A Comprehensive Survey on Human Facial Expression Detection

Archana Verma & Lokesh Kumar Sharma

Abstract

In the recent years recognition of Human's Facial Expression has been very active research area in computer vision. There have been several advances in the past few years in terms of face detection and tracking, feature extraction mechanisms and the techniques used for expression classification. This paper surveys some of the published work since 2001. The paper gives a time-line view of the advances made in this field, the applications of automatic face expression recognizers, the characteristics of an ideal system, the databases that have been used and the advances made in terms of their standardization and a detailed summary of the state of the art. The paper also discusses facial parameterization using FACS Action Units (AUs) and advances in face detection, tracking and feature extraction methods. It has the important role in the human-computer interaction (HCI) systems. There are multiple methods devised for facial feature extraction which helps in identifying face and facial expressions.

Keywords: Facial Expression, FACS, Fuzzy Inference System, Feature Extraction, HCI.

1. INTRODUCTION

Facial features and expressions are critical to everyday communication. Besides speaker recognition, face assists a number of cognitive tasks: for example, the shape and motion of lips forming visemes can contribute greatly to speech comprehension in a noisy environment. While intuition may imply otherwise, social psychology research has shown that conveying messages in meaningful conversations can be dominated by facial expressions, and not spoken words. This result has led to renewed interest in detecting and analyzing facial expressions in not just extreme situations, but also in everyday human–human discourse. A very important requirement for facial expression recognition is that all processes therein have to be performed without or with the least possible user intervention. This typically involves initial detection of face, extraction and tracking of relevant facial information, and facial expression classification. In this framework, actual implementation and integration details are enforced by the particular application. For example, if the application domain of the integrated system is behavioral science, real-time performance may not be an essential property of the system.

A Facial Expression is a visible manifestation of the affective state, cognitive activity, intention, personality and psychopathology of a person [1]. Facial Expression convey nonverbal communication cues in face-to-face inter actions. Ekman and Freisen[1] have produced FACS – Facial Action Coding System for describing visually distinguishable Facial movements [2] [32]. Using the FACS, Action Parameters is designated to each of the expressions which classify the Human Emotions [2]. Also, Mehrabian[2] indicated that the verbal part of a message contributes for 7% to the effect of the message; the vocal part contributes 38% while facial
expression contributes for 55% [3]. There are wide range of applications of Facial Expressions, some of which include image understanding, psychological studies, facial nerve grading in medicine [4], face image compression and synthetic face animation [5]. Previous studies on the automatic analysis of Facial Expressions have discussed the classification methods and the extraction methods [6] [7] [8]. Although, there are various studies on the applications of computers in areas related to our proposed intelligent sales systems in the form of e-commerce and m-commerce, personalized online product selection, and web-based shopping systems we have, so far, found limited conceptual or empirical studies in this field. This paper is conceptual in nature and therefore aimed at exploring and initiating the debate in this field.

2. RELATED WORK
It has been well researched that facial expressions do reflect cognitive behaviour and that individuals observe other’s facial expressions and then use these to regulate their own behavior in social interactions. We do find some studies relating facial expression to some aspects of marketing but none on facial expression and shopping behavior. Howard and Gengler [4] found that favourable facial expressions do have a positive bias on consumer product attitude. Sirakaya and Sonmez [5] used facial expression to study the gender images in Government tourism brochures. They describe a computerized intelligent sales assistant that gives sales personnel the ability to allocate their time where it will produce the best results, both for the customer, and for the business. Derbaix [6] also used facial expression to investigate the effect of television advertisement on attitude towards advertisement and brand attitude. They investigated 8- to 11-year-olds’ reactions to advertising communication with a specific focus on the construct validity of the scales used to measure the main variables studied. Yuasa et al. [7] developed a computer network negotiation support tool using facial expression to negotiate an agreement strategy between seller and buyer. They argued that, if players select a happy face, then there is a greater chance they will reach an agreement. Lee [8] developed computer software programs to implement an e-shopping authentication scheme that used facial recognition instead of user name and password to control access to the web site in a modern consumer e-shopping environment. Consumers would not need to remember their user name and password to get access to the web site. The computer, based upon the facial features of the user, would recognize faces and allow access to the web site to authorized users.

A recent study in the context of computer based tutoring systems has shed some light on the types and frequencies of facial expressions exhibited by students during interactions with a computer. These are the same as the results that were found in another recent study of human tutoring sessions in which it is suggested that the expressions displayed by students are not significantly affected by whether the tutor is human or artificial. However, much more complete studies would be required before firm conclusions could be reached on various aspects of the facial expressiveness of customers.

The outcome of this research has some useful applications in the field of marketing, especially in sales (online or physical) and the training of sales people. There is also an added advantage that, if sales people can quickly identify the potential customers from window shoppers with the intervention of the system, then they can spend more time on potential customers and convert these potential customers into buyers. It will save the sales people time and the customer will feel well attended to. This will also reduce the expenses on sales forces as the stores will employ less people as they can identify the potential customer and devote sufficient time on them to convert them into buyers instead of spending unreasonable time on window shoppers. Using the same technology, online shoppers can be directed to appropriate products which could result in increased sales.

3. DIFFERENT FACIAL EXPRESSIONS
Prior to the compilation of the FACS in 1977, most of the facial behaviour researchers were relying on the human observers who would observe the face of the subject and give their analysis. But such visual observations cannot be considered as an exact science since the
observers may not be reliable and accurate. Ekman et al. [1] questioned the validity of such observations by pointing out that the observer may be influenced by context. They may give more prominence to the voice rather than the face and furthermore, the observations made may not be the same across cultures; different cultural groups may have different interpretations. Following Figure 1 demonstrates the seven universal expressions of emotion. Each of these expressions is racially and culturally independent.

![Images of facial expressions](image1.png)

**FIGURE 1:** Some Sample Images from The JAFFE Database.

The limitations that the observers pose can be overcome by representing expressions and facial behaviors in terms of a fixed set of facial parameters. With such a framework in place, only these individual parameters have to be observed without considering the facial behavior as a whole. Even though, since the early 1920s researchers were trying to measure facial expressions and develop a parameterized system, no consensus had emerged and the efforts were very desperate [9] [10]. To solve these problems, in 1978, Ekman and Friesen [1] published the Facial Action Coding System (FACS), which, 30 years later, is still the most widely used method available. Through observational and electromyography study of facial behavior, they determined how the contraction of each facial muscle, both singly and in unison with other muscles, changes the appearance of the face. These changes in the face and the underlying (one or more) muscles that caused these changes are called Action Units (AU). The FACS is made up of several such action units. Figure 2 and 3 illustrates some of these Action Units.
For example:

- AU 1 is the action of raising the Inner Brow. It is caused by the Frontalis and Pars Medialis muscles,
- AU 2 is the action of raising the Outer Brow. It is caused by the Frontalis and Pars Lateralis muscles,
- AU 26 is the action of dropping the Jaw. It is caused by the Masetter, Temporal and Internal Pterygoid muscles, and so on [10]. However not all of the AUs are caused by facial muscles.
Some of such examples are:
- AU 19 is the action of ‘Tongue Out’,
- AU 33 is the action of ‘Cheek Blow’,
- AU 66 is the action of ‘Cross Eye’, and so on. The Face can be divided into Upper face [11] and Lower Face Action units [12] and the subsequent expressions are also identified. The Figures shows some of the combined action units.

4. STRUCTURE OF FACIAL ACTION

According to the manual of FACS [9] the facial expression analysis system estimates the measurement of actions and also classifies the actions. Various steps are involved in Facial Expression analysis i.e. 1. Face Acquisition 2. Facial Expression Extraction 3. Expression Recognition. Following Figure 4 shows the basic structure of facial expression analysis system.

Faces are detected from the input images or image sequences in the first step which is Face Acquisition phase. It can detect faces from input image or detect face form image sequence like detect face in the first frame and track the face in the remaining frames. After the face is located the facial features are extracted to identify the facial expression. Facial expression can be classified into two types namely Geometric or Intransient features and Appearance Features or Transient Features.

Geometric or Intransient Features: The features that are always present in the face but may be deformed due to any kind of facial expression. Eg) Eyes, Eyebrows, Mouth, Tissue Textures, Nose. The facial components or facial feature points are extracted to form a feature vector that represents the face geometry.

Appearance or transient Features: The features that appear temporarily in the face during any kind of Facial Expression. Eg) Different kinds of wrinkles, bulges, forehead, regions surrounding the mouth and eyes. With appearance-based method, image filters such as Gabor wavelets [14] are applied to either the whole-face or specific regions in a face image to extract a feature vector. Facial Expression Recognition is the last step in facial expression analysis where the extracted features are recognized based on the action units. The Recognizer identifies not only the basic emotions like anger, happy, surprise, sad[13] but also identifies the expression caused due to pain[14], temporal dynamics [15], Intensity of Expression[16], Spontaneous Expression [17].

![Basic Structure of Facial Expression analysis.](image)

**FIGURE 4:** Basic Structure of Facial Expression analysis.

4.1 Face Acquisition

The two basic face acquisition methods are to detect faces both in frontal view images and near frontal view images. To detect the faces, two methods are used namely face detection and head pose estimation.
4.1.1 Face Detection
Many detection methods have been employed to detect faces [18] [19] [20] [21] [22]. Some of the previous face detection methods since 2003 are summarized here. Yeasin et al. [9] had used robust and automated face detection to segment the face region which was based on the work of Rowley [9]. Bartlett et al. [3] had designed boosting techniques for face detection in a generative framework based on their own work [17][18]. Tong et al. [10] had drawn geometry of face by encoding the edges of the face in the graph based on the work of Wiskott [10]. Kotsia and Patras [13] had employed convolution neural network for detecting the face and the classification is performed using a rule based algorithm [19].

4.1.2 Head Pose Estimation
To handle the out-of-plane head motion, head pose estimation can be employed. The methods for estimating head pose can be classified as 3D model-based methods [20] [21] and 2D image-based methods [28]. In 3D Model based method Bartlett used a canonical wire-mesh face model to estimate face geometry and 3D pose from hand-labeled feature points. In 2D image based method to handle the full range of head motion for expression analysis, Tian et al. [11] detected the head instead of the face. The head is identified using the smoothed silhouette of the foreground object as a segment using background subtraction and computing the negative curvature minima (NCM) points of the silhouette.

4.2 Facial Expression Extraction
Feature extraction can be viewed as a pre-processing step which removes distracting variance from a dataset, so that downstream classifiers or regression estimators perform better. The area where feature extraction ends and classification, or regression, begins is necessarily murky: an ideal feature extractor would simply map the data to its class labels, for the classification task. On the other hand, a character recognition neural net can take minimally pre-processed pixel values as input, in which case feature extraction is an inseparable part of the classification process. The extraction is basically based on the type of features, Geometric Features and Appearance Features. The two basic concepts employed for extracting features are based on identifying facial deformation and facial motion. The deformation based features recognize the Action Units, and the classifier is trained to differentiate human emotional states based on identified Action Units.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Geometric Features</th>
<th>Appearance Features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deformation Extraction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Image Based</td>
<td>Gabor Filter[28]</td>
<td>Local Gabor Filter Bank [14], Fisher’s Linear Decomposition, Singular value Decomposition [21] [22]</td>
</tr>
<tr>
<td>Model Based</td>
<td>Point Distribution Model [20]</td>
<td>Feature point Tracking [25]</td>
</tr>
<tr>
<td><strong>Motion Extraction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frame Based</td>
<td>Active Contour (snake) [18]</td>
<td>Gabor Filter Bank [26]</td>
</tr>
<tr>
<td>Sequence Based</td>
<td>PCA [23] [24], Gabor Filter Bank &amp; AdaBoost [14]</td>
<td>Haar like feature [17], Multimodal facial feature Tracking [24], Candid Grid Node [16].</td>
</tr>
</tbody>
</table>

**TABLE 1**: Facial Expression Extraction Methods.

The deformation kind of extraction is applied to images and to image sequences [19]. The motion based features exploit the temporal correlation of facial expressions to identify variations within a probabilistic framework [15]. Image based models extract features from images, or reduced dimensional facial components [14]. Model based features are usually shape or texture models that fit human faces. The output of the feature extractor stage must contain separable and classifiable vectors. Active appearance models [22] and point distribution models [23] are used to fit on the shapes of interest. These shapes constitute the feature vectors. The expression
extraction methods are widely classified under two kinds namely deformation extraction and motion extraction. As for motion extraction techniques, some commonly used methods are dense optical flow [14], feature point tracking [5], and difference images [16]. The Various techniques under facial expression extraction methods are tabulated in the table.

4.2.1 Geometric Feature Extraction
Geometric Extraction is to detect and track changes of facial components in near frontal face images. Tian et al. [11] developed multi-state models to extract the geometric facial features. A three-state lip model describes the lip state: open, closed, tightly closed. A two-state model (open or closed) is used for each of the eyes. Each brow and cheek has a one-state model. Some appearance features, such as nasolabial furrows and crow’s-feet wrinkles (Figure 5), are represented.

Model Based: Automatic Active Appearance Model (AAM) mapping can be employed to reduce the manual preprocessing of the geometric feature initialization. Xiao et al. [11] performed the 3D head tracking to handle large out-of-plane head motion and track no rigid features. Once the head pose is recovered, the face region is stabilized by transforming the image to a common orientation for expression recognition [22].

Image Sequence: Given an image sequence, the region of the face and approximate location of individual face features are detected automatically in the initial frame. The contours of the face features and components then are adjusted manually in the initial frame. After the initialization, all face feature changes are automatically detected and tracked in the image sequence. The system groups 15 parameters for the upper face [11] and 9 parameters for the lower face [12], which describe shape, motion, and state of face components and furrows. To remove the effects of variation in planar head motion and scale between image sequences in face size, all parameters are computed as ratios of their current values to that in the reference frame.

4.2.2 Appearance Feature Extraction
Gabor wavelets are widely used to extract the facial appearance changes as a set of multiscale and multi-orientation coefficients [14][20][28]. The Gabor filter may be applied to specific locations on a face or to the whole face image [16] [21]. There are two types of features to recognize expressions, the geometric positions of 34 fiducial points on a face and 612 Gabor wavelet coefficients extracted from the face image at these 34 fiducial points.
Fiducial points marked by circles (global) and big black dots (local), and contour points marked by small black dots. Image Sequence: Techniques like Haar-like Feature, Facial Feature tracking are used to identify the facial features that produces the expressions. A multi-modal tracking [24] approach is required to enable the state switching of facial components during the feature tracking process. Twenty-six fiducial points and 56 contour points are used in the facial model. Using the Facial model the fiducial points are marked for an image sequence using feature tracking method. The marked features in an image sequence are shown in the figure 8.

4.3.1 Frame-Based Expression Recognition
Frame-based expression recognition does not use temporal information for the input images. It uses the information of current input image with/without a reference frame. The input image can be a static image or a frame of a sequence that is treated independently. Several methods can be found in the literature for facial expression recognition such as neural networks [5][6], support vector machines [20], linear discriminant analysis [21], and rule-based classifiers [17].

4.3.2 The Sequence-Based Recognition
This method uses the temporal information of the sequences to recognize the expressions of one or more frames. To use the temporal information, the techniques such as HMM [29], recurrent neural networks, and rule-based classifier [17] were employed in facial expression analysis. The article by Yunfeng Zhu, Fernando De la Torre, Jeffrey F. Cohn [12] have made comparative studies for FACS AU recognition in spontaneously occurring behaviour by using the same RU-FACS database. Several systems had tried to recognize AUs or expression in spontaneously occurring behaviour [17].
The Expression recognition classifiers are listed below

<table>
<thead>
<tr>
<th>Technique</th>
<th>Title</th>
<th>Year</th>
<th>Author(s)</th>
<th>Method</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame based</td>
<td>A New Facial Expression Recognition Method based on local Gabor filter bank and PCA plus LDA</td>
<td>2005</td>
<td>Deng et al. [13]</td>
<td>PCA plus LDA</td>
<td>JAFFE</td>
</tr>
<tr>
<td></td>
<td>A Region Based methodology for facial expression recognition</td>
<td>2006</td>
<td>Koutlas and Fotiadis [15]</td>
<td>Neural Networks</td>
<td>JAFFE</td>
</tr>
<tr>
<td></td>
<td>Automatic Recognition of Facial Actions in Spontaneous Expressions</td>
<td>2006</td>
<td>Bartlett et al. [16]</td>
<td>SVM &amp; ADABOost</td>
<td>RU-FACS</td>
</tr>
<tr>
<td></td>
<td>A Facial Expression Classification System Integrating Canny, Principal Component Analysis and Artificial Neural Network</td>
<td>2001</td>
<td>Thai et al. [17]</td>
<td>PCA &amp; ANN</td>
<td>JAFFE</td>
</tr>
<tr>
<td>Sequence Based</td>
<td>Recognition of Facial Expressions and Measurement of Levels of Interest From Video</td>
<td>2006</td>
<td>Yeasin et al. [18]</td>
<td>Hidden Markov Model</td>
<td>Cohn Kanade</td>
</tr>
<tr>
<td></td>
<td>Boosting encoded dynamic features for facial expression recognition</td>
<td>2009</td>
<td>Yang et al. [19]</td>
<td>Adaboost</td>
<td>Cohn Kanade</td>
</tr>
<tr>
<td></td>
<td>Robust facial feature tracking under varying face pose and facial expression</td>
<td>2007</td>
<td>Tong et al. [21]</td>
<td>Gabor filters &amp; Switching Hypothesis Measurement</td>
<td>FRGC 1.0</td>
</tr>
</tbody>
</table>
Dynamics of facial expression extracted automatically from video 2006 Littlewort et al. [4] SVM Cohn Kanade


**TABLE 2:** Expression Recognition Methods.

6. CONCLUSION

The objective of this paper is to show a survey on the structure of analyzing the facial expression. The steps involved in expression analysis like face acquisition, feature extraction and expression classification had been discussed. Each step is discussed with the approaches and methods that can be applied to attain the required goal. Facial expression recognition is the key to next generation human-computer interaction (HCI) systems. We have chosen a more integrated approach as compared to most of the general applications of FACS. Extracted facial features are used in a collective manner to find out ultimate facial expression.

7. REFERENCES


12. Yunfeng Zhu, Fernando De la Torre, Jeffrey F. Cohn, Associate Member, IEEE, and Yu-Jin Zhang, Senior Member, IEEE “Dynamic Cascades with Bidirectional Bootstrapping for Action Unit Detection in Spontaneous Facial Behavior”, Journal of LATEX Class Files, October 2010


