Facial Expression Recognition and Analysis

Miss. Bharati Chaudhari & Prof. R. R. Deshmukh
Computer Science and Engineering, H.V.P.M. COET, Amravati, India

Abstract—The face is one of the most powerful channels of nonverbal communication. Facial expression provides cues about emotion, intention, alertness, pain, personality, regulates interpersonal behavior, and communicates psychiatric and biomedical status among other functions. Within the past 15 years, there has been increasing interest in automated facial expression analysis within the computer vision and machine learning communities. This paper reviews fundamental approaches to facial measurement by behavioral scientists and current efforts in automated facial expression recognition.

Keywords—Facial expression analysis, Action unit recognition, Active Appearance Models, temporal clustering

1. Introduction

Studies on Facial Expressions and Physiognomy date back to the early Aristotelian era (4th century BC). Physiognomy is the assessment of a person's character or personality from their outer appearance, especially the face [1]. But over the years, while the interest in Physiognomy has been waxing and waning [1], the study of facial expressions has consistently been an active topic. The foundational studies on facial expressions that have formed the basis of today's research can be traced back to the 17th century. In 1667, Le Brun gave a lecture at the Royal Academy of Painting which was later reproduced as a book in 1734 [2]. It is interesting to know that the 18th century actors and artists referred to his book in order to achieve “the perfect imitation of 'genuine' facial expressions” [2].

Moving on to the 19th century, one of the important works on facial expression analysis that has a direct relationship to the modern day science of automatic facial expression recognition was the work done by Charles Darwin. In 1872, Darwin wrote a treatise that established the general principles of expression and the means of expressions in both humans and animals [3]. He also grouped various kinds of expressions into similar categories. The categorization is as follows:

- low spirits, anxiety, grief, dejection, despair
- joy, high spirits, love, tender feelings, devotion
- reflection, meditation, ill-temper, sulkiness, determination
- hatred, anger
- disdain, contempt, disgust, guilt, pride
- surprise, astonishment, fear, horror
- self-attention, shame, shyness, modesty

Furthermore, Darwin also cataloged the facial deformations that occur for each of the above mentioned class of expressions. For example: “the contraction of the muscles round the eyes when in grief”, “the firm closure of the mouth when in reflection”, “the depression of the corners of the mouth when in low spirits”, etc [3]. Traditionally (as we have seen), facial expressions have been studied by clinical and social psychologists, medical practitioners, actors and artists. However in the last quarter of the 20th century, with the advances in the fields of robotics, computer graphics and computer vision, animators and computer scientists started showing interest in the study of facial expressions.

The first step towards the automatic recognition of facial expressions was taken in 1978 by Suwa et al. Suwa and his colleagues presented a system for analyzing facial expressions from a sequence of images (movie frames) by using twenty tracking points.

2. Applications

Automatic face expression recognition systems find applications in several interesting areas. With the recent advances in robotics, especially humanoid robots, the urgency in the requirement of a robust expression recognition system is evident. As robots begin to interact more and more with humans and start becoming a part of our living spaces and work spaces, they need to become more intelligent in terms of understanding the human’s moods and emotions. Expression recognition systems will help in creating this
intelligent visual interface between the man and the machine.

Humans communicate effectively and are responsive to each other’s emotional states. Computers must also gain this ability. This is precisely what the Human-Computer Interaction research community is focusing on: namely, Affective Computing. Expression recognition plays a significant role in recognizing one’s affect and in turn helps in building meaningful and responsive HCI interfaces.

Apart from the two main applications, namely robotics and affective HCI, expression recognition systems find uses in a host of other domains like Telecommunications, Behavioral Science, Video Games, Animations, Automobile Safety, Educational Software, etc.

Another interesting application has been demonstrated by Anderson and McOwen, called the ‘EmotiChat’. It consists of a chat-room application where users can log in and start chatting. The face expression recognition system is connected to this chat application and it automatically inserts emoticons based on the user’s facial expressions.

3. Facial Parameterization

The various facial behaviors and motions can be parameterized based on muscle actions. This set of parameters can then be used to represent the various facial expressions. Till date, there have been two important and successful attempts in the creation of these parameter sets:

A. The Facial Action Coding System (FACS) developed by Ekman and Friesen in 1977 [4] and

B. The Facial Animation parameters (FAPs) which are a part of the MPEG-4 Synthetic/Natural Hybrid Coding (SNHC) standard, 1998 [5].

Let us look at each of them in detail:

A. The Facial Action Coding System (FACS)

Prior to the compilation of the FACS in 1977, most of the facial behavior 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. questioned the validity of such observations by pointing out that the observer may be influenced by context [5]. 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 [5].

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 disparate [5]. To solve these problems, Ekman and Friesen developed the comprehensive FACS system which has since then become the de-facto standard.

Facial Action Coding is a muscle-based approach. It involves identifying the various facial muscles that individually or in groups cause changes in facial behaviors. 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. 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,
- 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 [6].

AUs can be additive or non-additive. AUs are said to be additive if the appearance of each AU is independent and the AUs are said to be non-additive if they modify each other’s appearance [7]. Having defined these, representation of facial expressions becomes an easy job. Each expression can be represented as a combination of one or more additive or non-additive AUs. For example ‘fear’ can be represented as a combination of AUs 1, 2 and 26 [7]. Figs. 1 and 2 show some examples of upper and lower face AUs and the facial movements that they produce when presented in combination.
B. The Facial Animation Parameters (FAPs)

In the 1990s and prior to that, the computer animation research community faced similar issues that the face expression recognition researchers faced in the pre-FACS days. There was no unifying standard and almost every animation system that was developed had its own defined set of parameters. As noted by Pandzic and Forchheimer, the efforts of the animation and graphics researchers were more focused on the facial movements that the parameters caused, rather than the efforts to choose the best set of parameters [8]. Such an approach made the systems unusable across domains. To address these issues and provide for a standardized facial control parameterization, the Moving Pictures Experts Group (MPEG) introduced the Facial Animation (FA) specifications in the MPEG-4 standard. Version 1 of the MPEG-4 standard (along with the FA specification) became the international standard in 1999.

In the last few years, face expression recognition researchers have started using the MPEG-4 metrics to model facial expressions [9]. The MPEG-4 standard supports facial animation by providing Facial Animation Parameters (FAPs). Cowie et al. indicate the relationship between the MPEG-4 FAPs and FACS AUs: “MPEG-4 mainly focusing on facial expression synthesis and animation, defines the Facial Animation parameters (FAPs) that are strongly related to the Action Units (AUs), the core of the FACS” (page 125 from [9]). To better understand this relationship between FAPs and AUs, I give a brief introduction to some of the MPEG-4 standards and terminologies that are relevant to face expression recognition. The explanation that follows in the next few paragraphs has been derived from [5], [8], [9], [12], [13] (readers interested in a detailed overview of the MPEG-4 Facial Animation technology can refer to the survey paper by Abrantes and Pereira [9]. For a complete in-depth understanding of the MPEG-4 standards, refer to [5] and [8]. Raouzaiou et al. give a detailed note on FAPs, their relation to FACS and the modeling of facial expressions using FAPs [13]).

The MPEG-4 defines a face model in its neutral state to have a specific set of properties like a) all face muscles are relaxed; b) eyelids are tangent to the iris and so on. Key features like eye separation, iris diameter, etc are defined on this neutral face model [5].

The standard also defines 84 key feature points (FPs) on the neutral face [5]. The movement of the FPs is used to understand and recognize facial movements (expressions) and in turn also used to animate the faces. Fig. 3 shows the location of the 84 FPs on a neutral face as defined by the MPEG-4 standard.
Fig. 3: The 84 Feature Points (FPs) defined on a neutral face.

The FAPs are a set of parameters that represent a complete set of facial actions along with head-motion, tongue, eye and mouth control (we can see that the FAPs like the AUs are closely related to muscle actions). In other words, each FAP is a facial action that deforms a face model in its neutral state. The FAP value indicates the magnitude of the FAP which in turn indicates the magnitude of the deformation that is caused on the neutral model, for example: a small smile versus a big smile. The MPEG-4 defines 68 FAPs [5].

4. Emotions, Expressions And Features

One of the means of showing emotion is through changes in facial expressions. But are these facial expressions of emotion constant across cultures? For a long time, Anthropologists and Psychologists had been grappling with this question. Based on his theory of evolution, Darwin suggested that they were universal [12]. However the views were varied and there was no general consensus. In 1971, Ekman and Friesen conducted studies on subjects from western and eastern cultures and reported that the facial expressions of emotions were indeed constant across cultures [12]. The results of this study were well accepted. However, in 1994, Russell wrote a critique questioning the claims of universal recognition of emotion from facial expressions [13].

Cross cultural studies have shown that, although the interpretation of expressions is universal across cultures, the expression of emotions though facial changes depend on social context. For example, when American and Japanese subjects were shown emotion eliciting videos, they showed similar facial expressions. However, in the presence of an authority, the Japanese viewers were much more reluctant to show their true emotions through changes in facial expressions. In their cross cultural studies (1971), Ekman and Friesen used only six emotions, namely happiness, sadness, anger, surprise, disgust and fear. In their own words: “the six emotions studied were those which had been found by more than one investigator to be discriminable within any one literate culture” These six expressions have come to be known as the ‘basic’, ‘prototypic’ or ‘archetypal’ expressions.

Apart from the six basic emotions, the human face is capable of displaying expressions for a variety of other emotions. In 2000, Parrott identified 136 emotional states that humans are capable of displaying and categorized them into separate classes and subclasses.

5. Characteristics Of A Good System

We are now in a position where we can list down the features that a good face expression recognition system must possess:

- It must be fully automatic
- It must have the capability to work with video feeds as well as images
- It must work in real-time
- It must be able to recognize spontaneous expressions
- Along with the prototypic expressions, it must be able to recognize a whole range of other expressions too (probably by recognizing all the facial AUs)
- It must be robust against different lighting conditions
- It must be unobtrusive
- The images and video feeds do not have to be pre-processed
- It must be person independent
- It must work on people from different cultures and different skin colors. It must also be robust against age (in particular, recognize expressions of both infants, adults and the elderly)
- It must be invariant to facial hair, glasses, makeup etc.
• It must be able to work with videos and images of different resolutions
• It must be able to recognize expressions from frontal, profile and other intermediate angles.

6. Challenges And Future Work

The face expression research community is shifting its focus to the recognition of spontaneous expressions. As discussed earlier, the major challenge that the researchers face is the non-availability of spontaneous expression data. Capturing spontaneous expressions on images and video is one of the biggest challenges ahead. As noted by Sebe et al., if the subjects become aware of the recording and data capture process, their expressions immediately lose its authenticity.

A possible work for the future is the automatic recognition of microexpressions. Currently training kits are available that can train a human to recognize microexpressions. It will be interesting to see the kind of training that the machine will need.

The medical conditions that lead to ‘loss of facial expression’. Although it has no direct implication to the current state of the art, I feel that, as a face expression recognition researcher, it is important to know that there exist certain conditions which can cause what is known as ‘flat affect’ or the condition where a person is unable to display facial expressions. There are 12 causes for the loss of facial expressions namely: Asperger syndrome, Autistic disorder, Bell’s palsy, Depression, Depressive disorders, Facial paralysis, Facial weakness, Hepatolenticular degeneration, Major depressive disorder, Parkinson’s disease, Scleroderma, and Wilson’s disease.

7. Conclusion

This paper’s objective was to introduce the recent advances in face expression recognition and the associated areas in a manner that should be understandable even by the new comers who are interested in this field but have no background knowledge on the same. In order to do so, we have looked at the various aspects of face expression recognition in detail. Let us now summarize: We started with a time-line view of the various works on expression recognition. We saw some applications that have been implemented and other possible areas where automatic expression recognition can be applied. We then looked at facial parameterization using FACS AUs and MPEG-4 FAPs. Then we looked at some notes on emotions, expressions and features. The last section was the challenges and possible future work to be done.

Face expression recognition systems have improved a lot over the past decade. The focus has definitely shifted from posed expression recognition to spontaneous expression recognition.

8. References