Computers and Electrical Engineering 000 (2017) 1–12

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

A facial expression recognition system using robust face features from depth videos and deep learning[☆]

Md. Zia Uddin[a], Mohammed Mehedi Hassan[b,*], Ahmad Almogren[b], Mansour Zuair[b], Giancarlo Fortino[c], Jim Torresen[a]

[a] *Department of Informatics, University of Oslo, Norway*
[b] *College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia*
[c] *Department of Informatics, Modeling, Electronics, and Systems, University of Calabria, Italy*

A R T I C L E I N F O

Article history:
Received 13 October 2016
Revised 21 April 2017
Accepted 21 April 2017
Available online xxx

Keywords:
Depth image
Facial expression recognition
Modified local directional patterns
Generalized discriminant analysis
Deep belief network

A B S T R A C T

This work proposes a depth camera-based robust facial expression recognition (FER) system that can be adopted for better human machine interaction. Although video-based facial expression analysis has been focused on by many researchers, there are still various problems to be solved in this regard such as noise due to illumination variations over time. Depth video data in the helps to make an FER system person-independent as pixel values in depth images are distributed based on distances from a depth camera. Besides, depth images should resolve some privacy issues as real identity of a user can be hidden. The accuracy of an FER system is much dependent on the extraction of robust features. Here, we propose a novel method to extract salient features from depth faces that are further combined with deep learning for efficient training and recognition. Eight directional strengths are obtained for each pixel in a depth image where signs of some top strengths are arranged to represent unique as well as robust face features, which can be denoted as Modified Local Directional Patterns (MLDP). The MLDP features are further processed by Generalized Discriminant Analysis (GDA) for better face feature extraction. GDA is an efficient tool that helps distinguishing MLDP features of different facial expressions by clustering the features from the same expression as close as possible and separating the features from different expressions as much as possible in a non-linear space. Then, MLDP-GDA features are applied with Deep Belief Network (DBN) for training different facial expressions. Finally, the trained DBN is used to recognize facial expressions in a depth video for testing. The proposed approach was compared with other traditional approaches in a standalone system where the proposed one showed its superiority by achieving mean recognition rate of 96.25% where the other approaches could make 91.67% at the best. The deep learning-based training and recognition of the facial expression features can also be undertaken with cloud computing to support many users and make the system faster than a standalone system.

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. J. Wan.

* Corresponding author.
E-mail addresses: mdzu@ifi.uio.no (Md.Z. Uddin), mmhassan@ksu.edu.sa (M.M. Hassan), ahalmogren@ksu.edu.sa (A. Almogren), zuair@ksu.edu.sa (M. Zuair), g.fortino@unical.it (G. Fortino), jimtoer@ifi.uio.no (J. Torresen).

http://dx.doi.org/10.1016/j.compeleceng.2017.04.019
0045-7906/© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Human robot interactions (HRI) have attracted many researchers recently to contribute in developing smartly controlled healthcare systems [1,2]. Humans very often use nonverbal cues in their daily lives where the cues include hand gestures, facial expressions, and tone of the voice to express feelings. Hence, HRI systems should include these to make full benefit of natural interaction with the users. In a ubiquitous robotic healthcare system, HRI systems could be considerably improved if robots could understand peoples' emotions based on analyzing facial expressions and react as friendly as possible according to their current emotional states. Providing emotional healthcare support using robots could also be important to improve the quality of life. Humans' mental states are revealed through emotions in their daily situations where positive emotions represent healthy mental states by carrying positive facial expressions (e.g., pleasure and happiness). On the other hand, negative emotions can represent unhealthy mental states by carrying negative facial expressions such as anger and sadness. Thus, both positive and negative emotions can closely affect peoples' emotional health in their daily lives. To improve emotional health, a robust facial expression recognition (FER) system plays a key role in understanding mental states over time by the analysis of emotional behavior patterns.

Basically, vision-based FER systems can be categorized into two main types: pose-based and spontaneous. Pose-based FER systems usually recognize artificial facial expressions where expressions are produced by people by asking them to express a selection of expressions in sequence. On the contrary, spontaneous FER systems recognize the facial expressions that people does spontaneously in daily life such as during conservations and while watching movies. This work focuses on pose-based FER due to unavailability of pure spontaneous facial expression depth database. A video-based FER system consists of two types of classifications: frame-based and sequence-based. In the former one, only one frame is utilized to recognize different facial expressions. On the other hand, sequence-based methods apply temporal information in the frame sequences to recognize different facial expressions in videos. As image sequence-based FER systems contain more information than single frame-based one, this work utilizes face image sequences i.e., videos. A typical video-based FER system consists of three main parts: preprocessing, feature extraction, and recognition. In preprocessing, face area is detected in an image of a video. Feature extraction handles extracting salient features from each face to distinguish facial expressions. At last, face features are trained and used to recognize different facial expressions.

1.1. Organization of the paper

The rest of the paper is organized as follows. Section 2 discusses some significant research works related to the proposed system. Sections 3 and 4 illustrate the feature extraction process from depth images followed by expression modelling using deep learning, respectively. Furthermore, Section 5 describes experimental results using different approaches including the proposed one. Finally, Section 6 draws the conclusion of the work.

2. Related works

A huge amount of research works has been observed in developing reliable FER systems as it contributes to a large range of application domains in computer visions, image processing, and pattern classification [3–15]. A very challenging problem to solve in these is the ability of the computing systems to detect human faces to recognize underlying expressions in them such as angry, happy, neutral, sad, and disgust. Hence, accurate recognition of facial expressions is still considered to be a major challenge due to some parameters such as presence of noise from the environments due to illumination variation over time in the scene. Hence, robust FER still demands much attention to support various applications. The most important aspect for any FER approach is to find efficient feature representation in face images. Features in FER are considered as efficient when they can minimize within-class variations and maximize between-class variations. Hence, the main goal of feature extraction is to find a robust representation of face features which can provide robustness during the recognition process. Based on the features used in FER systems so far, feature extraction methods can be divided into two main categories i.e., geometric and appearance-based. In geometric feature-based FER, geometric feature vectors are formed considering geometric relationships such as angles and positions between different face parts such as eyes, ears, and nose. This has been a popular method but the effectiveness is highly dependent on accurate detection of facial components in the images, which can be a very challenging task in unconstrained environments and dynamic scenarios. Hence, researchers in this field moved their research directions over to appearance-based face analysis to obtain a more robust FER system. Appearance-based FER methods focus on facial appearance and try to do different analysis such as by applying filters on the whole face image. Among all the vision-based FER research, most of the work has been done using appearance-based methods. This work also describes an appearance-based FER system.

To represent facial expression features in a video, Principal Component Analysis (PCA) has been mostly applied in FER systems [3–6]. In [3], PCA was tried to recognize action units to represent and recognize different facial expressions. In [5], the authors applied PCA for providing a facial action coding system where different facial expressions were modelled and recognized. Independent Component Analysis (ICA), a higher level statistical approach than PCA was adopted later in FER works for statistically independent local face feature extraction [7–15]. In [10], ICA was adopted to obtain statistically independent features focusing on local face components (e.g., nose, lips, and eyebrows) in different facial expressions. In [11], ICA was adopted in FER to analyze facial action units in different facial expression images. For local feature extraction, some