EMOTION RECOGNITION BASED ON THE ANALYSIS OF FACIAL EXPRESIONS. A SURVEY

ANDONI BERISTAIN* and MANUEL GRAÑA[t]
Computational Intelligence Group
Computer Science Faculty, University of the Basque Country (UPV/EHU)
(2018) San Sebastian, Spain
**andoni.beristain@innovaevision.com*
[t]*manuel.grana@ehu.es*

Face expression recognition is an active area of research with several fields of applications, ranging from emotion recognition for advanced human computer interaction to avatar animation for the movie industry. This paper presents a review of the state-of-the-art emotion recognition based on the visual analysis of facial expressions. We cover the main technical approaches and discuss the issues related to the gathering of data for the validation of the proposed systems.

Keywords: Emotion recognition; Facial expression.

1. Introduction

In our daily communication, the verbal communication channel delivers only a fraction of the entire information transmitted. Components like voice intonation, body language and facial expressions provide much additional information[1, 11]. In the ~~last~~ past years, a big research effort has been addressed to enhance the Human Computer Interface (HCI) in several ways. New input and output devices and new modeling and computational approaches try to overcome the actual limitations in the HCI bottleneck. The underlying reasoning is that we humans cannot take full advantage of all our communicating abilities when dealing with computers because this communication process takes place in an artificial setting. It is conditioned by the computer and has nothing to do with the natural human communication processes we are used and adapted to. It is desirable to reduce the semantic gap between computer language and human language. To that end, it is broadly recognized that interaction must be multimodal[2], because humans may use several information channels during their communication process. Affective Computing [3] heads towards a more natural and human-like HCI. Rosalind Picard coined the expression and she postulated the need to integrate the emotional component into HCI. Emotion recognition from facial expression offers several advantages like non-intrusivity, generality and the use of already conventional hardware (CCD cameras).

The research on facial expression recognition did not begin until the 1990s[4]. Until then, the available CCD cameras were very expensive, and of low resolution for real-time processing. Also, the computers available at that time were not able to cope with the computational demands of the proposed algorithms. The research on automatic facial expression analysis has application into areas such as Ambient Intelligence, advanced manufacturing environments, able to offer services and tools proactively, human-robot interaction and HCI, reaching the level of emotional interaction in the Affective Computing paradigm (e.g. using virtual and emotional characters). The Emotional Mirror is one of the first applications defined in the Affective Computing paradigm. It would assist people training their facial expressions (e.g. actors). Other applications of emotion recognition through face expression recognition are the treatment of patients with psycho-affective illnesses (e.g. autism), and non-invasive measurement of the emotional response of subjects, under some specific stimuli, in psychological studies. A big industry potential lies in the marketing studies based on the non-invasive measurement of test subjects emotional response to new products, services, advertisements or web pages. Complete, reliable and precise emotion recognition requires placing the facial expressions into a context and situation. Employing additional multimodal information like user activity and voice intonation, the automatic interpretation of the emotion experienced by a person would be more reliable [2, 5]. However, we will restrict ourselves to face expression recognition in this paper.

The use of specific sensing hardware may ease some of the tasks involved in face expression recognition. For example, using infrared lighting and an infrared camera, human eye pupils are easily located.[6] Thermographic cameras have also been used[7] for robust face localization, as well as stereo vision which offers distance information[8]. Recently, developed 3D cameras and 3D sensors produce robust 3D information that can be used for face modeling and recognition. However, they are either expensive or more intrusive than conventional CCD cameras. Therefore, most of the research found in the literature is based on conventional CCD color digital video cameras.

A wide variety of computational techniques have been applied to facial expression recognition. They include Artificial Vision algorithms like Canny edge detection, optical flow, histogram equalization or the Kalman Filter, and machine learning algorithms like neural networks, support vector machines or Hidden Markov Models. Automatic facial expression recognition is often used as a benchmark to test new machine learning or artificial vision algorithms. Therefore, there are many papers presenting tentative systems, or following unfeasible approaches in practice, since the main purpose of those papers is to show the advantages of their techniques.

The remainder of the paper is structured as follows: In Sec. 1, facial expressions are introduced and described. In Sec. 2, some background ideas for automatic facial expression recognition are reviewed. In Sec. 3, several representative systems are described in detail. We concluded in Sec. 4.

2. Facial expressions

We begin reviewing how facial expressions are produced, how they can be analyzed objectively and which are~~the~~ the main problems when working with emotions~~ are~~. Facial expressions are produced due to face muscle movements that end up in temporary wrinkles in the face skin and the temporary deformation or displacement of facial features like eyebrows, eyelids, nose and mouth. In most cases, facial expression persistence is short in time; usually no more than a few seconds.[9] We can divide the facial expression generation process into three steps. First, a transition between the previous facial expression and the actual one is produced. During this phase the strength of the expression increases steadily. Second, the facial expression is sustained for some seconds, and finally there is a transition to the next facial expression, which implies a decrease in the strength of the facial expression. Facial expression strength refers to the easiness of recognizing it. A strong facial expression is one that is easily recognized by anybody.

There is a great variability of facial expressions, so a specific taxonomy is needed to describe and analyze them. It must be taken into account that each facial expression associated to each emotion has its own strength range. For example, the changes produced in a subject's face when showing a sad facial expression are more subtle than those related to happiness or surprise. Recognition systems need to bound and describe the face region where the facial expression takes place, its strength and the movement itself. Above all, some facial expressions are culture or ethnic specific, so their meaning varies depending on the subject population[10]. Furthermore, expressiveness (facial expression strength range) is also subject dependent. Finally, there are also differences between a spontaneous facial expression and a simulated one, both in appearance and strength. That introduces a methodological and ethical problem in the design and construction of emotion recognition systems. The actions needed to be done in order to obtain some spontaneous facial expressions for some emotions can violate ethical constraints. For this reason, researchers must work with video recordings of professional actor, even if actors cannot always perform all the nuances of a spontaneous facial expression.

Paul Ekman laid the ground for the systematic psychological works on emotions [11-15]. According to him there are only 6 emotions which have the same facial expressions for all human beings, regardless of culture, ethnic group or geographic location[16]. Those elemental universal emotions are happiness, sadness, anger, surprise, disgust and fear. Most of the researches on emotion recognition based on the analysis of facial expressions are focused~~ d~~ on recognizing the elemental universal emotions.

There are two general approaches to define the mapping between facial expressions and emotions[17]. One approach tries to map facial expressions as a whole into specific emotions. Some authors call this approach *message jud~~e~~gment*. We think that *holistic approach* is more appropriate. The research in this direction looks for transformations of the face image that provide optimal features for the ensuing classification. The advantage of this approach is that it is computationally cheap, able for real-time implementations. However, the systems built following this approach are difficult to expand, adding new emotions, because classifiers and even transformations must be retrained. Systems are not very robust to deviations from the training sample, because asymmetric changes in the face or subtle changes in specific sections of the face cannot be isolated and dealt with.

The other approach decomposes and encodes each facial expression into elemental face configurations. Emotion recognition is performed on these codifications. Some authors call this approach sign *vehicle*~~ this approach~~, we prefer to call it *coding approach.* To develop this approach, it is necessary to develop a coding system able to describe every visually perceptible and anatomically feasible change in the face that a human being can produce. Among the different coding systems defined in the literature[18], the most widely used are the Facial Action Coding System (FACS), defined by P~~.~~aul Ekman and W~~.~~allace V. Friesen for measurement used in psychological experiments, and the Facial Action Parameters (FAPS), which is part of the MPEG-4 standard. Both coding systems will be described later. These coding systems are akin to a face description language. Using a language to describe facial changes is easier for experts in emotions to point out which facial changes compose each specific facial expression, and helps more accurate descriptions. Moreover, building the recognition system tailored to building blocks of face expression, instead of whole facial expressions, makes it easier to introduce new expressions and emotion classes, which could be formed by some facial changes already recognizable by the system.

2.1. Reliability of ground truth coding

It is essential to have reliable data for both classifier creation and testing purposes. In this particular case, a video or picture database is required. But those media must be correctly labe~~l~~led with the right emotion for each data item, or they are useless. This labe~~l~~ling is done manually. Usually more than one person is involved because of the large amount of data gathered. Since facial expression interpretation is subjective, obtaining an objective and systematic labe~~l~~ling is difficult. To reduce the subjectivity to its minimum, it is of prime importance to define a set of rules to guide the recognition of each facial expression. Those rules are defined by psychologists or by emotion experts (e.g. FACS, explained later).

In artificial intelligence applications, we need objective data to design the system and to evaluate its performance (validation). Examples of available facial expression databases (some of them free for research purposes) are the following:

- The Cohn-Kanade AU-Coded Facial Expression Database is the most widely used. At present it contains video recordings of the facial behavio~~u~~r of 210 adults who are from 18 to 50 years old; 69% female and 31% male; per eth~~i~~nic group, 81% are Caucasian, 13% African, and 6% from other eth~~i~~nical groups. All image sequences have been FACS coded by certified FACS coders either for the entire sequence or for specific target Action Unions (AUs). Approximately 15% of these sequences were coded by two independent certified FACS coders to validate the accuracy of the coding. The first portion of the database DFAT-504 has been prepared for computer vision research. This database is still active and growing.
- The PIE (Pose, Illumination and Expression) Database developed by the Human ID Group at the Carnegie Mellon University is formed by a set of 41,368 images of 68 people. Each person was recorded showing 13 different face poses, under 43 different illumination conditions, and with 4 different face emotion expressions. The project was closed in 2000, but the database is still accessible.
- The Vision and Autonomous Systems Center's Image Database gathers several picture databases sorted by specific areas. Some face and facial expression databases can be found there. The PIE database is also included in this database.
- The FERET Database is also an important source for testing. Now this database includes both colo~~u~~r and graeyscale images of facial expressions. The last update of the database was in 2004.

- The AR Face Database from the Computer Vision Center (CVC) at the Universidad Autonoma de Barcelona contains pictures from 126 different people, 70 men and 56 women, amounting up to 4,000 colo~~u~~r images, showing different facial expressions, illumination conditions and some occlusions. There are two sessions per person, which took place in two different days, two weeks apart (14 days).

3. Automatic Facial Expression Analysis

Most of the existing literature about face expression recognition systems decomposes the image processing of facial expressions and the ensuing emotional state 27 estimation into the following steps:

- Face localization on the image.
- Facial feature vector extraction and representation.
- Facial expression recognition.
- Facial expression interpretation or emotion recognition.

These steps are illustrated in Fig. 1. Not all the systems follow them up to a point, [19, 20] but they are the building blocks for the recognition system.

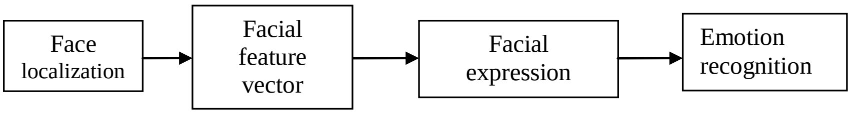

Fig. 1. Processing pipeline of facial expression recognition.

3.1. *System performance goals*

The design and construction of face expression recognition system have the following performance goals[1] :

(i) Automatic facial image acquisition. The system must be able to autonomously start the recognition process, deciding if a face is to be present in the video stream.

(ii) Subjects of any age, ethnicity and appearance. Facial expression recognition must be equally performing in diverse populations. Age may introduce visual noise in the form of wrinkles, or face shape deformations. Diverse ethnic groups have diverse face skin color and reflectivity, so that face shadows mayor may not appear and other face characteristics may or may not be highlighted.

(iii) Robust to variation in lightning. Light changes induce very strong changes in the face appearance. This is a classical problem in all the systems using Computer Vision techniques.

(iv) Robust to partially occluded faces. Occlusion of face regions may impede the recognition of emotions in diverse ways. Occluding the mouth, for instance, increases the difficulty to discriminate happiness from other emotions. Therefore, occlusion robustness implies modeling the relationship between face regions and emotion expression.

(v) No special markers/make-up required. Motion detection is a difficult problem in general. A way to solve it is the superposition of special markers on the face that can be easily segmented and tracked in real time. However, this setting is only valid for laboratory conditions, not for real-life conditions.

(vi) Dealing with rigid head motions, corresponding to natural motions of the human subject, in real time.

(vii) Automatic face detection. Face detection is a precondition for face processing. Although there have been a lot of work into this problem, it has not been solved in its more general setting.

(viii) Automatic and robust facial expression feature extraction. It is undesirable to have specific ~~hand picked~~handpicked features to characterize specific emotions or face expressions. The robustness of the classification process will be highly dependent on the feature extraction process.

(ix) Dealing with inaccurate facial expression data due to noise or the inaccurate labe~~l~~ling of the training database.

(x) Automatic facial expression classification. As the feature extraction module, the classification module must be independent of the specific emotions or expressions to be recognized. Its construction must follow some well-founded induction procedure from data.

(xi) Distinguish all possible expressions. The set of recognized expressions must not be a limitation or a precondition for the system performance. It is not desirable to have systems that perform well for a set of expressions and badly for another.

(xii) Deal with unilateral facial changes.

(xiii) Obeying anatomical rules, the system may be founded on solid modeling grounds.

(xiv) Adequate response time, which means a real-time behavio~~u~~r for most of the applications.

(xv) The systems may be incorporated into more complex systems, creating multimodal emotion recognition systems.[2,5,21]

3.2. *Face localization*

The face localization task involves detecting every face in the scene, obtaining its position and delimiting its area in some ways. The proposed methods should be independent of the position, size, rotation angle, partial occlusions~~,~~ and illumination of the face. In some cases, the result of the process is the bounding rectangle of the face,[22] while in others, a silhouette or a blob representing the face is obtained. [6,23] There are also 3D approaches which obtain not only the tri-dimensional position of the face, but also its 3D orientation in the scene.[24,25]

Some face localization algorithms take into account colo~~u~~r information.[26,27] Others are focused on locating eyes and other face specific characteristics. [22,28] There are also methods that use motion information. A very popular system uses Haar wavelets and Adaboost, which offer good performance at a reasonable computational cost.

Anyway, face detection is itself a research area, with partial solutions, although due to the difficulty of the task, the problem has not been solved yet at a feasible computational cost. There are still difficulties due to different lighting conditions, occlusions, picture size and robustness. For a complete review on face detection, refer Refs. 29 and 30. The web sites[31-33] offer recent information about face localization and tracking.

3.3. *Facial feature vector extraction and representation*

Facial feature extraction is usually the most difficult step in the facial expression recognition process, and the most computationally demanding in some approaches. It consists of obtaining the most significant information of a face that will allow recognizing the facial expression shown. The procedures found in the literature can roughly be categorized into feature-based, image-based and model-based methods. In the feature-based methods, the shapes and locations of some facial features (e.g. eyebrows, eyes, nose and mouth) are extracted to form the expression feature vectors. Usually the feature vector is a set of 2D or 3D points describing each detected facial feature. In the image-based methods, holistic or local spatial analyses are applied to recognize the facial expression. In the model-based methods, a statistical model is constructed from training images and used to recognize the facial expressions. From a practical point of view, we can distinguish still image-based methods from video sequence-based methods. The former are usually performed

Table 1. Common computational techniques for feature extraction

	Holistic	Local
Still image	-PCA [34, 35] -Edges -Colour [36] -Gabor wavelet [37-40]	-Active Contours [41-43] -Blobs [44] -Colour [42, 45] -Edges [42] -Gabor wavelet [46] -Local PCA [46] -Template [38, 47, 48]
Video based	-PCA [49] -2D Discrete Cosine Transform (DCT)[50] -Optical Flow [28, 51-53] -Image difference [54]	-Local PCA [55] -Local Optical Flow -Active Contours [45]

off-~~-~~line, while the latter aim to produce real-life real-time applications. This is the taxonomy that we will follow in the next sections.

3.3.1. *Still image based methods*

Still image systems work on single images. If applied to a video sequence, each frame is treated independently. In general, they are less computationally demanding than video sequence systems, which process several images to obtain motion information. Examples of still image systems using a feature-based approach are found in Refs. 34, 35, 38 and 56. There, in most cases, the eyebrows and the mouth are tracked. We have detected two different approaches in the literature. One approach tries to recognize a facial expression taking into account the absolute position of the facial features. This approach is neither very robust nor precise because different people have different face proportions. The other approach tries to obtain the position of facial features relative to a reference expressionless face image. They take a neutral expression picture of a person and obtain the position of the facial features. Each new picture is compared with the reference one by checking the difference in the position of the facial features in both pictures.

In most cases, facial appearance changes due to facial expressions (furrows) produce strong gradients in the spatial domain. Therefore, edge detection algorithms have been broadly used in the literature, and image-based methods to try to recognize them.[9] Colo~~u~~r information can also be useful to recognize certain facial features, as well as shape information. However, edge detection algorithms show poor performance when trying to segment some facial features. The mouth, and more precisely the lower lip, and the chin do not contain easily detectable edges. These are the most common techniques used by the still image systems applying an image-based method:

- Colo~~u~~r information analysis: Suitable to recognize most of the facial features but usual1y highly illumination sensitive. Nonetheless, it is one of the best choices for the segmentation of the mouth. Usually employed in conjunction with machine learning algorithms. [42,45,47].
- Edges: Edge information is also suitable for most facial features. It is difficult to distinguish between facial feature edges and furrows. Usually employed with machine learning

algorithms.[56] Snakes offer good results. [41,58] Snakes are deformable curves, which are fitted to specific features, like borders with strong gradient (e.g. eyelids). Snake fitting is performed minimizing an energy function that models the attraction of the snake to the strong gradient regions of the image.

- 2D and 3D face models: Facilitate facial feature detection adding constraints and help to avoid non-feasible facial feature localizations. On the other side, they are difficult to develop and, in most ~~of~~ cases, they are too rigid. The Candide 3D face mode1[59] and the Active Appearance Models (AAM)[8, 60, 61] are examples of these models.

3.3.2. *Video sequence based methods*

Video-based methods use motion information to recognize facial expressions. [53] It is necessary to distinguish between facial feature motion and head motion, which can be considered as a noise source. Therefore, a filtering process is required to isolate facial feature motion from the whole face motion. Many computer vision techniques have been used to recognize the motion produced by facial expressions. In image subtraction, the simplest way of motion detection offers really poor results. Consequently, most works in the literature apply more sophisticated methods of motion detection and modeling like computing the optical flow [51,53] or tracking algorithms like the Kalman Filter, particle filtering[57] and others. [62]

The computation of the optical flow provides the motion intensity and the motion direction for each image point. This can be useful, for example, to isolate head motion from facial feature motion, or to filter unfeasible motions, taking into account only anatomically possible motions. Unfortunately, optical flow requires a lot of computing power. Therefore, it is only used locally, around the facial features which are more significant for facial expression analysis.

The approach based on motion templates proposed[63] does not provide promising results. The ~~application of analytical motion models offer~~application of analytical motion models offers better results. [53, 64] ~~These~~ This kind~~s~~ of model~~s~~ ~~try~~ tries to recognize a temporal sequence of facial expression deformations, from a neutral facial expression to the peak intensity of the new facial expression and back to the neutral state. These kind of method~~s~~ deals with the motion generated around facial features during a facial expression.

Tracking algorithms focus on a set of facial feature points, which concentrate on information relating to the most important facial features for facial expression recognition. Kalman filter,[55,65] and the Condensation algorithm have been used[66].

Video-based algorithms usually have high computational requirements; thus, simplifications are imposed in many cases. Local approaches try to reduce to the minimum the regions of interest by means of still image-based methods. Holistic methods try to use the smallest image size and minimum number of colo~~u~~r channels that maintain the recognition performance sought. Markers have been applied to face expression analysis and characterization based on motion detection, but they are very uncomfortable and invasive, suited only for laboratory conditions.

The main critics to the mentioned computational methods are the following:

- Optical flow: This technique is computationally intensive, but it offers the greatest quantity of information: the direction and intensity of motion at each face pixel in the image. It requires filtering between rigid facial movements and facial feature motion. It can be computed either locally or holistically. For computational power limitations, it is usually done locally.
- Kalman filter: Use as point-tracking algorithm reported good results, but it has problems to deal with rapid motion, and requires that the tracked points are easily distinguishable from the surrounding points. It requires a correction procedure to be performed periodically to recover from the accumulated tracking error.

- Motion templates: Require~~s~~ less computer power than optical flow and consider specific motions, like opening of the mouth as a whole, in contrast to optical flow. Unfortunately, those templates are usually person specific and quite rigid.

3.3.3. *Local versus holistic approaches*

Holistic and local are two basic approaches to facial expression analysis. Holistic approaches process the face as a whole, while the local approaches focus on each element of the facial feature set independently. The former facilitates the recognition of the dominant facial expression while the local ones are able to recognize subtle changes in small areas of the face. Some authors[67] conclude that there is no significant improvement of the holistic approach over the local approach. In both cases, two kinds of facial features can be distinguished:

- Stationary facial features: Those facial features which are always visible in the face, but can show deformations due to facial expressions. Among them, eyelids, eyebrows, eyes and mouth are closely related to facial expressions. Skin furrows and the facial skin texture are also considered stationary facial features.
- Transient facial features: The wrinkles and lumps that appear in the face when showing a facial expression but disappear as soon as the face returns to its neutral facial expression. Most of them are shown in the forehead and around the eyes and mouth.

3.3.4. *Image versus model-based approaches*

Image-based methods do not need much *a priori* knowledge about the object of interest. Even when these kinds of methods are usually fast and simple, they are neither very reliable nor robust, especially when dealing with more than one different views of the same object because of the matching difficulties between the different views of the same object. They are unable to cope with motion and rotations taking place out of the image plane.

It seems that a more appropriate approach would consist ~~on~~ in the use of 2D and 3D face models[60,68-71] to cope with general rotations and motions. There are two important factors to consider defining a 3D model: ~~The~~ the complexity of the model, given by the vertex number and the flexibility of each vertex. More vertex points create better models but having taken into account that 3D models are usually hand crafted and that more vertices imply more computational cost, choosing the best vertex number requires a compromise between precision and efficiency. An excessive flexibility for each vertex increases the computational cost and lets bigger deformations and too low flexibility does~~n't~~ not permit to achieve an ideal adaptation of the model to the face. Dealing with facial movements at a level higher than the vertex level, i.e. defining muscles in the model makes the facial expression recognition task easier since it allows to define restrictions on anatomically possible facial motions.

The Active Appearance Models (AAM)[72] and Active Shape Models (ASM)[73] have been used by many researchers. These are well-known 2D and 3D parametric models, respectively, and can represent both the shape and the appearance of non-rigid objects such as a face. There are extensions to the AAM is like the stereo Active Appearance Model (STAAM),[8] which uses a geometric relationship between two tightly coupled views to speed up the model fitting process.

3.4. *Classification approaches to facial expression recognition*

The facial expression recognition process is a classification process. Usually classifiers are built to accept as input a feature vector computed from the raw input data (the images or the video sequence). Modern approaches are statistical pattern recognition-based systems endowed with some training (learning) algorithm that allows ~~to construct~~constructing the classifier by induction from the available data. Main categories of classifiers are dynamic and static ones. The first works on spatiotemporal features while the second works on static features. Obviously, they are related to static and dynamic feature extraction methods.

3.4.1. *Spatiotemporal classification features*

The main approach to spatiotemporal classifiers is the use of Hidden Markov Models (HMM), since they permit to model the facial expression dynamics along time. There are many examples of HMM applied to facial expression recognition in the literature. [44,52,74,75] Most times, they have been used with motion analysis methods for feature extraction.

Recurrent Neural Networks have also been used as an alternative to HMM, for facial expression classification.[76,77] Another way to consider the temporal evolution of facial expressions consists of using motion-energy templates. In this approach, the Euclidean distance can be used to determine the prominent facial expression. [78]

3.4.2. *Static spatial classification features*

The use of Neural Networks is broadly documented in the literature, using the original pictures as input[19,72, 79] or using dimensionally reduced data obtained by techniques like Principal Component Analysis (PCA),[8o] Independent Component Analysis (ICA) and Gabor filters. [39,81,82] Although Neural Networks have been used both in holistic and local approaches, most of the authors point out that using local approaches gives better results.

Neural Networks have well-known drawbacks. They are difficult to tune as the number of parameters grows. They require a long training period too, especially when using a coding approach, like FACS or FAPS, because it implies developing a Neural Network able to recognize all the Action Unit combinations defined in FACS, or all the Action Parameters combinations defined in FAPS. That implies a combinatorial explosion of the number of parameters as the number of classes grows while the straight classification of the image into emotions (holistic approach) needs only to perform the classification into the 6 universal emotion categories.

Support Vector Machines (SVM) are being used widely, and they achieve good results. [51,59] According to Ref. 83, the combination of AdaBoost with SVM improves the results obtained with other classifiers. And more recently, a variation of the SVM called Relevance Vector Machines (RVM) has shown promising results, since it produces similar results to SVM but with less computational and memory requirements. [84]

3.5. *Facial expression coding*

In the beginnings of facial expression recognition researches, each research group developed the rules and conditions to recognize each of the facial expressions separately. That added another task to the research process, and the works from different research groups were made incomparable. Since the creation of facial expression coding systems, by psychology and standardisation associations, and thanks to its ~~acceptation~~ acceptance by most of the research community, there are some *de facto* standards to decompose, analyse and recogniz~~s~~e facial expressions. The most frequently used coding systems are FACS (Facial Action Coding System) and FAPS (Facial Animation Parameters).

Facial Action Coding System (FACS)[85,86] is a facial expression coding system based on anatomical information about the facial muscles. FACS originally defined by ~~P.~~ Ekman and ~~W.~~ Friesen as a tool to study and measure objectively in psychological experiments. FACS decomposes a facial expression into a set of Action Units (AU), each of them representing a unique and specific movement of a muscle or a muscle group together. FACS defines more than 40 of these Action Units. There is a FACS coder certificate which proves that a person is able to manually code facial expressions objectively according to FACS. Those official FACS coders

Con formato: Fuente: Cursiva

are employed to perform manual and offline FACS coding of experiment sessions video recordings. The procedure is obviously time consuming.

The Facial Animation Parameters (FAPS)[87,88] is a coding system focused ~~in~~on describing the different kinds of movements that each facial feature can have. It is part of the MPEG-4 standard and its main purpose is to describe facial animations for animated characters. FAPS splits a facial expression into a set of facial feature movements, or Facial Animation Parameter (FAP). FAPS defines more than 60 different actions. A relevant difference with FACS is that FAPS defines the normalized movement of the facial features in a defined direction. All FAPs involving translational movement are expressed in terms of FAP Units (FAPUs). FAPUs correspond to fractions of distances between key facial features. For instance, Ref. 89 or more recently Refs. 43 and 90 use this coding system.

4. Representative Facial Expression Recognition Systems

In this section, we review some facial expression and emotion recognition systems that we consider representative of categories discussed previously in this document. At present, we do not know about any system able to achieve the ideal and desirable characteristic mentioned in the beginning of this article for a facial expression recognition system.

4.1. *Still image based systems*

The system that we select as representative of the still image-based systems is in Ref. 91. It reports a 78% ~~of~~ recognition rate. An original feature of the paper, in contrast to most of the literature, is that it does not use the Ekman's 6 elemental emotions. Instead, it defines an emotion's wheel, an activation-emotion classification space. The activation parameter is added to the 6 basic emotion set. Face localization is performed using non-parametric discriminant analysis with a Support Vector Machine (SVM), obtaining the bounding rectangle of the face. Then, the rectangle is segmented using static anthropometric rules into three overlapping rectangle regions of interest which include both facial features and facial background; these three feature-candidate areas include the left eye/eyebrow, the right eye/eyebrow and the mouth. Next, the feature extraction procedure is initiated on each region. First of all, eyes are localized using a feed-forward back propagation neural network with a sigmoid activation function. The localization of the detected eyes is used to restore the roll rotation of the face to an upright position, so the extraction of the remaining features can be performed. In many systems like the one being described, feature extraction is robust to roll rotation, at least when it is less than 30°. Pitch rotation is not considered, although it occurs often as well, and its negative effects for feature extraction are worse. Posterior eye position detection refinements include a modified Canny algorithm and a region-growing technique. Eyebrows are detected using morphological edge detection. Nose detection is based on nostril localization. Nostrils are easy to detect because they correspond to very low local minima of image intensity. Mouth detection is performed using three different procedures: a neural network, similar to the one used for eye detection, horizontal morphological gradient and thresholding. This paper uses a Coding approach, and employs the MPEG-4 FAPS coding system. After obtaining the representative points for each facial feature, facial muscle movements are obtained and translated to FAPs. A neurofuzzy network is trained and employed for the classification among the quadrant of emotion's wheel.

4.2. *Video based systems*

We have taken the article in Ref. 52 as an example of a video-based system. Its goal is to recognize Ekman's 6 basic emotions. They report 90.9% ~~of~~ average recognition rate on the

Cohn-Kanade facial expressions database. Face localization is performed by means of a neural network based approach, combined with a tracking algorithm. Then, the face region image is pre-processed to normalize its size and reduce environmental dependence. After that, optical flow is computed between consecutive frames of the sequence. The dimensionality of the optical flow is reduced so that only the most important information is kept using PCA. Next, the PCA reduced motion patterns are fed to a bank of linear classifiers to assign class labels from the set of universal expressions to each image of the sequence. The output of the linear classifiers over sequence of images is coalesced together to form a temporal signature. Then, the temporal signature generated is used to learn the underlying model of six universal facial expressions. Discrete HMMs are used in learning the models for facial expressions. Finally, recognized facial expression is mapped to compute levels of interest based on 3D affect spaces. The authors follow a holistic approach. Therefore, it has difficulties in distinguishing the little nuances of certain facial features and also with asymmetric facial changes and unilateral changes. In spite of that, the refinements introduced in the procedure produce a robust and precise system. Adding more emotions would require redoing all the training procedure.

4.3. *Multimodal frameworks*

The representative of multimodal recognition systems[92] reports an average combined recognition rate of 90%. It uses voice and facial appearance as input information. In this paper, the objective is not only to recognize Ekman's 6 universal emotions but also to recognize some cognitive/motivational states. Those are interest, boredom, confusion and frustration. The neutral state has been considered too counting eleven different affective states. This is an interesting improvement since it increases the possible applications of the developed system.

The audio features are three kinds of prosody features: The logarithm of energy, the syllable rate, and two pitch candidates together with their corresponding scores. Pitch extraction is performed using an autocorrelation-based pitch detector to extract two candidates of pitch frequency.

The face tracker uses a model-based approach. A 3D wireframe model formed by 16 Bezier volumes is constructed. It is adapted to the face manually by means of interactive selection of some landmark facial features such as the eye and mouth corners. After the manual fitting has been performed, the model is able to track the head motion and the local deformations of facial features like the eyebrows, eyelids and mouth. We think that the need of a completely manual calibration to start the face tracking is an important disadvantage. Moreover, the system ~~doesn't~~ does not really perform any face detection procedure. The tracking procedure employs 2D motion information obtained from the image to estimate the 3D motion of the facial features by solving an overdetermined system of equations, obtained from the projective motion models in the least squared sense. The recovered motions are represented in terms of magnitudes of some predefined motions of various facial features. Each feature motion corresponds to a simple deformation on the face, defined in terms of the Bezier volume control parameters. The obtained values represent the activation of a facial region, a direction and the intensity of the motion. Motion information is usually filtered to remove noise, so in the presence of a slow or fast enough facial expression change, the tracking algorithm would fail and the system would not be able to recover.

The features extracted from audio and visual inputs are used to feed the next classifying stage. The most innovative characteristic of this system is the way the information of both sources is combined. A Bayesian Network is used to merge the information from both sources previous to any classification. Other works make the fusion after each channel's information had been classified into an emotion category. The top node of the decision graph is the class variable (recognized emotional expression). It is affected by the recognized facial expressions, the recognized vocal expressions, and by the context in which the system operates (if that is available). Vocal emotions are recognized from audio features extracted from the person's audio track. Facial expressions are recognized by facial features tracked using video but the recognition is also affected by a variable that indicates whether the person is speaking or not. Recognizing whether a person is speaking uses both visual cues (mouth motion) and audio features. The parameters of the proposed network are learned from data. This approach is quite robust to noise and incomplete information.

5. Conclusion

From our experience and recent researches' trend, we consider that the most suitable approach to deal with emotion recognition based on facial expression analysis should have the following characteristics.

- It should use a video-based approach because facial expressions always involve motion and their subtle changes cannot be recognized otherwise. This choice reduces the application areas but the performance improvement obtained is worth it.
- It should use a 3D model of the face and a feature-based approach, where each facial feature is determined by a set of vertices in the 3D face model. The eyebrows, the eyes, the nose, the mouth and the chin should at least be tracked. The head silhouette or at least the face bounding itself should be mode~~l~~led and the orientation and position of the face should be tracked. This information can be used to improve the facial feature tracking procedure and to be used as an additional information source to recognize emotions.
- Regardless of the facial feature tracking procedure followed, feature-based tracking cannot recognize every subtle facial expression change. Therefore, optical flow or any other motion recognition procedure should be used locally in the surroundings of each facial feature in order to recognize them.
- Taking into account that facial expressions produce motion and they are finite and usually short in time, spatiotemporal classifiers are better fitted to them like the Hidden Markov Models.
- Finally, a coding approach ~~based approach~~ has many advantages, such as being easier to expand, adding more emotions or mental states, or easier to improve, since they can be done locally for each recognizable facial action (AU or FAP, depending on if we use FACS or FAPS).

Nowadays, systems meet some performance goals. Facial image acquisition is completely automatic. When dealing with every kind of people, most systems have recognition difficulties with particular ethnic groups (colo~~u~~r based techniques), babies (less texture), elder people (furrows). However, there are some systems able to cope with almost every kind of subject. Illumination is still a problem, but most of the systems are able to perform well if there is enough constant ambient light no matter its origin (fluorescent, daylight, incandescent, ~~...),~~ etc.), Occlusions have been successfully addressed by many researches, achieving good recognition results even under the presence of occlusions unless they occlude most critical regions like the whole mouth or both eyes. Special markers or make-up are hardly employed now, except in very specific researches. Rigid head motion is taken into account most of the times especially in systems based on video. Actual face localization algorithms offer good results, although the precision and reliability of the recognized face can be improved. Facial expression data extraction must be automatic and it is one of the requirements for every developed system. Since the information extracted from the face usually has some kind of errors, it is compulsory to deal with inaccurate facial expression data. Facial expression classification is completely automatic in all the reviewed systems, but they are only able to perform the claBsifica,tion int.o it ff~rlllr.erl set. of classes in most of the systems. Therefore, they are not able to recognize every anatomically feasible facial expression.

In order to deal with occlusion problems, systems are usually able to recognize unilateral facial changes. And finally, most of the researches following model based approaches are able to recognize and filter only anatomically possible facial expressions.

6. Acknowledgements

This work is supported by the Departamento de Industria of the Gobierno Vasco through the Ikertu grant appendix IV for the achievement of the PhD. studies at a R&D company which in this case is Innovae Vision.

7. References

1. M. Pantic, L.J.M. Rothkrantz., Automatic Analysis of Facial Expressions: The State of the Art. *J. IEEE Transactions on Pattern Analysis and Machine Intelligence*. **22**(12) (2000): pp. 1424-1445.
2. A. Jaimes, N. Sebe, Multimodal Human Computer Interaction: A Survey, *Computer Vision and Image Understanding,* **108**(1-2) (Elsevier Science Inc., New York, 2007) pp.116-134.
3. Picard, R.W., *Affective Computing*. (MIT Press., Cambridge, USA, 1997).
4. F. Dornaika, F. Davoine, Facial Expression Recognition using Auto-regressive Models, in *Proc. ICPR* **2**(0) (Hong Kong, China, 2006) pp.520-523.
5. L.S. Chen, T.S.Huang, T. Miyasato, R. Nakatsu Multimodal human emotion/expression recognition, in *Proc. of The International Conference on Automatic Face and Gesture Recognition*. (Nara, JAPAN, 1998) pp. 366-371.
6. X. Wei, Z. Zhu, L. Yin, Q. Ji. A Real Time Face Tracking and Animation System, in *Proc. 2004 Conference on Computer Vision and Pattern Recognition Workshop*. **27** (IEEE Computer Society, Laboratoire MAS MAS − Grande Voie des Vignes − 92290 Chatenay Malabry Cedex, 2004), pp.71.
7. L. Trujillo, et al. Automatic Feature Localization in Thermal Images for Facial Expression Recognition. in *Proc. 2005 Computer Vision and Pattern Recognition*. (IEEE Computer Society Washington, DC, USA ,2005), pp. 14.
8. S. Jaewon, L. Sangjae, K. Daijin. A Real-Time Facial Expression Recognition using the STAAM. in *18th International Conference on Pattern Recognition*. vol. 1 **(**IEEE Computer Society Washington, DC, USA, 2006), pp. 275-278.
9. B. Fasel, J. Luettin, Automatic Facial Expression Analysis: A Survey. *J. Pattern Recognition,* **36**(1) (2003) pp. 259-275.
10. P. Ekman, *Emotions Revealed* (Phoenix Press; New Ed edition, New York, 2004).
11. P. Ekman, and R. Davidson, *The Nature of Emotion: Fundamental Questions* (Oxford University Press, New York, 1994).
12. P. Ekman, Facial Expressions of Emotion: An Old Controversy and New Findings, J. *Philosophical Transactions: Biological Sciences*, **35** (1992) pp. 63-69.
13. P. Ekman, Facial expression and emotion. *J. American Psychologist*, **48**(4) (1993) pp. 384–392.
14. P. Ekman, Asymmetry in Facial Expression. *Science*, **209**(4458) (1980) pp. 833-836.
15. P. Ekman , Emotions Inside Out: 130 Years after Darwin's The Expression of the Emotions in Man and Animals. *Annals of the New York Academy of Sciences*, **1000** (2003) pp. 266-278.
16. P. Ekman, W.F., Constants Across Cultures in the Face and Emotion. *Journal of Personality*, **17**(2) (1971), pp. 124-129.
17. P. Ekman, *Emotions in the Human Face.* (Cambridge University Press, Cambridge, U.K, 1982).
18. K. R. Scherer, P. Ekman., *Handbook of Methods in Non-Verbal Behavior Research*, (Cambridge University Press, Cambridge, UK, 1982).
19. C. L. Lisetti, D.E.R. *Facial Expression Recognition using a Neural Network*, in *Proc. of the 11th International Flairs Conference*. (American Association for Artificial Intelligence (AAI) Press, California, USA, 1998) pp. 328 - 332.
20. W. A. Fellenz, J.G.T., N. Tsapatsoulis, S. Kollias. *Comparing Template-based, Feature-based and Supervised Classification of Facial Expressions from Static Images*. in *Proceedings of Circuits, Systems, Communications and Computers*. (IEEE Computer Society, Washington, USA, 1999) pp. 5331-5336.

21. Z.D., C. Busso, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, S. Narayanan. Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information. in *Proc. of the 6th international conference on Multimodal interfaces* (ACM Press, New York, USA, 2004) pp. 205-211.
22. Paul Viola, M. J. Jones., *Robust Real-Time Face Detection. International Journal of Computer Vision*, **57**(2) (2004) pp. 137-154.
23. G. R. M. J. Seow, D. Valaparla and K. V. Asari, A robust face recognition system for real time surveillance, in *Proc. International Conference on Information Technology: Coding and Computing*, vol. 1 (2004) pp. 631-635.
24. Q. J. Zhiwei Zhu, Real time 3D face pose tracking from an uncalibrated camera, in *Proc. of the 2004 Conference on Computer Vision and Pattern Recognition Workshop*, 27 (IEEE Computer Society, Washington, USA, 2004) pp.73.
25. Dornaika, F. and F. Davoine *On Appearance Based Face and Facial Action Tracking.* IEEE Transactions on Circuits and Systems for Video Technology 2006. 16(9): p. 1107- 1124.
26. T.-W. Yoo, I.-S. Oh, A fast algorithm for tracking human faces base chromatic histograms, *Pattern Recognition, Letters* **20**(10) (1999), pp. 967-978.
27. H. Stern and B. Efros, Adaptive color space switching for face tracking in multi-colored lighting environments, in *Proc. of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 21 (IEEE Computer Society Washington, DC, USA, 2002) pp. 249 - 254.
28. S. M. Guo et al., A key frame selection-based facial expression recognition system, in *First International Conference on Innovative Computing, Information and Control* (IEEE Computer Society Washington, DC, USA ,2006) pp. 341-344.
29. Erik, H. and L.B. Kee, *Face Detection: A Survey.* Computer Vision and Image Understanding, 2001. 3(3): p. 236-274.
30. Yang, M.-H., *Detecting Faces in Images: A Survey.* IEEE Transactions on Pattern Analysis and Machine Intelligence 2002. 24(1): p. 34-58.
31. Gorodnichy, D. *(Vision-based) Face Tracking*. (2005) [cited 2006 2006/12/04] http://synapse.vit.iit.nrc.ca/doc/facetracking.html.
32. Frischholz, D.R. *The Face Detection Homepage*. (1999) [cited 2007 01/08/2007] http://www.facedetection.com/facedetection/techniques.htm.
33. N. Ahuja, *Resources for Face Detection*. (2006) 2004/08/22 [cited 2006 2006/12/04] http://vision.ai.uiuc.edu/mhyang/face-detection-survey.html.
34. S. Dubussion, F. Devoine and M. Masson, A solution for facial expression representation and recognition, *Signal Processing: Image Communication* **17**(9) (2002) pp. 657-673.
35. A. J. Calder, A. M. Burton and P. Miller, A principal component analysis of facial expressions, *Vision Research* **41**(9) (2001) pp. 1179-1208.
36. P. Kakumanu and N. Bourbakis A local-global graph approach for facial expression recognition, in *Proc. 18th IEEE International Conference on Tools with Artificial Intelligence*, (IEEE Computer Society Washington, DC, USA, 2006) pp. 685-692.
37. Marian Stewart Bartlett, G.L., Ian Fasel, Javier R. Movellan, *Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction*, in *IEEE International Conference on Computer Vision and Pattern Recognition*. 2003.
37. G. L. Marian Stewart Bartlett, I. Fasel and J. R. Movellan, Real time face detection and facial expression recognition: Development and applications to human computer interaction, in *CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction* **5**(5) (2003) pp.53.
38. Y.-Z. Zhan et al., Facial expression recognition based on Gabor wavelet transformation and elastic templates matching, in *Proc. Third International Conference on Image and Graphics* 18 (IEEE Computer Society Washington, DC, USA, 2004) pp. 254-257.
39. M.-P. Loh, Y.-P. Wong and C.-O. Wong, Facial expression recognition for e-learning systems using Gabor wavelet & neural network, in *Proc. Sixth IEEE International Conference on Advanced Learning Technologies* (IEEE Computer Society Washington, DC, USA, 2006) pp. 523 - 525.
40. L. WeiPeng and W. ZengFu, Facial expression recognition based on fusion of multiple Gabor features, in *Proc. of the 18th International Conference on Pattern Recognition* (IEEE Computer Society Washington, DC, USA 2006) pp. 536-539.

41. B. A. Shafik Huq, A. Goshtasby and M. Abidi, Stereo matching with energy minimizing snake grid for 3D face modeling, in *Biometric Technology for Human Identification* 5404 (Orlando, FL, USA 2004), pp. 339-350.
42. T. Wakasugi, M. Nishiura and K. Fukui, Robust lip contour extraction using separability of multi-dimensional distributions, in *Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition* 17 (IEEE Computer Society Washington, DC, USA,2004), pp. 415 - 420.
43. Aleksic, P.S. and A.K. Katsaggelos, *Automatic facial expression recognition using facial animation parameters and multistream HMMs.* IEEE Transactions on Information Forensics and Security 2006. 1(1): p. 3-11.
44. N. Oliver, A.P., F. Berard, *LAFTER: A Real-Time Lips and Face Tracker with Facial Expression Recognition.* Pattern Recognition, 2000. 33: p. 1369-1382.
45. M. K. Moghaddam and R. Safabakhsh, TASOM-based lip tracking using the color and geometry of the face, in *Proc. Fourth International Conference on Machine Learning and Applications* (IEEE Computer Society Washington, DC, USA , 2005), pp. 63-68.
46. G. Neeharika and A. Yijayan, Gabor wavelet based modular PCA approach for expression and illumination invariant face recognition, in *Proc. of the 35th Applied Imagery and Pattern Recognition Workshop* (IEEE Computer Society Washington, DC, USA , 2006), pp. 13.
47. Denis Leimberg, M.V.-C., *Eye Tracking*, in *LYNGBY*. 2005, Technical University of Denmark.
48. B. U. Eamonn Boyle, D. Molloy and N. Murphy, Using facial features extraction to enhance the creation of 3D human models, in *6th International Workshop on Image Analysis for Multimedia Interactive Services* (Montreux, Switzerland, 2005).
49. T. K. James Jenn-Jier Lien, J. F. Cohn and C.-C. Li, Subtly different facial expression recognition and expression intensity estimation, in *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society Washington, DC, USA, 1998), pp. 853.
50. F. Wallhoff et al., Efficient recognition of authentic dynamic facial expressions on the feedtum database, in *IEEE International Conference on Multimedia and Expo* (Toronto, Canada, 2006), pp. 493-496.
51. Anderson, K. and P.W. McOwan, *A real-time automated system for the recognition of human facial expressions.* IEEE Transactions on Systems, Man and Cybernetics, Part B 2006. 36(1): p. 96-105.
52. Yeasin, M., B. Bullot, and R. Sharma, *Recognition of facial expressions and measurement of levels of interest from video.* Multimedia, IEEE Transactions on, 2006. 8(3): p. 500-508.
53. Y. Yacoob, L. S. Davis, Recognizing human facial expression from long image sequences using optical flow, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(6) (1996), pp. 636-642.
54. N. Tetsutani, S. Kawato, Detection and tracking of eyes for gaze-camera control, in *Proc. Image and Vision Computing*, **22**(12) (Elsevier Science Inc., New York, 2004), pp. 1031-1038.
55. D. Datcu, L. J. M. Rothkrantz, Automatic recognition of facial expressions using Bayesian belief networks, in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, 3 (IEEE Computer Society Washington, DC, USA, 2004), pp. 2209-2214.
56. Matsugu M., et al., *Subject independent facial expression recognition with robust face detection using a convolutional neural network.* Neural Networks, 2003. 16(5-6): p. 555-559.
57. I. Patras and M. Pantic, Particle filtering with factorized likelihoods for tracking facial features, in, *Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition* (IEEE Computer Society Washington, DC, USA, 2004), pp. 97 - 102.
58. D. Terzopoulos, K. Waters, Analysis of facial images using physical and anatomical models, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**(6) (1993), pp. 569 - 579.
59. Kotsia, I. and I. Pitas, *Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines.* Image Processing, IEEE Transactions on, 2007. 16(1): p. 172-187.
60. R. Gross, I. Matthews and S. Baker, Generic vs. person specific active appearance models, *Image and Vision Computing* **23**(11) (2005), pp. 1080-1093.
61. R. Gross, I. Matthews and S. Baker, Constructing and fitting active appearance models with occlusion, in *Conference on Computer Vision and Pattern Recognition Workshop*, (Washington, D.C., USA, 2004), pp. 72.
62. F. Dornaika, F.D., *On Appearance Based Face and Facial Action Tracking.* Circuits and Systems for Video Technology, IEEE Transactions on, 2006. 16(9): p. 1107- 1124.

63. Marco La Cascia, L.V., Stan Sclaroff. *Fully automatic, real-time detection of facial gestures from generic video*. in *IEEE 6th Workshop on Multimedia Signal Processing* 2004: IEEE.
64. BLACK, M.J., *Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion.* International Journal of Computer Vision, 1997. 25(1): p. 23-48.
65. W. Abd-Almageed, M. Sami and F. G. Bebis, A non-intrusive Kalman filter-based tracker for pursuit eye movement, in *Proc. American Control Conference*, (Anchorage AK, Alaska, 2002).
66. S. Hamlaoui and F. Davoine, Facial action tracking using an AAM-based condensation approach, in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing* 2 (IEEE Computer Society Washington, DC, USA, 2005), pp. 701- 704.
67. J. Whitehill and C. W. Omlin, Local versus global segmentation for facial expression recognition, in *Proc. 7th International Conference on Automatic Face and Gesture Recognition* (IEEE Computer Society Washington, DC, USA, 2006) pp. 357-362.
68. V. Blanz and T. Vetter, A morphable model for the synthesis of 3D faces. in *SIGGRAPH '99: Proc. 26th annual conference on Computer graphics and interactive techniques*. (ACM Press/Addison-Wesley Publishing Co, New York, Usa, 1999).
69. S. Baker, S. Koterba, I. Matthews, C. Hu, J. Xiao, J. Cohn and T. Kanade, Multi-view AAM fitting and camera calibration, in *Proc. Tenth IEEE International Conference on Computer Vision*, (IEEE Computer Society Washington, DC, USA, 2005), pp. 511-518.
70. Essa, I.A. and A.P. Pentland, *Coding, analysis, interpretation, and recognition of facial expressions.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997. 19(7): p. 757-763.
71. S. B. Gokturk et al., Model-based face tracking for view-independent facial expression recognition, in *Proc. Fifth IEEE International Conference on Automatic Face and Gesture Recognition* (IEEE Computer Society Washington, DC, USA, 2002), pp. 287.
72. Cootes, T.F., G.J. Edwards, and C.J. Taylor, *Active appearance models.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2001. 23(6): p. 681-685.
73. T. Cootes and C. Taylor, Active shape models smart snakes, in *British Machine Vision Conference* (Springer-Verlag, 1992), pp. 266-275.
74. J. Cohn, A. Zlochower J. J. Lien, Y. T. Wu and T. Kanade, Automated face coding: A Computer-vision based method of facial expression analysis, in *7th European Conference on Facial Expression measurement and meaning* (Salzbur g, Austria, 1997), pp. 329-333.
75. T. Otsuka, J. Ohya, Spotting segments displaying facial expression from image sequences using HMM, in *Proc. 3rd International Conference on Face & Gesture Recognition* (IEEE Computer Society Washington, DC, USA, 1998), pp. 442.
76. S. Kimura, M.Yachida, Facial expression recognition and its degree estimation, in *Computer Vision and Pattern Recognition,* (San Juan, Puerto Rico, 1997), pp. 295.
77. H. Kobayashi, F. Hara, Dynamic recognition of basic facial expressions by discrete-time recurrent neural network, in *Proc. of the International Joint Conference on Neural Network* **1**(25) (Nagoya, Japan, 1993), pp. 155-158.
78. I. A. Essa, Coding, analysis, interpretation and recognition of facial expressions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7) (1997) pp. 757-763.
79. M. Yoneyama, A. Ohtake and K. Shirai, Facial expression recognition using discrete Hopfield neural networks, in *Proc. of the International Conference on Image Processing* (IEEE Computer Society Washington, DC, USA, 1997), pp. 117.
80. L. Franco, A. Treves. A Neural Network Facial Expression Recognition System using Unsupervised Local Processing. in *Proc. 2nd International Symposium on Image and Signal Processing and Analysis*, (Pula, Croatia, 2001), pp. 628-632.
81. W. Liu, Z. Wang, Facial expression recognition based on fusion of multiple Gabor features, in *Proc. 18th International Conference on Pattern Recognition* (IEEE Computer Society Washington, DC, USA,2006), pp. 536-539.
82. Ma, L. and K. Khorasani, *Facial expression recognition using constructive feedforward neural networks.* Systems, Man and Cybernetics, Part B, IEEE Transactions on, 2004. 34(3): p. 1588-1595.
83. M. S. Bartlett et al., Recognizing facial expression: Machine learning and application to spontaneous behavior, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2 (San Diego, USA, 2005), pp. 568-573.
84. D. Datcu and L. J. M. Rothkrantz, Facial expression recognition with relevance vector machines, in *IEEE International Conference on Multimedia and Expo* (Amsterdam, Holland, 2005), pp. 193-196.

85. P. Ekman, W. Friesen., *Facial Action Coding System.* (Consulting Psychologist Press, Palo Alto CA, USA, 1978).
86. J. C. Hager, P. Ekman, W. Friesen, *Facial Action Coding System.* (A Human Face, Salt Lake City, UT, 2002).
87. A. M. A. M. Tekalp, J. Ostermann, Face and 2D mesh animation in MPEG-4, *Image Communication*, **15**(4-5) (2000) 387-421.
88. M. Pardas, A. Bonafonte, J.L. Landabaso, Emotion Recognition Based on MPEG-4 Facial Animation Parameters. in *Proc. Acoustics, Speech, and Signal Processing.*4 (IEEE Computer Society Washington, DC, USA, 2002), pp. 3624-3627.
89. N. Tsapatsoulis, A.R., S Kollias, R. Cowie, E. Douglas-Cowie, Emotion Recognition and Synthesis based on MPEG-4 FAPs. *MPEG-4 Facial Animation*, Igor S. Pandzic, Robert Forchheimer (John Wiley & Sons , Ltd, Hoboken, USA 2002), pp. 51-54.
90. M. Mufti and A. Khanam, Fuzzy rule based facial expression recognition, in *International conference on Computational Inteligence for Modelling Control and Automation and International Conference on Intelligent Agents Web Technologies and International Commerce* (Sydney, NSW, Australia, 2006), pp. 57.
91. S. Ioannou et al., Emotion recognition through facial expression analysis based on a neurofuzzy network, *Neural Networks* **18**(4) (Elsevier Science Ltd. Oxford, UK, 2005), pp. 423-435.
92. N. Sebe et al., Emotion recognition based on joint visual and audio cues, in *18th International Conference on Pattern Recognition* (Hong Kong, China, 2006), pp. 1136-1139.