



HUMAN FACIAL EXPRESSIONS RECOGNITION - A SURVEY

RATHINA X.A.^{1*}, PONNAVAIKKO M.², LAKSHMI C.² AND MEHATA K.M.¹

¹B.S. Abdur Rahman University Vandalur, Chennai- 600 048, TN, India.

²SRM University, Kattankulathur- 603203, TN India.

*Corresponding Author: Email- zarathna@gmail.com

Received: July 23, 2012; Accepted: December 20, 2012

Abstract- Recognition of human facial expressions is an active research area that finds potential applications in areas such as human-computer interfaces, image retrieval and human emotion analysis. The face is the most extraordinary communicator, capable of accurately signaling emotion in a bore link of a second. Moreover facial expressions convey not only emotions, but other mental activities, social interaction and physiological signals. In this paper we presented the evolution of the analysis of facial expressions and survey some of the published work in this area since 1990 to till date. This paper presents a time-line view of the advances made in this field, the facial parameterization using FACS Action Units (AUs) and MPEG-4 Facial Animation Parameters (FAPs), the applications, and the Databases.

Keywords- Survey of Facial Expressions Recognition, Facial Expression Techniques, Face databases, feature extraction and representation, motion extraction, facial expression classification, facial parameterization techniques, facial action coding system, facial animation parameters.

Citation: Rathina X.A., et al. (2012) Human Facial Expressions Recognition - A Survey. Journal of Signal and Image Processing, ISSN: 0976-8882 & E-ISSN: 0976-8890, Volume 3, Issue 4, 2012, pp. -130-142.

Copyright: Copyright©2012 Rathina X.A., et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Introduction

The human possesses superior expressive ability [1] and provides one of the most powerful, versatile and natural means of communicating motivational and affective state. We use facial expressions not only to express our emotions but also to provide important communicative cues during social interaction such as our level of interest, our desire to take a speaking tern and continuous feedback signaling understanding of the information conveyed. Facial expression constitutes 55% of the effect of a communicated message [2] and is hence a major modality in human communication.

Studies on Facial Expressions and Physiognomy date back to the early Aristotelian era (4th century BC). As quoted from Wikipedia, "Physiognomy is the assessment of a person's character or personality from their outer appearance, especially the face" [1]. But over the years, while the interest in Physiognomy has been waxing and waning, the study of facial expressions has consistently been an active topic. A detailed note on the various expressions and movement of head muscles was given in 1649 by John Bulwer in his book "Pathomyotomia". Another interesting work on facial expressions (and Physiognomy) was by Le Brun, the French academician and painter. In 1667, Le Brun gave a lecture at the Royal Academy of Painting which was later reproduced as a book in 1734.

The foundational studies on facial expressions that have formed the basis of today's research can be traced back to the 17th century. It is interesting to know that the 18th century actors and artists referred to Le Brun book in order to achieve "the perfect imitation of 'genuine' facial expressions" [2]. In 1872, Darwin wrote a treatise that established the general principles of expression and the means of expressions in both humans and animals [5]. He also grouped various kinds of expressions into similar categories. The categorization is as follows:

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

Moving on to the 19th century, one of the important works on facial expression analysis that has a direct relationship to the modern day science of automatic facial expression recognition was the work done by Charles Darwin [5]. A psychologist Paul Ekman and

his colleagues done a research on facial expressions since 1970. In 1971, Ekman and Friesen [18] 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. As we have seen, facial expressions have been studied by clinical and social psychologists, medical practitioners, actors and artists. However in the last quarter of the 20th century, with the advances in the fields of robotics, computer graphics and computer vision, animators and computer scientists started showing interest in the study of facial expressions. The first step towards the automatic recognition of facial expressions was taken in 1978 by Suwa et al. Suwa and his colleagues presented a system for analyzing facial expressions from a sequence of images (movie frames) by using twenty tracking points.

Although this system was proposed in 1978, researchers did not pursue this line of study till the early 1990s. This can be clearly seen by reading the 1992 survey paper on the automatic recognition of faces and expressions by Samal and Iyengar [6]. In this the 'Facial Features and Expression Analysis' section presents only four papers: two on automatic analysis of facial expressions and two on modeling of the facial expressions for animation. This paper also states that "research in the analysis of facial expressions has not been actively pursued" [6]. The reason for this is stated by Vinay Kumar Bettadapura [7] as the automatic recognition of facial expressions requires robust face detection and face tracking systems. These were research topics that were still being developed and worked upon in the 1980's, due to the capability of cheap com-

puting power in the year 1990, there was a development in the algorithms of face detection and face tracking. At the same time, Human computer interaction and Affective computing started gaining popularity. Thus, the interest in the development of recognition of human facial expressions is renewed in the year 1990. Since 1990's, research on facial expressions recognition has become very active.

The goal of this paper is to survey the work done in the automation of facial expressions recognition since 1990's to till date. The remainder of this paper will be organized as follows: in section 2, a detailed review of facial expression recognition techniques since 1990 to till date is given. In section 3, the different stages of automatic facial expression recognition system is addressed. The various facial parameterization techniques are discussed in section 4. Section 5 gives a note on the various databases that have been used is given. In section 6, we mentioned some of the applications of automatic facial expression recognition systems. Section 7 mentions the challenges and future work and the paper conclude with section 8.

Facial Expression Recognition Techniques

State of Art

Since it is almost impossible to cover all of the published work, we have selected 30 papers that we felt were important and very different from each other. A table was created that will give a details about all the surveyed papers. The details include the feature extraction technique, classifier, database, sample size, recognition performance and the important features of this technique. This table is included as [Annexure-1].

Annexure 1-

Reference	Feature Extraction	Classifier	Database	Sample size	Performance	Important Points
Tian et al., 2001 [10]	Permanent features: Optical Flow, Gabor Wavelets and Multistate Models. Transient features: Canny edge detection	2 ANNs, one for upper face and one for lower face	Cohn-Kanade and Ekman-Hager Facial Action Exemplars	Upper Face: 50 sample sequences from 14 subjects performing 7 AUs Lower Face: 63 sample sequences from 32 subjects performing 11 AUs	Recognition of Upper Face AUs: 96.4% Recognition of Lower Face AUs: 96.7% Average of 93.3% when generalized to independent Databases	Recognizes posed expressions. Real-time system. Automatic Face detection. Head motion is handled Invariant to scaling. Uses facial feature tracker to reduce processing time.
Bourel et al., 2001	Local spatiotemporal vectors obtained from the EKLT tracker	Modular classifier with data fusion. Local classifiers are rank weighted kNN classifiers. Combinatio is a sum scheme.	Cohn-Kanade	30 subjects 25 sequences for 4 expressions (total of 100 video sequences)	--	Deals with recognizing facial expressions in the presence of Occlusions. Proposes use of modular classifiers instead of monolithic Classifiers. Classification is done locally and then the Classifier outputs are fused.
Pardas and Bonafonte, 2002	MPEG-4 FAPs extracted using an improved Active Contour algorithm and motion estimation	HMM	Cohn-Kanade	Used the whole DB	Overall efficiency of 84% (across 6 prototypic expressions) Experiments with: joy, surprise and anger: 98%, joy, surprise and sadness 95%	Automatic extraction of MPEG-4 FAPs Proves that FAPs convey the necessary information that is required to extract the emotions
Cohen et al., 2003	A vector of extracted Motion Units (MUs) using PBVD tracker	NB, TAN, SSS, HMM, ML-HMM	Cohn-Kanade and own DB	Cohn-Kanade: 53 subjects Own DB: 5 subjects	---	Real-time system. Emotion classification from video. Suggests use of HMMs to automatically segment a video into different expression segments.
Cohen et al., 2003	A vector of extracted Motion Units (MUs) using PBVD tracker	NB, TAN, SSS.	Cohn-Kanade and Chen-Huang	---	---	Real-time system. Uses semi-supervised learning to work with some labeled data and large amount of unlabeled data.

Annexure 1- Continue

Reference	Feature Extraction	Classifier	Database	Sample size	Performance	Important Points
Bartlett et al., 2003	Gabor Wavelets	SVM, AdaSVM (SVM with AdaBoost)	Cohn-Kanade	313 sequences from 90 subjects. First and last frame used as training Images	SVM with Linear Kernel: Automatic face detection: 84.8%; Manual alignment: 85.3%. SVM with RBF Kernel: Automatic face detection: 87.5%; Manual alignment: 87.6%	Fully automatic system. Real-time recognition at high level of accuracy Successfully deployed on Sony's Aibo pet robot, ATR's RoboVie and CU Animator
Michel and Kaliouby, 2003	Vector of feature displacements (Euclidean distance between neutral and peak)	SVM	Cohn-Kanade	For each basic emotion, 10 examples were used for training and 15 examples were used for testing.	With RBF Kernel: 87.9%. Person independent: 71.8% Person dependent (train and test data supplied by expert): 87.5% Person dependent (train and test data supplied by 6 users during ad-hoc interaction): 60.7%	Real-time system. Does not require any preprocessing
Pantic and Rothkrantz, 2004	Frontal and Profile facial points	Rule based classifier	MMI	25 subjects	86% accuracy	Recognizes facial expressions in frontal and profile views. Proposed a way to do automatic AU coding in profile images. Not Real-time.
Buciu and Pitas, 2004	Image representation using Non negative Matrix Factorization (NMF) and Local Non negative Matrix factorization (LNMF).	Nearest neighbor Classifier using CSM and MCC	Cohn-Kanade and JAFFE	Cohn-Kanade: 164 Samples JAFFE: 150 samples	Cohn-Kanade: LNMF with MCC gave the highest accuracy of 81.4% JAFFE: Only 55% to 68% for all 3 methods	PCA was also performed for comparison purpose. LNMF outperformed both PCA and NMF whereas NMF produces the poorest performance. CSM classifier is more reliable than MCC and gives better recognition.
Pantic and Patras, 2005	Tracking a set of 20 facial fiducial points	Temporal Rules	Cohn-Kanade and MMI	Cohn-Kanade: 90 Images MMI: 45 images	Overall an average recognition of 90%	Recognizes 27 AUs. Invariant to occlusions like glasses and facial hair. Shown to give a better performance than the AFA system
Zheng et al., 2006	34 landmark points converted into a Labeled Graph (LG) using Gabor wavelet transform. Then a semantic expression vector built for each training face. KCCA used to learn the correlation between LG vector and semantic vector.	The correlation that is learnt is used to estimate semantic expression vector which is then used for classification	JAFFE and Ekman's Pictures of Affect	JAFFE: 183 images Ekman's: 96 images Neutral expressions were not chosen from either database	Using Semantic Info: On JAFFE DB: with Leave one image out (LOIO) cross validation: 85.79%, with Leave one subject out (LOSO) cross validation: 74.32%, On Ekman's DB: 81.25% Using Class Label Info: On JAFFE DB: with LOIO: 98.36%, with LOSO: 77.05%, On Ekman's B: 78.13%	Used KCCA to recognize facial expressions The singularity problem of the Gram matrix has been tackled using an improved KCCA algorithm.
Anderson and McOwen, 2006	Motion signatures obtained by tracking using spatial ratio template tracker and performing optical flow on the face using multichannel gradient model (MCGM)	SVM and MLP	CMUPittsburg AU coded DB and a Nonexpressive DB	CMU: 253 samples of 6 basic expressions. But these had to be preprocessed by reducing frame rate and scale Non-expressive: 10 subjects, 4800 frames long	Motion averaging using: co-articulation regions: 63.64%, 7x7 blocks: 77.92%, with ratio template algorithms, with MLP: 81.82%, with SVM: 80.52%	Fully automated, multistage system. Real-time system. Able to operate efficiently in cluttered scenes. Used motion-averaging to condense the data that is fed to the Classifier.
Aleksic and Katsaggelos, 2006	MPEG-4 FAPs, outer lip (group 8) and eyebrow (group 4) followed by PCA to reduce dimensionality	HMM and MS-HMM	Cohn-Kanade	284 recordings of 90 subjects	Using HMM: Only eyebrow FAPs: 58.8%, Only outer lip FAPs: 87.32%, Joint FAPs: 88.73% After assigning stream weights and then using a MS-HMM: 93.66% with outer lip having more weight than eyebrows.	Shown that performance improvement is possible by using MS-HMMs and proposed a way to assign stream weights. Used PCA to reduce the dimensionality of the features before giving it to the HMM.
Pantic and Patras, 2006	Mid level parameters generated by tracking 15 facial points using particle Filtering	Rule based classifier	MMI	1500 samples of both static and profile views (single and multiple AU activations)	86.6% on 96 test profile sequences	Automatic segmentation of input video into facial expressions. Recognition of temporal segments of 27 AUs occurring alone or in combination. Automatic recognition of AUs from profile images.
Sebe et al., 2007	MUs generated from the PBVD tracker	Bayesian nets, SVMs and Decision Trees. Used voting algorithms like bagging and Boosting to improve Results	Created Spontaneous emotions database. Also used Cohn-Kanade	Created DB: 28 subjects showing mostly neutral, joy, surprise and delight. Cohn-Kanade: 53 subjects	Using many different classifiers: Cohn-Kanade: 72.46% to 93.06%, Created DB: 86.77% to 95.57%. Using kNN with k = 3, best result of 93.57%	Recognizes spontaneous expressions Created an authentic DB where subjects are showing their natural facial expressions Evaluated several Machine Learning algorithms

Annexure 1- Continue

Reference	Feature Extraction	Classifier	Database	Sample size	Performance	Important Points
Kotsia and Pitas, 2007	Geometric displacement of Candide nodes	Multiclass SVM: For expression recognition: used six-class SVM, one for each expression. For AU recognition: used one class SVMs, one for each of the 8 chosen AUs used.	Cohn-Kanade	Whole DB	99.7% for facial expression recognition 95.1% for facial expression recognition based on AU Detection	Recognizes either the six basic facial expressions or a set of chosen AUs. Very high recognition rates have been shown
Wang and Yin, 2007	Topographic context (TC) expression descriptors	QDC, LDA, SVC and NB	Cohn-Kanade and MMI	Cohn-Kanade: 53 subjects, 4 images per subject for each expression. Total of 864 images MMI: 5 subjects, 6 images per subject for each expression. Total of 180 images	Person dependent tests: on MMI: with QDC: 92.78%, with LDA: 93.33%, with NB: 85.56%, on Cohn-Kanade: with QDC: 82.52%, with LDA: 87.27%, with NB: 93.29%. Person independent tests: on Cohn-Kanade: with QDC: 81.96%, with LDA: 82.68%, with NB: 76.12%, with SVC: 77.68%	Proposed a topographic modeling approach in which the gray scale image is treated as a 3D surface. Analyzed the robustness against the distortion of detected face region and the different intensities of facial expressions.
Dornaika and Davoine, 2008	Candide face model used to track features.	First head pose is determined using Online Appearance Models and then expressions are recognized using a stochastic approach	Created own data	Used several video sequences. Also created a challenge 1600 frame test video, where subjects were allowed to display any expression in any order for any duration	Results have been spread across different graphs and charts. Interested readers can refer [24] to view the same.	Proposes a framework for simultaneous face tracking and expression recognition. 2 AR models per expression gave better mouth tracking and in turn better performance. The video sequences contained posed expressions.
Kotsia et al., 2008	3 approaches: Gabor features, DNMF algorithm and by Geometric Displacement vectors extracted using Candide tracker	Multiclass SVM and MLP	Cohn-Kanade and JAFFE		Using JAFFE: with Gabor: 88.1%, with DNMF: 85.2% Using Cohn-Kanade: with Gabor: 91.6%, with DNMF:86.7%, with SVM: 91.4%	Developed a system to recognize expressions in spite of occlusions. Discusses the effect of occlusion on the 6 rototypic facial expressions.

Stages of Automatic Facial Expression Recognition System

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. Variations such as these have to be addressed at different stages of an automatic facial expression analysis system [Fig-1].

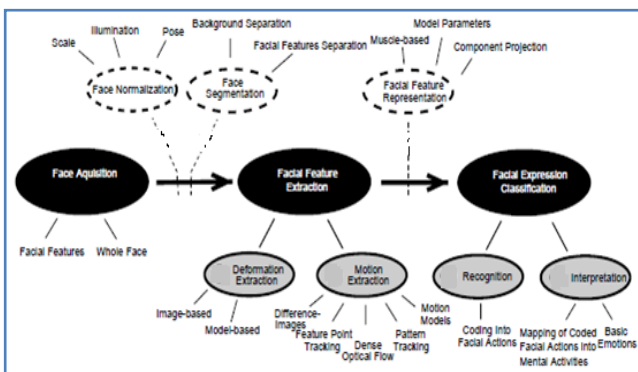


Fig. 1- Generic Facial Expression Analysis Framework

Face Acquisition

In Face acquisition stage, an automatic face detector is used to locate faces in complex scenes with cluttered backgrounds. Face analysis is complicated by face appearance changes caused by pose variations and illumination changes.

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.

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 eyes, 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, local changes of texture that cannot be predicted from antecedent frames.

Face segmentation allows isolating transient and intransient features within faces or can be used to separate faces of interest from the background.

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.

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 allows modeling facial features and faces based on their appearance, without attempting to recover the volumetric geometry of the scene.

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 based frameworks attempt to interfere muscle activities from visual information. This may be achieved e.g. by using 3D muscle models that allow mapping extracted optical flow into muscle actions.

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. They allow detecting line endings and edge borders over multiple scales and with different orientations. These features reveal much 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 [22,23] as well as in combination with labeled graphs [24,25]. We can distinguish local or holistic image-based deformation extraction approaches:

Holistic Image-based Approaches

Several authors have taken either whole faces as features [26] or Gabor wavelet filtered whole-faces [27]. 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 [26] as well as Cottrell and Metcalfe [27] 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 along segments [28] or by determining the density of high gradient components over windows of interest [29].

Model-based approaches constitute an alternative to image-based deformation extraction. Appearance-based model approaches allow separating fairly well different information sources such as facial illumination and deformation changes. Lanitis et al. [30] interpreted face images by employing active appearance models (AAM) [31,32]. Faces were analyzed by a dual approach, using both shape and texture models. Active shape models (ASM) allow simultaneously determining shape, scaling and posing by fitting an appropriate point distribution model (PDM) to the object of interest. 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 [33] used a point distribution model 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 [24, 34]. 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 [35] 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 [32] used a 2D point-based model. Facial feature representation using active shape models (ASM): The first row shows manually placed point models (PM) that were employed to create a point distribution model (PDM), 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 [37] of both frontal and a side views.

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

These approaches allow for whole-faces analysis and were employed e.g. are given in Refs. [28, 29]. Lien [27] 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 eigen flows both in horizontal and vertical directions.

Local Dense Optical Flow

Region-based dense optical flow was used by Mase and Pentland [38] in order to estimate the activity 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 [34] 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. 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. [27] employed Wu's [39] approach of optical flow to estimate facial motion by using scaling functions and wavelets from Cai and Wang [41] to capture both local and global facial characteristics. Essa and Pentland [24] used Simoncelli's [42] coarse-to-fine optical flow, while Yacoob and Davis [45] as well as Rosenblum et al. [43] employed Abdek-Mottaleb et al. [44] 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:

Holistic Motion Models

Terzopoulos and Waters [47] have used 11 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 [46] employed sophisticated 3D motion and muscle models for facial expression recognition and increased tracking stability by Kalman filtering. DeCarlo and Metaxas [48] 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 [49] 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 [50] as well as Yacoob and Davis [45] 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. [49] presented a convincing approach of how to track human lip motions by using 3D models. In contrast to low-level dense optical flow, there are also higher level variants that focus either on the movement on generic features points, patterns or markers:

Feature Point Tracking

Here, motion estimates are obtained only for a selected set of prominent features such as intransient facial features [29, 52, 53]. In order to reduce the risk of tracking loss, feature points are placed into areas of high contrast, preferably around intransient facial features. Hence, the movement and deformation of the latter can be measured by tracking the displacement of the 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 [53] 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. used different component models for the lips, eyes, eye 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. [43] 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 [54] or by affixing colored plastic dots to predefined locations on the subject's face. Markers allow rendering tissue motion visible and were employed as given in Refs. [55, 56]. 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. 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 [15, 28, 57, 58]. Choudhury and Pentland [59] 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 by using local receptive field histograms [60] in order to increase robustness with regard to rotation, translation and scale changes.

Facial Expression 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 interpretation in combination with judgment or sign dictionary-based frameworks. We can distinguish these as spatial and spatial-temporal classifier approaches:

Spatial-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 modeling the dynamics of facial actions. Several HMM-based classification approaches can be found in the literature [34] 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 [43]. Another way of taking temporal evolution of facial expression into account is so-called spatial-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 [46].

Spatial Approaches

Neural networks were often used for facial expression classification [32,28,29,26]. They were either applied directly on face images [22] or combined with facial features extraction and representation methods such as PCA Independent Component Analysis (ICA) or Gabor wavelet filters [22,23]. 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 [28] or locally, using mosaic-like patches extracted from small facial regions [23, 28]. Dailey and Cottrell [23] applied both local PCA and Gabor jets for the task of facial expression recognition and obtained quantitatively indistinguishable results for both representations. A problem is the great number of possible facial action combinations, about 7000 AU combinations have been identified within the FACS framework.

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 [50, 45]. 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. Therefore, another group of researchers have relied on facial expression coding schemes such as MPEG-4 [49] or FACS [61] Essa and Pentland [46] 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 also.

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 [30,63]. Only a few systems use rules or facial expression dictionaries in order to translate coded facial actions into emotion categories [50]. The latter approaches have not only the advantage of accurately describing facial expressions without resorting to interpretation, but allow also to animate synthetic faces, e.g. within the FACS coding framework [62]. This is of interest, as animated synthetic faces make a direct inspection of automatically recognized facial expressions possible.

Characteristics of a Good Facial Expression Recognition System

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

It must be fully automatic

It must have the capability to work with video feeds as well as images

It must work in real-time

It must be able to recognize spontaneous expressions Along with the prototypic expressions, it must be able to recognize a whole range of other expressions too (probably by recognizing all the facial AUs)

- It must be robust against different lighting conditions
- It must be able to work moderately well even in the presence of occlusions
- It must be unobtrusive
- The images and video feeds do not have to be pre-processed
- It must be person independent
- It must work on people from different cultures and different skin colors. It must also be robust against age (in particular, recognize expressions of both infants, adults and the elderly)
- It must be invariant to facial hair, glasses, makeup etc.
- It must be able to work with videos and images of different resolutions
- It must be able to recognize expressions from frontal, profile and other intermediate angles

From the past surveys (and this one), we can see that different research groups have focused on addressing different aspects of the above mentioned points. For example, some have worked on recognizing spontaneous expression, some on recognizing expressions in the presence of occlusions; some have developed systems that are robust against lighting and resolution and so on. However going forward, researchers need to integrate all of these ideas together and build systems that can tend towards being ideal.

Facial Parameterization Techniques & 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 5 s or less than 250 ms. Facial expression intensities may be measured by determining either the geometric deformation of facial features or the density of wrinkles appearing in certain face regions. Static images do not clearly reveal subtle changes in faces and it is therefore essential to measure also the dynamics of facial expressions. These streams stem directly from two major approaches to facial expression measurement in psychological research: 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 behaviour, such as facial movement or facial component shape. Most automatic facial expression analysis approaches found in the literature attempt to directly map facial expressions into one of the basic emotion classes introduced by Ekman and Friesen [12]. Facial action coding system (FACS), which was developed by Ekman and Friesen 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 action units (AUs) for the description of facial actions with regard to their location as well as their intensity. Individual expressions may be modeled by single action units or action unit combinations. Similar coding schemes are EMFACS [12], MAX [13] and AFFEX [12]. However, they are only directed towards emotions. Finally, the MPEG-4-SNHC [14] is a standard that encompasses analysis, coding [11] and animation of faces (talking heads) [15]. 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 [12,15]. Ekman et al. [16] presented also a database called facial action coding system affect interpretation database (FACSAID), which allows translating emotion related FACS scores into affective meanings. Emotion interpretations were provided by several experts, but only agreed affects were included in the database.

Facial Action Coding System (FACS)

Two main streams in the current research on automatic analysis of facial expressions consider facial affect (emotion) detection and facial muscle action (AU) detection [32]. Thus, an eye brow frown as shown in [Fig-2] can be judged as 'anger' 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. The most commonly used facial expression descriptors in message-judgment approaches are the six basic emotions (fear, sadness, happiness, anger, disgust and surprise) proposed by Ekman and discrete emotion theorists Keltner & Ekman, who suggest that these emotions are universally displayed and recognized from facial expressions.

Most 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. 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. The most commonly used facial action descriptors in sign judgement-approaches are the AUs

defined in the Facial Action Coding System (FACS; Ekman et al. 2002). FACS associates facial expression changes with actions of the muscles that produce them. It defines nine different AUs in the upper face, 18 in the lower face, five miscellaneous ones, 11 action descriptors (ADs) for head position, nine ADs for eye position and 14 additional descriptors for miscellaneous actions. Samples of Action units and their combinations are shown in [Fig-3].



Fig. 2- Facial Appearance of Corrugators' Muscle contraction.



Fig. 3- Examples of AUs and their combinations defined By FACS.

AUs are considered to be the smallest visually discernable facial movements. FACS also provides the rules for recognition of AUs' temporal segments (onset, apex and offset) in a face video. Using FACS, human coders can manually code nearly any anatomically possible facial expression, decomposing it into the specific AUs and their temporal segments that produced the expression. As AUs are independent of interpretation, they can be used for any higher order decision-making process, including recognition of basic emotions based on EMFACS rules; personality traits like extraversion and temperament [13], and social signals like emblems (i.e. culture-specific interactive signals like wink), regulators (i.e. conversational mediators like nod and smile) and illustrators (i.e. cues accompanying speech like raised eyebrows; Ekman & Friesen 1969; Ambady & Rosenthal 1992). AUs are very suitable for use as mid-level parameters in automatic facial behavior analysis, as the thousands of anatomically possible expressions defined in Cohn & Ekman 2005 can be described as combinations of 32 AUs and can be

mapped to any higher order facial display interpretation, including basic emotions, cognitive states, social signals and behaviours, and complex mental states like depression. It is not surprising, therefore, that automatic 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. [24], Lien et al. (1998) and Pantic et al [32]. 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 [32]. Several promising prototype systems that can recognize deliberately produced AUs in either (near-) frontal view face images (e.g. Bartlett et al. 1998 [24]; Tian et al. 2001; Pantic [14]) or profile view face images were reported. These systems employ ranges of approaches, including expert rules, machine learning methods such as neural networks, feature-based image representations i.e. using geometric features like facial points or shapes of facial components) or appearance-based image representations (i.e. using texture of the facial skin including wrinkles and furrows. 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 pieces of work have recently emerged on machine analysis of AUs in spontaneous facial expression data.

Facial Animation Parameters (FAP)

In the 1990s and prior to that, the computer animation research community faced similar issues that the face expression recognition researchers faced in the pre-FACS days. There was no unifying standard and almost every animation system that was developed had its own defined set of parameters. As noted by Pandzic and Forchheimer, the efforts of the animation and graphics researchers were more focused on the facial movements that the parameters caused, rather than the efforts to choose the best set of parameters [17]. To address these issues and provide for a standardized facial control parameterization, the Moving Pictures Experts Group (MPEG) introduced the Facial Animation (FA) specifications in the MPEG-4 standard. Version 1 of the MPEG-4 standard (along with the FA specification) became the international standard in 1999. In the last few years, face expression recognition researchers have started using the MPEG-4 metrics to model facial expressions [20]. The MPEG-4 standard supports facial animation by providing Facial Animation Parameters (FAPs). Cowie et al. indicate the relationship between the MPEG-4 FAPs and FACS AUs: “MPEG-4 mainly focusing on facial expression synthesis and animation, defines the Facial Animation parameters (FAPs) that are strongly related to the Action Units (AUs), the core of the FACS” [19]. To better understand this relationship between FAPs and AUs, I give a brief introduction to some of the MPEG-4 standards and terminologies that are relevant to face expression recognition. The detailed overview of the MPEG-4 Facial Animation technology is described in the survey paper by Abrantes and Pereira [20].

The MPEG-4 defines a face model in its neutral state to have a specific set of properties like a) all face muscles are relaxed; b) eyelids are tangent to the iris; c) pupil is 1/3rd the diameter of the iris and so on. Key features like eye separation, iris diameter, etc are defined on this neutral face model [21]. The standard also de-

finer 84 key feature points (FPs) on the neutral face [21]. The movement of the FPs is used to understand and recognize facial movements (expressions) and in turn also used to animate the faces. Fig-4] shows the location of the 84 FPs on a neutral face as defined by the MPEG-4 standard. The FAPs are a set of parameters that represent a complete set of facial actions along with head-motion, tongue, and eye and mouth control. In other words, each FAP is a facial action that deforms a face model in its neutral state. The FAP value indicates the magnitude of the FAP which in turn indicates the magnitude of the deformation that is caused on the neutral model. In order to use the FAP values for any face model, the MPEG-4 standard defines normalization of the FAP values. This normalization is done using Facial Animation Parameter Units (FAPUs). The 68 FAPs are grouped into 10 FAP groups. FAP group 1 contains two high level parameters: visemes (visual phonemes) and expressions. The MPEG-4 standard defines six primary facial expressions: joy, anger, sadness, fear, disgust and surprise. Although FAPs have been designed to allow the animation of faces with expressions, in recent years, the facial expression recognition community is working on the recognition of facial expressions and emotions using FAPs.

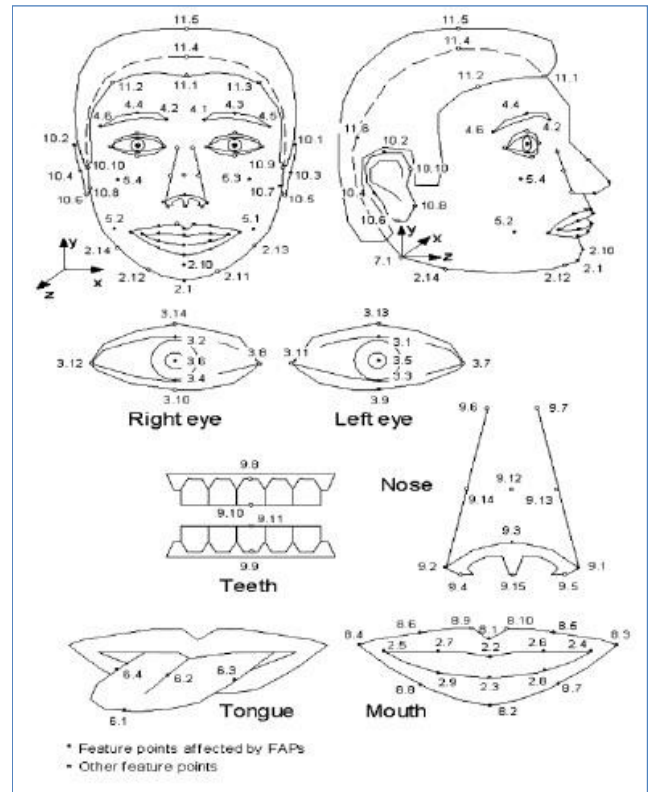


Fig. 4- Details of 84 FPs

Databases

The choice of the Database will play a vital role in developing any new recognition or detection system. Testing the new system, comparing it with the other state of the art systems and benchmarking the performance become a very easy and straightforward, if all Researchers are using common databases. However, building such a 'common' database that can satisfy the various requirements of the problem domain and become a standard for future research is a difficult and challenging task. With respect to face

recognition, this problem is close to being solved with the development of the FERET Face Database which has become a de-facto standard for testing face recognition systems. However the problem of a standardized database for face expression recognition is still an open problem. When compared to face recognition, face expression recognition poses a very unique challenge in terms of building a standardized database. This challenge is due to the fact that expressions can be posed or spontaneous. The initial work was carried out by Sebe et al, they listed the major problems that are associated with capturing spontaneous expressions. Their

main observations were as follows:

- Different subjects express the same emotions at different intensities
- If the subject becomes aware that he or she is being photographed, their expression loses its authenticity.
- Even if the subject is not aware of the recording, the laboratory conditions may not encourage the subject to display spontaneous expressions.

Annexure 2-

Database	Sample Details Expression Elicitation and Data Recording	Methods	Available Descriptions	Additional Notes
Cohn-Kanade Database (also known as CMUPittsburg database)	<ul style="list-style-type: none"> • 500 image sequences from 100 subjects • Age: 18 to 30 years • Gender: 65% female • Ethnicity: 15% African - Americans and 3% of Asians and Latinos 	<p>This DB contains only posed expressions. "The subjects were instructed to perform a series of 23 facial displays that included single action units (e.g. AU 12, or lip corners pulled obliquely) and action unit combinations (e.g. AU 1+2, or inner and outer brows raised). Each begins from a neutral or nearly neutral face. For each, an experimenter described and modeled the target display. Six were based on descriptions of prototypic emotions (i.e., joy, surprise, anger, fear, disgust, and sadness). These six tasks and mouth opening in the absence of other action units were annotated by certified FACS coders."</p>	"Annotation of FACS Action Units and emotion specified expressions"	"Images taken using 2 cameras: one directly in front of the subject and the other positioned 30 degrees to the subject's right. But the DB contains only the images taken from the frontal camera"
MMI Facial Expression Database	<ul style="list-style-type: none"> • 52 different subjects of both sexes • Gender: 48% female • Age: 19 to 6 years • Ethnicity: European, Asian, or South American • Background: Natural lighting and variable backgrounds (for some samples) 	<p>This DB contains posed and spontaneous expressions. "The subjects were asked to display 79 series of expressions that included either a single AU (e.g., AU2) or a combination of a minimal number of AUs (e.g., AU8 cannot be displayed without AU25) or a prototypic combination of AUs (such as in expressions of emotion). Also, a short neutral state is available at the beginning and at the end of each expression." For natural expressions: "Children interacted with a comedian. Adults watching emotion inducing videos"</p>	"Action Units, metadata (data format, facial view, shown AU, shown emotion, gender, age), analysis of AU temporal activation patterns"	"The emotions were determined using an expert annotator". A highlight of this DB is that it contains both frontal and profile view images.
Spontaneous Expressions Database	28 subjects	<p>This DB contains spontaneous expressions. The subjects were asked to watch emotion inducing videos in a custom built video kiosk. Their expressions were recorded using hidden cameras. Then, the subjects were informed about the recording and were asked for their consent. Out of 60, 28 gave consent.</p>	The database is self labeled. After watching the videos, the subjects recorded the emotions that they felt.	The researchers found that it is very difficult to induce all the emotions in all of the subjects. Joy, surprise and disgust were the most easy where as sadness and fear were the most difficult
The AR Face Database	<ul style="list-style-type: none"> • 126 people • Gender: 70 men and 56 women • Over 4000 color images are available 	<p>This DB contains only posed expressions. "No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. Each person participated in two sessions, separated by two weeks time. The same pictures were taken in both sessions"</p>		This database has frontal-faces with different expressions, illumination conditions and occlusions (scarf and sunglasses).
CMU Pose, Illumination, Expression (PIE) Database	<ul style="list-style-type: none"> • 41,368 images of 68 people • 4 different expressions. 	This DB contains only posed expressions.		This database provides facial images for 13 different poses 43 different illumination conditions.
The Japanese Female Facial Expression (JAFFE) Database	<ul style="list-style-type: none"> • 219 images of 7 facial expressions (6 basic facial expressions + 1 neutral) • 10 Japanese female models. 	<p>This DB contains only posed expressions. The photos have been taken under strict controlled conditions of similar lighting and with the hair tied away from the face.</p>	"Each image has been rated on 6 emotion adjectives by 92 Japanese subjects"	All the expressions are multiple AU expressions

Let us now look at some of the popular expression databases that are publicly and freely available. Thus, we will be covering only those databases that have mostly been used in the past few years.

Sebe and colleagues consulted with fellow psychologists and came up with a method to capture spontaneous expressions. They developed a video kiosk where the subjects could watch emotion inducing videos. The facial expressions of were recorded by a hidden camera. Once the recording was done, the subjects were notified of the recording and asked for their permission to use the captured images and videos for research studies. From their studies, they found out that it was very difficult to induce a wide range of expres-

sions among the subjects. In particular, fear and sadness were found to be difficult to induce. They also found that spontaneous facial expressions could be misleading. Strangely, some subjects had a facial expression that looked sad when they were actually feeling happy. As a side observation, they found that students and younger faculty were more ready in giving their consent to be included in the database whereas older professors were not. Sebe et al.'s study helps us in understanding the problems encountered with capturing spontaneous expressions and the round-about mechanisms that have to be used in order to elicit and capture authentic facial expressions.

Some of the popular expression databases that are publicly and freely available are listed in [Annexure-2] along with its brief details. In as [Annexure-2], we will be covering only those databases that have mostly been used in the past few years.

Challenges and Future Work

Recognition of Spontaneous Expression

The major challenge in this field is the recognition of spontaneous expressions. Capturing spontaneous expressions on images and video is one of the biggest challenges ahead. If the subjects become aware of the recording and data capture process, their expressions immediately loses its authenticity.

Elicit Emotions

The most common method that has been used to elicit emotions in the subjects is by the use of emotion inducing videos and film clips. While eliciting happiness and amusement is quite simple, eliciting fear and anger is the challenging part. The subjects need to become personally involved in order to display anger and fear. But this rarely happens when watching films and videos. The challenge lies in finding out the best alternative ways to capture anger and fear or in creating (or searching for) video sequences that are sure to anger a person or induce fear in him.

Recognition of all Expressions

Apart from the six prototypic expressions there are a host of other expressions that can be recognized. But capturing and recognizing spontaneous non-basic expressions is even more challenging than capturing and recognizing spontaneous basic expressions. This is still an open topic and no work seems to have been done on the same.

Existence of Differences in Facial Features and Facial Expressions

Differences do exist in facial features and facial expressions between cultures (for example, Europeans and Asians) and age groups (adults and children).

Microexpressions

A possible work for the future is the automatic recognition of micro expressions. A micro expression is a brief, involuntary facial expression shown on the face of humans according to emotions experienced. They usually occur in high-stakes situations, where people have something to lose or gain. Unlike regular facial expressions, it is difficult to fake micro expressions. Micro expressions express the seven universal emotions: disgust, anger, fear, sadness, happiness, surprise, and contempt. However in the 1990s Paul Ekman expanded his list of basic emotions, including a range of positive and negative emotions not all of which are encoded in facial muscles. These emotions are amusement, contempt, embarrassment, excitement, guilt, pride, relief, satisfaction, pleasure, and shame.

Loss of Facial Expression

It is important to know that there exist certain conditions which can cause what is known as '*flat affect*' or the condition where a person is unable to display facial expressions. There are 12 causes for the loss of facial expressions namely: Asperser syndrome, Autistic

disorder, Bells Palsy, Depression, Depressive disorders, Facial paralysis, Facial weakness, Major depressive disorder, Parkinson's disease, Scleroderma, Wilsons disease.

Applications of Automatic Facial Expression Recognition System

The Automation of Facial expression recognition systems find applications in several interesting areas. With the recent advances in robotics, especially humanoid robots, the urgency in the requirement of a robust expression recognition system is evident. As robots begin to interact more and more with humans and start becoming a part of our living spaces and work spaces, they need to become more intelligent in terms of understanding the human's moods and emotions. Expression recognition systems will help in creating this intelligent visual interface between the man and the machine. Humans communicate effectively and are responsive to each other's emotional states. Computers must also gain this ability. This is precisely what the Human-Computer Interaction (HCI) research community is focusing on: namely, Affective Computing. Expression recognition plays a significant role in recognizing one's affect and in turn helps in building meaningful and responsive HCI interfaces. Apart from the above two main applications, namely robotics and affect sensitive HCI, expression recognition systems find uses in a host of other domains like Telecommunications, Behavioral Science, Video Games, Animations, Psychiatry, Automobile Safety, Affect sensitive music juke boxes and televisions, Educational Software, etc. Practical real-time applications have also been demonstrated. Bartlett et al. [6] have successfully used their face expression recognition system to develop an animated character that mirrors the expressions of the user (called the *CU Animate*). They have also been successful in deployed the recognition system on Sony's *Aibo* Robot and ATR's *RoboVie*. Another interesting application has been demonstrated by Anderson and McOwen, called the '*EmotiChat*' [9]. It consists of a chat-room application where users can log in and start chatting. In this, the facial expression recognition system is connected to this chat application and it automatically inserts emoticons based on the user's facial expressions. As expression recognition systems become more real-time and robust, we will be seeing many other innovative applications and uses of this technique.

Conclusion

We started with a time-line view of the various works on Facial expression recognition. We saw some applications that have been implemented and other possible areas where automatic Facial expression recognition can be applied. We then looked at facial parameterization using FACS AUs and MPEG-4 FAPs. Then we looked at some notes on emotions, expressions and features followed by the characteristics of an ideal system. We then saw the recent advances in face detectors and trackers. This was followed by a note on the databases followed by a summary of the state of the art. A note on classifiers was presented followed by a look at the six prototypic expressions. The last section was the challenges and possible future work to be done. Facial expression recognition systems have improved a lot over the past decade. The focus has definitely shifted from posed expression recognition to spontaneous expression recognition. The next decade will be interesting since I think that robust spontaneous expression recognizers will be developed and deployed in real-time systems and used in build-

ing emotion sensitive HCI interfaces. This is going to have an impact on our day to day life by enhancing the way we interact with computers or in general, our surrounding living and work spaces.

References

- [1] Claude C. Chibelushi, Fabrice Bourel (2002) *Facial Expression Recognition: A Brief Overview*, 1-5.
- [2] Montagu J. (1994) *The Expression of the Passions: The Origin and Influence of Charles Le Brun's 'Conférence sur l'expression générale et particulière*, Yale University Press.
- [3] Darwin C. (1904) *The Expression of the Emotions in Man and Animals*, 2nd ed., J. Murray, London.
- [4] Samal A. and Iyengar P.A. (1992) *Pattern Recognition*, 25(1), 65-77.
- [5] Vinay Kumar Bettadapura (2012) *Face Expression Recognition and Analysis: The State of the Art*, Columbia University, 1-27.
- [6] Bartlett M.S., Littlewort G., Fasel I. and Movellan R. (2003) *CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction*, 5.
- [7] Anderson K. and McOwan P.W. (2006) *IEEE Trans. Systems, Man and Cybernetics Part B*, 36(1), 96-105.
- [8] Koenen R. (2000) *MPEG-4 Project Overview, International Organisation for Standardization*, La Baule.
- [9] Schwartz G., Fair P., Salt P., Mandel M., Klerman G. (1976) *Psychosomatic Med.* 38, 337-347.
- [10] Ekman P., Friesen W.V. (1978) *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press, Palo Alto.
- [11] Izard C. (1979) *The maximally discriminative facial movement coding system (MAX)*, Instructional Resource Center, University of Delaware, Newark, Delaware.
- [12] Tsapatsoulis N., Karpouzis K., Stamou G. (2000) *A Fuzzy System for Emotion Classification Based on the MPEG-4 Facial Emotion Parameter*, European Association on Signal Processing EUSIPCO
- [13] Ekman P., Rosenberg E., Hager J. (1998) *Facial Action Coding System Affect Interpretation Database*.
- [14] Pandzic I.S. and Forchheimer R. (2002) *MPEG-4 Facial Animation: The Standard, Implementation and Applications*, John Wiley and Sons.
- [15] Ekman P., Friesen W.V. (1971) *J. Personality Social Psychol.*, 17(2), 124-129.
- [16] Cowie R., Douglas-Cowie E., Karpouzis K., Caridakis G., Wallace M. and Kollias S. (2008) *Multimodal User Interfaces*, Springer Berlin Heidelberg.
- [17] Abrantes G.A. and Pereira F. (1999) *IEEE Trans. Circuits and Systems for Video Technology*, 9(2).
- [18] Fellenz W., Taylor J., Tsapatsoulis N., Kollias S. (1999) *Circuits, Systems, Communications and Computers*, Nugata, Japan, 5331-5336.
- [19] Dailey M., Cottrell G. (1999) *PCA Gabor for Expression Recognition*, Institution UCSD, Number CS-629.
- [20] Zhang Z., Lyons M., Schuster M., Akamatsu S. (1998) *Second International Conference on Automatic Face and Gesture Recognition*, IEEE, Nara, Japan, 454-459.
- [21] Lyons M., Budynek J., Akamatsu S. (1999) *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(12).
- [22] Padgett C., Cottrell G. (1996) *Advances in Neural Information Processing Systems*, 9, MIT Press, Cambridge, MA, 894-900.
- [23] Cottrell G.W., Metcalfe J. (1991) *Advances in Neural Information Processing Systems*, Morgan Kaufman, San Mateo, CA, 3, 564-571.
- [24] Bartlett M. (1998) *Face image analysis by unsupervised learning and redundancy reduction*, Ph.D. Thesis, University of California, San Diego.
- [25] Lien J. (1998) *Automatic recognition of facial expression using hidden Markov models and estimation of expression intensity*, Ph.D. Thesis, The Robotics Institute, CMU.
- [26] Lanitis A., Taylor C., Cootes T. (1997) *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7), 743-756.
- [27] Cootes T., Edwards G., Taylor C. (2001) *Active Appearance Models*, IEEE PAMI, 23(6), 681-685.
- [28] Edwards G., Cootes T., Taylor C. (1998) *Fifth European Conference on Computer Vision*, University of Freiburg, Germany, 2, 581-695.
- [29] Huang C., Huang Y. (1997) *J. Visual Comm. Image Representation*, 8(3), 278-290.
- [30] Otsuka T., Ohya J. (1998) *IEEE Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 442-447.
- [31] Kobayashi H., Hara F. (1997) *International Conference on Systems, Man and Cybernetics*, Orlando, FL, USA, 3732-3737.
- [32] Pantic M., Rothkrantz L. (2000) *Image Vision Comput. J.*, 18 (11), 881-905.
- [33] Mathews I. (1997) *Active Shape Model Toolbox*, University of East Anglia, Norwich, UK, Matlab Toolbox Version, 2.0.0.
- [34] Mase K., Pentland A. (1991) *IEICE Trans.*, E74(10), 3474-3483.
- [35] Wu Y., Kanade T., Cohn J., Li C. (1998) *IEEE Int. Conference on Computer Vision*, Mumbai, India, 992-998.
- [36] Cai W., Wang J. (1996) *Soc. Inc. Application Math*, 33(3), 937-970.
- [37] Simoncelli E. (1993) *Distributed Representation and Analysis of visual Motion*, Ph.D. Thesis, Massachusetts Inst. of Tech.
- [38] Rosenblum M., Yacoob Y., Davis L. (1996) *IEEE Trans. Neural Networks*, 7(5), 1121-1138.

- [39] Abdel Moltaleb M., Chellappa R., Rosenfeld A. (1993) *Computer Vision and Pattern Recognition*, IEEE, 321-327.
- [40] Yacoob Y., Davis L.S. (1996) *IEEE Trans. Pattern Anal. Mach. Intell.* 18(6), 636-642.
- [41] Essa I., Pentland A. (1997) *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7), 757-763.
- [42] Terzopular D., Waters K. (1994) *3rd International Conference on Computer Vision*, Osaka, Japan.
- [43] DeCarlo D., Metaxas D. (1996) *International Conference on Computer Vision and Pattern Recognition*, 231-238.
- [44] Eisert P., Girod B. (1997) *Picture Coding Symposium*, Berlin, Germany, 33-38.
- [45] Black M., Yacoob Y. (1997) *Internat. J. Comput. Vision*, 25(1), 23-48.
- [46] Basu N., Oliver A. Pentland (1998) *International Conference on Computer Vision*, Bombay, India.
- [47] Wang M., Iwai Y., Yachida M. (1998) *Second International Conference on Automatic Face and Gesture Recognition*, IEEE, Nara, Japan, 324-329.
- [48] Otsuka T., Ohya J. (1998) *First International Conference on Advanced Multimedia Content Processing*, Osaka, Japan, 442-453.
- [49] Bascle B., Blake A. (1998) *International Conference in Computer Vision*, Bombay, India.
- [50] Suwa M., Sugie N., Fujimora K. (1978) *Fourth International Joint Conference on Pattern Recognition*, Kyoto, Japan, 408-410.
- [51] Kaiser S., Wehrle T. (1992) *J. Nonverbal Behaviour*, 16(2), 67-83.
- [52] Donato G., Bartlett S., Hager C., Ekman P., Sejnowski J. (1999) *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(10), 974-989.
- [53] Choudhory T., Pentland A. (2000) *International Conference on Pattern Recognition*, Spain.
- [54] Schiele B., Chowley I. (2009) *Machine Intelligence*, 31(1), 39-58.
- [55] Arulampalam M.S., Maskell S., Gordon N. and Clapp T. (2002) *IEEE Trans. Signal Processing*, 50(2), 174-188.
- [56] Choi S. and Kim D. (2007) *Image Analysis and Recognition, Lecture Notes in Computer Science*, 4633, 548- 557.
- [57] Xu F., Cheng J. and Wang C. (2008) *Int. Conf. Automation and Logistics*, IEEE, 2252-2255.
- [58] Wang H., Chang S. (1997) *IEEE Trans. Circuits and Systems for Video Technology*, 7(4), 615 -628.