Chapter 11
Facial Recognition, Facial Expression and Intention Detection

Massimo Tistarelli, Susan E. Barrett, and Alice J. O'Toole

11.1 Introduction

Visual perception is probably the most important sensing ability for humans to enable social interactions and general communication. As a consequence, face recognition is a fundamental skill that humans acquire early in life and which remains an integral part of our perceptual and social abilities throughout our life span (Allison et al. 2000; Bruce and Young 1986). Not only faces provide information about the identity of people, but also about their membership in broad demographic categories of humans (including sex, race, and age), and about their current emotional state (Hornak et al. 1996; Tranel et al. 1988; Calder et al. 1996; Humphreys et al. 1993; Calder and Young 2005). Humans "sense" this information effortlessly and apply it to the ever-changing demands of cognitive and social interactions.

Face recognition has now also a well established place within the current biometric identification technologies. It is clear that machines are becoming increasingly accurate at the task of face recognition, but not only at this. While only limited information can be inferred from a fingerprint or the ear shape, the face itself conveys far more information than the identity of the bearer. In fact, most of the times humans look at faces not just to determine the identity, but to infer other issues which may or may not be directly related to the other person's identity. As an example, when speaking to someone we generally observe his/her face to either better understand the meaning of words and sentences, but mostly to understand the emotional

M. Tistarelli (✉)
Computer Vision Laboratory, University of Sassari, Sassari, Italy
e-mail: tista@uniss.it

S.E. Barrett
Lehigh University, Bethlehem, PA, USA

A.J. O'Toole
University of Texas at Dallas, Richardson, TX, USA

E. Mordini and D. Tzovaras (eds.), *Second Generation Biometrics: The Ethical, Legal and Social Context*, The International Library of Ethics, Law and Technology 11,
DOI 10.1007/978-94-007-3892-8_11, © Springer Science+Business Media B.V. 2012

state of the speaker. In this case, visual perception acts as a powerful aid to the hearing and language understanding capabilities (Hornak et al. 1996).

Even though most of past and current research has been devoted to ameliorate the performances of automatic face identification in terms of accuracy, robustness and speed in the assignment of identities to faces (a typical classification or data labeling problem) (Mika et al. 1999; Lucas 1998; Roli and Kittler 2002), the most recent research is now trying to address more abstract issues related to face perception (Humphreys et al. 1993; Haxby et al. 2000). As such, age, gender and emotion categorization have been addressed with some level of success (again, as a broader labeling or classification problem), but with limited application in real world environments. This paper analyzes the potential of *current* and *future* face analysis technologies as compared to the human perception ability. The aim is to provide a broad view of this technology and better understand how it can be beneficially deployed and applied to tackle practical problems in a constructive manner. Airports constitute a typical and interesting example where identification and personal categorization is often required. The adoption of the new electronic passport, encompassing biometric data, such as the face and fingerprint, constitutes a challenge for the use of biometric data without compromising the traveller's privacy. A simple solution is to use the data to reveal ancillary information about the traveller without even attempting to identify him/her.

Emotions are an integral and unavoidable part of our life. As such, should we be scared if machines acquire the ability to discern human emotions? Would it be an advantage or a limitation for the user if a Personal Computer or a Television is capable of sensing our mood and provide appropriate services accordingly? Should we perceive a "trespassing" of our privacy if a hidden camera in an airport lounge recognizes our tiredness and activates a signal to direct us to the rest facilities?

Nowadays, there is a considerable concern for biometric technologies breaking the citizen's privacy. These concerns have yet hindered the adoption of these technologies in many environments, thus depriving many of potential service improvements and benefits. Unfortunately, as in the past, also today misinformation and politically driven views impair technological advances which can be adopted for the wellness of the citizens. This chapter analyzes the impact of *current* and *future potential* face analysis technologies as compared to the human perception ability. The aim is to provide a broad view of this technology and better understand how it can be beneficially deployed and applied to tackle practical problems in a constructive manner. Yet, as any other technology, the way it is used, either for the good or bad, entirely depends on the intentions of the men behind its application.

11.2 Human Face Perception and Recognition

Human face perception seems effortless. We recognize people by their faces and make instantaneous judgments about their emotional state and intent scores of times per day. The human ability to extract this information quickly and efficiently from a face is an important pre-requisite for our survival as a species. In the next few sections, we discuss human face recognition and facial expression perception abilities.

We do this in a way that keeps in mind the question of whether or not machines are "up to the task" of stepping in for humans in automated surveillance applications. Most of us are entirely comfortable presenting an identification card to a border guard or customs official for verifying the identity of our face against the picture on the card. We are sometimes even comforted by the sight of a police officer or security guard in a deserted subway station or parking lot late at night. And yet, we are concerned at the possibility of handing our identification card over to a machine and standing in front of a camera, before being allowed to proceed on our way.

Indeed, some ethical concerns about allowing machines to aid or replace humans in security and surveillance tasks come from not knowing answers to the following questions. "How well do machines perform security and surveillance tasks?" "Are they going to make mistakes that humans would not make?" "If they do make a mistake, are we going to be able to talk to a person about it, to challenge the decision?" For the question of accurate face recognition, the first of these questions has been addressed continuously over the past decade with large scale evaluations of state-the-art face recognition algorithms (e.g. Phillips et al. 2005). We will discuss the results of some of these studies shortly. To the best of our knowledge, there have been no formal objective tests for the evaluation of machine-based expression recognition systems.

Implicitly, the questions we ask about trusting machines to do these tasks, depends on understanding how well humans recognize faces. Psychologists have long studied human strengths and weaknesses of human face recognition and expression perception. On this question, we will argue that computer scientists, engineers, and the policy makers, who make decisions about the application of automated surveillance systems, need to understand the competition for their machines–human beings.

First, we begin with a discussion of the basic issues involved in relating human and machine abilities to face recognition and expression perception. Second, we discuss human abilities to perceive facial expressions. We start with the influential categorical models of expression and emotion, discussing the universal human facial expressions defined by Ekman and Friesen in 1976 and how they are produced using specific sets of coordinated facial muscle movements. Third, because the ability to produce and perceive facial expression develops early in life, we briefly sketch out the course of facial expression development in infants and young children. This developmental course comes surprisingly close to offering a "to-do list" for machine recognition of expression. What do automatic expression recognition systems need to master before they are to be trusted in important applications? Fourth, expression and identity information coexist on the face and must be accessed simultaneously, with minimal cross-interference. Thus, we discuss the support for long-standing theories in the psychological and neuroscience literatures that posit a strong separation of the processes that underlie facial expression and face identity perception (Bruce and Young 1986; Calder and Young 2005). Although the idea of *strictly* separated systems has been challenged in recent years, the basic tenant of perceptual and neural separability in facial expression and identity processing remains a fundamental part of our understanding of human facial analysis. Finally, we will sum up by linking our understanding of human abilities to the factors that must be considered when we apply machines to the problems.

11.2.1 Face Recognition: Issues Relating Humans and Machines

Human performance on the task of face recognition has generally been considered the “gold standard” against which the performance of algorithms is judged. In fact, the available psychological and computational data suggest human face recognition performance is superior to even the best state-of-the-art algorithms in some, but not all (O’Toole et al. 2007b), cases. What is certain is that our ability to recognize the faces of people we know well (i.e., friends and family), displays a remarkable robustness to photometric variables such as changes in illumination, viewpoint, and resolution (cf., Burton et al. 1999). Face recognition algorithms have not yet achieved the level of robustness typical of human face recognition for highly familiar faces. Photometric variables, especially changes in viewpoint, remain highly challenging for face recognition algorithms and have a substantial negative impact on the performance of the algorithms (e.g., Phillips et al. 2005).

Less clear is the comparison between face recognition algorithms and humans on the task of recognizing relatively unfamiliar faces (Hancock et al. 1991). By unfamiliar faces, we mean those that we have limited experience with (e.g., have seen only once or twice from a photograph or live encounter). In this case, humans, like algorithms, perform less accurately when there are photometric changes between learning and test images (e.g., illumination Braje et al. (1999); Hill and Bruce (1991), viewpoint Moses et al. (1996); Troje and Bülthoff (1996)). In general, the kinds of face recognition and verification tasks done by security personnel are carried out almost exclusively with unfamiliar faces. Because human face perception skills are not at their best for unfamiliar faces, the difference between human and machine accuracy in real world applications might be less pronounced than one would fear.

In fact, under reasonably controlled viewing conditions, when there is a good match between image characteristics on viewpoint, the best face recognition algorithms now compete favorably with humans. In the Face Recognition Grand Challenge (FRGC), a recent U.S. Government sponsored test of face recognition algorithms, three out of the seven algorithms under evaluation surpassed humans matching identity in pairs of images that differed in illumination (O’Toole et al. 2007b). A follow-up experiment showed that there were qualitative differences between human and machine recognition strategies. This was demonstrated by fusing human and machine identification judgments, which can improve decisions only when the pattern of errors differs enough to combine the independent strengths of both strategies. In fact, the fusion of human judgments with the judgments of face recognition algorithms from the FRGC produce better performance than any of the systems operating alone (O’Toole et al. 2007a).

Combined, one important implication of this work is that we may over-estimate human performance on the task of face recognition. This suggests that the confidence we have in our own ability to recognize faces accurately may be misplaced. Ample evidence exists for the human ability to mistakenly identify someone with high confidence, a finding that is now well-established in the context of DNA-based exonerations of unjustly convicted individuals in the United States over the past few decades.

Mistaken identifications by witnesses figure prominently in many cases (Saks and Koehler 2005). When face recognition technologies are applied in surveillance situations, policy makers must make an effort to understand how these technologies can *compensate* for human weaknesses, while also understanding the limits of the technologies.

On this topic, there are analogous issues about the fairness of both human and machine recognition systems. For humans, it is well known that face recognition is not equally accurate for all faces. For example, *typical* faces are recognized less accurately than unusual faces (Light et al. 1979; Valentine 1991). The practical consequence of this finding is that people with typical faces, (i.e., few distinctive features) are more often mistakenly recognized than are people with more distinctive faces. It is unclear, to date, if machines show a similar bias, though it is highly likely. Certainly, the performance of machines on biometric recognition tasks has been characterized in terms of the recognizability of individuals (Doddington et al. 1998). One framework for understanding this issue in machine performance is referred to as the *Doddington Zoo*, which posits four categories of individuals based on the statistical pattern of recognition success. *Sheep* are individuals who are easy to recognize. *Goats* are difficult to recognize. *Lambs* are easy to imitate and *wolves* are good at imitating others. Recognition schemes can be adjusted to take into account aspects of a face that might make it a wolf or a lamb. This Doddington zoo characterization suggests potential inequities in the application of face recognition systems across individuals. The important point, however, is that these inequities may not differ substantially between humans versus machines.

More systematic inequities exist for humans recognizing faces of different races. The *other-race effect* has been established and replicated in many human face recognition studies. This effect refers to the finding that humans recognize faces of their own race more accurately than faces of other races (Malpass and Kravitz 1969; Meissner and Brigham 2001). The critical false identification problems that arise from this phenomenon have been well-documented in the eyewitness literature. Indeed, the recognition of a suspect by a witness of another race is less likely to be accurate than the recognition of the suspect by a person of the same race, though prosecutors, juries, and judges are often unwilling or unable to factor this into their confidence about an identification. Although there is still some controversy regarding the cause of the other-race in humans, most evidence relates this effect with the amount of experience we have with faces of our own race versus faces of other races (Bryatt and Rhodes 1998). Moreover, there is recent work suggesting the particular importance of learning faces early in life (Kelly et al. 2007). The idea in this case is that learning in infancy contributes to an optimization of the feature sets used to encode faces. Generally, this optimization is based on the statistical variations in the set of faces we learn first, usually faces of our own race.

Do automated face recognition systems show an other-race effect? In other words, does the geographic origin of an algorithm (i.e., where it was developed) affect how well it performs on faces of different races. Presumably, if the cause of the other-race effect is a bias in the selection of the features used to encode faces, there is reason to be concerned that algorithms will fare best with faces of the race(s) used to train them.

A recent study tested whether state-of-the-art face recognition systems show an other-race effect (O'Toole et al. 2008). The U.S. Government's Face Recognition Vendor Test 2006 (FRVT 2006) provided the face recognition algorithms used for the test. In the FRVT 2006, five algorithms from East Asian countries (China, Japan, and Korea) were compared with eight algorithms from Western Countries (France, Germany, and the United States) on the task of recognizing East Asian and Caucasian faces. This test revealed a clear other-race effect that was especially pronounced in the range of low false alarm rates required for most security applications. In other words, under the operating conditions that would be used for most applications, the algorithms developed in East Asia recognized East Asian faces more accurately than Caucasian faces and the Western algorithms recognized Caucasian faces more accurately than East Asian faces. In a direct comparison with humans of East Asian and Caucasian ethnicities (on a subset of the faces used in the first comparison), an other-race effect was seen for both the humans and algorithms. Humans performed somewhat more accurately than machines overall. Of note, however, the major difference between machine and human accuracy was that human performance was more stable across changes in the race of a face than was machine performance. Indeed, the face recognition algorithms showed much larger variations in performance across changes in the face race than did humans.

The cautionary note in the airport is that Big Brother is already there in human form, with his/her strengths and weaknesses. Regardless of whether a machine or human stands guard at the border, it is important to understand, anticipate, and mitigate the kinds of recognition errors that are likely to occur.

11.2.2 Facial Expression Perception: Human Perspectives That May Inform Machine Vision

An important difference between face identity and facial expression perception, is that identity has ground truth. Facial expressions are produced in response to felt emotions, or are in some cases, pretended emotions. Remarkably little is known about how accurately humans perceive the felt emotions of others from their facial expressions. Although it seems as if expressions provide us with direct access to the a person's emotional state, classic examples of the mismatch between felt emotions and perceived expressions abound (e.g., tears during the World Cup Finals and at weddings can be more than a bit difficult to interpret out of context). Thus, it's worth bearing in mind that facial expressions, which consist of non-rigid deformations of facial structure, provide only partial information about felt emotions. The rest of the information comes from the cognitive and social context of the episodes or events that elicit the expressions and emotions. More formally, psychologists have tried to put the definition of facial expressions onto solid ground by quantifying the patterns of facial muscle movements used to create (or portray) expressions that are universally recognizable across cultures. In the next section, we discuss ideas about the structure of facial expressions.

11.2.2.1 Universal Facial Expressions

Over the past several decades, the most influential theories have advocated a categorical view of expressions (Ekman and Friesen 1976). According to this view, each instance of an emotional expression is treated as a variation within a set of basic facial configurations. Facial expressions are part of a biologically based communication system for signaling emotions. Cross-cultural studies have looked at how people in different societies and with different cultural backgrounds perceive expressions (Ekman et al. 1969, 1972; Izard 1971). These studies have been used to argue that basic emotions are universally expressed and recognized. There is a striking degree of similarity in what elicits certain emotions and in how these emotions are expressed on the face. Although agreement about the perception of particular facial expressions is well above chance, it is not perfect. Notwithstanding, most psychologists agree that there are *universals* in emotional expression. Anger, disgust, fear, happiness, sadness and surprise have been offered as emotions that are associated with specific facial expressions and are basic to all human emotional life.

It is hard to overstate the importance of cross-cultural work in laying down the foundations of psychological and anthropological understandings of emotion. Darwin's insights gave rise to systematic studies of how emotions are expressed in people from widely different cultural backgrounds (Darwin 1898). Both Ekman and Izard and their colleagues drew on their own cross-cultural observations as they developed theories and coding systems for human facial expression. For example, Ekman and Friesen's Facial Action Coding System (FACS) defines a *prototypical* version of each of the six basic emotions (anger, disgust, fear, happiness, sadness and surprise). To do this, the system uses 40 separate and functionally independent muscle action units to classify a face as exemplifying a particular emotional expression. Timing, intensity, and laterality are also considered important variables for distinguishing among emotional expressions.

The status of these *prototypical expressions* as "ground truth" for the associated emotion, however, is unclear. Although Izard (1997) laments the "false notion that there is one full face prototypical expression for each of several basic emotions" (Izard 1997, p. 60), emotion prototypes are a driving force in both Ekman and Matsumoto's view: "When we perceive emotions in others, the process is analogous to template matching with the universal facial prototypes of emotion. Before a judgment is rendered, however, that stimulus is also joined by learned rules [that] may differ according to stable sociocultural dimensions such as IC [individualism vs. collectivism] and SD [behavioral differentiation based on status differences]" (Matsumoto 2001, p. 185).

One critical aspect of these rules is whether it is appropriate to show specific emotions in different contexts. Ekman and Friesen provided an early and powerful demonstration of how culture shapes emotional expression (Ekman et al. 1972). They found that both Japanese and American observers readily displayed negative emotions when viewing a stressful film in isolation but that Japanese observers attempted to conceal these negative emotions, often with a smile, when others were present. Smiles are popular masks and even children use social smiles to convey an emotional message they may not feel but are required to show. Feigning a negative emotion is more difficult as most individual are less experienced at falsifying emotions such as fear (Ekman 2009).

Although social conventions may lead individuals to try to cover up their feelings, Ekman believes that in most cases the message leaks through, especially if the feeling state is strong. Hence, with practice, someone can become skilled at detecting true expressions, which may be masked because of social conventions or because the expresser is trying to mislead an audience.“The stronger the emotion, the more likely it is that some sign of it will leak despite the liar’s best attempt to conceal it. Putting on another emotion, one that is not felt, can help disguise the felt emotion being concealed. Falsifying an emotion can cover the leakage of a concealed emotion” (Ekman 2009, p. 31–32).

When an emotion is masked, the original expression may be covered up after it is appears. Full facial expressions that last for only a fraction of their usual time are referred to as “micro expressions.” Micro expressions, however, do not occur very often. Emotions are often squelched, that is, the facial expression may be covered up before it is fully revealed in the face. Detecting deception may hinge on detecting full or partial expressions before they are masked by another emotion. Training programs have been developed to heighten sensitivity to “squelched emotions” and “micro expressions.” Differences in timing can also be used to discriminate feigned and felt emotions.

Individuals may be especially adept at reading emotions, or interpreting the “emotional dialect” of their culture. Matsumoto attributes these differences to learned cultural display rules, but perceptual and attentional factors may also play a role. Cultural differences in display rules affect which regions of the face are most informative (Matsumoto 2001). The mouth is likely to be a particularly informative region when individuals are emotionally demonstrative, whereas felt emotions are likely to be evident in the eyes when the emotion is masked. Whereas Western Caucasian observers distribute their gaze evenly across a face; East Asian observers focus on the eye region, at the expense of mouth region, and have difficulty when the eyes do not disambiguate the expression (Jack et al. 2009). Consequently, East Asian observers tend to confuse fear with surprise and disgust with anger. When asked to choose a label for less clearly expressed emotions, East Asian observers also have a tendency to choose the more socially desirable label. Hence, social desirability, cultural concepts, and gaze patterns (or sampling biases) all play an important role in how facial expressions are perceived (Jack et al. 2009; Russell 1991; Schyns 1998).

11.2.2.2 Development of Facial Expression Perception

In the past few decades, a revolution of sorts has taken place in the developmental literature as infancy researchers have uncovered hitherto unimagined competencies in young infants. Psychologists studying infants are faced with challenges similar to those one might encounter when testing expression recognition systems. Infants, like machines, can differentially respond to stimuli. But, we cannot ask them to explain how they reach their decisions and it is not obvious which aspects of the stimulus are driving their performance. A careful analysis of what infants can and cannot do shows that emotion perception, broadly defined, occurs on many levels. Infants may be able to discriminate among facial expressions, but this does not mean that they understand the "meaning" of the message. Infants are surprisingly competent social partners, even though they are not born with a system for recognizing basic universal emotions. Children are sensitive to emotional valence (positive or negative) and intensity, long before they are able to classify "prototypical examples" of the facial expressions that characterize the basic emotions). This suggests a potentially different path for developing emotion recognition systems and facial expression recognition systems.

The steps that infants go through in becoming competent perceivers of the emotions of others provide a kind of blue print for machine recognition systems. First, there is a need to discriminate the patterns of muscle movements that convey different expressions and to generalize these patterns across all faces. Next, natural displays of emotion include changes to facial expression, voice, and body posture. Being able to discriminate these displays reliably is a first step in understanding what they mean. Finally, emotional displays make us react with empathy or with an otherwise adaptive response. These adaptive responses are the ultimate measure for our understanding of the emotional displays of other people. Through these responses, emotional expression serves its function as a system of communication.

A prerequisite for identifying and labeling facial expressions is an ability to reliably discriminate different expressions (e.g., fear from anger, happy from surprise). Infancy researchers use simple behaviors to test how infants perceive the world. Changes in responses, such as increases in looking time, are taken as evidence that infants notice a stimulus change. In the habituation-dishabituation paradigm, infants are given the opportunity to inspect a stimulus (e.g., a happy face) until there is a decrease in looking time. If a new stimulus (e.g., a sad face) is presented, and infants detect the change, looking time increases. But does this increase in looking time really mean that infants discriminate facial expressions? Perhaps the infant only notices that the mouth has changed: tightly compressed lips have replaced bright teeth. Discriminating facial expressions and discriminating mouths are two different things. One way around this problem is to show the infant the same emotional expression by different people. Individual features (e.g., the "toothiness" of the smile) vary across individuals and infants do generalize smiles across individuals (Walker-Andrews and Bahrick 2001). If infants perceive facial expressions as wholes or "gestalts," the orientation of the face should matter, and it does. Infants smile at dynamic upright faces, but fail to engage emotionally with inverted faces (Muir and Nadel 1998). Thus, there seems good evidence from visual presentations of facial expressions that infants can reliably distinguish most facial expressions, and to generalize these expressions over the individuals who display them.

To ask the more difficult question about whether the ability to discriminate expressions is accompanied by an understanding of their meanings (i.e., the emotional states that underlie the expressions), researchers have made use of cross modal matching tests as a first step. These cross-modal tests are used to assess the extent to which infants can integrate emotionally similar information across different sensory modes of presentation. In studies of this kind, the modes are defined usually as visual presentations of faces and auditory presentations of voices. Indeed, caregivers communicate emotions with their voices and faces, as well as their bodies. Data from cross modal matching studies suggest that infants are sensitive to connections between facial and vocal expressions of emotion. In a cross-modal matching study, the infant might be presented with a happy voice while two silent films are presented on adjacent television screens. If infants are sensitive to the affective information conveyed in each modality, we would expect that they would look at the display that matches the sound track. Five to seven months old look at the matching display, and they make this connection even when the voice and video are not synchronized (see Walker-Andrews (1997) for a review). These findings suggest that infants are sensitive to the concordance of affective messages across sensory modalities.

Although cross-modal matching studies suggest that infants are sensitive to information that conveys the emotional meaning of the messages, these studies were not designed to test whether infants *react* in meaningful ways to specific facial expressions. Social-interactional theorists argue that if we are interested in whether infants recognize the meaning of facial expressions, we should focus on how infants respond to these messages (Lewis and Goldberg 1969; Stern 1985). Detailed observations of face-to-face interactions provide compelling evidence that infants and their caregivers are engaged in emotional dialogues (Sroufe 1995). At first, it is the parents who shoulder the responsibility for "meaning-making" as they interpret the infants actions as intentional and respond in meaningful ways to changes in the infants facial expressions. Infants engage and reengage with their partner's face. Surprisingly, which expression is posed does not seem to matter. A happy expression is nearly as disruptive as a sad or neutral one (D'Entremont and Muir 1997), although happy faces do elicit slightly more smiling, perhaps suggesting that infants expect their interactive bids will be more successful if they are directed at a smiling partner.

Stronger evidence that infants understand the emotional meaning of facial expressions comes from studies of social referencing. In social referencing studies, the infant is presented with an ambiguous situation, for example a "scary" visual cliff with a moderately steep drop-off. A visual cliff is made by juxtaposing two steps or ledges with a "drop-off" between the two. A clear plastic sheet of flooring is placed overtop to connect the ledges and to provide a safe bridge-like connection over the "drop-off". Babies can see the drop-off and are usually too frightened to cross over. An adult, often the mother, stands at the far side of the drop-off and poses a specific facial expression. If the mother poses a happy face, the infant will venture across the cliff; the infant is not likely to do so if the mother poses an angry, fearful, or sad face (Sorce et al. 1985).

One reason developmental psychologists are careful not to over interpret the infant data is that they are well aware of the difficulties older children have recognizing emotions. Toddlers can label emotions at above chance levels, but their performance is far from error-free. Although 2 year olds rarely place a happy face in the “angry” box, they place similar numbers of sad, fearful, angry, disgusted faces in this box (Widen and Russell 2008a). Preschoolers continue to have difficulty with sorting tasks even when the verbal requirements are minimized and the boxes are labeled with pictures. Children’s emotion categories and their ability to recognize facial expressions have a rather protracted developmental time course (Markham and Adams 1992).

For both humans and machines, ultimately, there is a need to label emotional expressions. Two-year-olds use a variety of emotion words both to describe their own feelings and to talk about other’s feelings (Dunn et al. 1987; Wellman et al. 1995). Children clearly differentiate positive feelings (happy, better, okay) from negative feelings (afraid, sad, angry) but they use these terms in a much less specific way than adults (Harris 1989). Children acquire emotion labels in a predictable order, although mastery of these labels is a gradual process. The first three emotion words children use are “happy”, “sad”, and “angry.” “Scared” and “surprised” are added next with “disgust” being added a bit later (Widen and Russell 2003). The frequency with which these labels are used also differs, the earliest acquired labels are also the most frequently used. This is a reflection of both the fuzzy nature of emotion categories and a pull toward more accessible labels (Widen and Russell 2008a).

One assumption that has been pervasive across the psychological literature on facial expression perception is that human expressions are categorically generated and perceived. In more recent theorizing, this assumption has been re-evaluated in a way that counters Ekman’s influential view of basic emotions that can be represented by discrete labels and conveyed with prototypical facial expressions. Specifically, dimensional models, such as the circumplex model, view affective experiences along a continuum. These experiences themselves are often ambiguous, or at least, best described by multiple labels that convey the interrelations among affective states. Variations in affective states are best captured by two dimensions, the first corresponds to something akin to valence and the second to arousal. Widen and Russell (2008b) argue that these two dimensions serve as a starting point for building emotional knowledge. Data from both free labeling and classification tasks support their view. Children use emotion words, such as angry, but these words map onto a broad fuzzy categories that only partially overlaps with what adults means when they use the label.

Widen and Russell (2008b) argue that the infant data do not make a compelling argument for the discrete-category account of emotional development. More specifically, they argue that while it is clear that infants distinguish between happy and negative expressions, there is little evidence that they differentiate among specific negative expressions, such as fear and anger. The problem, in their view, is that most studies focus on emotion pairs that differ in valence. Evidence for differential responding to specific negative emotion is scant, at best. For example, infants tend to avoid the apparent drop-off of the visual cliff when the mother poses both a fearful and angry face. In some studies, infants do respond differently to negative expressions, for example, sad expressions generate less interest than angry expressions (Haviland and Lelwica 1987) and infants will look in the appropriate direction when angry and sad faces are presented side by side in a cross-modal matching task (Soken and Pick 1999). But differences in infants' responses to sad and angry faces are taken as evidence that they are sensitive to second dimension, intensity, as well as valence.

Emotional labels together with the child's attempts to actively make sense of emotionally charged events, including antecedents and behavior consequences, drive emotional understanding. As children's perspective taking skills improve, they become increasingly tuned to the central role that appraisals play in emotion (Harris 1989). And as parents of preschoolers can attest, children ask many questions as they try to figure out the "why's" of emotion. Children face numerous challenges as they try to make connections between these explanations and experiences. Words, such as "angry" and "scared," map onto broad fuzzy categories and negative facial expressions are often confused, which can lead to incorrect inferences about causes and consequences. And as adults, we know that even if we could perfectly recognize and classify emotions, explaining the *why's* is never simple.

11.2.3 Separable Neural Processing Systems for Perceiving Facial Identity and Expression

Information about facial identity and facial expression coexist on a face. The simultaneous access of identity and expression from a face is computationally, but not perceptually, challenging. The computational difficulties occur because facial expressions consist of a set of complex non-rigid deformations of the features and shape of the face. Thus, the information that specifies identity is distorted to produce expressions. Concomitantly, facial expression movements are made with many different facial identities. In understanding how expression and identity analyses are coordinated neurally and psychologically, a look at the entire face processing system is helpful.

One of the first comprehensive models of human face processing, proposed by Bruce and Young in 1986, posited a separation of expression and identity processing. This was supported by evidence from neuropsychological case studies that indicated a *double dissociation* between expression and identity processing in brain-injured individuals. Indeed, evidence for the importance of particular local brain areas for processing face identity has been available for over a half of a century (Bodamer 1947). This evidence comes in the form of case studies of selective impairment of face recognition following brain damage to the inferior temporal area of the brain. Damage to this region of the brain can produce *prosopagnosia*, a neuropsychological impairment in which patients lose the ability to identify people by their faces, while retaining general visual object recognition skills.

Evidence for the first part of the double dissociation of face identity and expression processing comes from case studies of prosopagnosics who have impaired face recognition abilities, but spared expression perception (e.g., for a review Young et al. (1993)). The second part of the double dissociation comes from documented cases of brain damaged individuals who fail to perceive facial expressions, but have no difficulty recognizing faces (e.g., Adolphs et al. 1994). These individuals generally have damage to the amygdala, a sub-cortical brain area with importance for processing emotional stimuli. The amygdala connects to other subcortical structures that generate emotional responses, including increases in heart rate and blood pressure.

With the advent of more sophisticated functional brain imaging technologies, it has been possible to view the neural activity in intact brains in a way that highlights the importance of brain regions for specific perceptual and cognitive tasks. Functional magnetic resonance imaging (fMRI), for example, yields a high resolution three-dimensional map of brain activity. The application of this technology to human subjects actively engaging in face processing tasks has refined our knowledge of the face recognition and expression perception systems beyond what was known from neuropsychological case studies.

Based on functional neuroimaging data, neuropsychological case studies, and electrophysiological recordings from neurons in non-human primates, Haxby et al. (2000) proposed an updated functional model of the human face processing system. They proposed three brain regions as the core of a distributed neural network for processing faces: (a) the lateral fusiform gyrus (in the inferior temporal lobe), (b) the superior temporal sulcus (pSTS); and (c) the inferior occipital gyrus. According to this model, the lateral fusiform gyrus, sometimes called the fusiform face area (FFA) (Kanwisher et al. 1997), processes the invariant information from faces useful for identification. The pSTS processes the changeable information in faces that specifies facial expression, facial speech, and eye gaze. This separation between FFA and pSTS functions is linked to low level visual processing differences for static and dynamic (moving) stimuli. The FFA receives input from visual processing streams associated with high resolution form information and is thought to analyze the features and configuration of a face. The pSTS, on the other hand, receives input from motion sensitive neurons in the visual cortex and is thought to be involved in the perception of dynamic facial movements (i.e., for expression, gaze, speech). Haxby et al. note that a secondary set of brain structures continues processing the information from the core network. Among these structures is the amygdala, which is involved in the activation of emotional responses, including the increase in blood pressure.

One additional insight of the Haxby et al. model is that the pSTS as a processor of facial expression, facial speech, and gaze/head movement, by its nature, processes the *motions* of the face and body. A facial expression is created through movements of the face. Eye gaze shifts and head movement indicate changes in our focus of attention. Moreover, the areas of the pSTS that process facial motion are contiguous with areas that process gait (i.e., the way people walk) and a host of other body part motions including motions of the hands. It is through these movements that people express their intent to act.

In summary, a look at the neural processing of moving and static faces and bodies, is consistent with both a structural and functional separation of the information. The function of a moving face and body is often to communicate social or affective intent. The brain areas that process these messages are closely connected to the parts of the brain that can act and react appropriately. The function of the static information in faces and bodies may largely be more useful to cognitive processes that recognize and categorize people. Thus, the neural systems may suggest a logic for dividing the processing of people and their expressions/intents into components that focus on the invariant and changeable parts of the signal.

11.2.4 Lessons Learned: Human Facial Expression and Identity Processing

The primary lesson we learn from considering human face recognition and expression perception is that, like machines, we are not perfect. Indeed, there seems to be a bias in the machine recognition literature to believe that "If only, my machine could perform as well as humans, all would be well". A look at human performances suggests that we should beware of such simplistic goals. The fact that there is, as yet, no perfect system for these tasks should motivate us to set more realistic goals for developing machines that are aimed at identifying people and their intentions. The machine and human literatures seem to be moving toward a broader inclusion of multiple biometrics in establishing identity. The psychology literature on expression likewise points to the need to understand the full socio-cognitive and affective context to determine the intent of an individual. More basic research is needed to elucidate the many factors that may play a role in establishing this context. It is likely however that the interest of machine vision researchers in the problem of understanding human intent, will open the door to an interesting and productive dialog on these issues.

11.3 Face Recognition Technologies

First generation biometrics has been puzzled in devising a 100% error-free face recognition system. The large variations implied in face appearance make it rather difficult to be able to capture all relevant features in faces regardless of the environmental conditions. Therefore, the most successful face recognition systems are based on imposing some constraints on the data acquisition conditions. This is the case for the CASIA F1 system (Li et al. 2007), which is based on the active illumination by infrared illuminators and a near-infrared camera sensor. This is also the case for commercial products like the automated portal for automated border check developed by L1 Identity Solutions. This is not the way face recognition is going to revolutionarize today's use of identification technologies. We expect that face recognition systems will be more and more able to work in "open air", detecting and identifying people remotely (from a distance), in motion and with any illumination. The main breakthrough is expected to come from the use of high resolution images and exploiting the time-evolution of data rather than single snapshots.

In general, face recognition technologies are based on a two step approach:

- An off-line enrollment procedure is established to build a unique template for each registered user. The procedure is based on the acquisition of a pre-defined set of face images, selected from the input image stream, or a complete video, and the template is build upon a set of features extracted from the image ensemble;
- An on-line identification or verification procedure where a set of images are acquired and processed to extract a given set of features. From these features, a face description is built to be matched against the user's template.

Regardless of the acquisition devices exploited to grab the image streams, a simple taxonomy can be based on the computational architecture applied to: extract distinctive and possibly unique features for identification and to derive a template description for subsequent matching.

The two main algorithmic categories can be defined on the basis of the relation between the subject and the face model, i.e. whether the algorithm is based on a subject-centered (eco-centric) representation or on a camera-centered (ego-centric) representation. The former class of algorithms relies on a more complex model of the face, which is generally 3D or 2.5D, and it is strongly linked with the 3D structure of the face. These methods rely on a more complex procedure to extract the features and build the face model, but they have the advantage of being intrinsically pose-invariant. The most popular face-centered algorithms are those based on 3D face data acquisition and on face depth maps.

The ego-centric class of algorithms strongly relies on the information content of the gray level structures of the images. Therefore, the face representation is strongly pose-variant and the model is rigidly linked to the face appearance, rather than to the 3D face structure. The most popular image-centered algorithms are the holistic or subspace-based methods, the feature-based methods and the hybrid methods. Over these elementary classes of algorithms several elaborations have been proposed. Among them, the kernel methods greatly enhanced the discrimination power of several ego-centric algorithms, while new feature analysis techniques, such as the local binary pattern (LBP) representation, greatly improved the speed and robustness of Gabor-filtering based methods. The same considerations are valid for eco-centric algorithms, where new shape descriptors and 3D parametric models, including the fusion of shape information with the 2D face texture, considerably enhanced the accuracy of existing methods.

11.3.1 Face Analysis from Video Streams

When monitoring people with surveillance cameras at a distance it is possible to collect information over time. This process allows to build a rich representation than using a single snapshot. It is therefore possible to define a "dynamic template". This representation can encompass both physical and behavioral traits, thus enhancing the discrimination power of the classifier applied for identification or verification.

The representation of the subject's identity can be arbitrarily rich at the cost of a large template size. Several approaches have been proposed to generalize classical face representations based on a single-to-multiple view representations. Examples of this kind can be found in Lucas (1998), Lucas and Huang (2004) and Raytchev and Murase (2003, 2002, 2001), where face sequences are clustered using vector quantization into different views and subsequently fed to a statistical classifier. Recently, Zhou et al. (2003, 2004) proposed the "video-to-video" paradigm, where the whole sequence of faces, acquired during a given time interval, is associated to a class (identity). This concept implies the temporal analysis of the video sequence with dynamical models (e.g., Bayesian models), and the "condensation" of the tracking and recognition problems. Other face recognition systems, based on the still-to-still and multiple stills-to-still paradigms, have been proposed (Li et al. 2001, 2000; Howell and Buxton 1996). However, none of them is able to effectively handle the large variability of critical parameters, like pose, lighting, scale, face expression, some kind of forgery in the subject appearance (e.g., the beard). Other interesting approaches are based on the extension of conventional, parametric classifiers to improve the "face space" representation. Among them are the extended HMMs (Liu and Chen 2003), the Pseudo-Hierarchical HMMs (Bicego et al. 2006; Tistarelli et al. 2008) and parametric eigenspaces (Arandjelovic and Cipolla 2004), where the dynamic information in the video sequence is explicitely used to improve the face representation and, consequently, the discrimination power of the classifier.

11.4 Facial Expression and Emotion Recognition

Rosalind Picard in (2000) defined the notion of *affective computing* as the computational process that relates to, arises from, or deliberately influences emotions. This concept may be simply formulated as giving a computer the ability to recognize and express emotions (which is rather different from the question of whether computers can have 'feelings') and, consequently, develop the ability to respond intelligently to human emotion. Most human interactions are conveyed through emotional expression. Therefore it should not be surprising if computers are enabled to discern and interpret emotions. This should be regarded as an added feature which improves man machine interaction, and may facilitate the fruition of goods and services.

The most expressive (and the easiest to capture) way humans display emotion is through facial expressions (Sebe et al. 2005). Humans detect and interpret faces and facial expressions in a scene with little or no effort. Still, developing an automated system that accomplishes this task is rather difficult. There are several related problems: detection of an image segment as a face, extraction of the facial expression information, and accurate classification of the facial expression within a set of pre-assessed emotion categories. The achievement of this goal has been studied for a relatively long time with the aim of achieving human-like interaction between human and machine. Since the early 1970s Paul Ekman and his colleagues have performed extensive studies of human facial expressions (Ekman 1994). They found evidence for "universal facial expressions" representing happiness, sadness, anger, fear, surprise and disgust. They studied facial expressions in different cultures, including preliterate cultures, and found much commonality in the expression and recognition of emotions on the face. However, they observed differences in expressions as well and proposed that facial expressions are governed by "display rules" within different social contexts. In order to facilitate the analytical coding of expressions, Ekman and Friesen (1978) developed the Facial Action Coding System (FACS) code. In this system movements on the face, leading to facial expressions, are described by a set of action units (AUs). Each AU has some related muscular basis. This system of coding facial expressions is done manually by following a set of prescribed rules. The inputs are still images of facial expressions, often at the peak of the expression. Ekman's work inspired many researchers to analyze facial expressions by means of image and video processing. By tracking facial features and measuring the amount of facial movement, they attempt to categorize different facial expressions. Past work on facial expression analysis and recognition has used these "basic expressions" or a subset of them. Fasel and Luettin (2003), Pantic and Rothkrantz (2000) and more recently Gunes et al. (2008) and Zeng et al. (2009) provide an in-depth review of much of the research done in automatic facial expression recognition over the years since the early 1990s.

Among the different approaches proposed, Chen (2000) and Chen and Huang (2000) used a suite of static classifiers to recognize facial expressions, reporting on both person-dependent and person-independent results. Cohen et al. (2003) describe classification schemes for facial expression recognition in two types of settings: dynamic and static classification. In the static setting, the authors learned the structure of Bayesian network classifiers using as input 12 motion units given by a face tracking system for each frame in a video. For the dynamic setting, they used a multilevel HMM classifier that combines the temporal information and allows one not only to classify video segments with the corresponding facial expressions, as in the previous works on HMM-based classifiers, but also to automatically segment an arbitrary long sequence to the different expression segments without resorting to heuristic methods of segmentation (Sebe et al. 2005). These methods are similar in that they first extract some features from the images, then use these features as inputs into a classification system, and the outcome is one of the preselected emotion categories. They differ mainly in the features extracted from the video images and in the classifiers used to distinguish between the different emotions. An interesting feature of this approach is the commonality with some methods for identity verification from video streams (Shan and Braspenning 2009). On the other hand, a recurrent issue is the need to either implicitly or explicitly determine the changes in the face appearance. The facial motion is thus a very powerful feature to discriminate facial expressions and the correlated emotional states.

Recently Gunes et al. (Gunes and Piccardi 2006, 2009; Gunes et al. 2008) extended the concept of emotion recognition from face movements to a more holistic approach, including also the movements of the body. In this work also the range of detected emotions is increased, adding most common emotions such as Boredom and Anxiety. The integration of face and body motion allowed to improve the recognition performances as compared to the analysis of face alone.

11.4.1 Computation of Perceived Motion

The motion of 3D objects in space induces an apparent motion on the image plane. The apparent motion can either be computed on the basis of the displacement of few image features points or as a dense field of displacement vectors over the entire image plane. This vector field is often called optical flow. In general, there is a difference between the displacement field of the image points and the actual projection of the 3D motion of the objects on the image plane (the velocity field). This difference is due to the ambiguity induced by the loss in dimensionality when projecting 3D points on the image plane. A simple example is the motion of a uniformly colored disc, parallel to the image plane, rotating around its center. As there is no motion on the image plane the optical flow is zero, but the 3D motion of the object, and consequently its projection, the velocity field, is non-zero. Nonetheless, apart from some degenerate conditions, the optical flow is generally a good approximation of the velocity field. The optical flow can be computed by processing a set of image frames from a video. The process is based on the estimation of the displacement of the image points over time. In general terms, the displacements can be computed either explicitly or implicitly. In the first case, the motion of image patterns is computed by matching corresponding patches on successive image frames. In the latter case, the instantaneous velocity is determined by computing the differential properties of the time-varying intensity patterns. This generally implies the computation of partial derivatives of the image brightness, or filtering the image sequence with a bank of filters properly tuned in the spatial and temporal frequencies, and the fulfillment of some constraints on the apparent motion (Tretiak and Pastor 1984; Tistarelli 1996; Horn and Schunck 1981). The combination of multiple gradient-based constraints allows to compute the value of the flow vector, best approximating the true velocity field, at each image point. Some examples of flow fields computed from different image streams and different actions are presented in Figs. 11.1 and 11.2

11.4.2 Performances of Expression and Emotion Recognition Algorithms

Several algorithms to detect facial expressions and infer emotions have been developed and tested. The datasets used and the reported performances are quite varying, but all share some common denominators.

Firstly, no single system is capable of achieving 100% correct categorization, even with a very limited number of subjects (Sebe et al. 2006; Gunes and

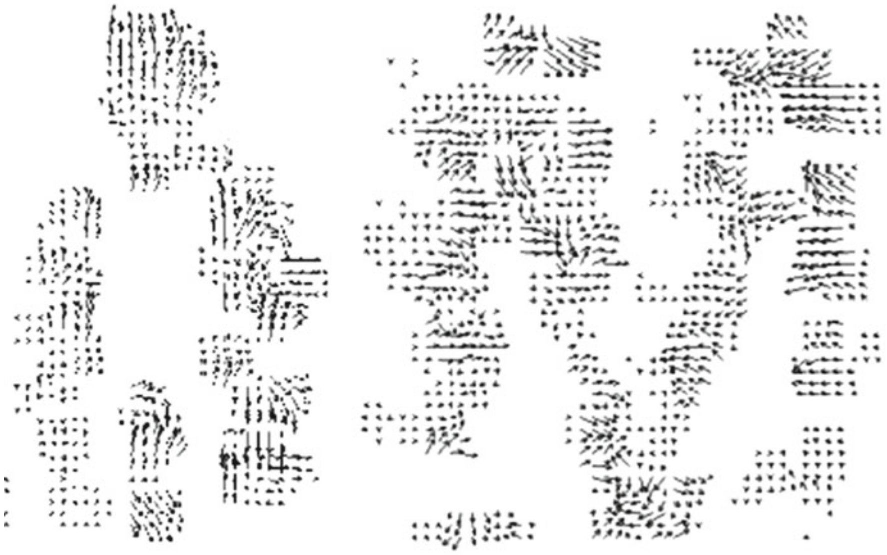

Fig. 11.1 (*Left*) Optical flow from an image sequence of a person jumping. (*Right*) Optical flow from an image sequence of two persons walking

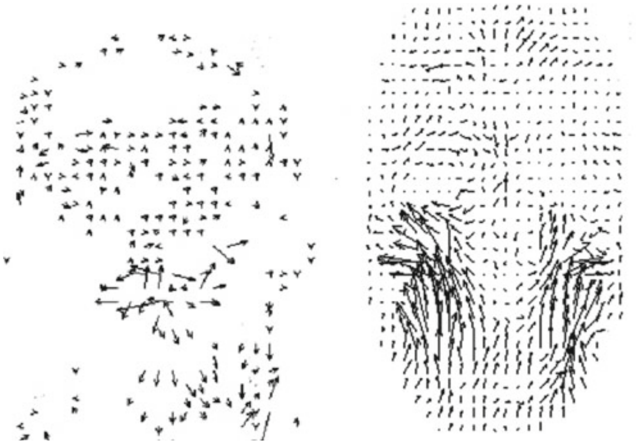

Fig. 11.2 (*Left*) Optical flow from an image sequence of a talking face. (*Right*) Optical flow from an image sequence of a face changing expression from neutral to hanger

Piccardi 2006). The datasets employed for the experiments are quite varied and include from 10 up to a maximum of 210 subjects. The Cohn Kanade AUCoded facial expression database (Kanade et al. 2000), being the largest data collection with 2105 videos from 100 subjects. Yet, the number of captured expression is not totally uniform and, more importantly, it is not fully representative of the variety of emotions which can be experienced and displayed by humans in different environments. The datasets and relative experimental results are limited to six basic expressions

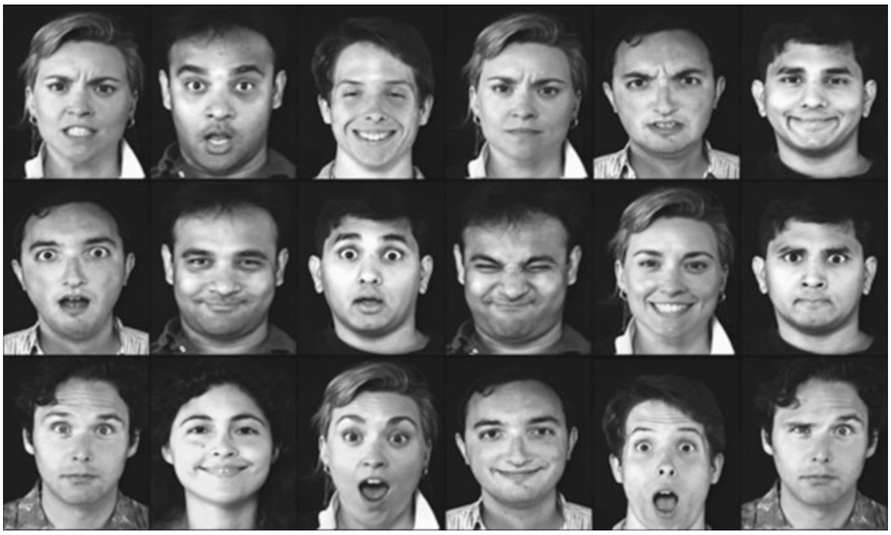

Fig. 11.3 Sample snapshots from a standard video database for facial expression analysis (Essa and Pentland 1995)

(Fear, Anger, Disgust, Happiness, Sadness, Surprise). More recently, these expression have been augmented to a more complete set including also Uncertainty, Anxiety and Boredom. These last three being most important to detect uncomfortable states of the subject, for example, while waiting in a long queue at the post office or at a ticket counter.

The second common denominator is the acquisition condition. All the datasets have been collected in indoor, controlled environments (generally research labs). As a consequence, most of expressed emotions are purposively constructed, rather than grabbed from a real scenery (see Fig. 11.3). Therefore, the datasets display a number of "expected" expressions, rather than genuine expressions as caused by either external (physical) or internal (emotional) factors. In a few cases the data was acquired from spontaneous emotions derived from external stimuli (Zeng et al. 2009). Even though, in most cases, humans can perceive a difference between the real and posed (or exaggerated) facial expression, devising an algorithm to automatically determine this difference requires an extensive test set.

Another similarity is the variation in the classification performances across different subjects. Not everybody displays the emotions exactly in the same way. Depending on the cultural background, there are several differences in the visible effects of emotions. Sometimes differences among emotions are just subtile movements of the eyes or the mouth. In other cases emotions are expressed by largely evident motions of the face parts. As a consequence, it is rather difficult to conceive a single system which is able of performing equally well over a large variety of subjects, especially in real world conditions.

Remarkably several systems have been recently proposed to improve the classification performances by using multiple visual and non visual cues (Zeng et al. 2009; Busso et al. 2004; Castellano et al. 2008; Datcu and Rothkrantz 2008; Ratliff and Patterson 2008; Paleari et al. 2009). The work by Sebe et al. (2006) and by Gunes et al. (Gunes and Piccardi 2006, 2009) are examples of how speech and the body motion can be exploited to improve the inference of displayed emotions. If emotional states are detected to facilitate the human computer interactions, as it is in many research projects, the available datasets and emotion categorizations are generally sufficient to describe the range of possible emotional states to be discovered by a computer to start the necessary actions. In the case of non-cooperative subjects, where a surveillance camera grabs images from freely moving people in the environment, the range of possible emotions and consequent motions and expressions is quite larger. Therefore, it is necessary to better understand the target environmental conditions first, and then understand the range of emotions which may be observed.

11.5 Emotion and Intention Recognition

There is strong relation between emotional states and intentions. Anxiety status can easily lead to an unpredictable action or a violent reaction. As discussed in the preceding sections, current technology can only discern among basic emotions through the categorization of facial motions. Still, tomorrow's technology may allow to discern more subtile expression changes to capture a wider spectrum of emotions, which may lead to detect dangerous intentions or possible future actions.

Nowadays, there is a considerable concern for biometric technologies breaking the citizen's privacy. The public view of the widespread adoption of surveillance camera is that the use of face identification systems may lead to a "state of control" scenario where people are continuously monitored against their supposed intentions. Science fiction and action movies greatly contributed to shape the public view of intelligent surveillance. Many people may see a "preemptive police corp" into action within the next few years. Even though we can't reasonably predict what the advance of technology will be in the next few years, these concerns have yet hindered the adoption of these technologies in many environments, thus depriving of many potential service improvements and benefits. Unfortunately, as in the past, also today misinformation and politically driven views impair technological advances which can be adopted for the wellness of the citizens. Yet, as any other technology, the way it is used, either for the good or bad, entirely depends on the intentions of the men behind its application. As stated above, the research on facial expression recognition has been largely lead by the aim to develop more intuitive and easy to use human computer interfaces (Picard 2000; Fasel and Luettin 2003). The overall idea being to "instruct" computers to adapt the delivered services or informations according to the "mood" and appreciation level of the users (Gunes et al. 2008). On the other hand, if security is the target application for an "emotion detection" surveillance system, at the basis there is a tradeoff between security and freedom. We already traded much of our personal freedom for air travel security.

Nobody can access any airport's gate area without being checked. We can't even carry a number of items onboard. Exceptions are quite limited and very hard to achieve. Still, millions of people travel by air every day (apparently) without any complaint. A more realistic scenario, where intelligent camera networks continuously check for *abnormal* and *potentially dangerous* behaviors in high risk areas or crowded places, may on the contrary increase the "sense of protection" and security of citizens. The same cameras may then be used in shops and lounges to determine the choice and satisfaction level of users to provide better customized services. In reality, there is no need to really impair personal privacy by coupling personal data with the identity of the bearer. This is far from being useful in the real practice.

We don't need to wait for the massive introduction of intelligent "emotionally driven" surveillance systems. Real privacy threats are already a reality. The widespread use of social networks and e-services, coupled with the misuse of personal data cross-check and fusion, is already a potential danger which may impair our privacy. This, unless the proper ruling actions are taken to hinder the mis-use (rather than the use) of personal data. Better spread of knowledge, at all levels, is the first step, then followed by an appropriate development of pan-national rules supported by technology standards is the only solution to defeat the privacy threats.

11.6 Conclusion

There is always a deep concern about the introduction of new technologies which may impact on the personal privacy. For this reason, all technologies which are devoted to either identify individuals or even for understanding some personal features, are looked suspiciously by the large public. Nonetheless, there are research projects which are now seriously looking not only into the identification of individuals, but indeed at the understanding of potentially hostile intentions. An example is the FAST (Future Attribute Screening Technologies) project of the US Department of Homeland Security.[1] A popular web magazine addresses this project in this way: *The Department of Homeland Security has been researching a sensor system that tries to predict"hostile thoughts" in people remotely for a while, but it's just spoken up about developments and renamed the system "Future Attribute Screening Technologies" FAST, which sounds really non-intimidating. It was called "Project Hostile Intent." But check out the technology's supposed powers for a re-think on how intimidating it sounds: it remotely checks people's pulse rate, breathing patterns, skin temperature and momentary facial expressions to see if they're up to no good.*

As any other technology, biometrics can be used for the good or bad of the citizens, but this is not a motivation to ban the technology per se. Rather this is a

[1] A description of the project can be found under the "Screening Technologies to detect intentions of humans" on the web page http://www.dhs.gov/xabout/structure/gc_1224537081868.shtm