IDIAP RESEARCH REPORT

Automatic Facial Expression Analysis: A Survey

Beat Fasel [a] and Juergen Luettin [b]

IDIAP–RR 99-19

November 2002

published in
Pattern Recognition, 36(1):259-275, 2003

Dalle Molle Institute for Perceptual Artificial Intelligence • P.O.Box 592 • Martigny • Valais • Switzerland

phone +41 – 27 – 721 77 11
fax +41 – 27 – 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

[a] IDIAP - Institut Dalle Molle d'Intelligence Artificielle Perceptive Rue du Simplon 4, CP592 - 1920 Martigny, Switzerland Beat.Fasel@idiap.ch
[b] Ascom Systec AG, Applicable Research & Technology, Gewerbepark CH-5506 Maegenwil, Switzerland, Juergen.Luettin@ascom.ch

IDIAP Research Report 99-19

Automatic Facial Expression Analysis: A Survey

Beat Fasel and Juergen Luettin

November 2002

published in
Pattern Recognition, 36(1):259-275, 2003

Abstract. Over the last decade, automatic facial expression analysis has become an active research area that finds potential applications in areas such as more engaging human-computer interfaces, talking heads, image retrieval and human emotion analysis. Facial expressions reflect not only emotions, but other mental activities, social interaction and physiological signals. In this survey we introduce the most prominent automatic facial expression analysis methods and systems presented in the literature. Facial motion and deformation extraction approaches as well as classification methods are discussed with respect to issues such as face normalization, facial expression dynamics and facial expression intensity, but also with regard to their robustness towards environmental changes.

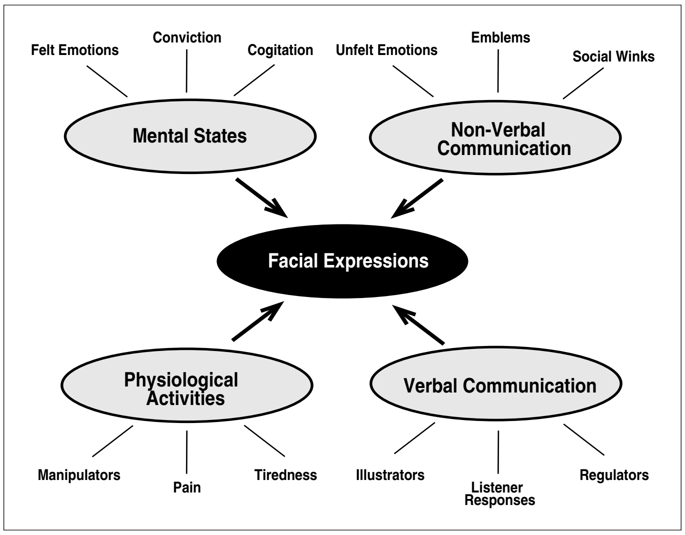

Figure 1: **Sources of Facial Expressions.**

1 Introduction

Facial expression analysis goes well back into the nineteenth century. Darwin [19] demonstrated already in 1872 the universality of facial expressions and their continuity in man and animals and claimed among other things, that there are specific inborn emotions, which originated in serviceable associated habits. In 1971, Ekman and Friesen [29] postulated six primary emotions that possess each a distinctive content together with a unique facial expression. These prototypic emotional displays are also referred to as so called *basic emotions*. They seem to be universal across human ethnicities and cultures and comprise happiness, sadness, fear, disgust, surprise and anger.

In the past, facial expression analysis was primarily a research subject for psychologists, but already in 1978, Suwa et al. [78] presented a preliminary investigation on automatic facial expression analysis from an image sequence. In the nineties, automatic facial expression analysis research gained much inertia starting with the pioneering work of Mase and Pentland [57]. The reasons for this renewed interest in facial expressions are multiple, but mainly due to advancements accomplished in related research areas such as face detection, face tracking and face recognition as well as the recent availability of relatively cheap computational power.

Various applications using automatic facial expression analysis can be envisaged in the near future, fostering further interest in doing research in different areas, including image understanding, psychological studies, facial nerve grading in medicine [23], face image compression and synthetic face animation [71], video-indexing, robotics as well as virtual reality.

Facial expression recognition should not be confused with human emotion recognition as is often done in the computer vision community. While *facial expression recognition* deals with the classification of facial motion and facial feature deformation into abstract classes that are purely based on visual information, human emotions are a result of many different factors and their state might or might not be revealed through a number of channels such as emotional voice, pose, gestures, gaze direction and facial expressions. Furthermore, emotions are not the only source of facial expressions, see Figure 1. In contrast to facial expression recognition, emotion recognition is an *interpretation attempt* and often demands understanding of a given situation, together with the availability of full contextual information.

2 Facial Expression Measurement

Facial expressions are generated by contractions of facial muscles, which results in temporally deformed facial features such as eye lids, eye brows, nose, lips and skin texture, often revealed by wrinkles and bulges. Typical changes of muscular activities are brief, lasting for a few seconds, but rarely more than five seconds or less than 250 ms. We would like to accurately measure facial expressions and therefore need a useful terminology for their description. Of importance is the *location* of facial actions, their *intensity* as well as their *dynamics*. Facial expression intensities may be measured by determining either geometric deformations of facial features or the density of wrinkles appearing in certain face regions. For example the degree of a smiling is communicated

by the magnitude of cheek and lip corner raising as well as wrinkle displays. Since there are inter-personal variations with regard to the amplitudes of facial actions, it is difficult to determine *absolute* facial expression intensities, without referring to the neutral face of a given subject. Note that measuring the intensity of spontaneous facial expressions is more difficult than measuring posed facial expressions, which are usually displayed with an exaggerated intensity and can thus be identified more realiably.
Not only the nature of the deformation of facial features conveys meaning, but also the relative timing of facial actions as well as their temporal evolution. Static images do not clearly reveal subtle changes in faces and it is therefore essential to measure also the dynamics of facial expressions. Although the importance of correct timing is widely accepted, only a few studies have investigated this aspect systematically, mostly for smiles [61]. Facial expressions can be described with the aid of three temporal parameters: *onset* (attack), *apex* (sustain), *offset* (relaxation). These can be obtained from human coders, but often lack precision. Few studies relate to the problem of automatically computing the onset and offset of facial expressions, especially when not relying on intruding approaches such as Facial Electromyography (EMG) [75].
There are two main methodological approaches of how to measure the afore mentioned three characteristics of facial expressions, namely *message judgment* based and *sign vehicle* based approaches [27]. The former directly associate specific facial patterns with mental activities, while the latter represent facial actions in a coded way, prior to eventual interpretation attempts.

2.1 Judgment-based Approaches

Judgment-based approaches are centered around the messages conveyed by facial expressions. When classifying facial expressions into a predefined number of emotion or mental activity categories, an agreement of a group of coders is taken as ground truth, usually by computing the average of the responses of either experts or non-experts. Most automatic facial expression analysis approaches found in the literature attempt to directly map facial expressions into one of the six basic emotion classes introduced by Ekman and Friesen [29].

2.2 Sign-based Approaches

With sign vehicle-based approaches, facial motion and deformation are coded into visual classes. Facial actions are hereby abstracted and described by their location and intensity. Hence, a complete description framework would ideally contain all possible perceptible changes that may occur on a face. This is the goal of FACS (Facial Action Coding System), which was developed by Ekman and Friesen [30] and has been considered as a foundation for describing facial expressions. It is appearance-based and thus does not convey any information about e.g. mental activities associated with expressions. FACS uses 44 AUs (Action Units) for the description of facial actions with regard to their location as well as their intensity, the latter either with three or five levels of magnitude. Individual expressions may be modeled by single action units or action unit combinations. Similar coding schemes are EMFACS (Emotional Facial Action Coding System) [36], MAX (Maximally Discriminative Facial Movement Coding System) [42] and AFFEX (System for Identifying Affect Expressions) [43]. However, they are only directed towards emotions. Finally, the MPEG-4 SNHC (MPEG-4: Object-based multimedia compression standard, SNHC: Synthetic Natural Hybrid Coding) [71] is a standard that encompasses analysis, coding [81] and animation of faces (talking heads) [39].
Instead of describing facial actions only with the aid of purely descriptive AUs, scores of sign-based approaches may be interpreted by employing *facial expression dictionaries*. Friesen and Ekman introduced such a dictionary for the FACS framework [37]. Ekman et al. [31] presented also a database called FACSAID (Facial Action Coding System Affect Interpretation Database), which allows to translate emotion related FACS scores into affective meanings. Emotion interpretations were provided by several experts, but only agreed affects were included in the database.

2.3 Reliability of Ground Truth Coding

The labeling of employed databases determines not only whether a given system attempts to recognize or interpret facial expressions, but may also influence the achievable recognition accuracy, especially when it comes to facial expression timing and intensity estimations. Furthermore, chosen classification schemes affect the design of facial expression classifiers, e.g. they have an influence on the number and nature of facial action categories that have to be treated. According to Ekman [28], there are several points that need to be addressed when measuring facial expressions: (a) a separate agreement index about the scoring of specific facial actions, as typically some actions are easier to recognize than others, (b) spontaneous rather than posed facial actions, (c) various subjects including infants, children, adults and aged populations, (d) limiting the disagreement in

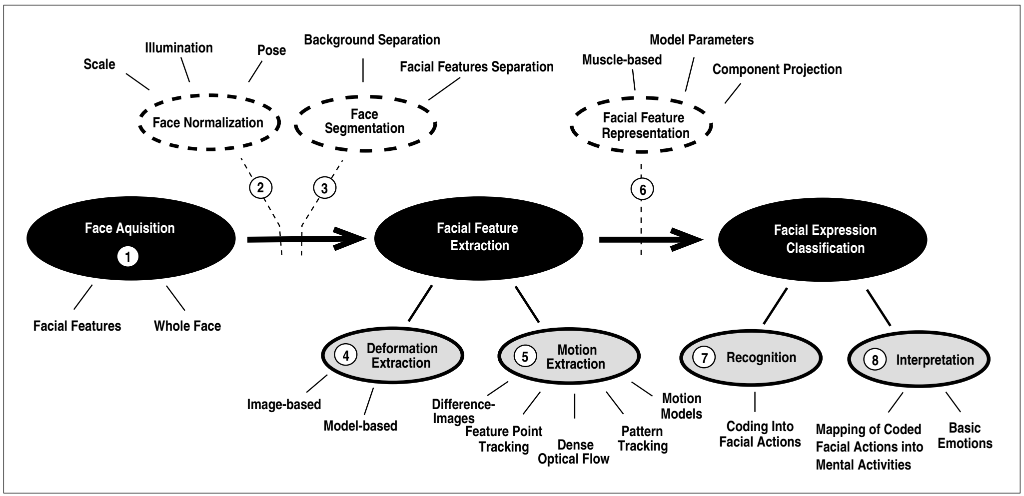

Figure 2: **Generic Facial Expression Analysis Framework.** The encircled numbers are used in the system diagrams presented in Section 4 and indicate relevant processing stages. Note that the face normalization, the face segmentation as well as the facial feature representation stages are only necessary in conjunction with some specific facial feature extraction and classification methods.

the judgment of facial actions by providing a minimal intensity threshold of facial actions, (e) inclusion of both expert and beginners for the measurement of facial actions and (f) the reliability should be reported not only for the type, but also the intensity and dynamics of facial actions. These points can probably be easier fulfilled with sign than with judgment-based approaches as the latter can only provided a limited labeling accuracy. For example, within a single basic emotion category, there is too much room for interpretation.
Cross-cultural studies have shown furthermore, that the judgment of facial expression is also culturally dependent and partially influenced by learned display rules [58]. Even though the afore mentioned basic emotions are universal across cultures, the assessment is hampered, if the encoder and decoder are of different cultures [59]. Sign-based coding schemes on the other hand increase objectivity, as coders are only required to record specific concerted facial components instead of performing facial expression interpretation. An advantage of sign-based methods is also the possibility of decomposing facial expression recognition and facial expression interpretation. Hence, the performance of the employed analysis methods may be evaluated directly with regard to their visual performance.

3 Automatic Facial Expression Analysis

Automatic facial expression analysis is a complex task as physiognomies of faces vary from one individual to another quite considerably due to different age, ethnicity, gender, facial hair, cosmetic products and occluding objects such as glasses and hair. Furthermore, faces appear disparate because of pose and lighting changes. Variations such as these have to be addressed at different stages of an automatic facial expression analysis system, see Figure 2. We have a closer look at the individual processing stages in the remainder of this chapter.

3.1 Face Acquisition

Ideally, a face acquisition stage features an automatic face detector that allows to locate faces in complex scenes with cluttered backgrounds. Certain face analysis methods need the exact position of the face in order to extract facial features of interest while others work, if only the coarse location of the face is available. This is the case with e.g. active appearance models [51]. Hong et al. Hong [40] used the PersonSpotter system by Steffens et al. [77] in order to perform realtime tracking of faces. The exact face dimensions were then obtained by fitting a labeled graph onto the bounding box containing the face previously detected by the PersonSpotter system. Essa and Pentland [33] located faces by using the view-based and modular eigenspace method of Pentland et al. [69]. Face analysis is complicated by face *appearance changes* caused by pose variations and illumination changes. It might therefore be a good idea to normalize acquired faces prior to their analysis:

- **Pose:** The appearance of facial expressions depends on the angle and distance at which a given face is being observed. Pose variations occur due to *scale changes* as well as *in-plane* and *out-of-plane rotations*

Deformation Extraction	Holistic Methods	Local Methods
Image-based	- Neural network [18][35][54]	- Intensity profiles [2]
	- Gabor wavelets [18][35]	- High gradient components [52]
		- PCA + Neural networks [66][17]
Model-based	- Active appearance model [51][16][24]	- Geometric face model [49]
	- Point distribution model [41]	- 2 view point-based models [68]
	- Labeled graphs [86][56][40]	
Motion Extraction	**Holistic Methods**	**Local Methods**
Dense Optical Flow	- Dense flow fields [52][2]	- Region-based flow [57][65][85]
Motion Models	- 3D motion models [33][25]	- Parametric motion models [9][84]
	- 3D deformable models [21]	- 3D motion models [5]
Feature Point Tracking		- Feature tracking [52][80][82][64][72]
Difference-Images	- Holistic diff.-imgs [2][22][35][34]	- Region-based difference-images [12]
Marker-based		- Highlighted facial features [4]
		- Dot markers [78][44]

Table 1: **Facial Feature Extraction Methods:** Overview of prominent deformation and motion extraction methods, which were used for the task of facial expression analysis.

of faces. Especially out-of-plane rotated faces are difficult to handle, as perceived facial expression are distorted in comparison to frontal face displays or may even become partly invisible. Limited out-of-plane rotations can be addressed by warping techniques, where the center positions of distinctive facial features such as the eyes, nose and mouth serve as reference points in order to normalize test faces according to some generic face models e.g. see [33]. Scale-changes of faces may be tackled by scanning images at several resolutions in order to determine the size of present faces, which can then be normalized accordingly [33][73].

- **Illumination:** A common approach for reducing lighting variations is to filter the input image with Gabor wavelets, e.g. see [35]. Difficult to solve is the problem of partly lightened faces, which has been addressed for the task of face recognition by Belhumeur et al. [6], but not yet sufficiently for facial expression analysis. Finally, specular reflections on eyes, teeth and wet skin may be encountered by using brightness models [8].

Note that even though *face normalization* may be a reasonable approach in conjunction with some face analysis approaches, it is *not mandatory*, as long as extracted feature parameters are normalized prior to their classification. Indeed, appearance-based model [51] and local motion model [9] approaches have dealt with significant out-of-plane rotations without relying on face normalization.

3.2 Facial Feature Extraction and Representation

Feature extraction methods can be categorized according to whether they focus on motion or deformation of faces and facial features, respectively whether they act locally or holistically. Table 1 gives an overview of methods that were employed by the computer vision community for the task of facial expression analysis.

3.2.1 Local versus Holistic Approaches

Facial feature processing may take place either holistically, where the face is processed as a whole, or locally, by focusing on facial features or areas that are prone to change with facial expressions. We can distinguish two types of facial features:

- **Intransient facial features** are always present in the face, but may be deformed due to facial expressions. Among these, the eyelids, eyebrows and the mouth are mainly involved in facial expression displays. Tissue texture, facial hair as well as permanent furrows constitute other types of intransient facial features that influence the appearance of facial expressions.

- **Transient facial features** encompass different kind of wrinkles and bulges that occur with facial expressions. Especially the forefront and the regions surrounding the mouth and the eyes are prone to contain transient facial features. Opening and closing of eyes and the mouth may furthermore lead to iconic changes [8], local changes of texture that cannot be predicted from antecedent frames.

Face segmentation allows to isolate transient and intransient features within faces or can be used to separate faces of interest from the background. Segmentation boundaries were often determined heuristically, where guidance was provided by a priori knowledge of human observers. Note that holistic feature extraction methods are good at determining prevalent facial expressions, whereas local methods are able to detect subtle changes in small areas. The latter are suitable especially with rule-based interpretation attempts, e.g. see [68].

3.2.2 Deformation versus Motion-based Approaches

Motion extraction approaches directly focus on facial changes occurring due to facial expressions, whereas deformation-based methods do have to rely on neutral face images or face models in order to extract facial features that are relevant to facial actions and not caused by e.g. intransient wrinkles due to old age. In contrast to motion-based approaches, deformation-based methods can be applied to both single images and image sequences, in the latter case by processing frames independently from each other. However, deformation-based feature extractors miss low-level directional flow information, i.e. they cannot reconstruct pixel motion. Nonetheless, high-level motion of intransient facial features may be inferred by using e.g. face and facial feature models that allow to estimate possible flow directions.

3.2.3 Image versus Model-based Approaches

Image-based methods extract features from images without relying on extensive knowledge about the object of interest. They have the advantage of being typically fast and simple. However, image-based approaches can become unreliable and unwieldy, when there are many different views of the same object that must be considered.
The facial structure can also be described with the aid of 2D or 3D face models. The former allow to model facial features and faces based on their appearance, without attempting to recover the volumetric geometry of the scene, e.g. see [51]. There are two types of 3D models, namely muscle and motion models [32][33]. The latter can significantly improve the precision of motion estimations, since only physically possible motion is considered. However, 3D models often require complex mapping procedures that generate heavy computational requirements. In addition, accurate head and face models have to be constructed manually, which is a tedious undertaking.

3.2.4 Appearance versus Muscle-based Approaches

In contrast to appearance-based image and 2D model approaches, where processing focuses on the effects of facial muscle activities, muscle-based frameworks attempt to interfere muscle activities from visual information. This may be achieved e.g. by using 3D muscle models that allow to map extracted optical flow into muscle actions [32][46]. Modeled facial motion can hereby be restricted to muscle activations that are allowed by the muscle framework, giving control over possible muscle contractions, relaxation and orientation properties. However, the musculature of the face is complex, 3D information is not readily present and muscle motion is not directly observable. For example, there are at least thirteen groups of muscles involved in the lip movements alone [38]. Mase and Pentland [57] did not use complex 3D models to determine muscle activities. Instead they translated 2D motion in predefined windows directly into a coarse estimate of muscle activity. Muscle-based approaches are not only well suited for the recognition of facial expression, but are also used to animate synthetic faces.

3.3 Deformation Extraction

Deformation of facial features are characterized by shape and texture changes and lead to high spatial gradients that are good indicators for facial actions and may be analyzed either in the image or the spatial frequency domain. The latter can be computed by high-pass gradient or Gabor wavelet based filters, which closely model the receptive field properties of cells in the primary visual cortex [70][20]. They allow to detect line endings and edge borders over multiple scales and with different orientations. These features reveal a lot about facial expressions, as both transient and intransient facial features often give raise to a contrast change with regard to the ambient facial tissue.
As we have mentioned before, Gabor filters remove also most of the variability in images that occur due to lighting changes. They have shown to perform well for the task of facial expression analysis and were used in image-based approaches [2][35][18] as well as in combination with labeled graphs [86][56][40].
For an illustration of Gabor filters applied to face images see Figure 3.

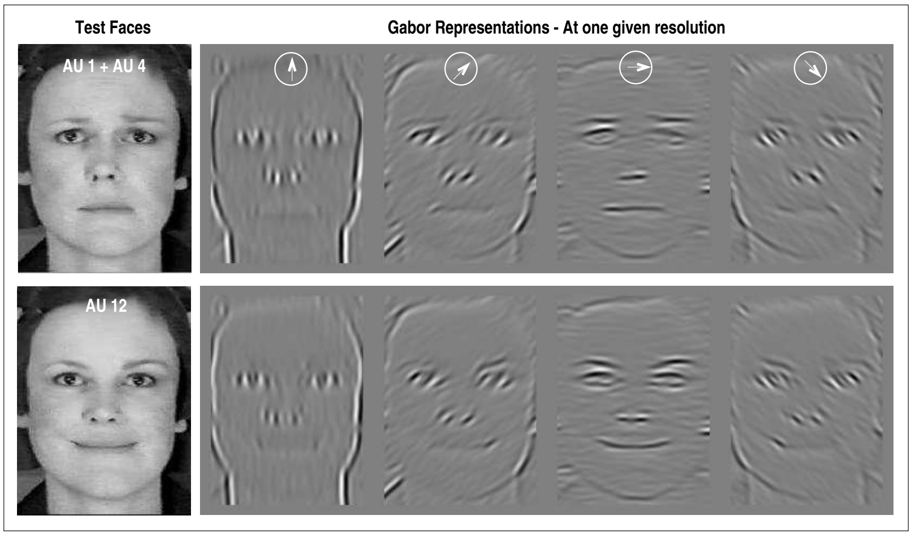

Figure 3: **Facial Feature Extraction using Gabor Wavelets:** Shown are two distinct facial expression displays on the left-hand side with the corresponding Gabor representations on the right-hand side. The latter were obtained by convoluting the face images on the left-hand side with four differently oriented wavelet kernels (at one given resolution).

We can distinguish holistic and local image-based deformation extraction approaches:

- **Holistic Image-based Approaches:** Several authors have taken either whole faces as features [35][54][18] or Gabor wavelet filtered whole-faces [35][18]. The main emphasis is hereby put on the classifier, which has to deal not only with face physiognomies, but - in the case of image domain based face processing - also with lighting variations. Common for most holistic face analysis approaches is the need of a thorough face-background separation in order to prevent disturbance caused by clutter.

- **Local Image-based Approaches:** Padgett and Cottrell [66] as well as Cottrell and Metcalfe [17] extracted facial expressions from windows placed around intransient facial feature regions (both eyes and mouth) and employed local Principal Component Analysis (PCA) for representation purposes. Local transient facial features such as wrinkles can be measured by using image intensity profiles along segments [2] or by determining the density of high gradient components over windows of interest [52].

Model-based approaches constitute an alternative to image-based deformation extraction. *Appearance-based model* approaches allow to separate fairly well different information sources such as facial illumination and deformation changes. Lanitis et al. [51] interpreted face images by employing Active Appearance Models (AAM) [16][24]. Faces were analyzed by a dual approach, using both shape and texture models. Active Shape Models (ASM) allow to simultaneously determine shape, scale and pose by fitting an appropriate Point Distribution Model (PDM) to the object of interest, see Figure 4.
A drawback of appearance-based models is the manual labor necessary for the construction of the shape models. The latter are based on landmark points that need to be precisely placed around intransient facial features during the training of the models. Huang and Huang [41] used a PDM to represent the shape of a face, where shape parameters were estimated by employing a gradient-based method.
Another type of holistic face models constitute the so called *labeled graphs*, which are comprised of sparsely distributed fiducial feature points [86][56][40]. The nodes of these feature graphs consist of *Gabor jets*, where each component of a jet is a filter response of a specific Gabor wavelet extracted at a given image point. A labeled graph is matched to a test face by varying its scale and position. The obtained graph can then be compared to reference graphs in order to determine the facial expression display at hand. Kobayashi and Hara [49] used a geometric face model consisting of 30 Facial Characteristic Points (FCP). They measured the intensity distribution along 13 vertical FCPs crossing facial lines with the aid of a neural network.
Finally, Pantic and Rothkrantz [68] used a 2D point-based model composed of both frontal and a side views. Multiple feature detectors were applied redundantly in order to localize contours of prominent facial features prior to their modeling.

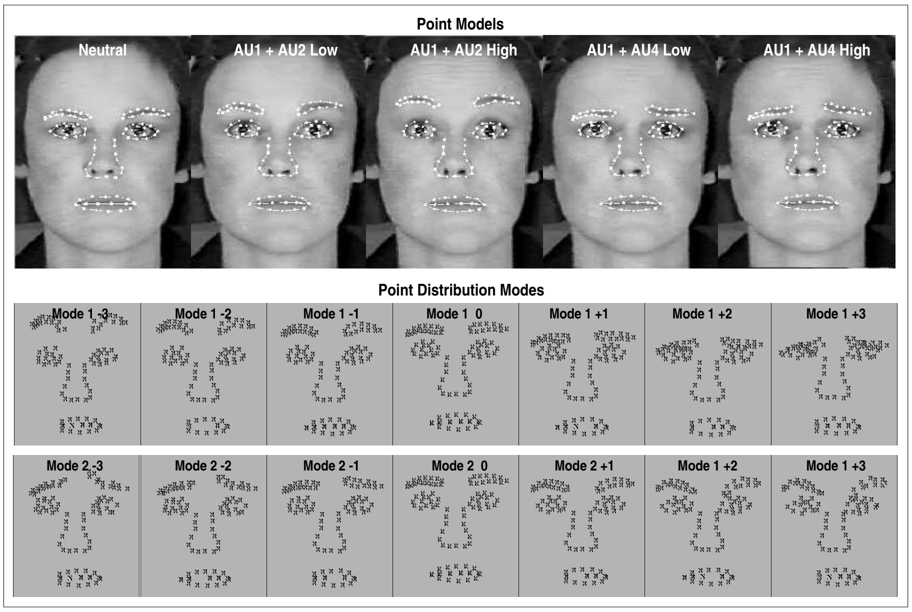

Figure 4: **Facial Feature Representation using Active Shape Models:** The first row shows manually placed Point Models (PM) that were employed to create a Point Distribution Model, represented by a few discrete instances of two point distribution modes shown in row two (mode 1) and three (mode 2), with intensities ranging from -3 to +3. The point distribution modes were computed using the active shape model toolbox implemented by Matthews [60].

3.4 Motion Extraction

Among the motion extraction methods that have been used for the task of facial expression analysis we find *dense optical flow, feature point tracking* and *difference-images*. Dense optical flow has been applied both locally and holistically:

- **Holistic Dense Optical Flow** approaches allow for whole-face analysis and were employed e.g. in [52][2]. Lien et al. [52] analyzed holistic face motion with the aid of wavelet-based, multi-resolution dense optical flow. For a compacter representation of the resulting flow fields they computed PCA-based eigenflows both in horizontal and vertical directions. Figure 5 shows sample dense optical flow fields, computed from two facial expression sequences.

- **Local Dense Optical Flow:** Region-based dense optical flow was used by Mase and Pentland [57] in order to estimate the activity of 12 of the totally 44 facial muscles. For each muscle, a window in the face image was defined as well as an axis along which each muscle expands and contracts. Dense optical flow motion was quantified into eight directions and allowed for a coarse estimation of muscle activity.
  Otsuka and Ohya [65] estimated facial motion in local regions surrounding the eyes and the mouth. Feature vectors were obtained by taking 2D Fourier transforms of the vertical and horizontal optical flow fields.
  Yoneyama et al. [85] divided normalized test faces into 8×10 regions, where local dense optical flow was computed and quantified region-wise into ternary feature vectors $(+1/0/-1)$, indicating upwards, none and downwards movements, while neglecting horizontal facial movements.

Different optical flow algorithms have been applied to facial motion analysis. For instance, Lien et al. [52] employed Wu's [83] approach of optical flow to estimate facial motion by using scaling functions and wavelets from Cai and Wang [10] to capture both local and global facial characteristics. Essa et al. [33] used Simoncelli's [76] coarse-to-fine optical flow, while Yacoob and Davis [84] as well as Rosenblum [72] employed Abdek-Mottaleb's [1] optical flow.
Apart from a certain vulnerability to image noise and non-uniform lighting, holistic dense optical flow methods often result in prodigious computational requirements and tend to be sensitive to motion discontinuities (iconic changes) as well as non-rigid motion. Optical flow analysis can also be done in conjunction with motion models that allow for increased stability and better interpretation of extracted facial motion, e.g. muscle activations:

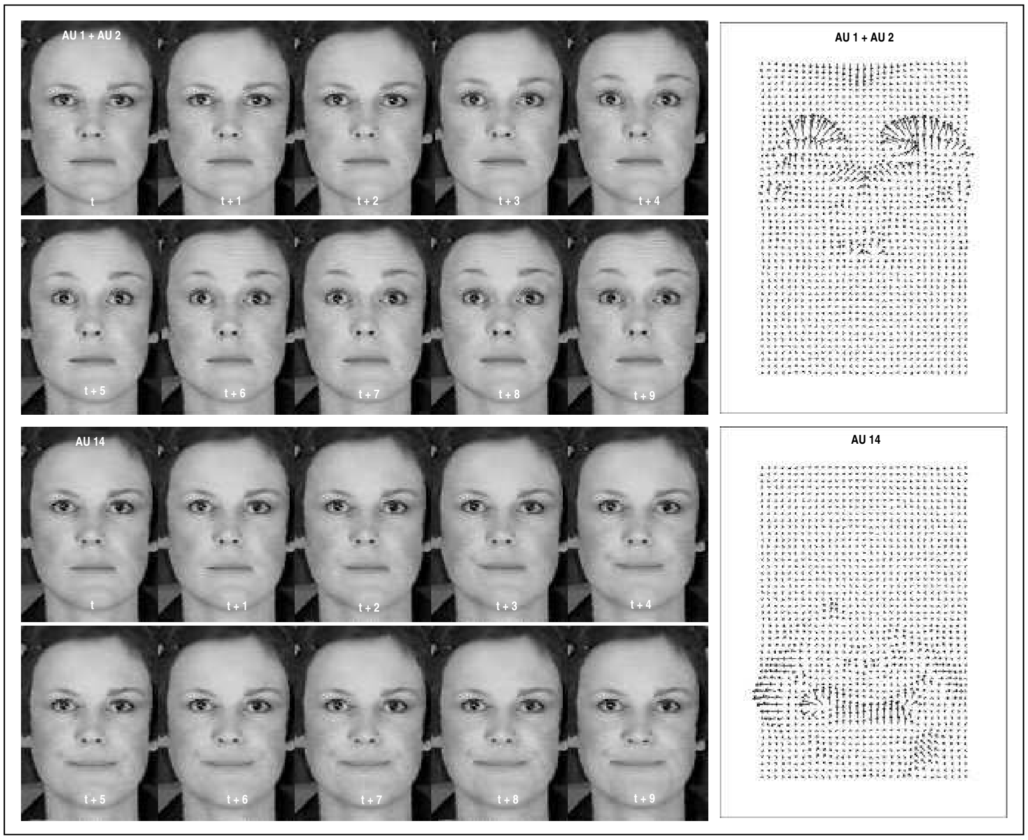

Figure 5: **Facial Motion Extraction using Dense Optical Flow:** Shown are two sample facial expression sequences on the left-hand side and the corresponding optical flow images on the right-hand side, which were computed with Nagel's algorithm [62]. Note the asymmetric facial action display in the lower facial expression sequence.

- **Holistic Motion Models:** Terzopoulos and Waters [79] have used eleven principal deformable contours (also known as “snakes”) to track lip and facial features throughout image sequences with the aid of a force field, which is computed from gradients found in the images. Only frontal faces were allowed and some facial make-up was used to enhance contrast.
  Essa and Pentland [33] employed sophisticated 3D motion and muscle models for facial expression recognition and increased tracking stability by Kalman filtering.
  DeCarlo and Metaxas [21] presented a formal methodology for the integration of optical flow and 3D deformable models and applied it to human face shapes and facial motion estimation. A relatively small number of parameters were used to describe a rich variety of face shapes and facial expressions.
  Eisert and Girod [25] employed 3D face models to specify shape, texture and motion. These models were also used to describe facial expressions caused by speech and were parameterized by FAPs of the MPEG-4 coding scheme.

- **Local Motion Models:** Black and Yacoob [9] as well as Yacoob and Davis [84] introduced *local parametric motion models* that allow, within local regions in space and time, to not only accurately model non-rigid facial motions, but to provide also a concise description of the motion associated with the edges of the mouth, nose, eyelids and eyebrows in terms of a small number of parameters. However, the employed motion models focus on the main intransient facial features involved with facial expressions (eyes, eye-brows and mouth) and the analysis of transient facial features, occurring in residual facial areas, was not considered. Last but not least, Basu et al. [5] presented a convincing approach of how to track human lip motions by using 3D models.

In addition to low-level dense optical flow, there are also higher level variants that focus either on the movement of generic features points, patterns or markers:

- **Feature Point Tracking:** Here, motion estimates are obtained only for a selected set of prominent facial features [52][80][82]. In order to reduce the risk of tracking loss, feature points are placed into areas of high contrast, preferably around intransient facial features as is illustrated on the right hand side of Figure 6. Hence, facial movement and deformation can be measured by tracking the displacement of corresponding feature points. Motion analysis is directed towards objects of interest and therefore does not have to be computed for extraneous background patterns. However, as facial motion is extracted only at selected feature point locations, other facial activities are ignored altogether.
  The automatic initialization of feature points is difficult and was often done manually. Otsuka and Ohya [64] presented a feature point tracking approach, where feature points are not selected by human expertise, but chosen automatically in the first frame of a given facial expression sequence. This is achieved by acquiring potential facial feature points from local extrema or saddle points of luminance distributions. Tian et al. [80] used different component models for the lips, eyes, brows as well as cheeks and employed feature point tracking to adapt the contours of these models according to the deformation of the underlying facial features. Finally, Rosenblum et al. [72] tracked rectangular, facial feature enclosing regions of interest with the aid of feature points.

- **Marker Tracking:** It is possible to determine facial actions with more reliability than with previously discussed methods, namely by measuring deformation in areas, where underlying muscles interact. Unfortunately, these are mostly skin regions with relatively poor texture. Highlighting is necessary and can be done by either applying color to salient facial features and skin [4] or by affixing colored plastic dots to predefined locations on the subject’s face, see the illustration on left hand side of Figure 6. Markers allow to render tissue motion visible and were employed in [44] and [78].

Note that even though the tracking of feature points or markers allows to extract motion, often only relative feature point locations, i.e. deformation information was used for the analysis of facial expressions, e.g. in [80] or [44].
Yet another way of how to extract image motion are **difference-images:** Specifically for facial expression analysis, difference-images are mostly created by subtracting a given facial image from a previously registered reference image, containing a neutral face of the same subject. However, in comparison to optical flow approaches, no flow direction can be extracted, but only differences of image intensities. In addition, accurate face normalization procedures are necessary in order to align reference faces onto the test faces. Holistic difference-image based motion extraction was employed in [2][22][35][34].
Choudhury and Pentland [12] used motion field histograms for the modeling of eye and eye brow actions. Motion was also extracted by difference-images, but taken from consecutive image frames and further processed

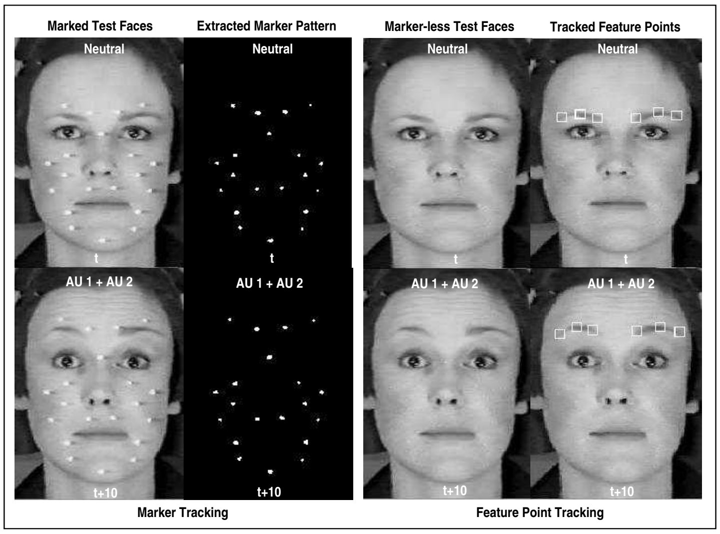

Figure 6: **Marker versus Feature Point Tracking:** On the left-hand side are shown two faces with affixed markers. The corresponding extracted marker patterns is depicted in the next column. Scale-normalized distances between marker points allow to determine underlying muscle activities [44]. A marker-less feature point tracking approach is shown on the right hand side. Here, six feature points (10×10 pixel windows) were used to determine the displacement of the eyebrows.

by using local receptive field histograms [74] in order to increase robustness with regard to rotation, translation and scale changes.

3.5 Classification

Feature classification is performed in the last stage of an automatic facial expression analysis system. This can be achieved by either attempting facial expression recognition using sign-based facial action coding schemes or by interpretation in combination with judgment or sign/dictionary based frameworks. We can distinguish spatial and spatio-temporal classifier approaches:

- **Spatio-Temporal Approaches:** *Hidden Markov Models (HMM)* are commonly used in the field of speech recognition, but are also useful for facial expression analysis as they allow to model the dynamics of facial actions. Several HMM-based classification approaches can be found in the literature [14][65][63] and were mostly employed in conjunction with image motion extraction methods.
  Recurrent neural networks constitute an alternative to HMMs and were also used for the task of facial expression classification [48][72]. Another way of taking temporal evolution of facial expression into account are so called spatio-temporal motion-energy templates. Here, facial motion is represented in terms of 2D motion fields. The Euclidean distance between two templates can then be used to estimate the prevalent facial expression [33].

- **Spatial Approaches:** *Neural networks* were often used for facial expression classification [66][2][86][87] [54][49][85] and were either applied directly on face images [35][54] or combined with facial features extraction and representation methods such as PCA (Principal Component Analysis), ICA (Independent Component Analysis) or Gabor wavelet filters [18][35]. The former are unsupervised statistical analysis methods that allow for a considerable dimensionality reduction, which both simplifies and enhances subsequent classification. These methods have been employed both in a holistic manner [2][22][34] as well as locally, using mosaic-like patches extracted from small facial regions [66][18][22][2].
  Dailey and Cottrell [18] applied both local PCA and Gabor jets for the task of facial expression recognition and obtained quantitatively indistinguishable results for both representations. Figure 7 shows an illustration of PCA and ICA components obtained from facial expression images.
  Unfortunately, neural networks are difficult to train if used for the classification of not only basic emotions, but unconstrained facial expressions. A problem is the great number of possible facial action combinations, about 7000 AU combinations have been identified within the FACS framework [28]. An

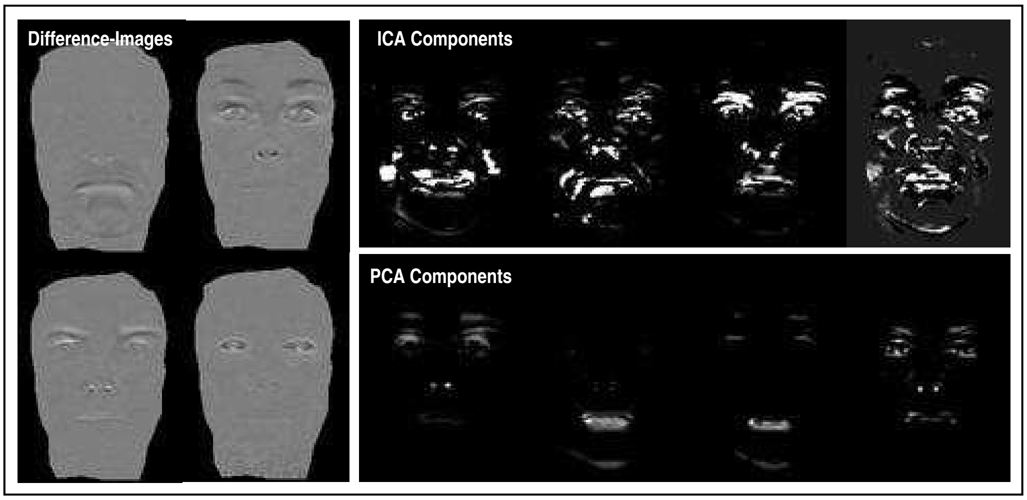

Figure 7: **Facial Feature Representation using Data-Driven Methods:** Sample difference-images are shown in the first row, corresponding holistic ICA components in the second and PCA components in the third row. The difference-images where computed by subtracting a neutral face reference image from face images displaying facial actions.

alternative to classically trained neural networks constitute compiled, rule-based neural networks that were employed e.g. in [44].

3.5.1 Facial Expression Recognition

Traditional approaches for modeling characteristics of facial motion and deformation have relied on hand-crafted rules and symbolic mid-level representations for emotional states, which have been introduced by computer scientists in the course of their investigations on facial expressions [57][84][9]. Human expertise is necessary to map these symbolic representations into e.g. emotions. However, facial signals consist of numerous distinct expressions, each with specific facial action intensity evolutions. In addition, individual realizations of facial expressions often differ only in subtle ways. This makes the task of manually creating facial expression classes rather difficult. Therefore, another group of researchers have relied on facial expression coding schemes such as MPEG-4 [25][46] or FACS [14][53][2][22][68][64][34][80]. Essa and Pentland [33] proposed an extension to FACS called FACS+, which consist of a set of control parameters using vision-based observations. In contrast to FACS, FACS+ describes also the dynamics of facial expressions.

3.5.2 Facial Expression Interpretation

Many automatic facial expression analysis systems found in the literature attempt to directly interpret observed facial expressions and mostly in terms of basic emotions [67][40][56][33][87][47][86][51][41][82][49]. Only a few systems use rules or facial expression dictionaries in order to translate coded facial actions into emotion categories [9][68]. The latter approaches have not only the advantage of accurately describing facial expressions without resorting to interpretation, but also allow to animate synthetic faces, e.g. within the FACS coding framework [39]. This is of interest, as animated synthetic faces make a direct inspection of automatically recognized facial expressions possible. See also [55] for an introduction to automatic facial expression interpretation.

4 Representative Facial Expression Recognition Systems

In this section we have a closer look at a few representative facial expression analysis systems. First, we discuss *deformation* and *motion-based* feature extraction systems. Then we introduce *hybrid* facial expression analysis systems, which employ several image analysis methods that complete each other and thus allow for a better overall performance. *Multi-modal frameworks* on the other hand integrate other non-verbal communication channels for improved facial expression interpretation results. Finally, *unified frameworks* focus on multiple facial characteristics, allowing for synergy effects between different modalities.

4.1 Deformation Extraction based Systems

Padgett and Cottrell [67] presented an automatic facial expression interpretation system that was capable of identifying six basic emotions. Facial data was extracted from 32×32 pixel blocks that were placed on the eyes as well as the mouth and projected onto the top 15 PCA eigenvectors of 900 random patches, which were extracted from training images. For classification, the normalized projections were fed into an ensemble of 11 neural networks. Their output was summed and normalized again by dividing the average outputs for each possible emotion across all networks by their respective deviation over the entire training set. The largest score for a particular input was considered to be the emotion found by the ensemble of networks. Altogether 97 images of 6 emotions from 6 males and 6 females were analyzed and a 86% generalization performance was measured on novel face images.
Lyons et al. [56] presented a Gabor wavelet based facial expression analysis framework, featuring a node grid of Gabor jets, similar to what was used by the Von der Malsburg group for the task of face recognition [50]. Hereby, each test image was convolved with a set of Gabor filters, whose responses are highly correlated and redundant at neighboring pixels. Therefore, it was only necessary to acquire samples at specific points on a sparse grid covering the face. The projections of the filter responses along discriminant vectors, calculated from the training set, were compared at corresponding spatial frequency, orientation and locations of two face images, where the normalized dot product was used to measure the similarity of two Gabor response vectors. Lyons et al. placed graphs manually onto the faces in order to obtain a better precision for the task of facial expression recognition. Experiments were carried out on subsets of totally 6 different posed expressions and neutral faces of 9 Japanese female undergraduates. A generalization rate of 92% was obtained for the recognition of new expressions of known subjects and 75% for the recognition of facial expressions of novel expressers.

4.2 Motion Extraction based Systems

Black and Yacoob [9] analyzed facial expressions with parameterized models for the mouth, the eyes and the eye brows and represented image flow with low-order polynomials [7]. A concise description of facial motion was achieved with the aid of a small number of parameters from which they derived mid- and high-level description of facial actions. The latter considered also temporal consistency of the mid-level predicates in order to minimize the effects of noise and inaccuracies with regard to the motion and deformation of the models. Hence, each facial expression was modeled by registering the intensities of the mid-level parameters within temporal segments (beginning, apex, ending).
Extensive experiments were carried out on 40 subjects in the laboratory with a 95-100% correct recognition rate and also with television and movie sequences resulting in a 60-100% correct recognition rate. Black and Yacoob proved that a recognition of basic emotions was possible also in presence of significant pose variations and head motion.
Essa and Pentland [33] presented a rather complete computer vision system featuring both automatic face detection and face analysis. Facial motion was extracted with the aid of holistic dense optical flow and coupled with 3D motion and muscle based face models. The latter allowed to describe the facial structure including facial tissue as well as muscle actuators and their force-based deformation. Essa and Pentland located test faces automatically by using a view-based and modular eigenspace method and determined also the position of facial features. The latter were then employed in order to warp face images to match canonical face meshes, which in turn allowed to extract additional feature points corresponding to fixed nodes on the meshes. After the initial model to image registration, Simoncelli's [76] coarse-to-fine optical flow was used to compute image motion. In addition, a Kalman filter based control framework was applied in order to prevent chaotic responses of the physical system. The employed dynamic face model allowed not only to extract muscle actuations of observed facial expressions, but it was also possible to produce noise corrected 2D motion fields via the control-theoretic approach. The latter where then classified with motion energy templates in order to extract facial actions. Experiments were carried out on 52 frontal view image sequences with a correct recognition rate of 98% for both the muscle and the 2D motion energy models.

4.3 Hybrid Systems

Hybrid facial expression analysis systems combine several facial expression analysis methods. This is most beneficial, if the individual estimators produce very different error patterns. Bartlett et al. [3] proposed a system that integrates holistic difference-image based motion extraction coupled with PCA, feature measurements along predefined intensity profiles for the estimation of wrinkles and holistic dense optical flow for whole-face

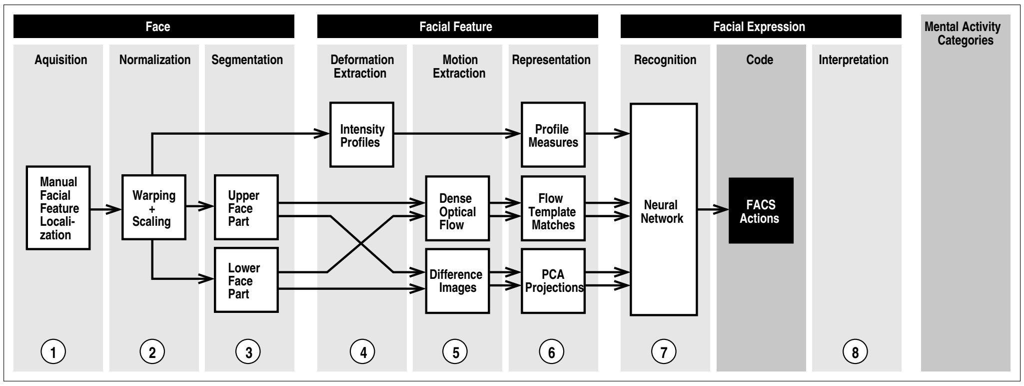

Figure 8: **Hybrid facial expression analysis system** proposed by Bartlett et al. [3]. The three analysis methods employed showed to produce different error patterns and thus allow for an improved recognition performance if combined.

motion extraction, see Figure 8. These three methods were compared with regard to their contribution to the task of facial expressions recognition. Bartlett et al. estimated that without feature measurement, there would be a 40% decrease of the facial expression recognition improvement gained by all methods combined. Faces were normalized by alignment through scaling, rotation and warping of aspect ratios. However, eye and mouth centers were located manually in the neutral face frame, each test sequence had to start with.
Facial expression recognition was achieved with the aid of a feed-forward neural network, made up of 10 hidden and 6 output units. The input of the neural network consisted of 50 PCA component projections, 5 feature density measurements and 6 optical flow based template matches. A Winner Takes it All (WTA) judgment approach was chosen to select the final AU candidates.
Initially, Bartlett et al.'s hybrid facial expression analysis system was able to classify 6 upper FACS action units on a database containing 20 subjects, correctly recognizing 92% of the AU activations, but no AU intensities. Later it was extended to allow also for the classification of lower FACS action units and achieved a 96% accuracy for 12 lower and upper face actions [2][22].
Cohn et al. [14] and Lien et al. [53] introduced systems that were based on holistic image flow analysis, feature point tracking and high-gradient component analysis methods, which were integrated into a spatio-temporal framework and combined with an HMM recognition stage, see Figure 9. Local face motion was estimated by feature point tracking. Hereby, a pyramidal multi-scale search approach was employed that is sensitive to subtle feature motion and also allowed to track large displacements of feature motion with sub-pixel accuracy. Holistic facial motion on the other hand was estimated by employing Wu's multi-resolution wavelet-based optical flow [83]. Forehead, cheek and chin regions were analyzed for transient facial features by using high-gradient component analysis based on horizontal, vertical and diagonal line and edge detectors in the spatial domain as well as frame comparisons in the temporal domain. The latter allowed to separate transient from intransient facial features and hair occlusion. Face tracking and face alignment were manually initialized by selecting three facial feature points in the first frame of each test image sequence. Lien et al.'s system was trained to analyze both the activity and intensity of 15 FACS AUs situated in the brow, eye and mouth regions. The holistic dense optical flow approach gave the best average AU recognition rates, followed by feature point tracking and the high gradient component analysis approach, see also Table 2. Lien et al.'s facial expression analysis system performed well even with difficult sequences such as those containing baby faces. The latter differ from adult faces both in morphology and tissue texture. Unfortunately, heavy computational requirements arised with the use of optical flow.

4.4 Unified Frameworks

Facial expression recognition may be improved by considering not only facial actions but also face characteristics such as identity, gender, age and ethnicity.
Lanitis et al. [51] proposed a unified framework that employs 2D active shape models which yield, once aligned to a test face, face appearance parameters from which it is possible to estimate 3D pose, identity, gender and facial expressions. During training, shape models are derived from a set of images by statistical analysis of the landmark point positions. These represent main facial features and are placed manually on training images prior to the shape model creation process, see Figure 10. During testing, facial features are located in test

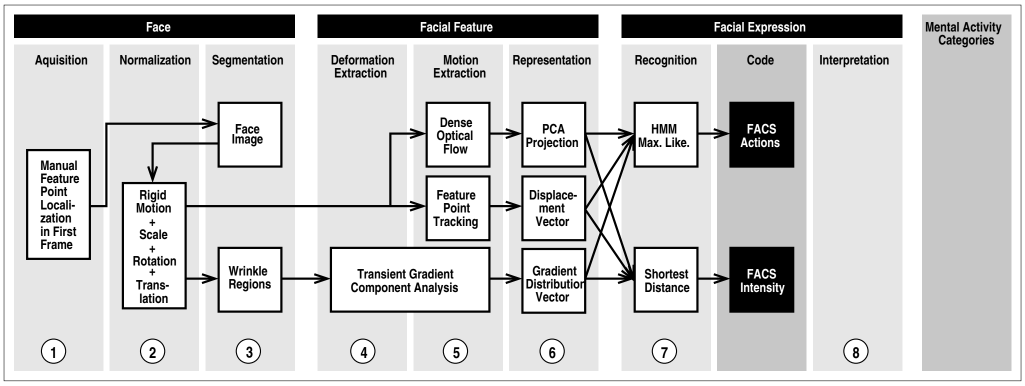

Figure 9: **Hybrid facial expression analysis system** proposed by Lien et al. [53]. The whole facial expression analysis framework is situated in the spatio-temporal domain, including the classification stage that is driven by Hidden Markov Models (HMM).

Authors	**# Subjects** Test	Train	**# Faces** Test	Train	**# FACS AUs**	**Extraction Methods**	**Classification Methods**	**Rec. Rate**
Bartlett [2]	4-20	4-20	111*5	111*5	$\mathbf{6}^a+\mathbf{6}^b$	Diff-Img.+G. Jets	Near. Neigh.	**96%**
	4-20	4-20	111*5	111*5	$\mathbf{6}^a+\mathbf{6}^b$	Diff. Img.	ICA+N. Neigh.	**96%**
	4-20	4-20	111*5	111*5	$\mathbf{6}^a+\mathbf{6}^b$	Optical Flow	Motion Templ.	**86%**
Fasel [34]	1	1	45	182	$\mathbf{9}^c$	Diff.-Img.	ICA+Eucl. Dist.	**83/41%**
Cohn [14]	30	N/A	N/A	N/A	**15**	Feat. Point Track.	HMM	**86%**
Lien [53]	N/A	N/A	75*20	60*20	$\mathbf{3}^a$	Feat. Point Track.	HMM	**85%**
	N/A	N/A	150*20	120*20	$\mathbf{6}^b$	Feat. Point Track.	HMM	**88%**
	N/A	N/A	75*20	44*20	$\mathbf{3}^a$	Optical Flow	PCA+HMM	**93%**
	N/A	N/A	160*20	100*20	$\mathbf{4}^a$	High Grad. Comp.	HMM	**85%**
	N/A	N/A	80*20	50*20	$\mathbf{2}^b$	High Grad. Comp.	HMM	**81%**
Pantic [68]	8	N/A	496	N/A	**31**	Multi Feat. Detect.	Expert Rules	**89%**

Table 2: **Selected Facial Expression Recognition Systems:** Classification of Facial Actions using FACS. ^{a}Upper-Face FACS coding. ^{b}Lower-Face FACS coding. ^{c}Nine asymmetric FACS classes, each with 5 intensity levels on 2 face sides.

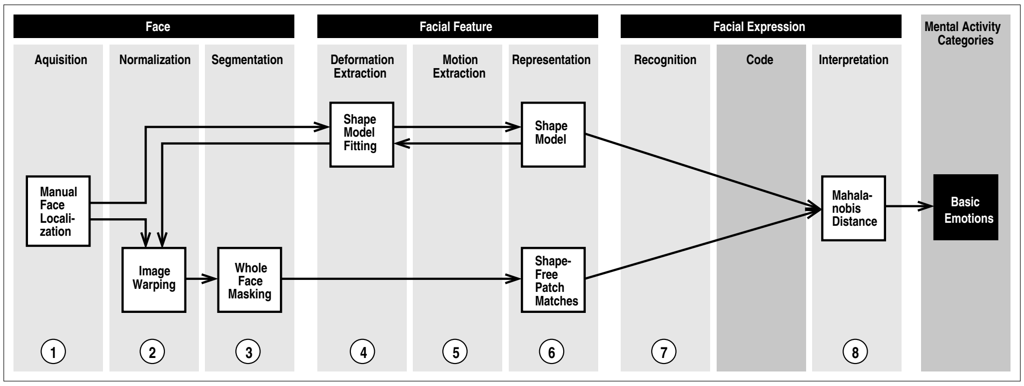

Figure 10: **Unified facial expression analysis framework** proposed by Lanitis et al. [51]. It is based on active shape models (ASM). Shape information is also used to extract shape-free whole-face patches that represent texture information.

image using ASM search [15], guided by the flexible shape models obtained during training. Gray-level profile information is collected at each model point and used to determine the best fitting shape model. Lanitis et al. then deformed test faces to the mean face shape by using the previously registered shape information in order to extract holistic shape-less patches, which account for facial texture.

Finally, Hong et al. [40] proposed an online facial expression recognition system, which is based on personalized galleries and uses identity information in conjunction with facial expression analysis. Faces of interest are detected and tracked in live video sequences with the aid of the PersonSpotter system [77] and recognition of facial expressions is achieved by performing elastic graph matching. The nodes of the employed graphs are comprised of Gabor jets and are fitted to given test faces. The obtained graph is first used to determine the identity of a given subject by choosing the closest matching graph. In a second stage, the closest matching graph found in the personalized gallery of the identified person is used to determine the displayed facial expression, thus allowing for a better focus on intra-personal variations.

4.5 Multimodal Frameworks

Today, most facial expression analysis systems are of the *unimodal* type, as they focus only on facial expressions when determining mental activities. However, the evaluation of multiple communication channels may foster robustness as well as improve correct interpretation of facial expressions in ambiguous situations. At present, most attempts of channel fusion are of the *bimodal* type and integrate voice in addition to facial expressions. Vocal expressions are conveyed by prosodic features, which include the fundamental frequency, intensity and rhythm of the voice. Cohn and Katz [13] as well as Chen et al. [11] focused on the fundamental frequency, as it is an important voice feature for emotion recognition and can be easily extracted.

5 Discussion

In this survey on automatic facial expression analysis we have discussed automatic face analysis with regard to different motion and deformation based extraction methods, model and image based representation techniques as well as recognition and interpretation based classification approaches.

It is not possible to directly compare facial expression recognition results of face analysis systems found in the literature due to varying facial action labeling and different test beds that were used for the assessment. Recently, a rather complete facial expression database has been presented by Kanade et al. [45]. However, there is a lack of publicly available, all encompassing facial expression databases that would allow for testing facial expression analysis methods in a more transparent way. Nonetheless, we tried to characterize a few selected systems with regard to the employed feature extraction and classification methods. Table 2 lists systems that perform facial expression recognition by classifying facial actions, while Table 3 presents systems that attempt either direct interpretation of emotional facial displays or indirect interpretation via facial expression dictionaries.

The application of currently available automatic facial expression recognition systems is often very restricted due to the limited robustness and hard constraints imposed on the recording conditions. Many systems assume faces to be centered in the input image and seen from a near frontal view throughout the whole test sequence.

Authors	# Subjects		# Faces		# Em.	Extraction	Classification	Rec.
	Test	Train	Test	Train	Class.	Methods	Methods	Rate
Lyons [56]	9	N/A	193	N/A	**7**	G.Wav.+El.Gr.	LDA+PCA+Cl.	**75-92%**
Kobayashi [49]	15	N/A	90	N/A	**6**	-	Feed Forw. NN	**85%**
Rosenblum [72]	32	20/14	34 sq.	20/14 sq.	**2**	Optical Flow	RBF NN	**88%**
Padgett [66]	12	N/A	N/A	N/A	**6**	-	PCA + NN	**86%**
Essa [33]	8	8	8 sq.	8 sq.	**6**	Opt.F.+3D M.	Motion Templ.	**98%**
Lanitis [51]	30	30	300	390	**7**	Appear. Mod.	Mahal. Distance	**74%**
Black [9]	40	N/A	70 sq.	N/A	**6**	Motion Mod.	Expert Rules	**83-100%**
Pantic [68]	8	N/A	496	N/A	**6**	Mul. Feat. Det.	Expert Rules	**91%**

Table 3: **Selected Facial Expression Interpretation Systems:** Classification of Emotional Displays. Note that the systems presented in the last two rows perform facial action recognition prior to using dictionaries in order to interpret facial actions.

Also, it is often taken for granted that there are only small rigid head motions between any two consecutive frames. In addition, most facial expression analysis systems require important *manual intervention* for the detection and accurate normalization of test faces, during the initialization of facial feature tracking approaches or for warping video sequences. Most facial expression analysis systems are limited to the analysis of either static images or image sequences. However, an ideal system should be capable of analyzing both static images as well as image sequences, as there are sometimes no image sequence available, respectively if there are, motion should be extracted in order to obtain directional information of skin and facial feature deformation.
The measurement of *facial expression intensities* has only be addressed by a few systems [52][54][34]. It is of importance for the interpretation of facial expressions, especially when attempting to analyze the temporal evolution and timing of facial actions.
Out-of-plane rotated faces are difficult to tackle and only a few approaches found in the literature were able to deal with this problem: active appearance models [51][24], local parametric models [9][84], 3D motion models [21][5] and to some degree also feature point tracking approaches [80][52].
Hybrid facial expression analysis systems are also of interest, as they combine different face analysis methods and may thus give better recognition results than the individual methods applied on their own [2]. This is true, if the employed extraction algorithms focus on different facial features or if the combined extraction, representation and recognition stages produce different error patterns.

6 Conclusion

Today, most facial expression analysis systems attempt to map facial expressions directly into basic emotional categories and are thus unable to handle facial actions caused by non-emotional mental and physiological activities. FACS may provide a solution to this dilemma, as it allows to classify facial actions prior to any interpretation attempts. So far, only marker-based systems are able to reliably code all FACS action unit activities and intensities [44].
More work has to be done in the field of automatic facial expression interpretation with regard to the integration of other communication channels such as voice and gestures. Although facial expressions often occur during conversations [26], none of the cited approaches did consider this possibility.
If automatic facial expression analysis systems are to be operated autonomously, current feature extraction methods have to be improved and extended with regard to robustness in natural environments as well as independence of manual intervention during initialization and deployment.

References

[1] M. Abdel-Mottaleb, R. Chellappa, and A. Rosenfeld. Binocular Motion Stereo using MAP Estimation. *IEEE CVPR*, pages 321–327, 1993.

[2] M. S. Bartlett. *Face Image Analysis by Unsupervised Learning and Redundancy Reduction.* PhD thesis, University of California, San Diego, 1998.

[3] M. S. Bartlett, P. A. Viola, T. J. Sejnowski, B. A. Golomb, J. Larsen, J. C. Hager, and P. Ekman. Classifying Facial Action. In *Advances in Neural Information Processing Systems*, volume 8. MIT Press, 1996.

[4] B. Bascle and A. Blake. Separability of Pose and Expression in Facial Tracking and Animation. In *Proceedings of the International Conference on Computer Vision*, 1998.

[5] S. Basu, N. Oliver, and A. Pentland. 3D Modeling and Tracking of Human Lip Motions. In *Proceedings of ICCV 98*, Bombay, India, 4-7January 1998.

[6] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[7] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical Model-based Motion Estimation. In G. Sandini, editor, *Preceedings of the Second European Conference on Computer Vision, ECCV-92*, volume 588 of *LNCS-Series*, pages 237–252. Springer-Verlag, May 1992.

[8] M. J. Black, D. J. Fleet, and Y. Yacoob. A Framework for Modeling Appearance Change in Image Sequences. In *Sixth International Conference on Computer Vision, (ICCV'98).* IEEE, Computer Society Press, January 1998.

[9] M. J. Black and Y. Yacoob. Recognizing Facial Expressions in Image Sequences using Local Parameterized Models of Image Motion. *International Journal of Computer Vision*, 25(1):23–48, 1997.

[10] W. Cai and J. Wang. Adaptive Multiresolution Collocation Methods for Initial Boundary Value Problems of Nonlinear PDEs. *Society for Industrial and Applied Mathematics*, 33(3):937–970, June 1996.

[11] L. S. Chen, T. S. Huang, T. Miyasato, and R. Nakatsu. Multimodal Human Emotion / Expression Recognition. In *Proceedings of the Secong International Conference on Automatic Face and Gesture Recognition, (FG'98)*, Nara, Japan, April 14-16 1998. IEEE.

[12] T. Choudhury and A. Pentland. Motion Field Histograms for Robust Modeling of Facial Expressions. In *Proceedings of the International Conference on Pattern Recognition (ICPR 2000)*, Barcelona, Spain, September 3-8 2000.

[13] J. Cohn and G. S. Katz. Bimodal Expression of Emotion by Face and Voice. In *Workshop on Face/Gesture Recognition and Their Applications, The Sixth ACM International Multimedia Conference*, September 1998.

[14] 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*, pages 329–333, July 1997.

[15] T. F. Cootes, C. J. Taylor, and A. Lanitis. Multi-Resolution Search using Active Shape Models. In *12th International Conference on Pattern Recognition*, volume 1, pages 610–612, Los Alamitos, California, 1994. IEEE CS Press.

[16] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active Appearance Models. In *University of Manchester, Wolfson Image Analysis Unit, Department of Medical Biophysics*, 1998.

[17] G. W. Cottrell and J. Metcalfe. EMPATH: Face, Gender and Emotion Recognition using Holons. In R. P. Lippman, J. Moody, and D. S. Touretzky, editors, *Advances in Neural Information Processing Systems*, volume 3, pages 564–571, 1991.

[18] M. N. Dailey and G. W. Cottrell. PCA = Gabor for Expression Recognition. Technical Report CS-629, UCSD, 1999.

[19] C. Darwin. *The Expression of the Emotions in Man and Animals.* J. Murray, London, 1872.

[20] J. Daugman. Complete Discrete 2D Gabor Transform by Neural Networks for Image Analysis and Compression. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36:1169–1179, 1988.

[21] D. DeCarlo and D. Metaxas. The Integration of Optical Flow and Deformable Models with Applications to Human Face Shape and Motion Estimation. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR '96)*, pages 231–238, 1996.

[22] G. Donato, S. Bartlett, C. J. Hager, P. Ekman, and J. T. Sejnowski. Classifying Facial Actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):974–989, October 1999.

[23] P. Dulguerov, F. Marchal, D. Wang, C. Gysin, P. W. Gidley, B. J. Gantz, J. T. Rubinstein, S. R. Seiff, L. K. Poon, K. S. Lun, and Y. M. Ng. Review of Objective Topographic Facial Nerve Evaluation Methods. *The American Journal of Otology*, 20(5):672–678, September 1999.

[24] G. J. Edwards, T. F. Cootes, and C. J. Taylor. Face Recognition using Active Appearance Models. In *Proceedings of the Fifth European Conference on Computer Vision (ECCV)*, volume 2, pages 581–695, University of Freiburg, Germany, June 1998.

[25] P. Eisert and B. Girod. Facial Expression Analysis for Model-based Coding of Video Sequences. In *Picture Coding Symposium*, pages 33–38, September 1997.

[26] P. Ekman. About Brows: Emotional and Conversational Signals. In J. Aschoff, M. Con Carnach, K Foppa, W. Lepenies, and D. Plog, editors, *Human Ethology*, pages 169–202. Cambridge University Press, 1979.

[27] P. Ekman. Emotions in the Human Face. *Cambridge University Press*, 1982.

[28] P. Ekman. Methods for Measuring Facial Actions. In K. R. Scherer and P. Ekman, editors, *Handbook of Methods in Nonverbal Behaviour Research*, pages 45–90. Cambridge University, 1982.

[29] P. Ekman and W. V. Friesen. Constants Across Cultures in the Face and Emotion. *Journal of Personality and Social Psychology*, 17(2):124–129, 1971.

[30] P. Ekman and W. V. Friesen. Facial Action Coding System: A Technique for the Measurement of Facial Movement. *Consulting Psychologists Press, Palo Alto*, 1978.

[31] P. Ekman, E. Rosenberg, and J. C. Hager. Facial Action Coding System Affect Interpretation Database (FACSAID). http://nirc.com/Expression/FACSAID/facsaid.html, July 1998.

[32] I. Essa and A. Pentland. Facial Expression Recognition using a Dynamic Model and Motion Energy. In *IEEE Proceedings of the 5th International Conference on Computer Vision (ICCV 1995)*, pages 360–367, Cambridge, MA, 1995.

[33] I. Essa and A. Pentland. Coding, Analysis, Interpretation and Recognition of Facial Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, 1997.

[34] Beat Fasel and Juergen Luettin. Recognition of Asymmetric Facial Action Unit Activities and Intensities. In *Proceedings of the International Conference on Pattern Recognition, (ICPR 2000)*, Barcelona, Spain, 2000.

[35] W. A. Fellenz, J. G. Taylor, N. Tsapatsoulis, and S. Kollias. Comparing Template-based, Feature-based and Supervised Classification of Facial Expressions from Static Images. In *Proceedings of Circuits, Systems, Communications and Computers (CSCC'99)*, pages 5331–5336, 1999.

[36] W. V. Friesen and P. Ekman. Emotional Facial Action Coding System. Unpublished manual, 1984.

[37] W. V. Friesen and P. Ekman. Dictionary - Interpretation of FACS Scoring. Unpublished manuscript, 1987.

[38] W. J. Hardcastle. *Physiology of Speech Production.* Academic Press, New York, NY, 1976.

[39] M. Hoch, G. Fleischmann, and B Girod. Modeling and Animation of Facial Expressions based on B-Splines. *The Visual Computer*, pages 87–95, November 1994.

[40] Hai Hong, Hartmut Neven, and Christoph Von der Malsburg. Online Facial Expression Recognition based on Personalized Galleries. In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, (FG'98)*, pages 354–359, Nara, Japan, April 14-16 1998. IEEE.

[41] C. L. Huang and Y. M. Huang. Facial Expression Recognition using Model-based Feature Extraction and Action Parameters Classification. *Journal of Visual Communication and Image Representation*, 8(3):278–290, 1997.

[42] C. Izard. The Maximally Descriminative Facial Movement Coding System (MAX). Available from Instructional Resource Center, University of Delaware, Newark, Delaware, 1979.

[43] C. E. Izard, L. M. Dougherty, and E. A. Hembree. A System for Indentifying Affect Expressions by Holistic Judgments. Unpublished manuscript, 1983.

[44] S. Kaiser and T. Wehrle. Automated Coding of Facial Behavior in Human-Computer Interactions with FACS. *Journal of Nonverbal Behavior*, 16(2):67–83, 1992.

[45] T. Kanade, J. Cohn, and Y. L. Tian. Comprehensive Database for Facial Expression Analysis. In *IEEE Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition (FG'00)*, March 2000.

[46] K. Karpouzis, G. Votsis, G. Moschovitis, and S. Kollias. Emotion Recognition using Feature Extraction and 3-D Models. In *Proceedings of IMACS International Multiconference on Circuits and Systems Communications and Computers, (CSCC'99)*, pages 5371–5376, Athens, Greece, 1999.

[47] S. Kimura and M. Yachida. Facial Expression Recognition and its Degree Estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 295–300, June 1997.

[48] H. Kobayashi and F. Hara. Dynamic Recognition of Basic Facial Expressions by Discrete-time Recurrent Neural Network. In *Proceedings of the International Joint Conference on Neural Network*, pages 155–158, 1993.

[49] H. Kobayashi and F. Hara. Facial Interaction between Animated 3D Face Robot and Human Beings. In *Proceedings of the International Conference on Systems, Man and Cybernetics*, pages 3732–3737, 1997.

[50] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, and C. Von der Malsburg. Distortion Invariant Object Recognition in the Dynamic Link Architecture. In *IEEE Transactions on Computers*, volume 42, pages 300–311, 1993.

[51] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic Interpretation and Coding of Face Images using Flexible Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.

[52] J. J. Lien. *Automatic Recognition of Facial Expression using Hidden Markov Models and Estimation of Expression Intensity*. PhD thesis, The Robotics Institute, CMU, April 1998.

[53] J. J. Lien, T. Kanade, J. F. Cohn, and C. C. Li. Automated Facial Expression Recognition based on FACS Action Units. In *IEEE Proceedings of the Second International Conference on Automatic Face and Gesture Recognition (FG'98)*, Nara, Japan, April 14-16 1998.

[54] C. L. Lisetti and D. E. Rumelhart. Facial Expression Recognition using a Neural Network. In *Proceedings of the 11th International Flairs Conference*. AAAI Press, 1998.

[55] C. L. Lisetti and D. J. Schiano. Automatic Facial Expression Interpretation: Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect. *Pragmatics and Cognition (Special Issue on Facial Information Processing: A Multidisciplinary Perspective)*, 8(1):185–235, 2000.

[56] M. J. Lyons, J. Budynek, and S. Akamatsu. Automatic Classification of Single Facial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12), December 1999.

[57] K. Mase and A. Pentland. Recognition of Facial Expression from Optical Flow. *IEICE Transactions*, E 74(10):3474–3483, October 1991.

[58] D. Matsumoto. Cultural Similarities and Differences in Display Rules. *Motivation and Emotion*, 14(3):195–214, 1990.

[59] D. Matsumoto. Ethnic Differences in Affect Intensity, Emotion Judgments, Display Rules, and Self-reported Emotional Expression. *Motivation and Emotion*, 17:107–123, 1993.

[60] I. Matthews. *Active Shape Model Toolbox.* University of East Anglia, Norwich, UK, 21 July 1997. Matlab Toolbox version 2.0.0.

[61] D. Messinger, A. Fogel, and K. L. Dickson. What's in a Smile? *Developmental Psychology*, 35(3):701–708, 1999.

[62] H. H. Nagel. On the Estimation of Optical Flow: Relations between Different Approaches and some New Results. *Artificial Intelligence*, 33:299–324, 1987.

[63] N. Oliver, A. Pentland, and F. Berard. LAFTER: A Real-time Lips and Face Tracker with Facial Expression Recognition. In *Proceedings of the IEEE Conference on Computer Vision, (CVPR97)*, S. Juan, Puerto Rico, 1997.

[64] T. Otsuka and J. Ohya. Extracting Facial Motion Parameters by Tracking Feature Points. In *Proceedings of First International Conference on Advanced Multimedia Content Processing*, pages 442–453, November 1998.

[65] T. Otsuka and J. Ohya. Spotting Segments Displaying Facial Expression from Image Sequences using HMM. In *IEEE Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, (FG'98)*, pages 442–447, Nara, Japan, April 14-16 1998.

[66] C. Padgett and G. W. Cottrell. Representing Face Image for Emotion Classification. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems*, volume 9, pages 894–900, Cambridge, MA, 1997. MIT Press.

[67] C. Padgett, G. W. Cottrell, and R. Adolphs. Categorical Perception in Facial Emotion Classification. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, 1996.

[68] M. Pantic and L. J. M. Rothkrantz. Expert System for Automatic Analysis of Facial Expression. *Image and Vision Computing Journal*, 18(11):881–905, 2000.

[69] A. Pentland, B. Moghaddam, and T. Starner. View-based and Modular Eigenspaces for Face Recognition. In *IEEE Conference of Computer Vision and Pattern Recognition*, pages 84–91, 1994.

[70] D. Pollen and S. Ronner. Phase Relationship between Adjacent Simple Cells in the Visual Cortex. *Science*, 212:1409–1411, 1981.

[71] Rob Koenen. *Mpeg-4 Project Overview.* International Organisation for Standartistion, ISO/IEC JTC1/SC29/WG11, La Baule, October 2000.

[72] M. Rosenblum, Y. Yacoob, and L. Davis. Human Expression Recognition from Motion using a Radial Basis Function Network Architecture. *IEEE Transactions on Neural Networks*, 7(5):1121–1138, 1996.

[73] H. A. Rowley, S. Baluja, and T. Kanade. Neural Network-Based Face Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, January 1998.

[74] B. Schiele and J. Crowley. Probabilistic Object Recognition using Multidimensional Receptive Field Histograms. In *Proceedings of the International Conference on Pattern Recognition, (ICPR 1996)*, Vienna, Austria, August 1996.

[75] G. E. Schwartz, P. L. Fair, P. Salt, M. R. Mandel, and G. L. Klerman. Facial Expression and Imagery in Depression: An Electromyographic Study. *Psychosomatic Medicine*, 38(337-347), 1976.

[76] E. P. Simoncelli. *Distributed Representation and Analysis of Visual Motion.* PhD thesis, Massachusetts Institute of Technology, 1993.

[77] J. Steffens, E. Elagin, and H. Neven. PersonSpotter - Fast and Robust System for Human Detection, Tracking and Recognition. In *Proceedings of the Second International Conference on Face and Gesture Recognition, (FG'98)*, pages 516–521, Nara, Japan, April 1998.

[78] M. Suwa, N. Sugie, and K. Fujimora. A Preliminary Note on Pattern Recognition of Human Emotional Expression. In *Proceedings of the 4th International Joint Conference on Pattern Recognition*, pages 408–410, 1978.

[79] D. Terzopoulos and K. Waters. Analysis of Facial Images using Physical and Anatomical Models. In *Proceedings of the Third International Conference on Computer Vision*, pages 727–732, 1990.

[80] Y. Tian, T. Kanade, and J. F. Cohn. Recognizing Action Units for Facial Expression Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), February 2001.

[81] N. Tsapatsoulis, K. Karpouzis, and G. Stamou. A Fuzzy System for Emotion Classification based on the MPEG-4 Facial Definition Parameter. *European Association for Signal Processing (EUSIPCO)*, 2000.

[82] M. Wang, Y. Iwai, and M. Yachida. Expression Recognition from Time-Sequential Facial Images by use of Expression Change Model. In *IEEE Proceedings of the Second International Conference on Automatic Face and Gesture Recognition (FG'98)*, pages 324–329, Nara, Japan, April 14-16 1998.

[83] Y. T. Wu, T. Kanade, J. F. Cohn, and C. C. Li. Optical Flow Estimation using Wavelet Motion Model. In *IEEE International Conference on Computer Vision*, pages 992–998, Bombay, India, January 1998.

[84] Y. Yacoob and L. S. Davis. Recognizing Human Facial Expression from Long Image Sequences using Optical Flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6):636–642, 1996.

[85] M. Yoneyama, Y. Iwano, A. Ohtake, and K. Shirai. Facial Expression Recognition using Discrete Hopfield Neural Networks. In *Proceedings of the International Conference on Image Processing (ICIP)*, volume 3, pages 117–120, 1997.

[86] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu. Comparison between Geometry-based and Gabor-Wavelets-based Facial Expression Recognition using Multi-Layer Perceptron. In *IEEE Proceedings of the Second International Conference on Automatic Face and Gesture Recognition (FG'98)*, pages 454–459, Nara, Japan, April 14-16 1998.

[87] J. Zhao and G. Kearney. Classifying Facial Emotions by Backpropagation Neural Networks with Fuzzy Inputs. In *Proceedings of the International Conference on Neural Information Processing*, volume 1, pages 454–457, 1996.