Computers and Electrical Engineering 38 (2012) 1299–1309

Contents lists available at SciVerse ScienceDirect

Computers and Electrical Engineering

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

Localized discriminative scale invariant feature transform based facial expression recognition ☆

Hamit Soyel [a,*], Hasan Demirel [b]

[a] Department of Computer Science, Queen Mary University of London, London E1 4NS, United Kingdom
[b] Department of Electrical and Electronic Engineering, Eastern Mediterranean University, Famagusta, Mersin 10, Turkey

ARTICLE INFO

Article history:
Available online 3 December 2011

ABSTRACT

This paper presents a discriminative scale invariant feature transform (D-SIFT) based feature representation for person-independent facial expression recognition. Keypoint descriptors of the SIFT features are used to construct distinctive facial feature vectors. Kullback Leibler divergence is used for the initial classification of the localized facial expressions and weighted majority voting based classifier is employed to fuse the decisions obtained from localized rectangular facial regions to generate the overall decision. Experiments on the Bosphorus and BU-3DFE databases illustrate that the D-SIFT is effective and efficient for facial expression recognition.

1. Introduction

Feature extraction is one of the most vital aspects in facial expression recognition. Deriving an effective facial representation from original expression images is an important process for successful facial expression recognition [1,2]. There are two common approaches to extract facial features: geometric feature-based methods and appearance-based methods.

Geometric features present the shape and locations of facial components, which are extracted to form a feature vector that represents the face geometry. As demonstrated in [3–6] geometric feature-based methods provide similar or better performance than appearance-based approaches by using primitive surface face descriptors. However, the geometric feature-based methods usually require accurate and reliable facial feature detection which is difficult to accommodate in many applications. The state-of-art approaches like SIFT [7] extend to the detection of localized regions of interest and improve performance by using region analysis. The SIFT aims to overcome many of the practical problems in low-level feature extraction in matching images. In [8], SIFT is used to detect and represent salient points in multiple range images derived from face models for the purpose of face recognition. SIFT descriptors have been also used to perform expression recognition from frontal and non-frontal face images [9,10].

This paper presents a new feature representation, called discriminative scale invariant feature transform (D-SIFT), which is invariant to image scaling and rotation, and partially invariant to changes in illumination. Keypoint descriptors of the SIFT features are used to construct distinctive facial feature vectors. Kullback Leibler divergence (KLD) [11] is used for the initial classification of the localized facial expressions and weighted majority voting (WMV) [12] based classifier is employed to fuse the decisions obtained from localized rectangular facial regions to generate the overall decision. Experiments on the Bosphorus [13] and BU-3DFE [14] illustrate that the D-SIFT is effective and efficient for facial expression recognition. We achieved the highest level of accuracy, 100% for surprise, where 96.2% average accuracy is achieved for all of the expressions including anger, disgust, fear, happy neutral, sadness and surprise.

☆ Reviews processed and approved for publication by Editor-in-Chief Dr. Manu Malek.
* Corresponding author.
E-mail address: hsoyel@eecs.qmul.ac.uk (H. Soyel).

0045-7906/$ - see front matter
doi:10.1016/j.compeleceng.2011.10.016

The rest of this paper is organized as follows: in Section 2, we explain the procedures applied during D-SIFT methodology. In Section 3, the proposed system is compared with conventional methods available in the literature. Finally, we give conclusions in Section 4.

2. Methodology

2.1. Overview of the proposed method

SIFT and WMV have been successfully applied to various tasks in computational face processing. In this paper, we apply a novel combination to the facial expression recognition task. The facial expression recognition system consists of three modules as shown in Fig. 1. In the first module of the training phase, PDF descriptors are calculated by using regular SIFT-feature extraction. In the second module, a KLD-based feature matching computation is processed among the facial expression images and a ranking procedure among the PDF descriptors, according to a discriminative criterion, is deployed. Once, this ranking is computed, we use the most discriminating PDF descriptors. Finally in the third module a WMV-based classifier is adopted to enhance the performance of the facial expression recognition.

2.2. Feature representation

SIFT features are extracted for any given image (i.e. a face or a subimage of a face) as its feature vector for classification. Scale and orientation invariant interest points are located by using the local extrema of Difference of Gaussian (DoG) pyramid [7]. To build the DoG pyramid the input image is convolved iteratively with a Gaussian kernel of σ. The last convolved image is down-sampled in each image direction by factor of 2, and the convolving process is repeated. This procedure is repeated as long as the down sampling is possible as depicted in Fig. 2.

Each collection of images of the same size is called an octave. All octaves build together the so-called Gaussian pyramid, which is represented by a 3D function $L(x, y, \sigma)$. The DoG pyramid $D(x, y, \sigma)$ is computed from the difference of each two nearby images in Gaussian pyramid. The local extrema (maxima or minima) of DoG function are detected by comparing each pixel with its 26 neighbors in the scale-space (8 neighbors in the same scale, 9 corresponding neighbors in the scale above and 9 in the scale below). The search for extrema excludes the first and the last image in each octave because they do not have a scale above and a scale below, respectively. To increase the number of extracted features the input image is doubled before it is treated by SIFT algorithm, which however increases the computational time significantly. In the method presented in this paper, the image doubling is avoided but the search for extrema is performed over the first three octaves and the pixel comparing is carried out only with available neighbors. The detected local extrema are good candidates for keypoints. However, they need to be exactly localized by fitting a 3D quadratic function to the scale-space local sample point. The quadratic function is computed using a second-order Taylor expansion having the origin at the sample point. Then, local extrema with low contrast and such that correspond to edges are discarded because they are sensitive to noise. Fig. 3 shows

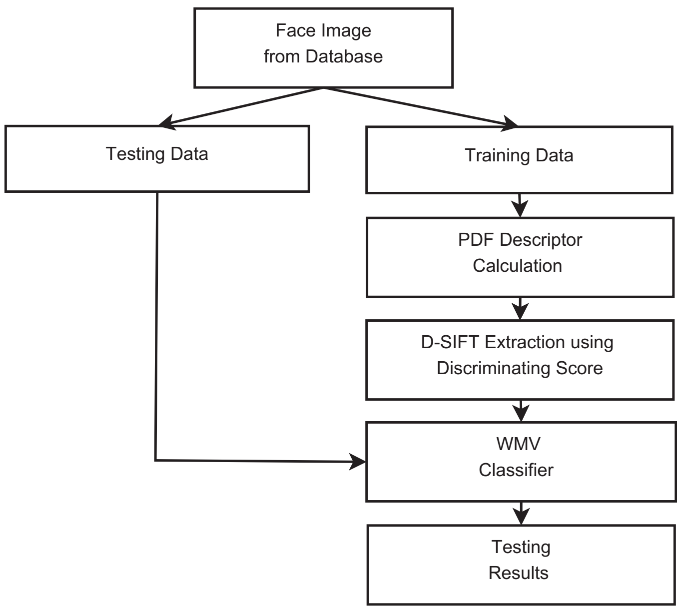

Fig. 1. Proposed facial expression recognition system based on D-SIFT.