J. Vis. Commun. Image R. 35 (2016) 1–14

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

# 2D facial expression recognition via 3D reconstruction and feature fusion ☆

Ali Moeini [a,*], Karim Faez [a], Hamid Sadeghi [a], Hossein Moeini [b]

[a] Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran
[b] Electrical Engineering Department, Semnan University, Semnan, Iran

ARTICLE INFO

*Article history:*
Received 4 June 2015
Accepted 15 November 2015
Available online 2 December 2015

*Keywords:*
Facial expression recognition
Facial expression generic elastic model
3D local binary pattern
Local normal binary pattern
Expression-invariant 3D face reconstruction
Local binary pattern
Feature fusion
2D and 3D features

ABSTRACT

In this paper, a novel feature extraction method is proposed for facial expression recognition by extracting the feature from facial depth and 3D mesh alongside texture. Accordingly, the 3D Facial Expression Generic Elastic Model (3D FE-GEM) method is used to reconstruct an expression-invariant 3D model from the human face. Then, the texture, depth and mesh are extracted from the reconstructed face model. Afterwards, the Local Binary Pattern (LBP), proposed 3D High-Low Local Binary Pattern (3DH-LLBP) and Local Normal Binary Patterns (LNBPs) are applied to texture, depth and mesh of the face, respectively, to extract the feature from 2D images. Finally, the final feature vectors are generated through feature fusion and are classified by the Support Vector Machine (SVM). Convincing results are acquired for facial expression recognition on the CK+, CK, JAFFE and Bosphorus image databases compared to several state-of-the-art methods.



## 1. Introduction

Person-independent facial expression recognition that has generalization towards other images or databases is the most difficult and challenging task in computer vision. Automatic facial expression recognition has various significant applications, for instance, visual observation, intelligent user interface, human–computer interaction and so on. Facial image analysis from static images is a more challenging problem than dynamic images because less information for expression actions is available. Nonetheless, information in a single image is sometimes adequate and also useful for expression recognition. Generally, facial expression recognition encompasses two main steps: feature extraction and classification [1]. Hence, there are many methods to extract the features that robustly recognize facial expression under the restriction of one 2D single training sample for each class as static methods.

Using the 3D facial geometric information and extracting the 3D features has so far not been heavily studied for facial expression recognition. The 3D images give a large amount of data to be captured (3D and 2D), including out-of-plane movement which 2D cannot record, whilst also handling the problems of facial illumination and pose to 2D data. For this reason, some work has begun to employ 3D facial geometry data for facial expression recognition [2]. Researches performed in the context of 3D facial expression are still in the infant phase, with a large number of expected projects coming soon as the existing technological advances permit the affordable and simple acquisition for elevating the quality of 3D face information. Therefore, attempts yielded in this context are performed on 3D databases that have been pointed out in [2]. In this paper, the 3D face reconstruction from 2D facial expression databases is used that has so far not been heavily studied. It can resolve several issues that remain unsolved in this field in terms of 3D face acquisition.

The method to extract the features from 3D images includes mapping the 3D information onto a 2D representation. This representation can then be utilized for alignment, for the division of the mesh image prior to extracting the 3D features, or for the direct application of conventional 2D methods [40,45,46]. However, there are several scenarios where 3D captures are unavailable during the initial enrollment and only a single 2D face image is available from people. Therefore, reconstructing a 3D facial image from a single image is one of the fundamental problems of 3D face reconstruction in computer vision. This is due to several requests, containing pose-invariant face recognition [3–6], synthesizing new face images from a frontal 2D image [5], facial appearance modeling and a 3D facial expression analysis [7], age-insensitive face recognition [8], etc. A comprehensive study has been proposed in [3] for

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author at: Amirkabir University of Technology, Department of Electrical Engineering, Tehran, Iran. Fax: +98 2166406469.
*E-mail addresses:* ali.moeini1989@gmail.com, ali.moeini.ir@ieee.org (A. Moeini).

http://dx.doi.org/10.1016/j.jvcir.2015.11.006
1047-3203/

3D and multi-modal 3D + 2D face recognition. In this context, Huang et al. [7] proposed an efficient method to represent facial shapes from face scans through the reconstruction of face models based on regional data and a generic model. The efficiency of using 3D models has motivated us to use the 3D models for facial expression recognition. Thus, this paper uses a 3D reconstructed model for 2D facial expression recognition.

The Local Binary Patterns (LBPs) method is widely applied to the problem of facial expression recognition [9]. The LBPs have also been presented as: Local Phase Quantisers (LPQ) [10], Local Gradient Orientation Binary Patterns (LGOBPs) [11], and Local Gabor Binary Patterns (LGBPs) [12] for static images. Also, LBPs, LPQs and LGBPs have been extended to the dynamic problem in the form of LPQ-TOP [13], LBP-TOP [14] and V-LGBPs [15]. Moreover, some works have recently begun to apply descriptors of these types to the 3D problem by proposing the 3DLBP, the traditional LBP descriptor applied to the depth map of a facial mesh [16]. Furthermore, some works recently attempted to encode the shape of the mesh for action recognition that are: Local Azimuthal Binary Patterns (LABPs) [17], Local Azimuthal Phase Quantisers (LAPQs) [17], Local Normal Binary Patterns (LNBPs) [18], Local Depth Monogenic Binary Patterns (LDMBPs) [17], Local Depth Gabor Binary Patterns (LDGBPs) [17], etc.

In this paper, a combination of feature extraction approaches is proposed for person-independent facial expression recognition. An overview of the proposed method is illustrated in Fig. 1. Accordingly, a 3D model is initially reconstructed from 2D frontal facial images. For expression-invariant 3D face reconstruction from each human frontal face, the Facial Expression Generic Elastic Model (FE-GEM) [6] method is used. The FE-GEM method is the extension of the GEM [19] method to resolve the drawback of handling facial expression in 3D face reconstruction. Then, the texture, depth and mesh images are extracted from reconstructed models. Afterwards, for feature extraction, the LBP, proposed 3D High-Low Local Binary Pattern (3DH-LLBP) and LNBPs are applied to texture, reconstructed depth and reconstructed mesh images, respectively. Finally, the obtained features are combined through feature fusion and then the final feature vectors are generated and classified by the Support Vector Machine (SVM) [20]. To discover the impact of adding the depth and mesh images through 3D face reconstruction, the proposed method is compared with several state-of-the-art methods. In these comparisons, the results are initially achieved when each of the feature descriptors are used alone. Then, the impact of adding the depth and mesh images is investigated for fusion feature extraction.

In summary, the main contributions of this paper are as follows:

(1) The FE-GEM method is utilized to handle facial expression changes for the purpose of facial expression recognition.
(2) The 3DH-LLBP is proposed for feature extraction from facial depth images in a task of facial expression recognition.
(3) 2D facial expression recognition is performed by expression-invariant 3D face reconstruction and a combination of 2D and 3D features through feature fusion.
(4) The impact of adding the 3D features to 2D features through 3D face reconstruction is investigated for facial expression recognition.
(5) The proposed method improves the rate of facial expression recognition on five public face databases including MMI [21], CK+ [22], CK [23], JAFFE [24] and Bosphorus [25] compared to the state-of-the-art methods.
(6) The generalization power of the proposed method is evaluated across five public face databases: MMI [21], CK+ [22], CK [23], JAFFE [24] and Bosphorus [25].

This paper is organized as follows: Section 2 discusses the related work. Section 3 describes the 3D FE-GEM method for 3D face modeling from a single frontal face image. In Section 4, the feature extraction manner is proposed by feature fusion for facial expression recognition. Experimental evaluations are given in Section 5 and conclusions are presented in Section 6.

## 2. Related works

Due to its different applications, spontaneous facial expression recognition has engrossed much attention in facial image analysis by 2D and 3D images [2]. In this section, a brief review of the existing work on facial expression recognition that used 2D data as the input image is presented. Two common feature extraction methods used in facial expression recognition systems are: appearance feature-based methods and geometric feature-based methods. In the geometric feature-based methods, the geometric shape of facial components, such as the mouth and eyes, is utilized for face geometry representation. In the appearance feature based methods, facial expression appearance features are locally or globally extracted from either the whole face or specified regions of the face using such image filters as the Gabor wavelet.

There are several studies which use the geometric feature-based method for facial expression recognition [26–28]. For example, multiple feature detection approaches were used to obtain face

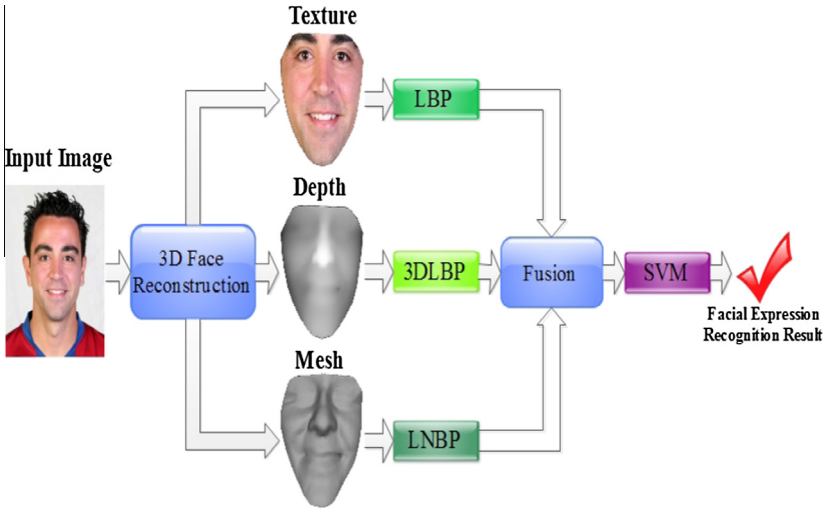


**Fig. 1.** An overview of the proposed method. The features are extracted from 2D images of texture, depth and mesh of face and are combined through feature fusion.