# Chapter 9
# Facial Expressions to Emotions: A Study of Computational Paradigms for Facial Emotion Recognition

**Dipti Deodhare**

## 9.1 Introduction

The crown jewel of nonverbal communication is facial expression. Facial expressions are reflective of the emotional state of a person's mind and provide nonverbal cues for effective everyday communication. Apart from indicating affective state, they also lend insight into a person's personality and psychopathology. Facial expression cues usually complement speech, enabling a listener to better elicit the intended meaning of spoken words.

Facial emotion recognition is one of the final frontiers of the man–machine interface. Computational methods that enable this capability in machines are being actively researched and are gaining significant importance particularly when taken in context to the notion of *technological singularity.* Technological singularity or simply singularity is currently a much discussed and hotly debated topic by philosophers, computer scientists, physicists, etc. Simply speaking, technological singularity is a theoretical prediction that artificial intelligence (AI) would have progressed to a greater than human intelligence, and machines driven by AI would have radically changed human civilization and even biology and intelligence as the way we know it. Irrespective of whether one agrees with this view or not, a machine's ability to read and interpret human emotions using visual cues clearly creates an important new paradigm in computational advancement. Cognitive systems that respond to human behavior, and intelligent systems that attempt to interact with humans, need to incorporate the ability to read and interpret human facial expressions. Use of computational technologies for recognition and interpretation

D. Deodhare (✉)
Centre for Artificial Intelligence and Robotics, Bangaluru, India
e-mail: dipti.deodhare@gmail.com



M.K. Mandal and A. Awasthi (eds.), *Understanding Facial Expressions in Communication*, DOI 10.1007/978-81-322-1934-7_9

of facial emotions is a thriving area of research and has drawn substantial attention from researchers for over a decade now. Literature survey reveals a vast body of research in the field. Various methods and tests have been formulated by researchers over the last few years, and today, these methods demonstrate some ability to perform interpretations of human expressions and emotions. In what follows, we touch upon some of the landmark contributions in this rigorously researched field.

This chapter is organized as follows: In this, the Introduction Section, besides making some general contextual comments, we also introduce one of the most popular methods for detection of faces in an image. Needless to say, detection of faces in an image is the first important step towards detection of emotions from facial expressions in a computational process. In Sect. 9.2, we give a brief overview of the various approaches discussed in recent literature for automatic detection of emotions from facial expressions. Section 9.3 discusses mechanisms for recognizing emotions from faces, as proposed by psychological and neurological studies. Tools and constructs from computer science that maybe relevant in realizing these theories on a computer are briefly discussed. In Sect. 9.4, we delve into the algorithmic and mathematical details of how typical automatic algorithms are constructed to obtain emotions from images of faces. These technologies are heavily anchored in advanced AI techniques such as neural networks, machine learning, particle swarm optimization, genetic algorithms, and principal component analysis. For the sake of completeness, simple and intuitive descriptions of these techniques have been included where relevant. Section 9.5 presents a specific algorithm from the literature that describes a mechanism of identifying emotions from faces in videos. Finally, Sect. 9.6 offers some concluding comments.

Several technological approaches have been proposed for facial emotion recognition to classify human emotions successfully. Most of these approaches focus on *seven* basic emotions owing to their constancy across culture, age, and other identities. These emotions are as follows: joy, sadness, anger, surprise, disgust, fear, and neutrality. Facial hair, eye glasses, and headwear, etc, affect computational analysis. Further complexities emerge when the subject has a facial injury, or the face is unnaturally constricted due to tight turbans or scarves. Finally, the context of the emotion being expressed also becomes a critical aspect of computational classification. The computational classification may not show the actual emotions in case the subject becomes aware of scrutiny and surveillance and if the subject tries to hide or mask his or her emotions. It is, therefore, critical that such computational analysis techniques and algorithms be developed in close concert with experts in the field of psychology and psychiatrics to arrive at an accurate analysis.

Adolphs (2002) showed in his work that recognition requires some knowledge about the world; it thus requires memory of some sort. One of the simplest forms of recognition is recognition memory, which basically involves the ability to retain information about the perceptual properties of a visual image, to which another image can be compared. This form of recognition may be sufficient to discriminate between two faces that are presented at separate points in time. For an AI-based system to "recognize" emotions, it needs a large memory bank of correlational database. The efficiency with which visual cues are detected, compared,

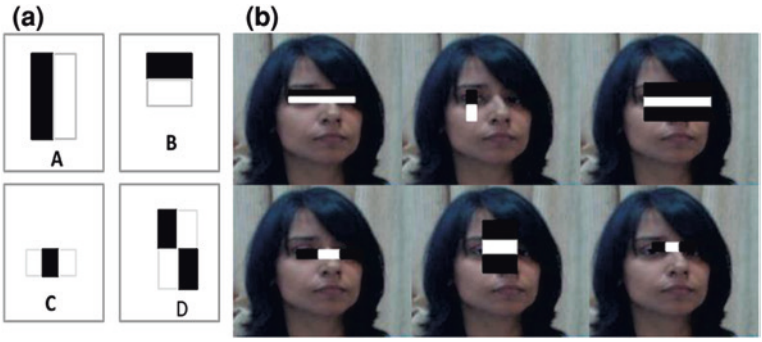



**Fig. 9.1** **a** Four basic Haar-like feature extractors. **b** Feature extraction from an image

and lead to an inference is a direct function of the size of the database. Unless large volumes of data are made available to the algorithm for training and correlation, computational intelligence algorithms and reasoning methods cannot produce accurate results. Every instance of an analysis revealed as accurate can be added to the database and gradually help the system "learn." Emotion recognition involves three stages of computation.

Step 1: Face Detection—The Viola–Jones face detection algorithm is of importance and is discussed below.
Step 2: 3D modeling of the face—The active appearance method described by Cootes and Taylor is popular, though other methods are also described here.
Step 3: Computational intelligence-based analysis, using a large databank (at least 2,000 annotated images).

The Viola–Jones algorithm is a popular algorithm for face detection that can detect faces in an image in real time. Proposed by Viola and Jones (2001), the approach, in fact, describes a generic framework for object detection. However, it was first motivated by the problem of face detection in a digital image. The Viola–Jones algorithm works by looking for Haar-like features. All the features are rectangles. Four basic feature extractors are shown in Fig. 9.1a. These rectangular Haar features are positioned at various locations in a detection window. The sum of the pixel intensities in the region below the rectangles is computed. The difference between the sums of adjacent rectangles is then calculated to obtain the feature. The algorithm works on the premise that a human face has contrast patterns organized in a particular manner. For instance, the hairline is darker than the forehead; the eyes are darker than the cheeks, etc. Each of these rectangular patterns is compared with a portion of the image underneath (see Fig. 9.1b) to check how much that particular part of the image matches that pattern. This way the algorithm basically looks for a signature contrast pattern that might be one particular part of the face. To begin with, the tests are coarse so that parts of the image that have no contrast are eliminated very quickly. The feature detection mechanism iterates as a cascade so that more matching requirements are added in more specific spaces.

If a combination of matches passes a threshold comparison, that portion of the image is assumed to potentially contain a face. Each Haar-like feature gives a very weak cue to the potential existence of a face in the image. Therefore, a very large number of Haar features are organized into a classifier cascade to form a strong classifier. A special representation of the image sample, called the integral image, makes feature extraction faster. Typically, a basic classifier operates on $24 \times 24$ sub-windows. To detect faces at different scales, the detectors are scaled usually by factors of 1.25. Hence, features are easily evaluated at any scale. The detectors are moved around the image, by one pixel increments, for example. As a result, a real face may result in multiple nearby detections. To suppress these multiple detections into a single detection, the detected sub-windows are post-processed to combine overlapping detections.

A computational technique for facial emotion recognition, essentially, maps key points of the face and derives information from the geometry of these points. The features for emotion recognition can be static, dynamic, point-based (geometric), or region-based (appearance). Geometric features are extracted from the shape of important facial components, such as the mouth and eyes, using salient point locations. Facial states, such as the state of eyes, shape of key points of the mouth and eyebrows, and head orientation, are some of the aspects that are used for analysis. The computational approach needs large volumes of data, essentially graphic images with tags and metadata, which can then be used to correlate the expressional characteristics of the target subject. The expression identification is complicated by possibly small inter-subject variations. Various imaging parameters, such as aperture, exposure time, and lens aberrations, can compound the problem by increasing the intra-subject variability. All these factors are confounded in the image data so that variations between the images of same face are almost always larger than image variations due to the change in face identity. There are a number of face databases available which incorporate these variations in images, for example, Japanese Female Facial Expression (JAFFE) and Cohn—Kanade AU coded facial expression database. Care has to be taken to group data phylogenetically, based on gender, age, ethnicity, culture, geographical origin, literacy levels, and background of the subject, etc.

In the next section, we proceed to give an overview of various approaches that have been attempted to establish algorithms for recognition of emotions from faces in an image.

## 9.2 An Overview of Computational Algorithms for Emotion Recognition

There is a vast body of literature on the topic of face recognition and analysis. Zeng et al. (2009) classified facial features for the purpose of feature recognition under two broad categories: (i) geometric features and (ii) appearance-based features. Geometric features are extracted from the shape or salient point locations

of some important facial components, such as mouth and eyes. Suwa et al. (1978) presented an early attempt to automatically analyze facial expressions by tracking the motion of twenty identified spots on an image sequence. In the work of Chang et al. (2006), 58 landmark points were used to construct an active shape model. These landmark points were then used to obtain valid facial emotions across faces.

Appearance-based features, such as *texture,* have been employed in the work of Lyon and Akamatsu (1998) using Gabor filters. In computer programming, a filter is a program or a section of code that is designed to examine each input for certain qualifying criteria and then process or forward it accordingly. Gabor filters are band-pass filters that allow signals between two specific frequencies to pass, but discriminate against signals at other frequencies. These filters are used in image processing for feature extraction, texture analysis, and stereo disparity estimation. A set of Gabor filters with different frequencies and orientations can be helpful for extracting useful features from an image. Gabor filters gained importance when it was discovered that simple cells in the visual cortex of mammalian brains can be modeled by Gabor functions (Daugman 1985). Thus, image analysis by the Gabor functions is similar to perception in the human visual system. The effort proposed a methodology for coding facial expressions with multi-orientation and multi-resolution set of Gabor filters that are ordered topographically and aligned approximately with the face. Similarity values and semantic ratings were calculated using the Gabor coding. The degree of correlation obtained was significantly high, without any parameter fitting. The work of Gu et al. (2012) is also motivated by some characteristics of the human visual cortex (HVC). The authors proposed a new scheme for facial expression recognition, which involved the statistical synthesis of hierarchical classifiers. The method subjects images to multi-scale, local Gabor filter operations. The resulting images are encoded using radial grids. This is similar to the mapping structure displayed in the HVC. The coded images are further processed by local classifiers to result in global features, which represent facial expressions. This hybrid approach that combines the HVC mapping structure with a hierarchy of classifiers has been experimentally demonstrated to improve accuracy on a large number of databases. Lyons et al. (1998) coded facial expression images using a multi-orientation, multi-resolution set of Gabor filters, which are aligned approximately with the face. They derived the similarity space from this representation and compared it with the similarity space derived from the semantic ratings of images by human observers. This representation using Gabor filters shows a noteworthy consistency with human psychology, which can prove very relevant for human–computer interface design.

Feature geometry is an explicit and precise function of facial deformation occurring due to expression, but it does not capture any textural changes. Although addition of grid points enhances the performance of a geometric measure, the overall computational complexity increases exponentially. Azcarate et al. (2005) devised an algorithm based on the piecewise Bezier volume deformation tracker along with a Haar face detector, which was used to initially locate the human face automatically. Experiments in this work were conducted with Naive Bayes and the Tree-Augmented-Naive Bayes (TAN) classifiers in person-dependent and

person-independent tests on the Cohn—Kanade database. The results obtained through this approach were better for person-dependent experiments as compared to those for person-independent experiments.

Habibizad and Mirnia (2012) proposed an algorithm to classify emotions through eyes and lips using particle swarm optimization. Particle swarm optimization (PSO) is an artificial intelligence technique to find optimal solutions to really complex problems. In this approach, a population of candidate solutions is represented by a swarm of particles. These particles move around in the search space based on simple mathematical formulae. These formulae take into cognisance a particle's best-known position so far in the search space and guide the particle toward better-known positions in the search space as updated using results contributed by other particles in the swarm. Convergence capabilities of PSO algorithms are demonstrated through empirical results, since there is really no sound mathematical analysis of their convergence properties. PSO algorithms do not require the objective function of the optimization problem to be differentiable. They can search very large spaces and hence can be used in domains where the problems are noisy, change over time, etc. This approach was employed to optimize eyes and lips elliptical characteristics. The results obtained in this approach suggest a high success rate and a high processing speed. Ioannou et al. (2005) proposed an algorithm to recognize emotions using a neuro-fuzzy approach. This approach is robust to facial expression variations among different users. Mpiperis et al. (2008) present a novel approach for expression recognition, which was inspired by the advances in an ant colony and particle swarm optimization techniques. Anatomical correspondence between faces was established using a genetic 3D face model, which was deformed elastically to match the facial surfaces. They achieved a recognition rate of 92.3 % with the BU-3DFEDB database. Kaushik and Mohamed (2012) in their latest research proposed a new lip boundary localization scheme using game theory to elicit lip contour accurately from a facial image. They applied a feature subset selection scheme based on particle swarm optimization to select the optimal facial features. They could achieve recognition rates of 93 % on the JAFFE database. Ghandi et al. (2009) presented an approach which was based on tracking the movements of facial action units placed on the face of a subject. They defined some swarm particles, so that they have a component around the neighborhood of each action unit.

Huang et al. (2012) proposed an algorithm for emotion recognition by a novel triangular feature extraction method that uses statistical analysis and genetic algorithms to extract a set of optimal triangular facial features. In artificial intelligence, genetic algorithms are a type of evolutionary algorithms that help obtain solutions to complex optimization problems, through mechanisms inspired by nature's process of natural selection. The method starts off with generating a set of feasible solutions that acts as the initial population. The objective function of the optimization problem acts as what is described as the fitness function and helps assess if particular samples in this parent population are good. The best are selected to generate a new set of offspring by selecting two parents and taking part of the solution they represent and combining them to create a new possible solution. This

mechanism is described as crossover or recombination. Sometimes, the elements in the offspring are slightly changed to represent the mechanism of mutation. Mutation is caused in nature by errors in copying genes from parents. If the new offspring are feasible solutions to the objective function, they are accepted. The process starts again with a suitably selected new population. The algorithm stops if the end criterion is met or a preset number of iterations are completed. The best fitting sample in the current population is returned as the solution. The emotion recognition algorithm proposed is claimed to be robust against noisy features and feature rotations. It also shows a significant dimension reduction in facial features.

Londhe and Pawar (2012) proposed a statistics-based approach combined with artificial neural network (ANN) techniques for analyzing facial expressions. An artificial neural network is a computing model that seeks to emulate the style of computing of the human brain. Its architecture comprises a massively parallel interconnection of simple units called "neurons." These interconnections are weighted, and the long-term knowledge of the network is encoded in the strengths of these connections. Depending on their architecture, neural networks can be classified into two basic types: (i) *feedforward* neural networks and (ii) *recursive* or *recurrent* neural networks. In feedforward networks, signals flow in only one direction, and hence, the network can be represented by an acyclic graph. On the other hand, in recursive networks, a unit may be influenced by its own output directly or indirectly through other units. The *multilayer perceptron* or the MLP is the most widely used feedforward network. It falls in the category of static networks, i.e., their output is a function only of the current input and not of past and future inputs or outputs. In the referred paper (Londhe and Pawar 2012), the authors have studied the changes in the curvatures on the face and the intensities of corresponding pixels of images. They have used statistical parameters to compute these changes, and the computed results were recorded as feature vectors. An ANN was used to classify these features into six universal emotions such as anger, disgust, fear, happiness, sadness, and surprise. A two-layered feedforward neural network was trained and tested using the scaled conjugate gradient back-propagation algorithm to obtain a 92.2 % recognition rate.

Saudagare and Chaudhari (2012) gave an overview of facial expression recognition techniques using neural networks. They used neural networks for face recognition, feature extraction, and categorization. Karthigayan et al. (2008) used the eye and lip regions for the study of emotions. They performed their study on a Southeast Asian face database. They applied genetic algorithms to get the optimized value of the minor axis of an irregular ellipse corresponding to the lips and the minor axis of a regular ellipse related to the eye. Their successful classification went to a maximum of 91.42 %. In order to classify six basic facial expressions of emotions, Dailey et al. (2002) proposed a simple yet plausible neural network model. Their model matched a variety of psychological data on categorization, similarity, reaction times, and recognition difficulty without any parameter tuning. Kobayashi and Hara (1992) investigated the methods of machine recognition of human expressions and their strength. They used back-propagation algorithm for neural network learning and obtained a correct recognition ratio of 90 %. Anam et al. (2009) proposed

a face recognition system for personal identification and verification using genetic algorithm and back-propagation neural network. They applied some preprocessing on the input images followed by facial features extraction. These features were taken as the input to the neural network and genetic algorithm for classification. Agrawal et al. (2011) proposed a highly efficient facial expression recognition system using PCA, optimized by a genetic algorithm. They could achieve reduced computational time and comparable efficiency. Yen and Nithianandan (2002) proposed an automatic feature extraction method that was based on edge density distribution of the image. The face is approximated to an ellipse in the preprocessing stage. Consequent to this, a genetic algorithm is applied to search for the best ellipse region match.

Busso et al. (2008) describe visual-feedback-based emotion detection for natural man–machine interaction. Their paper introduces an emotion detection system realized with a combination of a Haar cascade classifier and a contrast filter to detect and localize facial features. The detected feature points are then used to estimate the emotional state using the action units. Based on the exact position of these features, a probability for each of the six basic emotions—fear, happiness, sadness, disgust, anger, and surprise—is assigned. The final test experiments with reference images show correct detection rates of about 60 % for the emotions happiness, sadness, and surprise.

This section has presented some of the more discussed methods in the literature for emotion recognition. The literature landscape is indicative of the state of the art. A large variety of mechanisms have been tried out, and there are several researchers across the globe working in this domain. However, computational paradigms still remain fragile and much more needs to be done to establish robust paradigms. It is not difficult to motivate the conjecture that mechanisms for emotion recognition will also have to include large knowledge bases and fast learning and reasoning mechanisms that operate on them in real time to provide the necessary background and contextual knowledge that would be important to arrive at the correct classification of the emotion through a computational paradigm. The next section discusses this aspect briefly.

## 9.3 Mechanisms for Recognizing Emotion from Faces

While in an attempt to engineer robust computational paradigms for detecting emotions from facial expressions, we are not necessarily committed to biological plausibility; literature from psychological and neurological studies can contribute effectively in the generation of algorithms. Mechanisms for recognizing facial emotions are tied to specific neural structures and their interconnections. A given brain structure typically participates in multiple strategies. Thus, a recognition task needs disparate strategies and, hence, disparate sets of neural structures. The strategies suggested by Adolphs (2002) are outlined below with brief companion comments on the relevant computational mechanisms that could be potentially used and even effectively combined to realize these mechanisms on the computer.

### 9.3.1 Recognition as a Part of Perception

The first strategy is to consider recognition as a part of perception. Recognition of basic topographies of a stimulus, and thus recognition that one stimulus differs from another, is fundamentally a matter of perception. To recognize an emotion, we need to be able to discriminate, categorize, and identify emotions on the basis of the geometric visual properties of a stimulus image.

Computer models of psychological studies are evidence that meaningful processing can be carried out from the information present in the geometric properties of a human stimulus. Mathematical analyses reveal that the structure present in images of facial expressions is sufficient in principle to generate some of the structure of the emotion categories that humans perceive (Calder et al. 2001). Network models can judge a sharp perceptual difference between different expressions, even when the expressions are structurally very similar, provided they straddle the boundary of an emotion category (analogous to the way in which we segment a rainbow into bands of color despite linear changes in wavelength). Categorization of morphed images generated from the expressions of two different emotions has been explored in normal subjects (Calder et al. 1996; de Gelder et al. 1997; Etcoff and Magee 1992) and has been investigated in a neural network model trained to classify faces (Cottrell et al. 2001; Padgett and Cottrell 1998).

### 9.3.2 Recognition Through Associated Knowledge

Recognition typically involves more than just perceptual information. It has associated information. When we see a facial expression, we associate it with a particular type of event or occurrence—past, present, or future (expected). This knowledge is not present in the topographical structure of the stimulus; it is retrieved from our past experience with the emotion (and to a limited extent, may even be present innately). To obtain comprehensive knowledge of the emotion being experienced, we need the means to train the network to store such associated knowledge either as metadata with the emotion data, or in any other form, and retrieve the same when the emotion is encountered. The means to reconstruct with accuracy the knowledge associated with an emotion is a complex aspect of AI-based systems and deals with representation of knowledge.

Knowledge representation and reasoning are important branches of symbolic artificial intelligence and aim to design intelligent computer systems that can reason on machine interpretable representations of knowledge and arrive at conclusions autonomously. Knowledge representation is a substitute representation of the real world that enables determining consequences by “thinking” rather than “acting”, i e., by reasoning about the world rather than taking action in it. It is a set of ontological commitments that embody what is important and what can be ignored. Philosophically, ontologies are specifications of what exists or what can be said

about the world. They are discussed in the context of the science of being. Modern AI and natural language processing mostly define ontologies to be hierarchical knowledge structures that represent relations between various entities and their combinations, parts/wholes, sets, and individuals. Going back to our discussion on recognition through associated knowledge, ontological representations are best suited for such a task. The advantage of an ontological representation is its ability to build associations across different categories of parameters, which may be geometric, lexical, or semantic, involving varying data structures. This provides a neural scheme for implementing the above representation mechanisms, which can bind information between separate neural representations, so that they can be processed as components of knowledge about the same concept. Extensive feedback connections as well as feedforward connections between different neural regions are needed for integrating the neural representations that are spatially separated in the brain. Another advantage is the ability to build on associated knowledge through "learning" and "training." The representation of the stimulus and its associated knowledge evolves abreast. One continuously modulates and is simultaneously influenced by the other, and perception and recognition become coupled parts of the same process.

### *9.3.3 Recognition Via Generation of a Simulation*

The above two mechanisms are direct methods, which categorize and link together the various components of perceptual and conceptual knowledge about an emotion, signaled by a stimulus. This provides all the information, and all that is now required to complete the recognition task is to reconstruct the perceptual and conceptual knowledge and provide a categorized inference. But imagine a situation where the explicit knowledge obtained from an expressed emotion is itself insufficient to trigger recognition. This may be because the particular emotion has never been encountered before, or the associated knowledge available in the network is insufficient to reconstruct the emotion. An indirect method—*simulation*—is found to succeed in such instances.

Simulation uses the concept of inverse mapping to generate some conceptual knowledge and thereby trigger the states normally antecedent to producing a given facial expression. This is also the concept of *synthesis*, explained below. By simulating an emotional state (based on a partially informed presumption) and generating the motor representations associated with that emotion, this mechanism attempts to trigger conceptual knowledge within the network and complete the task of recognition. Simulation thus provides a mechanism to trigger conceptual knowledge, but the trigger is not the motor stimulus of an easily recognizable emotion, but a conceptual presumption (based on a superficial recognition) that may trigger a full-blown recognition.

This undoubtedly is a hugely complex and advanced recognition task, and it requires extensive inherent conceptual knowledge to be contained within the

network. The simulation hypothesis is actually modeled on the biological model (as many computational models are), wherein a suggestion of an emotion triggers the emotion. In the experimental findings of Rizzolatti et al. (1996), they have shown that in the pre-motor cortex of monkeys, neurons not only respond when the monkey prepares to perform an action itself but also respond somatotopically, when the monkey observes the same visually presented action performed by someone else (Gallese et al. 1996; Rizzolatti et al. 1996).

The theory of confabulation, as described in an article at www.scholarpedia.org, offers a detailed comprehensive explanation of the mechanism of thought, e.g., vision, reasoning, language, cognition, planning, origin of thought process, and hearing in humans and other vertebrates and also potentially in some invertebrates, such as bees and octopi. This theory estimates that the gray matter of a human cerebral comprises roughly 4,000 largely mutually disjoint, localized modules. Each module has an area of roughly 45 $mm^2$. It is further conjectured that a process of genetic selection connects pairs of these modules through knowledge bases of which humans have roughly 40,000 in number. These are rough estimates and would of course vary from individual to individual. The individual module and knowledge base also include a uniquely dedicated, small zone of thalamus. These modules and knowledge bases constitute the thought hardware. The principle of computation is called confabulation and is designed to *maximize cogency*. Simply speaking, confabulation theory works by processing lots of information and then from this information finds out which symbols belong together. Those symbols that are often seen together constitute the information contained within the network. The proponent of this theory Hecht-Nielsen (2007a, b) describes this as the duck test—if a duck-sized creature quacks like a duck, walks like a duck, swims like a duck, and flies like a duck, then we accept it as a duck, because duck is the symbol that most strongly supports the probability of these assumed facts being true; there is no logical guarantee that this creature is a duck; but maximization of cogency makes the decision that it is and moves on. Confabulation can now produce new associations by continuously generating possible symbols based on the context that is seen prior to that. In effect, the confabulation theory uses an architecture that can produce entirely new associations, which are plausible in the context by maximizing cogency.

Computational approaches to interpret facial emotions are based on the mechanisms of *analysis* and *synthesis*. Broadly speaking, in analysis, given a facial expression, computational mechanisms identify what the underlying emotion is. This is a direct classification task based on available data sets. There is a need to examine each aspect of classification and organize data before computational techniques can be applied. In the synthesis approach, soft embodied agents are created, which generate the facial expression based on the emotions presented. The synthesis approach acquires importance when the explicit information from stimulus cannot be recognized by the network. There is, therefore, a need to build a trigger to enable further evaluation. An important aspect of the synthesis approach is to give weights to each of the multiple emotions the subject may be expressing, as humans rarely feel just one emotion at a given point in time. They might

be blended based on the weights given to each of the basic emotions from a predefined set. In both approaches, when a face is presented, the system identifies its closeness to a cluster of expressions in a generated bank of the same to infer the blend of expression. Computational solutions that can accurately identify the emotion of the target subject may need a hybrid of the classification and synthesis approach. Some of the mechanisms are discussed in the following sections.

## 9.4 Basic Computational Processes

In this section, we plunge into the actual computational mechanisms that are the current state of the art for recognition of emotions from facial expressions. The basic computational process of recognizing emotions consists of two phases—**Training** the algorithm and **Classification** of data. Training consists of *labeling* and *modeling*, while Classification consists of *model-fitting* and *emotion classification*. In labeling, facial images of different emotions are collected from databases, and the landmark points of face are hand-labeled in a prescribed manner. These set of landmark points are then fed into the modeling stage where shape models for each class of emotions are constructed. The shape models of each class of emotions, its corresponding mean, and the eigenvectors are stored in files for further reference. The mean face and the eigenvectors for each class of emotions are then read by the model-fitting module, and a test image is fed into this module. Once an appropriate representation is obtained from the model-fitting stage, the final emotion classification stage is carried out.

Cootes (2000) devised the active shape model for modeling. The Cootes method for modeling used in conjunction with the Euclidean distance model for final emotion recognition is reproduced in Fig. 9.2.

### *9.4.1 Training*

Training consists of labeling, shape modeling, alignment of shapes, and model extraction using principal component analysis. Each of these steps is discussed in some detail below.

**Labeling**—Labeling is a stage in which each landmark point of a face is annotated manually (A landmark is a point of correspondence on each object that matches between and within populations). Figure 9.3 is an example of a hand-labeled facial image. A face is represented as a set of $n$ landmark points defined in (usually) two or three dimensions. It is any shape that is defined as the quality of the configuration of points and is invariant over the Euclidean similarity transformation.

Shape is all the geometrical information that remains when location, scale, and rotational effects are filtered out from an object. Mathematically, a shape is defined by $n$ landmark points in $k$-dimensional space and is represented by a $nk$

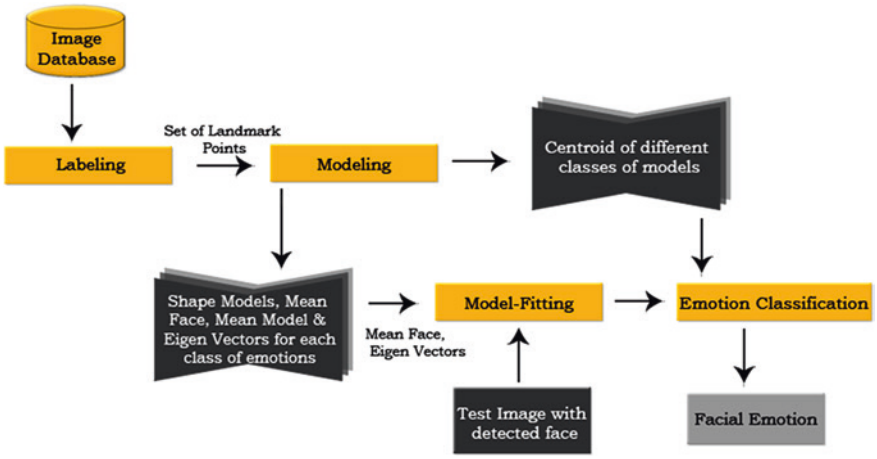
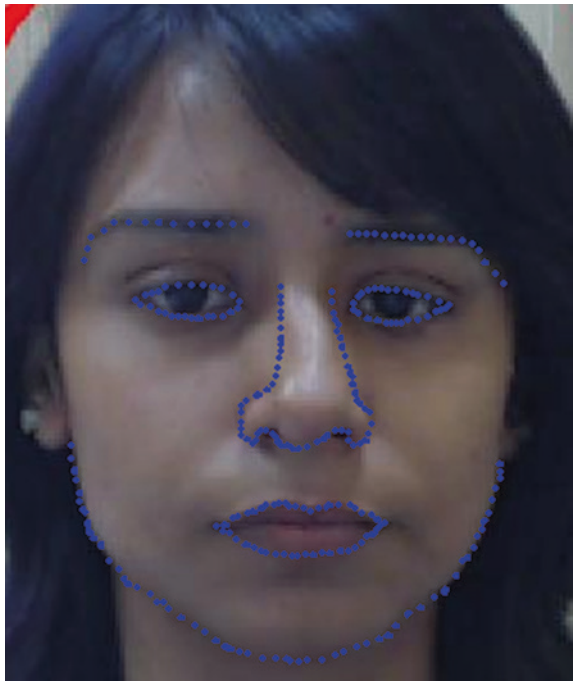



**Fig. 9.2** Cootes methodology for emotion recognition

**Fig. 9.3** Example of a hand-labeled image



vector. In two-dimensional images ($k = 2$), $n$ landmarks, $\{(x_i, y_i): i = 1,\ldots, n\}$, define the 2 $n$ vector $\boldsymbol{x}$ as follows:

$$\boldsymbol{x} = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)'$$

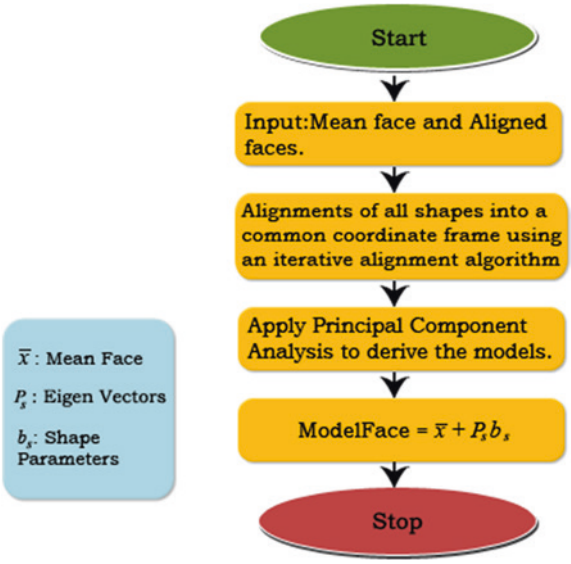



**Fig. 9.4** Flowchart of shape modeling

**Shape Modeling**—In shape modeling, face models for different emotions are constructed. Models for each emotion are varied in terms of shape parameters. Different landmark points are selected for different emotions and are varied between specific ranges to derive models for each emotion. The algorithm for shape modeling is as follows (Fig. 9.4):

**Alignment of Shapes**—The alignment of two shapes comprises of finding the similarity parameters (scale, rotation, and translation) that best maps one shape to another by minimizing a given metric. A classical solution to align two shapes is the **procrustes analysis** method. Procrustes analysis is the most popular approach to aligning shapes in a common reference frame. This aligns each shape such that the sum of the distances of each shape to the mean ($D = \sum |\boldsymbol{x}_i - \bar{\boldsymbol{x}}|^2$) is minimum. An iterative approach for aligning shapes into a common coordinate frame is depicted in the flowchart below (Fig. 9.5):

Convergence is declared if the estimate of the mean does not change significantly after a single iteration. On convergence, all the examples are aligned in a common coordinate frame and can be analyzed for shape change. After the alignment of faces, principal component analysis is performed on the aligned faces.

**Model Extraction Using Principal Components Analysis (PCA)**—Principal component analysis (PCA) is a statistical technique for data dimensionality reduction. In this method, the directions in the data that have the largest variance are searched for. The data are projected along directions of large variance resulting in a linear transformation of data into a new coordinate system that orients the axes based on the spread of the data in the high dimensional space. Coordinates along the axes with high variance are taken and the remaining are ignored. This results in dimensionality reduction.

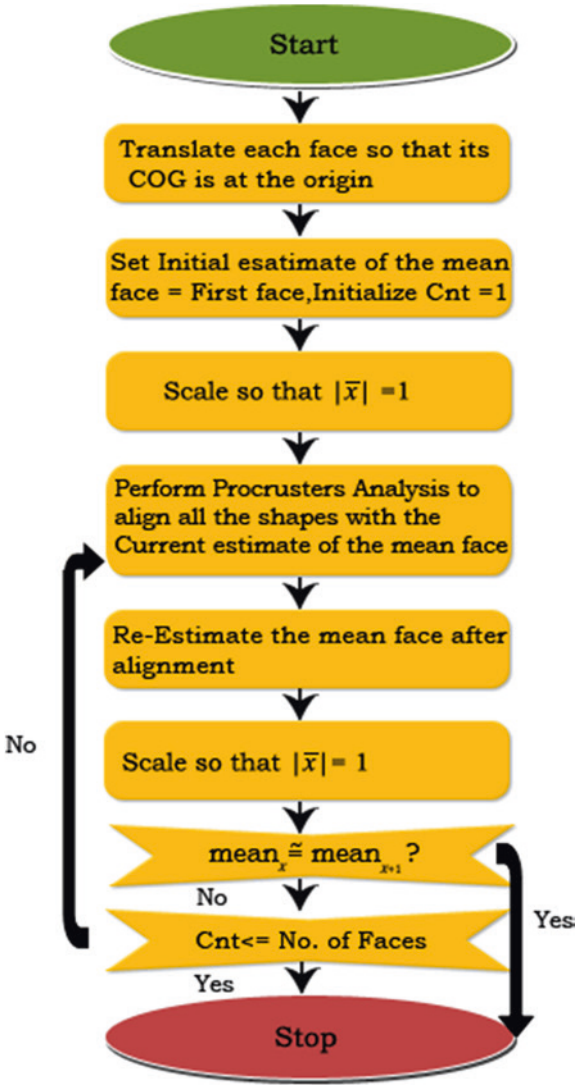



**Fig. 9.5** Flowchart for alignment of shapes

Consider a data set with $N$ vectors $\boldsymbol{x}_i$: $i = 1,\ldots, N$, where each $\boldsymbol{x}_i$ is an $n$-dimensional vector. PCA is performed in the following manner:

(i) Compute the $N$ vectors average,

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i$$

(ii) Subtract the mean from each data vector to form the matrix **D** as follows:

$$\mathbf{D} = ([\boldsymbol{x}_1 - \bar{\boldsymbol{x}}]| \ldots | \ldots |[\boldsymbol{x}_N - \bar{\boldsymbol{x}}])$$

(iii) Calculate the covariance

$$\boldsymbol{S} = \frac{1}{N}\mathbf{D}'\mathbf{D}$$

$\boldsymbol{S}$ is an $N \times N$ matrix. Calculate the eigenvectors $\boldsymbol{P_s}$ of $\boldsymbol{S}$. $\boldsymbol{P_s}$ should be orthonormal.

(iv) Calculate the shape parameters using the following equation

(v) Model face can be calculated using the following equation

$$\boldsymbol{b_s} = \boldsymbol{P_s}(\boldsymbol{x} - \bar{\boldsymbol{x}})$$

$$\boldsymbol{x}_{\text{model}} = \bar{\boldsymbol{x}} + \boldsymbol{P_s} * \boldsymbol{b_s}$$

Variations in shapes are usually incorporated by varying each element of $\boldsymbol{b_s}$ between [±3] (Fig. 9.6).

### *9.4.2 Classification*

The Classification process includes model-fitting and emotion classification.

**Model-Fitting**—Model-fitting is performed to obtain a suitable representation of a new image for further classification. In this stage, every test image fed into the module is hand-labeled to obtain the geometrical features or the landmark points of the face. The labeling of landmark points is performed in an orderly fashion and in a consistent manner. These landmark points are then aligned with the mean face using procrustes analysis. Using this mean face, eigenvectors of the models, and the landmark points obtained after alignment of the shapes, the shape parameters are obtained. Then, using an iterative approach, a model is created using the equation described in step (v) in the previous section. The process of model-fitting is depicted in the Fig. 9.7:

**Emotion Classification**—This is the final stage, in which the model representation obtained from the model-fitting stage is used for further classification under one of the 5 classes of universal emotions, namely *neutral, joy, sadness, surprise,* and *anger.* Emotion classification is carried out by calculating the Euclidean distance between the centroid of each group and the model obtained from the model-fitting stage. The image is placed under the class of emotion where the Euclidean distance between the representation of the test image and the mean model of that class is a minimum. In cartesian coordinates, if $\boldsymbol{p} = (p_1, p_2, \ldots, p_n)$ and $\boldsymbol{q} = (q_1, q_2, \ldots, q_n)$ are two points in Euclidean $n$-space, then the distance from $\boldsymbol{p}$ to $\boldsymbol{q}$ or from $\boldsymbol{q}$ to $\boldsymbol{p}$ is given by:

$$d(\boldsymbol{p}, \boldsymbol{q}) = d(\boldsymbol{q}, \boldsymbol{p}) = \sqrt{(\boldsymbol{q}_1 - \boldsymbol{p}_1)^2 + (\boldsymbol{q}_2 - \boldsymbol{p}_2)^2 + \cdots + (\boldsymbol{q}_n - \boldsymbol{p}_n)^2}$$

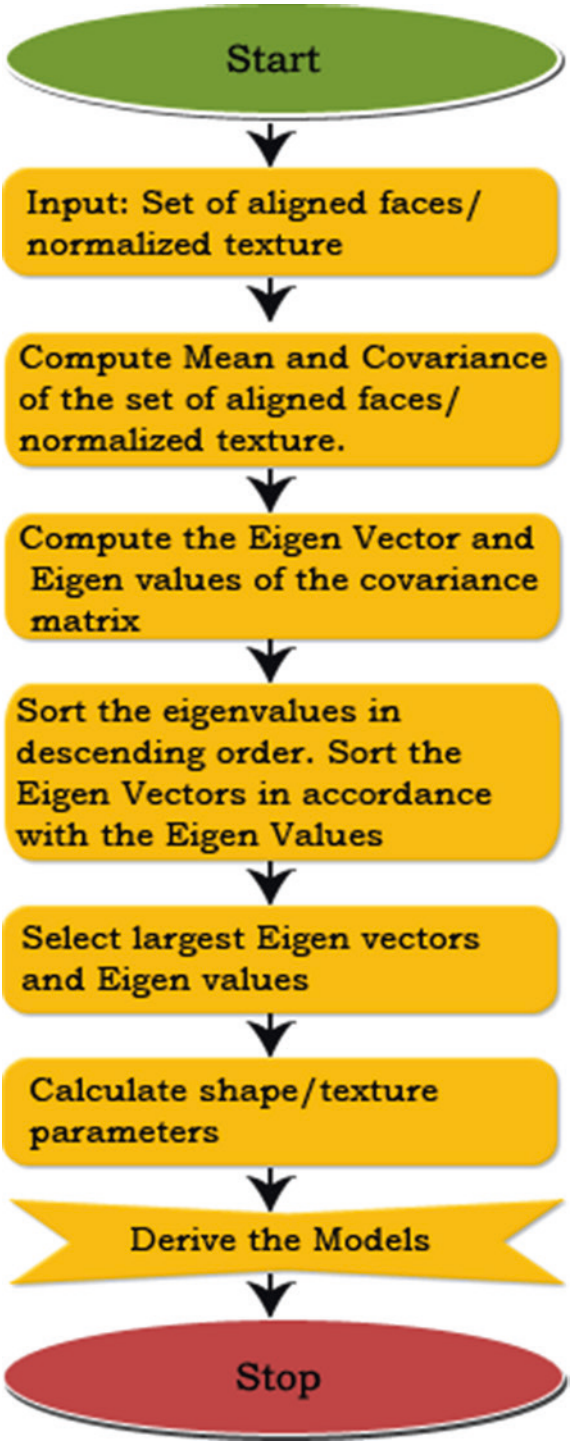



**Fig. 9.6** Flowchart depicting the process of PCA

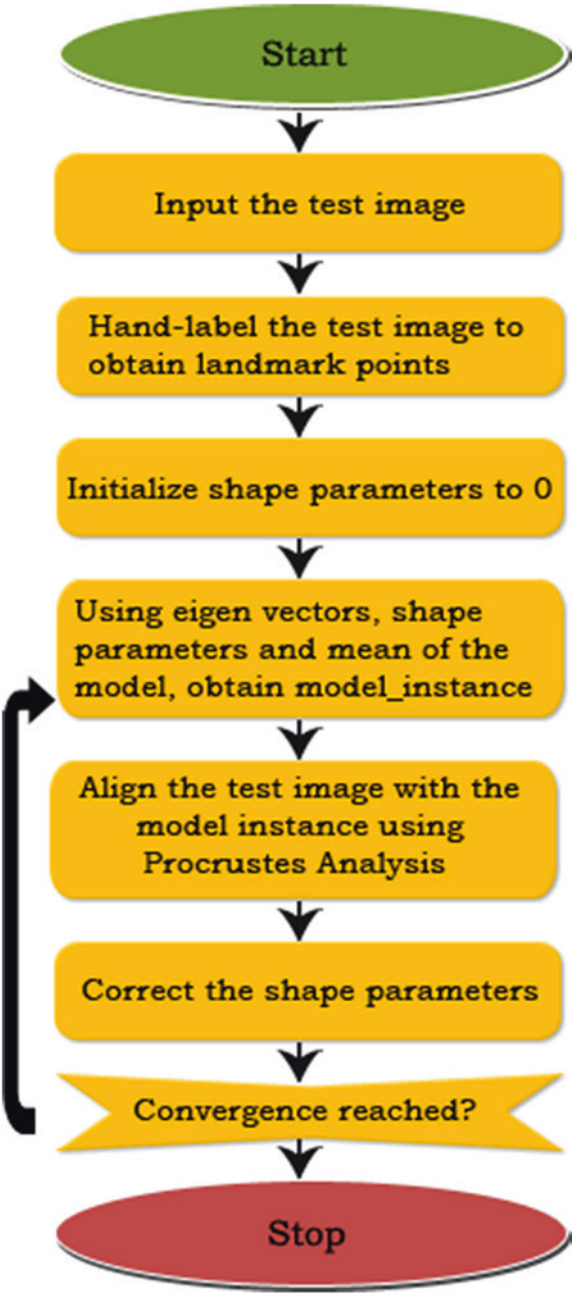



**Fig. 9.7** Flowchart depicting the procedure of model-fitting

The new image is classified under that class in which the distance is the minimum. The process of emotion classification is further elucidated in the algorithm below:

(i) Input the model representation obtained from model-fitting stage.
(ii) Read the mean models, i.e., centroids of a set of models for each class of emotions.
(iii) Determine the Euclidean distance between each of the 5 centroids and the model representation.
(iv) Classify the test image under that class in which the Euclidean distance is the minimum.

The approach discussed in this section deals with methods for handling static data repositories. The classifiers are trained a priori on labeled data and used later for classifying new data. The next section discusses a recent algorithm that demonstrates an online processing in the context of videos.

## 9.5 A Computational Technique Using Static and Dynamic Information

This section discusses a recent mechanism for detecting emotions from facial expressions in videos.

### *9.5.1 The Approach*

Metallinou et al. (2010) investigate the role of static and dynamic information conveyed by the face during speech for emotion recognition. Their focus is twofold: first, to compute compact facial representations by capturing usable information from the facial shape and facial movements and second, to model and recognize emotions by conditioning on knowledge of speech-related lip movements (*visemes*), which occur in parallel.

Their use of direct facial marker data enables the overcoming of some of the present challenges in feature processing from video data and focuses on establishing feasibility bounds for emotion classification using visual features. They rightly premise that facial information obtained from multiple markers across the face is redundant; neighboring markers tend to be highly correlated because they are controlled by the same underlying muscle movements. As the human face has a specific configuration, the possible range of physical movement of each facial marker is also limited.

They apply principal component analysis (PCA) for dimensionality reduction. In an alternative method, they select face markers using either Principal Feature Selection (PFA), a recently proposed technique motivated by PCA, or apply Fisher criterion in order to select features that better discriminate between different emotional classes. In order to constrain the speech-related variability of facial movement, they use the concept of *viseme*, which represents the lip shape during the articulation of a phoneme. Visemes are widely used for speech analysis and

audio–visual recognition of speech, especially under noisy conditions and animation. While averaging can be used to smooth the speech-related face movements, in contrast, they incorporate these movements in their analysis by modeling the evolution of emotional visemes. Dynamic modeling of information streams using HMMs has been shown to be a powerful method for audio–visual recognition.

In their work, they use a multi-speaker database and perform speaker-independent cross-validations. Facial features resulting from averaged, de-correlated, and normalized marker information (PFA features) achieve good performance. Happiness is the most recognized emotion using facial cues, with a recognition performance of the order of 75 %, in leave-one-speaker-out cross-validation experiments. Anger and happiness have performance of the order of 50–60 %, while neutrality has performance of the order of 35 %.

The Interactive Emotional Dyadic Motion Capture (IEMOCAP) database has been used in these experiments (Busso et al. 2008). This database contained approximately 12 h of audio–visual data from five mixed gender pairs of actors—male and female. IEMOCAP contained detailed facial information is obtained from motion capture as well as video, audio, and transcripts of each session. In comparison to other acted emotion databases where actors are asked to read out sentences displaying a specific emotion, in IEMOCAP, two techniques of actor training are used in order to elicit emotional displays—scripts and improvisation of hypothetical scenarios. The sessions are approximately 5 min in length. During these sessions, actors displayed various emotions according to the content of the session and the course of the interaction. The sessions were later manually segmented into utterances and annotated into categorical (anger, happiness, neutrality, etc.) and dimensional tags (valence, activation, and dominance). This study uses facial motion capture data, as well as the transcripts from all 10 speakers used in the corpus. Classes of anger, happiness, excitation, neutrality, and sadness were examined.

The IEMOCAP data contain detailed facial marker coordinates from the actors during their emotional interaction. For details on the layout of the face markers and the actual setup used for creating the corpus refer to (Busso et al. 2008). A total of 53 markers were attached to the faces of the subjects during the recordings. The markers were normalized for head rotation and translation. The nose marker is defined as the local coordinate center of each frame. There were five nose markers, and these were excluded from the computation because of their limited movement. In total, information in the form of $(x, y, z)$ coordinates from 46 facial markers was used. This resulted in a 138-dimensional facial representation, which tends to be redundant because it does not exploit the correlations of neighboring marker movements and the structure of the human face.

### *9.5.2 Feature Extraction*

Four feature extraction approaches were examined in order to find compact facial representations well suited for emotion recognition applications in terms of recognition accuracy.

**Speaker Face Normalization**—While examining various speakers, individual speaker face characteristics that were not related to emotion were smoothed out. The speaker normalization approach consists of finding a mapping from the individual average face to the general average face. The mean value of each marker coordinate of each speaker is shifted to the mean value of that marker coordinate across all speakers to achieve the normalization. The mean of each face feature (marker coordinate) is computed across all emotions $m_{ij}$ (where "$i$" is the speaker index and "$j$" is the feature index) for each speaker. The mean of each feature is also computed across all speakers and all emotions, $M_j$ (where "$j$" is the marker coordinate index). To obtain the set of normalized features, each feature is further multiplied with the coefficient $c_{i,j} = M_j m_{ij}$.

**Principal Component Analysis**—As already discussed in Sect. 9.4 above, PCA is a widely used method for dimensionality reduction. This method finds the projection of data into a lower dimensional linear space in which the variance of the projected data is maximized. The application of PCA for facial emotion recognition is inspired by the technique of eigenfaces. In eigenfaces, a feature vector is constructed from pixel values of facial image. PCA finds the principal faces, which can be linearly combined to reconstruct any face. Similarly, in this approach, the feature vector consists of the facial marker coordinates, and the principal projections can be interpreted as the directions of facial movement along which the variance is the maximum.

After performing the PCA, the face is reconstructed from the first 30 principal components, as they encode more than 95 % of the total variance. Some projections correspond to recognizable directions of facial movement, which affects either the lower or the upper facial parts or both. The PCA transformation matrix is computed, using data from all available speakers. Therefore, individual speaker characteristics are indirectly taken into account. Speaker normalization, either prior to or after the PCA transformation, does not improve recognition performance, and, therefore, it is not done. The window used for feature extraction is 25 ms with an overlap of about 16 ms. The choice of a short window enables further dynamic modeling of the visemes (as the average phoneme lasts about 100 ms).

**Principal Feature Selection**—The transformation space in PCA is a linear combination of the initial space of face marker coordinates. It has no inherent intuitive interpretation. The projections can be interpreted as directions of the specific face gestures and movements, but it is difficult to find meaning behind these projections. Principal feature analysis can be used to find more meaningful facial representations. In this method, the PCA transformation matrix is computed and used to cluster together highly correlated facial marker coordinates. After this, a representative feature is selected from each cluster, which performs feature selection while using similar criteria as PCA.

Normalization smoothes out the individual face characteristics that are unrelated to emotion and focus on emotional modulations. As in PCA, about 30 features are selected for this analysis. Principal feature analysis shows that the facial features are clustered together in a meaningful way. For example, same coordinates of neighboring or mirroring markers, as in left and right cheek, are clustered together. After 100 repetitions of PFA, it was found that, on an average, 28 % x

coordinates, 39 % y coordinates, and 33 % z coordinates were selected. This is indicative of the fact that all the 3 coordinates demonstrate significant variability in the context of emotional speech.

The jaw movements are mainly in the vertical direction which explains the comparatively high percentage of selected y coordinates selected. On an average, 22 % of the selected y coordinates are from mouth markers, while only 14 % of the initial markers are placed around the mouth. The z coordinates come from lip protrusion during articulation. The distribution of initial markers across the face regions is (chin, mouth, cheeks, eyebrows, and forehead) = (11, 14, 28, 36, and 11 %), wherein the distribution of the selected markers is (13, 23, 25, 31, and 8 %). This clearly shows a bias toward selecting lower face marker coordinates (especially mouth). This is expected because the movement of the jaw conveys a great amount of variability. Since the mouth can be automatically tracked more reliably than other face regions, such as cheeks and forehead, this is a useful result.

**Feature Selection Using Fisher Criterion**—The features described before are selected so as to capture maximum variance in the data. Such a set of features do not necessarily separate the different emotion classes well. To overcome this, Fisher criterion is used to extract a set of features, which maximizes the between-class variability and minimizes the within-class variability. Ad hoc averaging of neighboring markers is performed on these features to reduce from 46 to 28. After this, speaker face normalization is performed. In the final stage, 30 best marker coordinates are selected according to this criterion. The Fisher criterion value of each feature is computed on the training set, where the emotion classes are known. The Fisher criterion values are slightly different in each fold, so the features selected in different folds may vary slightly. The 30 ad hoc features are chosen so that this feature set is comparable with the previous two sets. From the selected features, across the 10-fold, on an average, 29 % are *x*, 34 % are *y,* and 37 % are *z* coordinates. On an average, about 34 % of the markers come from upper face including eyebrows and forehead, and 66 % come from lower face. Similar tendencies with PFA concerning the feature selection are observed, in general.

### *9.5.3 Viseme Information*

The lip shape during the articulation of a phoneme is called a viseme. The viseme is conditioned to constrain the variability related to speech, which recognizes the underlying emotion better. Visemes provide a reasonable time unit for HMM training, besides incorporating speech-related information and associated dynamical models of the facial movement. The phoneme-to-viseme mappings are many to one, and various such mappings exist in the literature depending on the desired detail. The authors here used 14 visemes. They have the word transcription for each utterance, and through forced alignment, they obtain the phoneme-level transcription. They use this transcription to group facial data corresponding to each viseme.

Through their experiment, Metallinou et al. (2010) find that emotion recognition accuracy is highly speaker dependent. Also, the lower face seems to convey more information as compared to the upper face. Explicitly modeling articulation movements improves recognition for anger, happiness, and neutrality, but decreases performance for sadness.

## 9.6 What the Future Holds

This research has utilization in the domains of human–computer interaction, medical applications, nonintrusive interrogation, etc. It can be used for building cognitive systems that read and respond to human behavior and actions (including intent), for military and security applications such as surveillance and analysis, intelligent robotic systems, and medical diagnosis. Health professionals can develop better rapport with patients and make the right diagnosis by obtaining more complete information.

Computational analysis based on video and image analysis is used in numerous real-world applications particularly those that are security-related. Findings can be based on either silent surveillance or analysis of video frames/images during a face-to-face interaction. Typical examples are cited below:

(a) Airport surveillance using surveillance cameras can help detect any suspicious behavior in a passenger prior to boarding the aircraft. The most relevant analysis will be when the subject is clearing a security check or the subject's baggage is being screened. Subjects communicating on phone or on the Internet will also reveal emotions in case some information of value is received that affects the subject either adversely or positively.
(b) During questioning or interrogation, a subject will experience numerous emotions that a video analysis can interpret. The polygraph lie-detector test which uses heart rate monitors for emotion interpretation is highly inaccurate and not admissible as evidence usually. However, video analysis-based advanced algorithms for emotion detection may eventually come to be admitted, if they can prove their accuracy.

There are many potential directions for future work. One immediate step is to include multiple modalities, such as speech and gestures, to improve emotion recognition performance. The dynamic statistical modeling of multiple modalities and their effective fusion is an interesting and challenging problem and needs more work. *Affective computing* is a modern evolving interdisciplinary domain that consolidates work in this area. Affective computing is about inducing empathy in a machine, so that it understands human emotions and adapts and responds to these perceived emotions in an empathetic manner. Although facial expressions of emotion are presently categorized into discrete categories, and although there is even evidence for categorical perception of such facial expressions, it is also clear that expressions are typically members of multiple emotion categories and that the

boundaries between categories are fuzzy at the level of recognition (Russell and Bullock 1986). Further, it is evident that the categorization of an emotional facial expression depends to some extent on the contextual relation to other expressions with which it may be compared (Russell and Fehr 1994). Some mathematical models further argue that emotions shown in facial expressions could be thought of as exhibiting features both of discrete emotions and continuous dimensions (Calder et al. 2001).

As researchers across the world evolve techniques for recognizing emotions from facial expressions, one fact is abundantly clear. Any computational technique for recognising emotions will, to begin with, be limited by our own understanding of human emotions. Even though humans themselves have enormous exposure to the world and can therefore form concepts around explicit facial expressions, they still remain handicapped in their understanding of human emotions. But as computational systems evolve, they will slowly assist humans in their perception of emotions. And then a new, more complex frontier of machine-based human emotion recognition may be breached.

## References

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioural and Cognitive Neuroscience Reviews, 1*, 21–62.

Agrawal, M., Giripunje, S. D., & Bajaj, P. R. (2011). Recognizing facial expressions using PCA and genetic algorithm. *International Journal of Computer & Communication Technology, 2*(7), 32–35.

Anam, S., Islam, M. S., Kashem, M. A., Islam, M. N., Islam, M. R., & Islam, M. S. (2009). Face recognition using genetic algorithm and back propagation neural network. *Proceedings of the International MultiConference of Engineers and Computer Scientists, I.*

Azcarate, A., Hageloh, F., Sande, K., & Valenti, R. (2005). *Automatic facial emotion recognition*. Technical report of Universiteit van Amsterdam.

Busso, C., Bulut, M., Lee, C. C., Kazemzadeh, A., Mower, E., Kim, S., et al. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Journal of Language Resources and Evaluation, 42*(4), 335–359.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3*, 81–117.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research, 41*, 1179–1208.

Chang, Y., Hu, C., Feris, R., & Turk, M. (2006). Manifold based analysis of facial expression. *Image and Vision Computing, 24*, 605–614.

Cootes, T. (2000). An introduction to active shape models. In R. Baldock & J. Graham (Eds.), *Image Processing and Analysis,* (pp. 223–248). Oxford: Oxford University Press. http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/Papers/asm_overview.pdf

Cottrell, G. W., Dailey, M. N., Padgett, C., & Adolphs, R. (2001). Is all face processing holistic? In M. J. Wenger & J. T. Townsend (Eds.), *Computational, Geometric, and Process Perspectives on Facial Cognition* (pp. 347–396). Mahwah: Lawrence Erlbaum.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorises facial expressions. *Journal of Cognitive Neurosciences, 14*(8), 1158–1173.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America, 2*(7), 1160–1169.

de Gelder, B., Teunisse, J.-P., & Benson, P. J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition and Emotion, 11*, 1–24.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44*, 227–240.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593–609.

Ghandi, B. M., Nagarajan, R., & Desa, H. (2009). Particle Swarm Optimization algorithm for facial emotion detection. *IEEE symposium on Industrial Electronics and applications, 2*, 595–599.

Gu, W. F., Xiang, C., Venkatesh, Y. V., Huang, D., & Lin, H. (2012). Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. *Pattern Recognition, 45*, 80–91.

Habibizad, A., & Mirnia, M. K. (2012). A new algorithm to classify face emotions through eye and lip feature by using particle swarm optimization. *Proceedings of 4th International Conference on Computer Modelling and Simulation, 22*, 268–274.

Hecht-Nielsen, R. (2007). Confabulation theory (computational intelligence). *Scholarpedia*, *2*(3),1763. http://www.scholarpedia.org/article/Confabulation_theory_%28computational_intelligence%29

Hecht-Nielsen, R. (2007b). *Confabulation theory: The mechanism of thought*. Heidelberg: Springer.

Huang, K. C., Kuo, Y. M., & Horng, M. F. (2012). Emotion recognition by a novel triangular facial feature extraction method. *International Journal of Innovative Computing, Information and Control, 8*(11), 7729–7746.

Ioannou, S. V., Raouzaiou, A. T., Tzouvaras, V. A., Mailis, T. P., Karpouzis, K. C., & Kollias, S. D. (2005). Emotion recognition though facial expression analysis based on a neurofuzzy network. *Neural Networks, 18*(4), 423–435.

Karthigayan, M., Rizon, M., Nagarajan, R., & Yaacob, S. (2008). Genetic algorithm and neural network for face emotion recognition. In Jimmy (Ed.), *Affective Computing* (pp. 57–68). http://www.intechopen.com/download/get/type/pdfs/id/5178

Kaushik, R., & Mohamed, S. K. (2012). Facial expression recognition using game theory and particle swarm optimization. *Lecture Notes in Computer Science, 7594*, 2133–2136.

Kobayashi, H., & Hara, F. (1992). Recognition of six basic facial expression and their strength by neural network. *Proceeding of IEEE International workshop on Robot and Human Communication* (pp. 381–386).

Londhe, R. R., & Pawar, V. P. (2012). Analysis of facial expression and recognition based on statistical approach. *International Journal of Soft Computing and Engineering, 2*(2), 391–394.

Lyon, M., & Akamatsu, S. (1998). Coding facial expression with Gabor wavelets. *Proceedings of 3rd IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 200–205), Nara, Japan.

Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 200–205). Nara, Japan.

Metallinou, A., Busso, C., Lee, S., & Narayanan, S. (2010). Visual emotion recognition using compact facial representations and viseme information. *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. pp. 2474–2477.

Mpiperis, I., Malassiotis, S., Petridis, V., & Strintzis, M. G. (2008). 3D facial expression recognition using swarm intelligence. *IEEE International conference on Acoustics, Speech and Signal Processing* (pp. 2133–2136).

Padgett, C., & Cottrell, G. W. (1998). A simple neural network models categorical perception of facial expressions. *Paper presented at the Proceedings of the 20th Annual Cognitive Science Conference.* Madison, WI.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.

Russell, J. A., & Bullock, M. (1986). Fuzzy concepts and the perception of emotion in facial expressions. *Social Cognition, 4*, 309–341.

Russell, J. A., & Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: Varieties of anger. *Journal of Personality and Social Psychology, 67*, 186–205.

Saudagare, P. V., & Chaudhari, D. S. (2012). Facial expression recognition using neural network-an overview. *International Journal of Soft Computing and Engineering, 2*(1), 224–227.

Suwa, M., Sugie, N., & Fujimora, K. (1978). A preliminary note on pattern recognition of human emotional expression. *Proceedings of the 4th International Joint Conference on Pattern Recognition* (pp. 408–410).

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features, Computer Vision and Pattern Recognition, 2001. CVPR 2001. *Proceedings of the 2001 IEEE Computer Society Conference on*, 1, (pp. I-511–I-518).

Yen, G. G., & Nithianandan, N. (2002). Facial feature extraction using genetic algorithm. *Proceedings of the 2002 Congress on Evolutionary Computation*, *2*, (pp. 1895–1900).

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods, audio, visual and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(1), 39–58.