Image and Video Processing for Affective Applications

Maja Pantic and George Caridakis

Abstract Recent advances in the research area of affective computing have broadened the range of application areas of its findings, and additionally, as the state of the art advances in affective computing, other related research areas (computer vision, pattern recognition, etc.) discover new challenges that are related to image and video processing related to the task of automatic affective analysis. Although humans cope, relatively easily, with the task of perceiving facial expressions, gestural expressivity, and other visual cues involved in expressing emotion the automatic counterpart of the task is far from trivial. This chapter summarizes current research efforts in solving these problems and enumerates the scientific and engineering issues that arise in meeting these challenges toward emotion-aware systems.

1 The Problem Domain

Because of its practical importance and the theoretical interest of cognitive and medical scientists (Ekman et al., 2002; Pantic, 2005; Chang et al., 2006), machine analysis of facial expressions attracted the interest of many researchers. For exhaustive surveys of the related work, readers are referred to Samal and Iyengar (1992) for an overview of early works, Tian et al. (2005) and Pantic and Bartlett (2007) for surveys of techniques for detecting facial muscle actions, and Pantic and Rothkrantz (2000, 2000) for surveys of facial affect recognition methods. However, although humans detect and analyze faces and facial expressions in a scene with little or no effort, development of an automated system that accomplishes this task is rather difficult.

M. Pantic (✉)
Department of Computing, Imperial College, London, UK; Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands
e-mail: M.Pantic@imperial.ac.uk

P. Petta et al. (eds.), *Emotion-Oriented Systems*, Cognitive Technologies,
DOI 10.1007/978-3-642-15184-2_7,

1.1 Level of Description: Action Units and Emotions

Two main streams in the current research on automatic analysis of facial expressions consider facial affect (emotion) detection and facial muscle action (action unit) detection. These two streams stem directly from two major approaches to facial expression measurement in psychological research (Cohen, 2006): message and sign judgment. The aim of message judgment is to *infer* what underlies a displayed facial expression, such as affect or personality, while the aim of sign judgment is to *describe* the "surface" of the shown behavior, such as facial movement or facial component shape. Thus, a brow furrow can be judged as "anger" (Ekman, 2003; Kapoor et al., 2003) in a message-judgment and as a facial movement that lowers and pulls the eyebrows closer together in a sign-judgment approach. While message judgment is all about interpretation, sign judgment attempts to be objective, leaving inference about the conveyed message to higher order decision making.

FACS (Ekman and Friesen, 1969, 1978) provides an objective and comprehensive language for describing facial expressions and relating them back to what is known about their meaning from the behavioral science literature. Because it is comprehensive, FACS also allows for the discovery of new patterns related to emotional or situational states. For example, what are the facial behaviors associated with driver fatigue? What are the facial behaviors associated with states that are critical for automated tutoring systems, such as interest, boredom, confusion, or comprehension? Research based upon FACS has also shown that facial actions can show differences between those telling the truth and lying at a much higher accuracy level than naive subjects making subjective judgments of the same faces (Cohn and Schmidt, 2004; Fasel et al., 2004).

It is not surprising, therefore, that automatic Action Units (AU) coding in face images and face image sequences attracted the interest of computer vision researchers. Historically, the first attempts to encode AUs in images of faces in an automatic way were reported by Bartlett et al. (2006), Lien et al. (1998), and Pantic et al. (1998). These three research groups are still the forerunners in this research field. The focus of the research efforts in the field was first on automatic recognition of AUs in either static face images or face image sequences picturing facial expressions produced on command. Several promising prototype systems were reported that can recognize deliberately produced AUs in either (near-) frontal-view face images (Anderson and McOwan, 2006; Samal and Iyengar, 1992; Pantic and Rothkrantz, 2003) or profile-view face images (Pantic and Rothkrantz, 2003; Pantic and Patras, 2005). These systems employ different approaches including expert rules and machine learning methods such as neural networks and use either feature-based image representations (i.e., use geometric features like facial points, see Sect. 2.3) or appearance-based image representations (i.e., use texture of the facial skin including wrinkles and furrows, see Sect. 2.3).

One of the main criticisms that these works received from both cognitive and computer scientists is that the methods are not applicable in real-life situations, where subtle changes in facial expression typify the displayed facial behavior rather than the exaggerated changes that typify posed expressions. Hence, the focus of the research in the field started to shift to automatic AU recognition in spontaneous

facial expressions (produced in a reflex-like manner). Several works have recently emerged on machine analysis of AUs in spontaneous facial expression data (e.g., Cohn, 2006; Bartlett et al., 1999; Valstar and Pantic, 2006). These methods employ probabilistic, statistical, and ensemble learning techniques, which seem to be particularly suitable for automatic AU recognition from face image sequences (see, e.g., Tian et al., 2001; Lien et al., 1998).

1.2 Facial Expression Configuration and Dynamics

When it comes to research on automatic AU coding, automatic recognition of facial expression configuration (in terms of AUs constituting the observed expression) has been the main focus of the research efforts in the field. However, both the configuration and the dynamics of facial expressions (i.e., the timing and the duration of various AUs) are important for interpretation of human facial behavior. The body of research in cognitive sciences, which argues that the dynamics of facial expressions are crucial for the interpretation of the observed behavior, is ever growing (Ekman et al., 1993; Lee and Kim, 1999). Facial expression temporal dynamics are essential for categorization of complex psychological states like various types of pain and mood; they represent a critical factor for interpretation of social behaviors like social inhibition, embarrassment, amusement, and shame, and they are a key parameter in differentiation between posed and spontaneous facial displays (Ekman et al., 1993). For instance, spontaneous smiles are smaller in amplitude, longer in total duration, and slower in onset and offset time than posed smiles (e.g., a polite smile) (Cohn and Schmidt, 2004). Another study showed that spontaneous smiles, in contrast to posed smiles, can have multiple apexes (multiple rises of the mouth corners – AU12) and are accompanied by other AUs that appear either simultaneously with AU12 or follow AU12 within 1 s (Cohn et al., 2004). Similarly, it has been shown that the differences between spontaneous and deliberately displayed brow actions (AU1, AU2, AU4) are in the duration and the speed of onset and offset of the actions and in the order and the timing of actions' occurrences (Valstar and Pantic, 2006).

In spite of these findings, the vast majority of the past work in the field does not take dynamics of facial expressions into account when analyzing shown facial behavior. Some of the past work in the field has used aspects of temporal dynamics of facial expression such as the speed of a facial point displacement or the persistence of facial parameters over time (e.g., Lien et al., 1998). However, only three recent studies analyze explicitly the temporal dynamics of facial expressions. These studies explore automatic segmentation of AU activation into temporal segments (neutral, onset, apex, offset) in frontal- (Pantic and Bartlett, 2007; Tian et al., 2005) and profile-view (Pantic and Patras, 2005) face videos.

1.3 Facial Expression Intensity and Context Dependency

Facial expressions can vary in intensity. By intensity we mean the relative degree of change in facial expression as compared to a relaxed, neutral facial expression.

It has been experimentally shown that the expression-decoding accuracy and the perceived intensity of the underlying affective state vary linearly with the physical intensity of the facial display (Gu and Ji, 2004). Hence, explicit analysis of expression intensity variation is very important for accurate expression interpretation and is also essential to the ability to distinguish between spontaneous and posed facial behavior discussed in the previous sections. While FACS provides a 5-point intensity scale to describe AU intensity variation and enable manual quantification of AU intensity (Ekman and Friesen, 1978), fully automated methods that accomplish this task are yet to be developed. However, first steps toward this goal have been made. Automatic coding of intensity variation was explicitly compared to manual coding in Bartlett et al. (1999). They found that the distance to the separating hyperplane in their learned classifiers correlated significantly with the intensity scores provided by expert FACS coders.

Rapid facial signals do not usually convey exclusively one type of messages. For instance, squinted eyes may be interpreted as sensitivity of the eyes to bright light if this action is a reflex (a manipulator), as an expression of disliking if this action has been displayed when seeing someone passing by (affective cue), or as an illustrator of friendly anger on friendly teasing if this action has been posed (in contrast to being unintentionally displayed) during a chat with a friend, to mention just a few possibilities. As already mentioned in Sect. 1.3, to interpret an observed facial expression, it is important to know the context in which the observed expression has been displayed – where the expresser is (outside, inside, in the car, in the kitchen, etc.), what his or her current task is, are other people involved, and who the expresser is. Knowing the expresser is particularly important as individuals often have characteristic facial expressions and may differ in the way certain states (other than the basic emotions) are expressed. Since the problem of context-sensing is extremely difficult to solve (if possible at all) for a general case, pragmatic approaches (e.g., activity/application- and user-centered approach) should be taken when learning the grammar of human facial behavior (Pantic et al., 1998; Pantic and Patras, 2006). However, except for a few works on user-profiled interpretation of facial expressions like those of Fasel et al. (2004) and Pantic and Rothkrantz (2004a), virtually all existing automated facial expression analyzers are context insensitive.

1.4 Facial Expression Databases

To develop and evaluate facial behavior analyzers capable of dealing with different dimensions of the problem space as defined above, large collections of training and test data are needed (Pantic and Rothkrantz, 2000; Tian et al., 2001).

A complete overview of existing, publicly available data sets that can be used in research on automatic facial expression analysis is given by Pantic and Bartlett (2007). We will provide here a description of two relevant facial expression databases: the Cohn–Kanade database (Juslin and Scherer, 2005), which is the most widely used database in research on automated facial expression analysis, and the MMI facial expression database (Pantic et al., 2005a; Pantic, 2006), which

represents the most comprehensive, online reference set of face images and videos of both deliberate and spontaneously displayed facial expressions.

2 The State of the Art

Although humans detect and analyze faces and facial expressions in a scene with little or no effort, development of an automated system that accomplishes this task is rather difficult. There are several related problems (Pantic et al., 2006). The first is to find faces in the scene independent of clutter, occlusions, and variations in head pose and lighting conditions. Then, geometric facial features such as facial salient points (e.g., the mouth corners) or parameters of an appearance-based facial model (e.g., parameters of a fitted active appearance model) should be extracted from the regions of the scene that contain faces. The system should perform this accurately, in a fully automatic manner and preferably in real time. Eventually, the extracted facial information should be interpreted in terms of facial signals (winks, blinks, smiles, affective states, cognitive states, moods) in a context-dependent (personalized, task-, situation-, and application-dependent) manner. This section summarizes current research efforts in solving these problems and enumerates the scientific and engineering issues that arise in meeting these challenges.

2.1 Face Detection

The first step in facial information processing is face detection, i.e., identification of all regions in the scene that contain a human face. The problem of *finding faces* should be solved regardless of clutter, occlusions, and variations in head pose and lighting conditions. The presence of non-rigid movements due to facial expression and a high degree of variability in facial size, color, and texture make this problem even more difficult. Numerous techniques have been developed for face detection in still images (Wierzbicka, 1993; Larsen and Diener, 1992). Arguably the most commonly employed face detector in automatic facial expression analysis is the real-time face detector proposed by Viola and Jones (2004).

2.2 Facial Feature Extraction

After the presence of a face has been detected in the observed scene, the next step is to extract the information about the displayed facial signals. The problem of *facial feature extraction* from regions in the scene that contain a human face may be divided into at least three dimensions (Pantic et al., 2006):

(a) Is temporal information used?
(b) Are the features holistic (spanning the whole face) or analytic (spanning subparts of the face)?
(c) Are the features view- or volume-based (2D/3D)?

Most of the existing facial expression analyzers are directed toward 2D spatiotemporal facial feature extraction. The usually extracted facial features are either *geometric features* such as the shapes of the facial components (eyes, mouth, etc.) and the locations of facial fiducial points (corners of the eyes, mouth, etc.) or *appearance features* representing the texture of the facial skin including wrinkles, bulges, and furrows. Typical examples of geometric feature based methods are those of Gokturk et al. (2002), who used 19-point face mesh; Chang et al. (2006), who used a shape model defined by 58 facial landmarks; and Pantic et al. (Pantic and Rothkrantz, 2003; Pantic and Bartlett, 2007; Pantic and Patras, 2005; Tian et al., 2005), who used a set of facial characteristic points visible in either frontal or profile view of the face. Typical examples of *hybrid*, geometric and appearance feature based, methods are those of Tian et al. (2005), who used shape-based models of eyes, eyebrows, and mouth and transient features like crows-feet wrinkles and nasolabial furrow, and of Zhang and Ji (2005), who used 26 facial points around the eyes, eyebrows, and mouth and the same transient features as Tian et al. (2005). Typical examples of appearance feature based methods are those of Bartlett et al. (1999), Anderson and McOwan (2006) and Lien et al. (1998), who used Gabor wavelets; Anderson and McOwen (2006), who used a holistic, monochrome, spatial-ratio face template; and Valstar et al. (2006), who used temporal templates.

To illustrate geometric facial feature detection and tracking, the methods developed by Vukadinovic and Pantic (2005) for automatic point localization and by Patras and Pantic (2004) for facial point tracking will be shortly explained.

It has been reported that methods based on geometric features are often outperformed by those based on appearance features using, e.g., Gabor wavelets or eigenfaces (Anderson and McOwan, 2006). Certainly, this may depend on the classification method and/or machine learning approach which takes the features as input. Recent studies like that of Pantic and Patras (2005) and Valstar and Pantic (2006) show that in some cases geometric features can outperform appearance-based ones. Yet, it seems that using both geometric and appearance features might be the best choice in the case of certain facial expressions (Pantic and Patras, 2005).

2.3 Facial Muscle Action Coding

As already mentioned in Sect. 2.1, two main streams in the current research on automatic analysis of facial expressions consider facial affect (emotion) detection and facial muscle action detection such as the AUs defined in FACS (Ekman and Friesen, 1969, 1978). Although FACS provides a good foundation for AU coding of face images by human observers, achieving AU recognition by a computer is not an easy task. A problematic issue is that AUs can occur in more than 7,000 different complex combinations, causing bulges (e.g., by the tongue pushed under one of the lips) and various in- and out-of-image plane movements of permanent facial features (e.g., jetted jaw) that are difficult to detect in 2D face images. Historically, the first attempts to encode AUs in images of faces in an automatic way were reported by Bartlett et al. (2006), Lien et al. (1998), and Pantic et al. (1998). These three research groups are still the forerunners in this research field.

Pantic and her colleagues reported on multiple efforts aimed at automating the analysis of facial expressions in terms of facial muscle actions that constitute the expressions. The majority of this previous work concerns geometric feature based methods for automatic FACS coding of face images. Early work was aimed at AU coding in static face images (Pantic and Rothkrantz, 2003) while more recent work addressed the problem of automatic AU coding in face video (Pantic and Bartlett, 2007; Tian et al., 2005; Pantic and Patras, 2005; Valstar and Pantic, 2006). Based upon the tracked movements of facial characteristic points, as discussed in Sect. 2.3, Pantic and her colleagues mainly experimented with rule-based (Pantic and Bartlett, 2007; Pantic and Patras, 2005) and support vector machine based methods (Tian et al., 2005; Valstar and Pantic, 2006), for recognition of AUs in either near frontal-view or near profile-view face image sequences. As already mentioned in Sect. 2.2, automatic recognition of facial expression configuration (in terms of AUs constituting the observed expression) has been the main focus of the research efforts in the field. In contrast to the methods developed elsewhere, which thus focus onto the problem of spatial modeling of facial expressions, the methods proposed by Pantic and her colleagues address the problem of temporal modeling of facial expressions as well. In other words, these methods are very suitable for encoding temporal activation patterns (onset → apex → offset) of AUs shown in an input face video. This is of importance for there is now a growing body of psychological research that argues that temporal dynamics of facial behavior (i.e., the timing and the duration of facial activity) is a critical factor for the interpretation of the observed behavior (see Sect. 2.2). Black and Yacoob (1997) presented the earliest attempt to automatically segment prototypic facial expressions of emotions into onset, apex, and offset components. To the best of our knowledge, the only systems to date for explicit recognition of temporal segments of AUs are the ones by Pantic and colleagues (Pantic and Bartlett, 2007; Tian et al., 2005; Pantic and Patras, 2005; Valstar and Pantic, 2006). A short explanation of the methods will follow.

Appearance-based approaches to AU recognition such as the ones by Kapoor et al. (2003), Valstar et al. (2006), and Bartlett and colleagues (e.g., Anderson and McOwan, 2006; Lien et al., 1998; Bartlett et al., 1999) differ from those of Pantic and colleagues (e.g., Pantic and Rothkrantz, 2003; Pantic and Bartlett, 2007) and Tian et al. (2005), in that learning the appearance of any AU is based on a set of labeled training data. Hence the limiting factor in appearance-based machine learning approaches is having enough of various labeled examples for a robust system. Previous explorations of this idea showed that, given accurate 3D alignment, at least 50 examples are needed for moderate performance (in the 80% range), and over 200 examples are needed to achieve high precision (Pantic, 2006). An example of appearance-based approaches to AU recognition is the system of Bartlett et al.

3 Facial Affect Recognition

To interpret someone's behavioral cues, including emotional states, people rely mainly on shown facial expressions (Ambadar et al., 2005; Kapoor et al., 2003),

and it is not surprising, therefore, that the majority of efforts in affective computing concern automatic analysis of facial displays. For exhaustive surveys of studies on machine analysis of facial affect, readers are referred to Pantic et al. (2006), Pantic and Rothkrantz (2000), Tian et al. (2001) and Pantic (2006). These surveys indicate that the capabilities of currently existing facial affect analyzers are rather limited. More specifically, current facial affect analyzers

- handle only a small set of volitionally displayed prototypic facial expressions of six basic emotions,
- do not perform a context-sensitive analysis (either user-, or environment-, or task-dependent analysis) of the observed facial behavior,
- do not analyze extracted facial expression information on different time scales (i.e., short pre-segmented videos are only handled) – consequently, inferences about the expressed mood and attitude (larger time scales) cannot be made, and
- adopt strong assumptions (i.e., the systems can handle only portraits or nearly frontal views of faces with no facial hair or glasses, recorded under constant illumination and displaying exaggerated prototypic expressions of emotions).

Automatic detection of the six basic emotions under these assumptions, that is, in posed, controlled displays, can be done with reasonably high accuracy. However, detecting these facial expressions in the less constrained environments of real applications is a much more challenging problem which is just beginning to be explored. There have been just a few such tentative efforts aimed at detection of cognitive and psychological states like interest (Ekman and Rosenberg, 2005), pain (Bartlett et al., 1999), and fatigue (Goleman, 1995). An example is the pain detector of Bartlett et al.

Also an attempt to discern spontaneous from volitionally displayed facial behavior has been reported (Valstar and Pantic, 2006). Description of the method will be provided.

3.1 Machine Analysis of Facial Expressions: Challenges

Automating the analysis of facial expressions is important to realize more natural, context-sensitive (e.g., affective) human–computer interaction, to advance studies on human emotion and affective computing, and to boost numerous applications in fields as diverse as security, medicine, and education. Although most of the facial expression analyzers developed so far target human facial affect analysis and attempt to recognize a small set of prototypic emotional facial expressions like happiness and anger (Pantic et al., 1998; Pantic, 2006), some progress has been made in addressing a number of other scientific challenges that are considered essential for realization of machine understanding of human facial behavior. First of all, the research on automatic detection of facial muscle actions, which produce facial expressions, witnessed a significant progress in the past years. A number of promising prototype systems have been proposed recently that can recognize up to

27 AUs (from a total of 44 AUs) in either (near-) frontal-view or profile-view face image sequences (Tian et al., 2001; Pantic, 2006). Further, although the vast majority of the past work in the field does not make an effort to explicitly analyze the properties of facial expression temporal dynamics, a few approaches to automatic segmentation of AU activation into temporal segments (neutral, onset, apex, offset) have been recently proposed (e.g., Pantic and Bartlett, 2007; Pantic and Patras, 2005; Tian et al., 2005). Also, even though most of the past work on automatic facial expression analysis is aimed at the analysis of posed (deliberately displayed) facial expressions, a few efforts were recently reported on machine analysis of spontaneous facial expressions (e.g., Cohn, 2006; Bartlett et al., 1999; Valstar and Pantic, 2006). In addition, exceptions from the overall state of the art in the field include a few works toward detection of attitudinal and non-basic affective states such as interest (Ekman and Rosenberg, 2005), pain (Bartlett et al., 1999), and fatigue (Goleman, 1995); a few works on context-sensitive (user-profiled) interpretation of facial expressions (El Kaliouby and Robinson, 2004; Pantic and Rothkrantz, 2004); and an attempt to explicitly discern in an automatic way spontaneous from volitionally displayed facial behavior (Valstar and Pantic, 2006). However, many research questions raised in Sect. 2.2 remain unanswered and a lot of research has yet to be done.

When it comes to automatic AU detection, existing methods do not yet recognize the full range of facial behavior (i.e., all 44 AUs defined in FACS).

Existing methods for machine analysis of facial expressions discussed throughout this chapter assume that the input data are near frontal- or profile-view face image sequences showing facial displays that always begin with a neutral state. In reality, such assumption cannot be made.

If we consider the state of the art in face detection and facial feature localization and tracking, noisy and partial data should be expected. A facial expression analyzer should be able to deal with these imperfect data and to generate its conclusion so that the certainty associated with it varies with the certainty of face and facial point localization and tracking data.

Another related issue that should be addressed is how to include information about the context (environment, user, user's task) in which the observed expressive behavior was displayed so that a context-sensitive analysis of facial behavior can be achieved.

4 Machine Analysis of Body Gestures

Gesture and sign language recognition has gathered abundant attention in the recent years and the research area has developed an adequate relevant literature. Several approaches have been proposed and tested on a variety of data sets. An extensive review of these techniques is presented in Ong and Ranganath (2005) and Ying and Huang (2001). The first focuses mainly on sign language recognition and classification issues while examining closely hand localization and tracking, and various

feature extraction related to automatic analysis of manual signing. In addition to the previous, they examine the linguistic aspect of sign language and non-manual signals and how they would be incorporated in the sign language recognition chain. On the other hand, Wu and Huang (Ying and Huang, 2001) delve more into publications related to hand modeling (shape analysis, kinematics chain and dynamics) and computer vision and pattern recognition issues associated to hand localization and extracting features from image sequences.

One of the most common approaches is to extract features from the input signal and use these features as input to a fine-tuned HMM. Perrin et al. (2004) track finger gestures using laser light and compute three features based on the finger's coordinates. Starner et al. (1998) present two systems based on head- and desk-mounted cameras feeding HMMs with uniform architecture. For each case two experiments were performed by varying the feature set used for classifying the gestures. Vogler and Metaxas (1998) developed a system based on parallel HMMs to recognize American SL gestures with a 3D camera system.

Also variations of the previous group have been widely presented. Hossain and Jenkin (2005) present two variations of HMM, implicit and explicit temporal information encoded, in order to recognize a single gesture type (hand-raise). Lee et al. (1998) propose another HMM variation enhanced by a gesture spotting network for calculating the likelihood threshold for "pick a winner" situations. Lee implemented the PowerGesture which recognizes 10 PowerPoint continuous commands with an average detection rate of 98% and recognition rate of 93%. Ozer et al. (2005) utilize one HMM per articulated body part. Their main focus is on the real-time aspect of the overall system: their image processing modules feed a graph matching module; this forms the input of the system. Wilson and Bobick (1999) introduce parametric HMMs to cope with gesture variations.

Other approaches have adopted other machine learning and artificial intelligence techniques. Juang and Ku (2005) present FTRFN (fuzzified TSK-type recurrent fuzzy network) and they test the proposed system on 10 trajectories achieving an average of 92% recognition rate. Mu-Chun (2000) presents a fuzzy rule based on hyperrectangular composite neural networks (HRCNNs) for selecting models. Hong et al. (2000) perform 2D gesture recognition using manually constructed finite state machines and have promising results reaching an average of 92% on two gesture data sets, one based on hand gestures and one based on mouse gestures, although the data sets seem quite small (three and four classes, respectively). Wong and Cipolla (2006) achieved 80–93% recognition rate over nine quite elementary gestures by adopting a sparse Bayesian classifier. As inputs to the classifier they utilized motion gradient orientation features over continuous video streams. Yang et al. (2002) base their algorithm on motion trajectories or ensembles of them which feed a time delay neural network. Huang et al. (1998) present an isolated gesture recognition system, which uses as input the monocular sequence of dynamic single-handed gesturing by dynamic time warping.

Finally, there have been some approaches combining more than one technique. Mantyla et al. (2000) present a system for static gestures recognition using a self-organizing mapping scheme of Kohonen while a hidden Markov model is used for

recognizing dynamic gestures. The input of the two systems was acquired by acceleration sensors attached to a mobile device. These two systems (SOM and HMM) are not combined in any way but each one is utilized in different gesture types. Black and Jepson (1998) present an extension of the "condensation" algorithm in which gestures are modeled as temporal trajectories of the velocity of the tracked hands. Fang et al. (2001) present an additional layer enhancing the HMM architecture with SOFM and improving their recognition rate by 5%. In a more recent work, the same group of researchers introduced a fuzzy decision tree in an attempt to reduce the search space of recognized classes without loss of accuracy.

References

Ambadar Z, Schooler J, Cohn JF (2005) Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. Psychol Sci 16(5):403–410

Ambady N, Rosenthal R (1992) Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. Psychol Bull 111(2):256–274

Anderson K, McOwan PW (2006) A real-time automated system for recognition of human facial expressions. IEEE Trans Syst Man Cybern B 36(1):96–105

Bartlett MS, Hager JC, Ekman P, Sejnowski TJ (1999) Measuring facial expressions by computer image analysis. Psychophysiology 36(2):253–263

Bartlett MS, Littlewort G, Frank MG, Lainscsek C, Fasel I, Movellan J (2006) Fully automatic facial action recognition in spontaneous behavior. In: Seventh IEEE international conference on automatic face and gesture recognition (FG 2006), April 10–12, Southampton, UK, pp 223–230

Bartlett MS, Viola PA, Sejnowski TJ, Golomb BA, Larsen J, Hager JC, Ekman P (1996) Classifying facial actions. Adv Neural Inf Process Syst 8:823–829

Black MJ, Jepson AD (1998) Recognizing temporal trajectories using the condensation algorithm. In: Proceedings of the 3rd international conference on Face and Gesture Recognition FG, IEEE Computer Society, Washington, DC

Black M, Yacoob Y (1997) Recognizing facial expressions in image sequences using local parameterized models of image motion. Comput Vis 25(1):23–48

Bobick AF, Wilson AD (1997) [a]A state-based approach to the representation and recognition of gesture. IEEE Trans Pattern Anal Mach Intell 19(12):1325–1337

Chang Y, Hu C, Feris R, Turk M (2006) Manifold based analysis of facial expression. J Image Vis Comput 24(6):605–614

Cohen MM (2006) Perspectives on the face. Oxford University Press, Oxford, UK

Cohn JF (2006) Foundations of human computing: facial expression and emotion. In: Proceedings of the ACM international conference on Multimodal Interfaces, Banff, Canada, November 2–4, pp 233–238

Cohn JF, Reed LI, Ambadar Z, Xiao J, Moriyama T (2004) Automatic analysis and recognition of brow actions in spontaneous facial behavior. In: Proceedings of the IEEE international conference on systems, man and cybernetics, The Hague, Netherlands, October 10–13, pp 610–616

Cohn JF, Schmidt KL (2004) The timing of facial motion in posed and spontaneous smiles. J Wavelets Multiresolution Inf Process 2(2):121–132

Ekman P (2003) Darwin, deception, and facial expression. Ann N Y Acad Sci 1000:205–221

Ekman P, Friesen WF (1969) The repertoire of nonverbal behavioral categories – origins, usage, and coding. Semiotica 1:49–98

Ekman P, Friesen WF (1978) Facial action coding system. Consulting Psychologist Press, Palo Alto, CA

Ekman P, Friesen WF, Hager JC (2002) Facial action coding system. A Human Face, Salt Lake City

Ekman P, Huang TS, Sejnowski TJ Hager JC (eds) (1993) NSF understanding the face.A Human Face Store, Salt Lake City, (see Library)

Ekman P, Rosenberg EL, (eds) (2005) What the face reveals: basic and applied studies of spontaneous expression using the FACS. Oxford University Press, Oxford, UK

El Kaliouby R, Robinson P (2004) Real-time inference of complex mental states from facial expressions and head gestures. Proc Int Conf Comput Vis Pattern Recogn 3:154

Fang G, Gao W, Ma J (2001) Signer-independent sign language recognition based on SOFM/HMM, Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 2001. In: Proceedings of the IEEE ICCV Workshop on RATFG-RTS'01, Vancouver, Canada, July 13, pp 90–95

Fasel B, Monay F, Gatica-Perez D (2004) Latent semantic analysis of facial action codes for automatic facial expression recognition. In: Proceedings of the 6th ACM SIGMM international workshop on Multimedia Information Retrieval, New York, NY, October 10–16, pp 181–188

Fridlund AJ (1997) The new ethology of human facial expression. In: Russell JA, Fernandez-Dols JM (eds) The psychology of facial expression. Cambridge University Press, Cambridge, MA, pp 103–129

Gokturk SB, Bouguet JY, Tomasi C, Girod B (2002) Model-based face tracking for view independent facial expression recognition. In: 5th IEEE international conference on automatic face and gesture recognition (FGR 2002), Washington, DC, May 20–21, pp 272–278

Goleman D (1995) Emotional intelligence. Bantam Books, New York, NY

Gu H, Ji Q (2004) An automated face reader for fatigue detection. In: Sixth IEEE international conference on automatic face and gesture recognition (FGR 2004), IEEE Computer Society, Seoul, Korea, May 17–19, pp 111–116

Hong P, Turk M, Huang TS (2000) Gesture modeling and recognition using finite state machines. In: Proceedings of the 4th IEEE international conference and Gesture Recognition, Mar 2000, Grenoble

Hossain M, Jenkin M (2005) Recognizing hand-raising gestures using HMM. In: Proceedings of the 2nd Canadian conference on Computer and Robot Vision (CRV′05) – vol 00 (9–11 May 2005). CRV, IEEE Computer Society, Washington, DC, pp 405–412

Huang Y, Zhu Y, Xu G, Zhang H (1998) Spatial-temporal features by image registration and warping for dynamic gesture recognition. In: IEEE international conference on Systems, Man, and Cybernetics, vol 5, 11–14 Oct 1998, San Diego, pp 4498–4503

Juang CF, Ku K-C (2005 Aug) A recurrent fuzzy network for fuzzy temporal sequence processing and gesture recognition. IEEE Trans Syst Man Cybern B 35(4):646–658

Juslin PN, Scherer KR (2005) Vocal expression of affect. In: Harrigan J, Rosenthal R, Scherer K (eds) The new handbook of methods in nonverbal behavior research. Oxford University Press, Oxford

Kanade T, Cohn JF, Tian Y (2000) Comprehensive database for facial expression analysis. In: 4th IEEE international conference on automatic face and gesture recognition (FGR 2000), IEEE Comptuer Society, Grenoble, France, March 26–30, pp 46–53

Kapoor A, Qi Y, Picard RW (2003) Fully automatic upper facial action recognition. In: Proceedings of the IEEE international workshop on Analysis and Modeling of Faces and Gestures, Nice, France, pp 195–202

Larsen RJ, Diener E (1992) Promises and problems with the circumplex model of emotion. Emotion 13:25–59. In: Clark MS (ed) Review of personality and social psychology. Sage, Newbury Park, CA

Lee C, Ghyme S, Park C, Wohn K (1998) The control of avatar motion using hand gesture. In: Proceedings of the ACM symposium on Virtual Reality Software and Technology (Taipei, Taiwan, 2–5 Nov 1998). VRST ′98, ACM Press, New York, NY, pp 59–65

Lee H-K, Kim JH (1999) [a]An HMM-based threshold model approach for gesture recognition[o]. IEEE Trans Pattern Anal Mach Intell 21(10):961–973

Li SZ, Jain AK, (eds) (2005) Handbook of face recognition. Springer, New York, NY
Lien JJJ, Kanade T, Cohn JF, Li CC (1998) Subtly different facial expression recognition and expression intensity estimation. In: Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition, Research Triangle Park, North Carolina, pp 853–859
Mantyla VM, Mantyjarvi J, Seppanen T, Tuulari E (2000) Hand gesture recognition of a mobile device user. In: IEEE international conference on Multimedia and Expo (ICME 2000), New York, NY, vol 1, pp 281–284
Mu-Chun S (2000) A fuzzy rule-based approach to spatio-temporal hand gesture recognition. IEEE Trans Syst Man Cybern C Appl Rev 30(2):276–281
Ong SCW, Ranganath S (2005) Automatic sign language analysis: a survey and the future beyond lexical meaning. IEEE Trans Pattern Anal Mach Intell 27(6):873–891
Ozer IB, Lu T, Wolf W (2005) Design of a real-time gesture recognition system: high performance through algorithms and software. Signal Process Mag IEEE 22(3):57–64
Pantic M (2005) Affective computing. In: Pagani M (ed) Encyclopedia of multimedia technology and networking, vol 1. Idea Group Reference, Hershy, PA, pp 8–14
Pantic M (2006) Face for ambient interface. Lect Notes Artif Intell 3864:35–66
Pantic M, Bartlett MS (2007) Machine analysis of facial expressions. In: Kurihara K (ed) Face recognition. Advanced Robotic Systems, Vienna, Austria
Pantic M, Patras I (2005) Detecting facial actions and their temporal segments in nearly frontal-view face image sequences. In: Proceedings of the IEEE international conference on systems, Man and Cybernetics, Salerno, Italy, pp 3358–3363
Pantic M, Patras I (2006) Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. IEEE Trans Syst Man Cybern B 36(2):433–449
Pantic M, Pentland A, Nijholt A, Huang TS (2006) Human Computing and machine understanding of human behaviour: a Survey. In: Proceedings of the ACM international conference Multimodal Interfaces, Banff, Canada, pp 239–248
Pantic M, Rothkrantz LJM (2000) Automatic analysis of facial expressions – the state of the art. IEEE Trans Pattern Anal Mach Intell 22(12):1424–1445
Pantic M, Rothkrantz LJM (2003) Toward an affect-sensitive multimodal human-computer interaction. Proc IEEE 91(9):1370–1390
Pantic M, Rothkrantz LJM (2004a) Facial action recognition for facial expression analysis from static face images. IEEE Trans Syst Man Cybern B 34(3):1449–1461
Pantic M, Rothkrantz LJM (2004b) Case-based reasoning for user-profiled recognition of emotions from face images. In: Proceedings of the 2004 IEEE international conference on multimedia & expo, ICME 2004, IEEE, Taipei, Taiwan, June 27–30, pp 391–394
Pantic M, Rothkrantz LJM, Koppelaar H (1998) Automation of non-verbal communication of facial expressions. In: Proceedings of the conference on Euromedia '98, De Montfort University, Leicester, UK, January 5–7, pp 86–93
Pantic M, Sebe N, Cohn JF, Huang TS (2005) Affective multimodal human-computer interaction. In: Proceedings of the ACM international conference on Multimedia, Hilton, Singapore, pp 669–676
Pantic M, Valstar MF, Rademaker R, Maat L (2005) Web-based database for facial expression analysis. In: Proceedings of the IEEE international conference on Multimedia and Expo, pp 317–321. www.mmifacedb.com
Patras I, Pantic M (2004) Particle filtering with factorized likelihoods for tracking facial features. In: Sixth IEEE international conference on automatic face and gesture recognition (FGR 2004), IEEE Computer Society, Seoul, Korea, May 17–19, pp 97–102
Perrin S, Cassinelli A, Ishikawa M (2004) Gesture recognition using laser-based tracking system. In: Sixth IEEE international conference on automatic face and gesture recognition (FGR 2004), IEEE Computer Society, Seoul, Korea, May 17–19, pp 541–546
Russell JA (1994) Is there universal recognition of emotion from facial expression? Psychol Bull 115(1):102–141

Samal A, Iyengar PA (1992) Automatic recognition and analysis of human faces and facial expressions: A survey. Pattern Recogn 25(1):65–77
Starner T, Weaver J, Pentland A (1998) Real-time American sign language recognition using desk and wearable computer-based video. IEEE Trans Pattern Anal Mach Intell 20(12):1371–1375
Tian YL, Kanade T, Cohn JF (2001) Recognizing action units for facial expression analysis. IEEE Trans Pattern Anal Mach Intell 23(2):97–115
Tian YL, Kanade T, Cohn JF (2005) Facial expression analysis. In: Li SZ, Jain AK (eds) Handbook of face recognition. Springer, New York, NY, pp 247–276
Valstar MF, Pantic M (2006) Fully automatic facial action unit detection and temporal analysis. Proc IEEE Int Conf Comput Vis Pattern Recogn 3:149
Valstar MF, Pantic M, Ambadar Z, Cohn JF (2006) Spontaneous vs. posed facial behavior: automatic analysis of brow actions. In: Proceedings of the ACM international conference on Multimodal Interfaces, Banff, Alberta, Canada, pp 162–170
Valstar MF, Pantic M, Patras I (2004) Motion history for facial action detection from face video. Proc IEEE Int Conf Syst Man Cybern 1:635–640
Viola P, Jones M (2004) Robust real-time face detection. J Comput Vis 57(2):137–154
Vogler C, Metaxas D (1998) ªASL recognition based on a coupling between HMMs and 3D Motion Analysis.º In: Proceedings of the 6th IEEE international conference on Computer Vision (ICCV-98), Narosa Publishing House, Bombay, India, January 4–7, pp 363–369
Vukadinovic D, Pantic M (2005) Fully automatic facial feature point detection using Gabor feature based boosted classifiers. In: Proceedings of the IEEE international conference on Systems, Man and Cybernetics, The Big Island, Hawaii, pp 1692–1698
Wierzbicka A (1993) Reading human faces. Pragmat Cogn 1(1):1–23
Wilson I, Bobick A (1999) Parametric hidden Markov models for gesture recognition. IEEE Trans Pattern Anal Mach Intell 21(9):884–900
Wong S, Cipolla R (2006) Continuous gesture recognition using a sparse Bayesian classifier. In: Proceedings of the 18th international conference on Pattern Recognition – vol 01, 2006. ICPR, IEEE Computer Society, Washington, DC, pp 1084–1087
Yang MH, Ahuja N, Tabb M (2002) Extraction of 2d motion trajectories and its application to hand gesture recognition. IEEE Trans Pattern Anal Mach Intell 24(8):1061–1074
Yang MH, Kriegman DJ, Ahuja N (2002) Detecting faces in images: a survey. IEEE Trans Pattern Anal Mach Intell 24(1):34–58
Ying W, Huang TS (2001) Hand modeling, analysis and recognition. Signal Process Mag IEEE 18(3):51–60
Zhang Y, Ji Q (2005) Active and dynamic information fusion for facial expression understanding from image sequence. IEEE Trans Pattern Anal Mach Intell 27(5):699–714