies on the observation that the vectors formed by the landmark point coordinates belong to a different manifold for each of the expressions. In addition experimental measurements validate the hypothesis that each of these manifolds can be decomposed to a small number of linear subspaces of very low dimension. This yields a parameterization of the manifolds that allows for computing the distance of a feature vector from each subspace and consequently from each one of the six manifolds. Two alternative classifiers are next proposed that use the corresponding distances as input: the first one is based on the minimum distance from the manifolds, while the second one uses SVMs that are trained with the vector of all distances from each subspace. The proposed technique is tested for two scenarios, the subject-independent and the subject-dependent one. Extensive experiments for each scenario have been performed on two publicly available datasets yielding very satisfactory expression recognition accuracy.

1. Introduction

Facial expression recognition from images or image sequences has attracted the attention of computer vision scientists for the last two decades. Recognition of facial expressions is important for various applications which range from interactive human–computer interfaces to psychological tests.

Some expression recognition algorithms attempt to measure facial activity based on the Facial Action Coding System [1,2], while other attempt to recognize a small set of prototypic expressions. The most commonly used such set includes the six basic facial expressions (anger, disgust, fear, happiness, sadness, surprise) defined by Ekman [3].

Significant in facial expression recognition is the extraction of relevant image features describing changes in the appearance of the face. These features can be either holistic or anatomical. Holistic features (e.g. the image texture or a transformation of it) are used in [4–8]. Many approaches though are based on anatomical features (e.g. distance measurements, geometrical or appearance deformations of anatomical features/areas). For example, anatomical features are used in [9–14].

Facial expressions are dynamic phenomena where the features vary from frame to frame. For example, Lien in [15] uses optical flow fields, while Choudhury and Pentland in [16] use motion field histograms. In order to benefit from motion dynamics, Hidden Markov Models are used in many approaches [10,15–17].

A different class of techniques that tries to learn motion dynamics is based on the manifolds formed in the feature space. Various features can be used in these approaches such as raw image pixels, transformed texture, facial landmark points or the parameters of a model. These approaches use embedding techniques that depict non-linearly the feature vectors into a low-dimensional space, while keeping the initial relative distances between the feature vectors.

Manifold-based analysis of facial expressions was introduced by Chang et al. in [18,19]. They claim that although

* Corresponding author.
E-mail addresses: naif@iti.gr (N. Aifanti), adelo@eng.auth.gr (A. Delopoulos).

0923-5965/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.image.2013.10.004

images of a subject lie in a very high-dimensional space, a class of images generated by latent variables lies on a manifold in this space. Thus, they consider that expression images of one subject form a manifold in the feature space. They represent images by landmark points and they apply non-linear dimensionality reduction methods in order to embed the feature vectors representing the images of a subject into a low-dimensional space, while keeping the main structure of the subject's manifold. They use a database of two subjects with multiple image sequences and they recognize the six basic expressions only of these two subjects.

A major difficulty in manifold-based analysis is that the path formed by the images of a subject performing an expression usually does not match with the path formed by the images of another subject performing the same expression. Actually, they can be quite different. Shan et al. [20] use appearance features for 96 subjects from the Cohn–Kanade database and develop an alignment method in order to align the paths that correspond to same expressions in the different subjects' manifolds and achieve expression recognition.

On the other hand, a recently published approach [21] does not perform any alignment. In this approach the shape and appearance parameters of an Active Appearance Model (AAM) are used. They compute feature deformations by subtracting the parameters of every expression image of a subject from the parameters of a neutral image of the subject. Authors claim that this representation does not require any alignment process for a unified manifold embedding because they assume that all people have similar patterns of facial expressions and consequently similar patterns of feature deformations. This enables the direct application of k-Isomap framework in order to compute one manifold embedding space for all persons.

In [22] a multiple manifold approach is presented. For each expression one manifold embedding space is computed using Locality Preserving Projection (LPP). The intrinsic features of each expression are learned separately and then a Genetic Algorithm (GA) is employed to obtain the nearly optimal dimensionality of each expression manifold. The proposed method uses appearance features (raw image data, local binary patterns, Gabor features) and is tested on the six basic expressions.

Chen and Huang in [23] develop a different approach for expression recognition that performs clustering of images corresponding to the same expression. This method uses image pixel values as features and applies PCA only for dimensionality reduction prior to clustering. Afterwards, Clustering based Discriminant Analysis (CDA) is applied. Chen and Huang develop CDA, which is a modification of Linear Discriminant Analysis, in order to handle data that form different clusters per class. CDA can be applied to two-class experiments. Thus, Chen and Huang use the one-to-many method to classify using CDA three expressions (smile, anger, neutral).

In our approach expression classification is performed on the basis of shape information alone. Shape information is expressed by means of the 2D coordinates of a number of p landmark points. In practice, AAM methodology is used for the automatic detection of these landmark positions. Afterwards, the subject dependent offset of landmark coordinates at neutral state is subtracted. The remaining deviation-from-the-neutral vector is considered to contain the whole useful information. Thus, we represent each facial expression image by these $2p$-dimensional vectors, which we call expression vectors.

Using the expression vectors we eliminate to an extend the identity information inherent in the position of the landmark points, and we calculate the displacement of landmark points due to expression. Furthermore, we have relatively low-dimensional feature vectors.

Since not the entire $\mathbb{R}^{2p}$ space corresponds to valid expression vectors, we adopt the manifold approach. We assume that expression vectors that correspond to the same expression belong to a manifold. We avoid though to directly map the manifold of each expression to some low-dimensional subspace.

We make instead – and later validate – the assumption that the manifold of each expression can be parameterized by a small number of low-dimensional linear subspaces. These subspaces one-to-one correspond to clusters of the expression paths. An expression path is formed by expression vectors that belong to the image sequence of an expression instantiation. Although the identity information in expression vectors is removed to an extend, expression paths form clusters due to the different ways of performing an expression.

Clustering is achieved using the principal directions of the expression paths. The principal direction of a path is a linear representation of the path. We show that despite its linearity, principal direction is able to represent an expression.

Based on the clustering of the expression paths belonging to the same expression, we define a low-dimensional basis for each subspace, which is able to span the expression vectors that lie in the corresponding subspace. This yields a parameterization of the manifolds that allows for computing the distance of a feature vector from each subspace and consequently from each one of the six manifolds.

The proposed method uses simple and fast techniques, such as k-means and PCA, on a relatively small number of samples with low dimensionality. This means that it has a very low computational cost. Moreover, the proposed method, in contrast to the other manifold-based approaches, does not use embedding techniques which require many training samples and they do not generalize well to new samples.

Furthermore, it does not pose any limitation either to the number of subjects or to the number of expressions. It is able to classify expressions in images that depict mild to peak expression intensity and it takes into account the fact that there are different ways that a facial expression is performed. For this reason, it is expected to be efficient not only with posed expressions but also with spontaneous expressions.

Next section presents firstly the expression vectors and paths in the landmark points' space. Then, it describes the structure of expression manifolds and in the last subsection it explains how the manifold structure is learned. In Section 3 the two classifiers based on distances of the manifolds are presented. The next section describes a method for the calculation of the expression vector at the