Original research article

Facial expression recognition on real world face images using intelligent techniques: A survey

Sajid Ali Khan [a,b,*], Ayyaz Hussain [c], Muhammad Usman [a]

[a] Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan
[b] Department of Software Engineering, Foundation University Islamabad, Pakistan
[c] Department of Computer Science and Software Engineering, International Islamic University, Islamabad, Pakistan

ARTICLE INFO

ABSTRACT

Most of the research in facial expression recognition system has been done on the datasets which are acquired in predetermined lab environment. Acquiring real world data is crucial for the benchmarking of algorithms because facial expression recognition system will eventually have to work on the data from real world in an uncertain environment. In the recent years, the local descriptor-based facial expression recognition has been shown to be more robust to pose changes, misalignment and illumination effects. Local Binary Pattern (LBP) and Weber Local Descriptor (WLD) are commonly used local descriptors used to effectively represent local micro texture information of the face. Classification of facial expression images became more challenging due to different norms and gestures used among different races and ethnic groups and the change between different expression classes is very subtle. It necessitates storing not only micro but also macro texture information in case of real world face images to acquire better accuracy results. Traditional local descriptors are not capable to extract the macro texture information. The purpose of the proposed work is to present a novel framework to recognize facial expressions with high accuracy rate. The proposed framework intends to be time and memory efficient, and unlike LBP and WLD it would not only store micro but also macro texture information. Performance of the proposed technique will be measured using both synthetic (i.e. lab-based) and real world face images datasets.

© 2016 Elsevier GmbH. All rights reserved.

1. Introduction

Machine learning is a field of computer science which deals with developing such algorithms that may automatically learn without human intervention. Due to this learning ability machine can perform some tasks automatically like making accurate predictions, suggesting remedies for the disease, recognizing human face expressions, etc.

There are different ways to express the body emotions. The most important way among all is the facial expressions recognition. Recent research has proven that facial expressions and other emotions are an excellent source of conveying the non-verbal communication cues in face to face interactions. Sometimes a facial expression can be more useful instead of speaking a lot of words.

---

* Corresponding author at: Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Street # 09, Plot # 67, Islamabad 44000, Pakistan.
E-mail address: sajidalibn@gmail.com (S. Ali Khan).

Facial expression recognition is the process of classifying the human facial expressions like happiness, sadness and anger from the face images. Facial expressions recognition can be an important component of natural human-machine interfaces; it may also be used in various other fields of behavioral and medical sciences. Although humans can recognize facial expression quickly and precisely but reliable expression recognition by machine is still a challenge.

1.1. What is facial expression recognition?

Facial expression recognition involves identification of cognitive activity, deformation of facial feature and facial movements. This is done using static images and their sequences or videos. The purpose is to categorize them into different abstract classes based on the visual facts only. Naturally, human faces generally reflect the inner feelings/emotions and hence facial expressions are susceptible to changes in the environment. This makes a human face index of mind; therefore, expression recognition assists in interpreting the states of mind and distinguishes between various facial gestures.

1.1.1. Natural similarity of facial expression

The process of recognition and identification is important due to the similarity of facial expressions. The deformation occurs due to expressed emotions on the human faces. Despite the variation in age, ethnicity, gender there is a similarity in facial expressions. It was first stated by Darwin in 1872 [1]. After almost 100 years later, Ekman and Friesen [2] suggested six expressions named as *sadness, happiness, anger, surprise, fear and disgust*. These six emotions are known as basic emotions with their own and different nature. This natural similarity in facial expressions of human is utilized by every facial expression recognition system.

1.2. Facial expression recognition as a research topic in computer vision

The motivation behind any research area is its capability to resolve a problem and its applications. The main purpose of all images processing and computer vision algorithms is to make the visual data useful. Hence in the domain of computer vision, the facial expression recognition begins with the same purpose. The importance and requirement of this research area are enhanced because of its applications in the HCI (Human Computer Interaction) where visual look of human, sight and touch sensations (also known as moodes) and voice are utilized at the same time. Besides, social psychology states that facial expressions supports in coordinating conversation. Facial expression produces 55% of the result of spoken message if it delivered along with visual information [3]. The contribution of spoken words is 7% and vocal tone contributes 38%. Therefore facial expression is the most significant among all of HCI modes. This makes the research on facial expression recognition necessary. In addition to this domain robot vision, facial animation and virtual reality also needs analysis of facial expression. These systems are constructed with the purpose to eliminate human error in vehicle control. The recognition of basic emotion and recognition of Aus (action units) are the two categories on which research has been conducted. The purpose of basic emotion recognition is to track the states of mind while the recognition of "Aus" focuses on representation of muscular deformation of facial muscles.

1.3. Realistic facial expression database overview

A database gives a consistent platform in order to test the algorithms. The algorithms are usually examined against particular image or video databases. These databases consist of images and their respective ground truths. There could be some variations in the content of images like different subjects, age, illuminations, gender, ethnicities and textures due to uncontrolled and varying circumstances. Existing facial expression databases are classified into two classes: Lab-based database, where the emotions are intentionally expression under controlled environment and realistic database, where the emotions naturally occur in an uncontrolled environment (i.e. real-world conditions). Most of the existing facial expression database belong to the first class which includes JAFFE, BU-3DFE, CK+, Semaine, SAL, MMI, AAI and NVIE. In contrast to lab-based emotions, realistic facial expression contains big variations in illumination face pose, size and facial occlusions etc and thus they are more challenging to categories and have more importance in real-world applications. Examples of realistic databases are shown in Table 1. Belfast [4] and luggage [5] lost early attempt at collecting realistic data in unconstrained environment such as conversion between subjects and interviews. Recent studies collected data from sources like World Wide Web and TV broadcasts to represent the facial expression in real situations and produced databases like Gv [6], (HAPPEI) [7], static facial expressions in the wild (SFEW) [8], happy people images GENKI–4 K [9], acted facial expressions in the wild (AFEW) [10], HUMAINE [11], TV data [12] and VAM [13]. Basic emotion categories are not directly annotated in VAM and HUMAINE video frames. Yeasin et al. [12], presented TV broadcast data but details are not provided. Compared with other data, AFEW is new dataset which consist of images and videos collected from different movies. SFEW database extracts 700 images from AFEW videos. Flickr website is used to collect 4600 happiness images for the creation of HAPPEI database. Similarly GENKI–4 K database consists of smile images, while detail is not provided for Gv images.