Neurocomputing 82 (2012) 99–108

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

# A remarkable standard for estimating the performance of 3D facial expression features

Xiaoli Li [a,b,*], Qiuqi Ruan [a,b], Yue Ming [a,b]

[a] Institution of Information Science, Beijing Jiaotong University, Beijing 100044, China
[b] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China

ARTICLE INFO

Article history:
Received 23 December 2010
Received in revised form
28 September 2011
Accepted 8 October 2011
Communicated by D. Tao
Available online 30 December 2011

Keywords:
3D facial expression recognition
Relative entropy
KL divergence
Discrimination power
Performance evaluation

ABSTRACT

All the previous work on 3D facial expression recognition is always based on different feature extraction algorithms and different classifiers, so there is no uniform standard for us to identify which features are the "best". This paper investigates KL divergence (relative entropy) for discrimination power computation to determine the "best" features in this field. From experiments, we can conclude that local facial expression features in flow-matrix form are more beneficial to 3D facial expression recognition than in geometry-matrix form; the feature points in local expression regions can be more discriminative than points in face contour; and the slope and angle features are more powerful than distance features. Above all, this paper verifies that the KL divergence can definitely be considered as the standard for determining the "best" features to recognize 3D facial expressions. This is the first exploration on BU-3DFE (Binghamton University 3D Facial Expression) database to find a standard for evaluating the extracted facial expression features, and all of these results are remarkable for 3D facial expression feature extraction.



## 1. Introduction

Facial expression recognition is a challenging field in numerous psychology researches. People's mental activities can be identified by their facial expressions, thus it is significant in security systems, entertainment places, etc. Also the researches on facial expression recognition will open a broad new way for changing traditional human–machine interactive mode. All these requirements inspired a number of researchers dedicated to facial expression recognition.

Facial expression recognition has been researched for decades. Traditionally, it is based on 2D static images [1–5] or 2D video sequences [6–8] with a wide variety of features, (e.g. LBP [1], image ratio feature [2], Gabor wavelets [4], etc.), using previous approaches (e.g. LPP [5], HMM [6,7], etc.). It is impossible to name all of them here (for further discussion, one can refer to the excellent survey [9]). However, the common theme in these current researches is that facial models are treated as flat patterns, and thus expression variations are considered on the picture plane. Therefore, it is hard to detect subtle in-depth skin motions, which could significantly affect the performance of facial expression recognition [10]. Moreover, large variations of head pose and illuminations deteriorate the effectiveness of facial features tracking in 2D facial expression recognition. Expression analysis on the 3D facial surface can overcome this challenging issue and decrease the influence of large pose variations.

The facial expressions are closely related to the shape changes of the whole 3D facial surface, and the previous literature has pointed out the distinctive advantages of expression processing in multi-dimensional space [11]. Some recent work [12–14] has expressed that the approaches based on 3D facial models can perform better than those based on 2D images. The work in [15] provided a positive answer to construct multimodal probabilistic graphical models for tensor format data, which bridged the relationship between tensor descriptor and the vector-based feature representation. All these exploratory researches enable us to target facial expression recognition in 3D space.

To foster 3D facial expression research, Binghamton University created a 3D facial expression database, named BU-3DFE database [10], which included 100 subjects with 2500 facial models. The research has made great improvements recently, and the following list illustrates the most related work based on BU-3DFE database:

- Wang et al. [16] are pioneers in 3D facial expression recognition using BU-3DFE database. They constructed seven local expression regions based on 64 primary points and extracted 12 primitive facial features from each of the local regions and with the LDA classifier, the highest recognition rate was higher than 80%.
- Soyel and Demirel [17] extracted six distance vectors, which were based on the facial shape information for facial expression recognition. Then they got the average facial expression recognition rate up to 91.3% with a previously trained neural

* Corresponding author at: Institution of Information Science, Beijing Jiaotong University, Beijing 100044, China

E-mail addresses: 09112087@bjtu.edu.cn, lily69915@gmail.com (X. Li), qqruan@center.njtu.edu.cn (Q. Ruan), 08112075@bjtu.edu.cn (Y. Ming).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.10.029

network. Meanwhile in [18], they used another six distance vectors and the average recognition rate of 87.8% was reached.

- Tang and Huang [19] worked for facial expression recognition based on a regularized multi-class AdaBoost classification. The selected features were composed of the fusion of 24 normalized Euclidean distances and they achieved the average recognition rate of 95.1%. In addition, they extracted a set of 96 distinguishing features to describe the properties of line segments for facial expression recognition in [20].
- Venkatesh et al. [21] proposed the modified PCA to extract the discriminative features on a limited set of feature points. Moreover, they only employed spectral flow matrices as features to recognize facial expressions based on the uniformly sampled 3D matrix structure [22]. The proposed method achieved the average recognition rate of 85.56% based on the *k*-means clustering algorithm, but the uniformly sampled 3D matrix structure is remarkable.
- Tekguc et al. [12] used NSGA II to determine the optimal set of facial features from the entire feature space, which was produced by normalized distance vectors based on BU-3DFE database. The average classification rate of the proposed strategy reached up to 88.18%, which was better than the 2D methods, such as Gabor Wavelet and Topographic Context Method, and the method proposed in [16].
- With the properties of high relevance and low redundancy, NCBF [23] was introduced to extract the facial expression features, then PCA was performed on the selected features to obtain the most discriminative information. This two-step approach outperformed the conventional methods.

According to the previous work, recent researches have mainly focused on solving the challenging issues of BU-3DFE database. However, with the different classifiers for facial expression recognition, there is no uniform standard that can be used to determine whether the extracted features better performed than other features. An attractive scheme is to look for an automatic algorithm, which is able to identify the "best" features from a large feature pool.

To select the "best" features, this paper calls for an automatic algorithm based on the Kullback–Leibler divergence (KLD) (or named relative entropy), which is able to identify whether a feature is the "best". The metric of relative entropy was already used in [24,25] to measure distances between models.

According to the previous work both on 3D facial expression recognition and KL divergence applications, this paper extracts different sets of facial expression features from BU-3DFE database detailed in Section 2. Kullback–Leibler divergences as the discrimination power for all kinds of the features are computed in Section 3. Detailed experimental results are described in Section 4. Section 5 and Section 6 present the discussion and concluding remarks, respectively.

## 2. Candidate feature pool generation

In BU-3DFE database, 83 feature vertices were marked on each cropped facial model. Given the set of the labeled feature points, the feature regions on the face surface can be easily detected, and these features could be used as baseline for performance evaluation of the algorithms for 3D facial expression analysis. Their distribution and the corresponding numbers are presented in Fig. 1 using VrmlPad.[1]

[1] VrmlPad is currently one of the strongest VRML source editing tools. It was developed by Parallel Graphics, and supports real-time preview of the VRML source files.

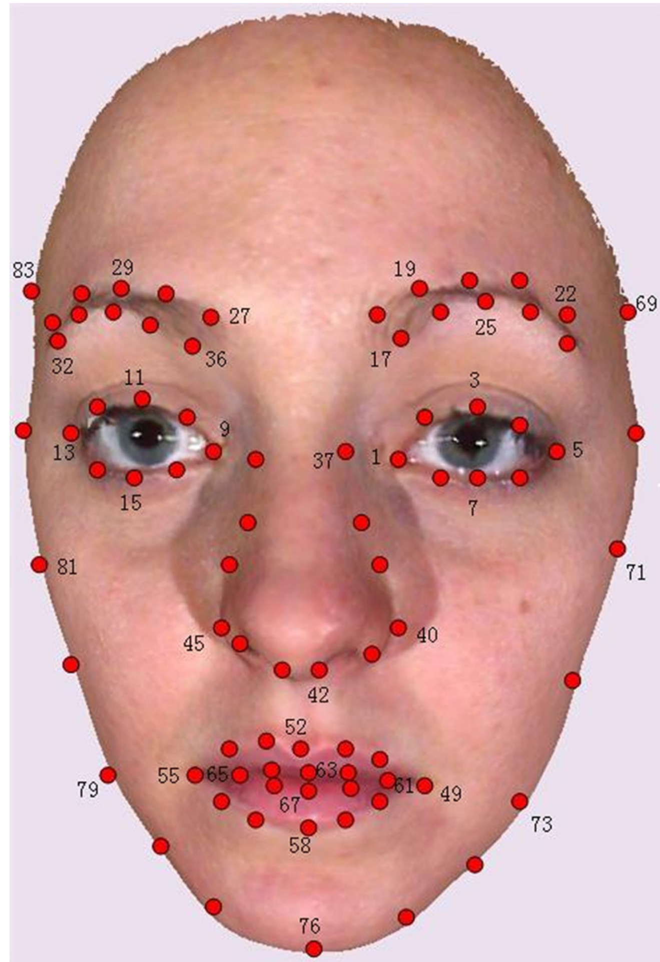


**Fig. 1.** 83 feature points on each cropped facial model and each point is numbered in turn. The numbers of the points are ordered according to eyes, eyebrows, nose, mouth, and face contour, from left to right, up to down.

Based on these labeled feature points, a series of geometry features are calculated referring to recent publications on 3D facial expression recognition [12,16–22,29], including distance features, slope features, and angle features. All these features will be described in the following sections, and thus each set of these features, and their fusions are used to generate a large candidate feature pool.

### 2.1. Facial model normalization

The facial models in BU-3DFE database are obtained independently, so they are based on the different coordinates and result in the different scales for different subjects. To obtain a better performance of the feature vectors, all the facial models should be normalized to the same scale, from the same view point and in the same coordinate system. The normalization is described in the following steps:

The 83 feature vertices defined on each facial model can be described as follows:

$$\alpha_i = (x_i, y_i, z_i), \quad i = 1, 2, \ldots, 83 \tag{1}$$

The nose tip is assigned to be the origin of *X–Y–Z* coordinate system, since it is the least impacted point by facial expressions. It can be defined by the 42th feature point shown in Fig. 1, and it has been defined as follows:

$$origin = \alpha_{42} = (x_{42}, y_{42}, z_{42}) \tag{2}$$