Volume : 2 | Issue : 11 | Nov 2013 • ISSN No 2277 - 8160

Research Paper | Engineering

# Automatic Human Emotion Recognition Using Particle Swarm Optimization Based Facial Action Coding System


| | |
|---|---|
| Nehru P | B.S.Abdur Rahman University, Chennai |
| S. Revathi | B.S.Abdur Rahman University, Chennai |
| X. Arputha Rathina | B.S.Abdur Rahman University, Chennai |



## ABSTRACT

*The interaction between human beings and computers will be more natural if computers are able to perceive and respond to human non-verbal communication such as emotions. Although several approaches have been proposed to recognize human emotions based on facial expressions or speech, relatively limited work has been done to fuse these two, and other, modalities to improve the accuracy and robustness of the emotion recognition system. In this paper, to propose an particle swarm optimization (PSO) algorithm based facial action coding system (FACS) for automatic human emotion recognition. The method proposed here enables the detection of a much larger range of facial behavior by recognizing facial muscle actions [action units (AU5)] that compound expressions. AUs are agnostic, leaving the inference about conveyed intent to higher order decision making (e.g., emotion recognition). The proposed fully automatic method not only allows the recognition of 22 AUs but also explicitly models their temporal characteristics (i.e., sequences of temporal segments: neutral, onset, apex, and offset). To do so, it uses a facial point detector based on Gabor-feature-based boosted classifiers to automatically localize 20 facial fiducial points.*


**KEYWORDS : detection; particle optimization; PSO; facial emotions; expressions; facial action units**

## . I. INTRODUCTION

Affective computing has gained enormous research interest in the development of human computer interaction over the past decades. With the increasing power of emotion recognition, an intelligent computer system can provide a more friendly and effective way to communicate with users in areas such as video surveillance, interactive entertainment, intelligent automobile system and medical diagnosis [1]. The computer vision based pattem recognition techniques largely focuses on facial features due to the importance of human perception of emotion through visual facial expressions [2]. Each emotion corresponds to unique facial expression, whose characteristics serves as the inpnt to the classification system, and thc outpnt gives the machine-perceived emotion statc based on various classification algorithms [3].

Differently from previous versions, thc current version is fully automated and aimed at tht recognition of 22 AUs and their temporal activatior models. The system described in this work is the firsi that is able to explicitly model the temporal dynamics of AUs in terms of its temporal phases. Also, thu work describes extensive tests on databases of posec facial expression data as well as on spontaneow facial expression data.

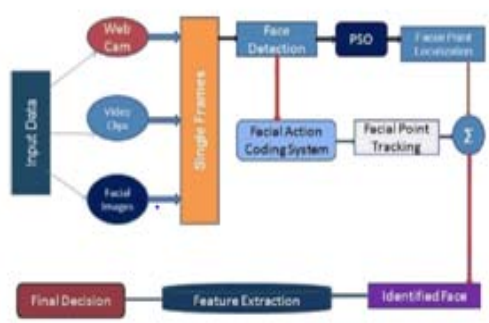


Figure 1. The Proposed System Diagram

Two main streams in the current research on automatic analysis of facial expressions consider facial affect (emotion) detection and facial muscle action detection [4]—[10]. These streams stem directly from the two major approaches to facial expression measurement in psychological research [11]: message and sign judgment. The aim of the former is to mfer what underlies a displayed facial expression, such as affect or personality, while the aim of the latter is to describe the "surface" of the shown behavior, such as facial movement or facial component shape. Thus a frown can be judged as "anger" in a message judgment approach and as a facial movement that lowers and pulls the eyebrows closer together in a sign-judgment approach. While message judgment is all about interpretation, sign judgment is agnostic

independent from any interpretation attempt leaving the inference about the conveyed message to higher order decision making.

Most facial expression analyzers developed so far adhere to the message judgment stream and attempt to recognize a small set of prototypic emotional facial expressions such as the six basic emotions proposed by Ekman [7]—[9] & [12]. Even though the automatic recognition of the six basic emotions from face images and image sequences is considered largely solved, reports on novel approaches are published even to date (e.g., [13]—[16]) Exceptions from this overall state of the art in machine analysis of human facial affect include few tentative efforts to detect cognitive and psychological states like interest [17], pain [18], [19], and fatigue [20]

In sign judgment approaches [21], a widely used method for manual labeling of facial actions is the Facial Action Coding System (FACS) [22]. FACS associates facial expression changes with actions of the muscles that produce them. It defines 9 different action units (AU5) in the upper face, 18 in the lower face, and 5 AUs that cannot be classified as belonging to either the upper or the lower face.

FACS also provides the rules for the recognition of AUs' temporal segments (onset, apex, and offset) in a face video. Using FACS, human coders can manually code nearly any anatomically possible facial expression, decomposing it into the specific AUs and their temporal segments that produced the expression. As Aus are independent of any interpretation, they can be used as the basis for any higher order decision making process including the recognition of basic emotions [22], cognitive states like (dis)agreement and puzzlement [23], psychological states like pain [24], and social signals like emblems (i.e., culture-specific interactive signals like a wink, coded as left or right AU26), regulators (i e, conversational mediators like the exchange of a look, coded by AUs for eye position), and illustrators (i e, cues accompanying speech like raised eyebrows, coded as AU1+AU2) *[25]*. Hence, AUs are extremely suitable to be used as midlevel parameters in an automatic facial behavior analysis system as they reduce the dimensionality of the problem [26] (thousands of anatomically possible facial expressions *[25]* can be represented as combinations of 32 AU5).

## II. RELATED WORKS

Edge detection is the first stage for many computer vision tasks, which roughly can be understood as detecting the abrupt changes of gray level intensity in a digital image A broad selection of edge detectors are available, simplest ones are the Roberts, Sobel and Prewitt edge detector, however they work poorly under complex background with extreme conditions. The Canny detector has been an industry standard edge detector. The essence of this method is to achieve maximum signal to noise ratio (SNR), good localization, and single response However it's a first derivative based operator such that it's less sensitive to subtle facial lines. Moreover non-maxima suppression as part of this algorithm works poorly under random noise.

The edge information plays a critical role in a way they represent the geometric of an image such as the size and shape of a facial component As compared to gradient based operators, a second derivative based edge detecting mechanism called Marr-Hildreth detector has caught our attention due to its greater response to weak, non-rigid, deformable facial muscle movements. However, any Laplacian based edge detector has drawbacks such as higher response to unwanted high frequency details, the appearance of double edge and the so called 'spaghetti effect' when improper zero -crossing detection is used. Certainly these undesired natures must be treated carefully.

## III. METHODOLOGY

### *A. Face Region Localization*

The first step in any facial information extraction process is face detection, i.e., the identification of all regions in the scene that contain a human face. The second step in facial expression analysis is to extract *geometric features* (facial points and shapes of facial components) andlor *appearance features* (descriptions of the texture of the face such as wrinkles and furrows). The work presented here is a typical example of a geometric-feature-based method.

The facial region is detected using the face detector with local normalization and optimal adaptive correlation [12]. Compare to Viola and Johns' algorithm [12], this method is adaptive to the normalized input image and designed to complete the segmentation in a single iteration.

### *B. Particle Swarm Optimization (PSO)*

PSO is a population-based search algorithm initially designed to simulate the social behavior of birds in a flock as they fly in search of food. A PSO algorithm maintains a swarm of particles, where each particle represents a potential solution [7]. Particles are "flown" through a multi -dimensional search

space where the position of a particle is adjusted according to two factors

• Its own successful experience

•The successful experiences of its neighbors.

The experiential knowledge of the Particle is referred to as the *cognitive* component, and is proportional to the distance of the particle from its own best position [18]. The socially exchanged information is referred to as the *social* component of the velocity equation. Originally, two PSO algorithms were developed, which differ in the size of their particle neighborhoods.

```
Create and initialize an n dimensional swarm;
repeat
    for each particle i = 1,      n do
        //set the personal best position
        if f(xi) < f(yi) then
            yi = xi;
        end
        !! set the golobal best position
        if gyi)cfiy)then
            '=3%
        end

    for each particle i = 1 . n do
        update the velocity using equation (2);
        update the position using equation (1);

until stopping condition is true
```

Figure. 2. PSO (Global best) algorithm.

*C PFFL For Facial Point Tracking*

After the fiducial facial points are found in the first frame, we track their positions in the entire image sequence. Standard optical flow techniques are commonly used for facial point tracking in facial expression analysis.

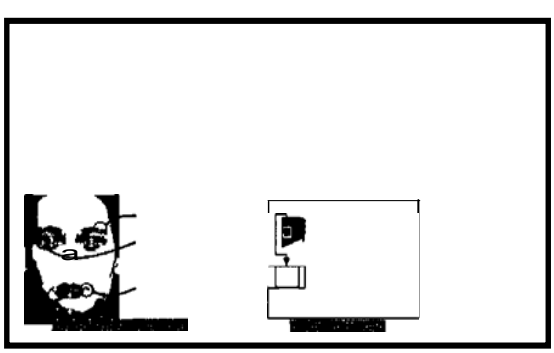

Figure. 3. Outline of the fully automated fiducial facial point detection method.

The tracking scheme that we adopt is based on particle filtering The main idea behind particle filtering is to maintain a set of solutions that are an efficient representation of the conditional probability *p(aY),* where a is the state of a temporal event to be tracked given a set of noisy observations *Y= {yl,* y—, y) up to the current time instant. This means that the distribution p(aY) is represented by a set of pairs *(sic, ffk}* such that, if *sic* is chosen with probability equal to *nk,* then it is as if *sic* was drawn *fromp(a* Y) By maintaining a set of solutions instead of a single estimate (as is done by Kalman filtering), particle filtering is able to track multimodal conditional probabilities *p(a Y* ), and it is therefore robust to missing and inaccurate data and particularly attractive for estimation and prediction in nonlinear non-Gaussian systems. In the particle filtering framework, our knowledge about the a *posterzori* probabi1ityp(aY) is updated in a recursive way

Several researchers used the condensation algorithm to track facial features in face image sequences (e.g., [28] and [29]). However, the algorithm has three major drawbacks. The first is that a large amount of particles that result from sampling from the proposal density *p(a Y —)* might be wasted because they are propagated into areas with small likelihood. Second, the scheme ignores the fact that while a particle *sk = skl, sk2 skN* might have low likelihood, parts of it might be close to the correct solution. Finally, the estimation of the particle weights does not take into account the interdependences between the different parts of a Neither do they encode any information about the rate of change of the values of these features in consecutive frames (e.g., the velocity of a point)

*D. Emotion Detection Evaluation*

The detection of six basic emotions in posed facial expression databases is considered to be largely solved, particularly when the subject being tested is known and was part of the training data. However for optimal comparability with existing automatic facial expression recognition works, we evaluate our six-basic-emotion detection system on *50* Images and some videos taken from the CK-db. The videos were selected with the criterion that two coders were able to attain a consensus on what emotion was shown in that video. This is a stricter ground truth criterion than using the ground truth provided with the CK db This strategy was used to reduce the label error in the data set.

Emotions are detected per video, i.e , the table shows event detection results. From the results we have to conclude that it is very hard to distinguish between the emotions angry and sadness. The reason

for this is that both expressions often incur similar brow movements

## IV. EXPERIMENTS AND RESULTS

### *A Face Detection*

Because of its practical importance and relevance to face recognition and, in turn, for security, face detection received a lot of attention. Numerous techniques have been developed. However, virtually all of them can detect only (near-) upright faces in (near-) frontal view. Most of these methods emphasize statistical learning techniques and use appearance

### *B Facial Edge Detection*

Methods for facial feature point detection can be classified as either *texture based methods* (modeling local texture around a given facial point) or *shape-based methods* (which regard all facial points as a shape that is learned from a set of labeled faces). A typical texture-based methods is that of Holden and Owens [27j, who used log-Gabor filters.

The point detection algorithm is tolerant to changes in illumination as long as they remain locally constant.

## V. CONCLUSIONS

Accurate fully automatic facial expression analysis would have many real-world applications In this paper, we have shown that not only fully automatic highly accurate AU activation detection based on geometric features is possible but also that it is possible to detect the four temporal phases of an AU with high accuracy and that geometric features are very well suited for this task. The proposed system was tested extensively on multiple databases and was shown to generalize well when trained on data from one database and tested on data from another. This being said, generalization to completely novel data is not possible yet without some loss of accuracy. At this point, a major limitation of the system is that it can only recognize facial expressions as long as the face is viewed from a pseudo frontal view. If the head has an out-of-plane rotation greater than 200, the system will fail. This is something that we wish to address in our future research.

## REFERENCES

Y.Tie L. Guan "Human Emotion Recognition Using a Deformable 3D Facial Expression Model", ISCAS, 1115- 1118, 2012. | [2] P. Ekman and E. L. Rosenberg, 14iat the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System. Oxford, U.K.: Oxford Univ Press, 2005. | [3] J. Russell and J. Fernandez-Dols, The | Psychology of Facial Expression. New York | Cambridge Univ. Press, 1997. | [4] B. Golomb and T. Sejnowski, "Benefits of machine understanding of facial expressions," in NSF Report—Facial Expression Understanding P. Ekman, T. Huang, T. Sejnowski, and J. Hager Eds. Salt Lake City, UT, 1997, pp. 55—71. | [5] M. Pantic, "Face for ambient interface," in | Ambient Intelligence in Everyday Life, vol. 3864 | Lecture Notes on Artificial Intelligence. Berlin Germany: Springer-Verlag, 2006, pp. 32—66