

# Recognizing Facial Expressions Automatically from Video

Caifeng Shan and Ralph Braspenning

## 1 Introduction

Facial expressions, resulting from movements of the facial muscles, are the face changes in response to a person's internal emotional states, intentions, or social communications. There is a considerable history associated with the study on facial expressions. Darwin (1872) was the first to describe in details the specific facial expressions associated with emotions in animals and humans, who argued that all mammals show emotions reliably in their faces. Since that, facial expression analysis has been a area of great research interest for behavioral scientists (Ekman, Friesen, and Hager, 2002). Psychological studies (Mehrabian, 1968; Ambady and Rosenthal, 1992) suggest that facial expressions, as the main mode for non-verbal communication, play a vital role in human face-to-face communication. For illustration, we show some examples of facial expressions in Fig. 1.

Computer recognition of facial expressions has many important applications in intelligent human-computer interaction, computer animation, surveillance and security, medical diagnosis, law enforcement, and awareness systems (Shan, 2007). Therefore, it has been an active research topic in multiple disciplines such as psychology, cognitive science, human-computer interaction, and pattern recognition. Meanwhile, as a promising unobtrusive solution, automatic facial expression analysis from video or images has received much attention in last two decades (Pantic and Rothkrantz, 2000a; Fasel and Luetin, 2003; Tian, Kanade, and Cohn, 2005; Pantic and Bartlett, 2007).

This chapter introduces recent advances in computer recognition of facial expressions. Firstly, we describe the problem space, which includes multiple dimensions: level of description, static versus dynamic expression, facial feature extraction and

---

Caifeng Shan  
Philips Research, Eindhoven, The Netherlands, e-mail: caifeng.shan@philips.com

Ralph Braspenning  
Philips Research, Eindhoven, The Netherlands, e-mail: ralph.braspenning@philips.com



**Fig. 1** Facial expressions of Tony Blair.

representation, expression subspace learning, facial expression recognition, posed versus spontaneous expression, expression intensity, controlled versus uncontrolled data acquisition, correlation with bodily expression, and databases. Meanwhile, the state of the art of facial expression recognition is also surveyed from different aspects. In the second part, we present our recent work on recognizing facial expressions using discriminative local statistical features. Specially Local Binary Patterns are investigated for facial representation. Finally, the current status, open problems, and research directions are discussed.

The remainder of this chapter is organized as follows. We first review the problem space and the state of the art in Section 2. Our recent work on facial expression recognition using Local Binary Patterns are then described in Section 3. Finally Section 4 concludes the chapter with a discussion on challenges and future opportunities.

## 2 Problem Space and State of the Art

### 2.1 Level of Description

Facial expressions can be described at different levels (Tian, Kanade, and Cohn, 2005). Two mainstream description methods are facial affect (emotion) and facial muscle action (action unit) (Pantic and Bartlett, 2007). Psychologists suggest that some basic emotions are universally displayed and recognized from facial expressions (Ekman and Friesen, 1971), and the most commonly used descriptors are the six basic emotions, which includes anger, disgust, fear, joy, surprise, and sadness (see Fig. 2 for examples). This is also reflected by the research on automatic facial expression analysis; most facial expression analysis systems developed so far target facial affect analysis, and attempt to recognize a set of prototypic emotional facial

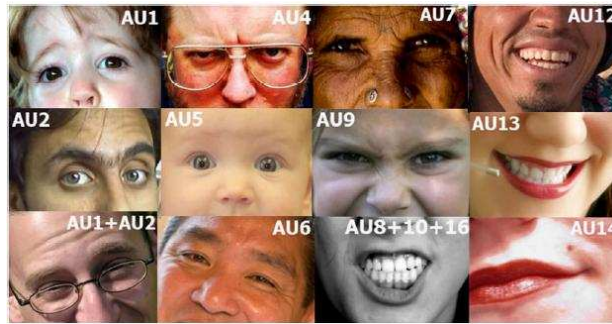
expressions (Pantic and Rothkrantz, 2000a, 2003). There have also been some tentative efforts to detect cognitive and psychological states like interest (Kaliouby and Robinson, 2004), fatigue (Gu and Ji, 2005), and pain (Littlewort, Bartlett, and Lee, 2006b). To describe subtle facial changes, the Facial Action Coding System (FACS) (Ekman, Friesen, and Hager, 2002) has been widely used for manually labeling of facial actions in behavioral science. FACS associates facial changes with actions of the muscles that produce them. See Fig. 3 for some examples of action units (AUs). It is possible to map AUs onto the basic emotions using a finite number of rules (Ekman, Friesen, and Hager, 2002). Automatic AU detection has been studied recently (Donato, Bartlett, Hager, Ekman, and Sejnowski, 1999; Tian, Kanade, and Cohn, 2001; Pantic and Rothkrantz, 2004; Zhang and Ji, 2005; Littlewort, Bartlett, Fasel, Susskind, and Movellan, 2006a). The major problem of AU-related research is the need of highly trained experts to manually perform FACS coding frame by frame. Approximately 300 hours of training are required to achieve minimal competency of FACS, and each minute of video tapes takes around two hours to code comprehensively (Braathen, Bartlett, Littlewort, Smith, and Movellan, 2002). Another possible descriptor is the bipolar dimensions of *Valence* and *Arousal* (Russell, 1994). Valence describes the pleasantness, with positive (pleasant) on one end (e.g. happiness), and negative (unpleasant) on the other (e.g. disgust). The other dimension is arousal or activation, for example, sadness has low arousal, whereas surprise has a high arousal level. Different emotional labels can be plotted at various positions on a two-dimensional plane spanned by these two axes (as shown in Fig. 4).



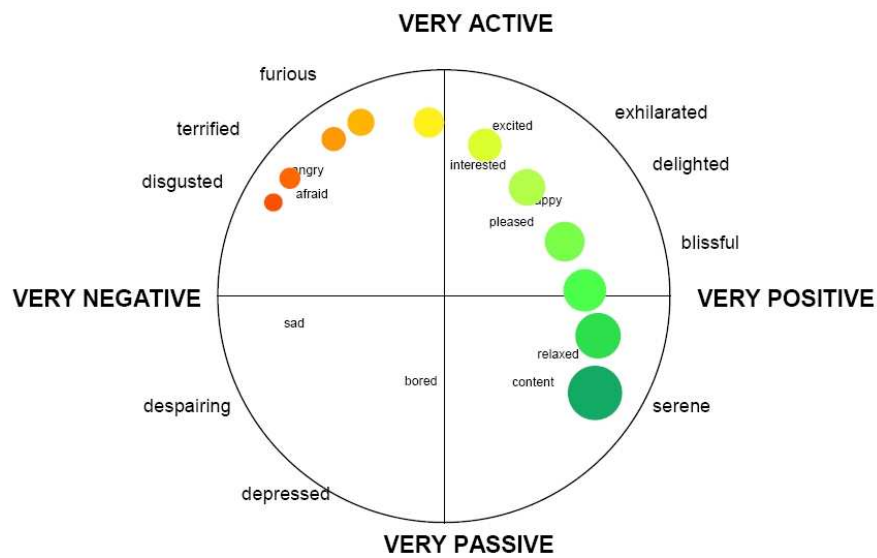
**Fig. 2** Prototypic emotional facial expressions: Anger, Disgust, Fear, Joy, Sadness, and Surprise (from left to right). From the Cohn-Kanade database (Kanade, Cohn, and Tian, 2000).

## 2.2 *Static Versus Dynamic Expression*

Facial expressions can be detected and recognized in a snapshot. So many existing works attempt to analyze facial expressions in each image (Lyons, Budynek, and Akamatsu, 1999; Donato, Bartlett, Hager, Ekman, and Sejnowski, 1999; Pantic and Rothkrantz, 2004; Littlewort, Bartlett, Fasel, Susskind, and Movellan, 2006a). For example, Cohen et al (2003) adopted Bayesian network classifiers to classify a frame in video sequences to one of the basic emotional expressions. However, psychological experiments (Bassili, 1979) suggest that the dynamics of facial expressions are



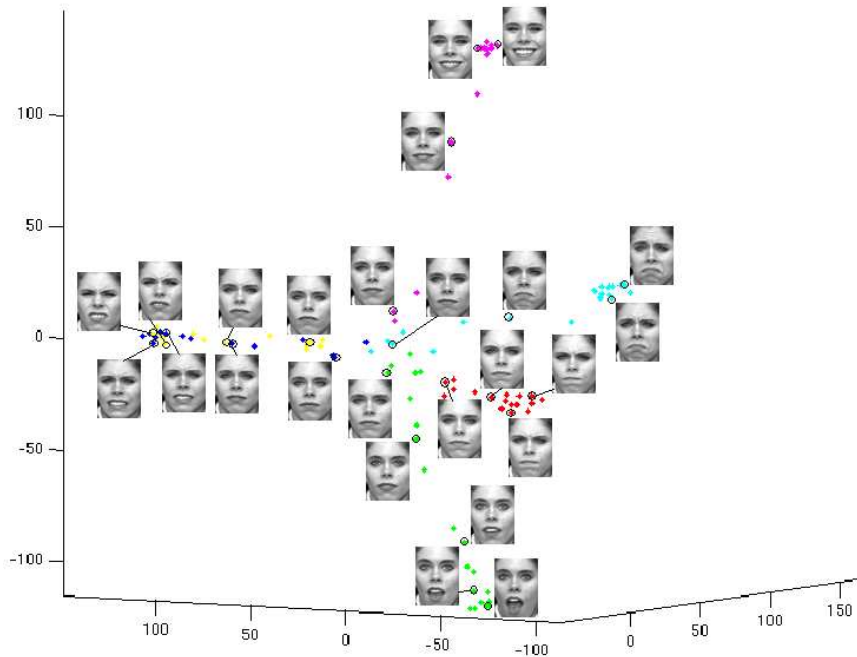
**Fig. 3** Examples of facial action units and their combination defined in FACS (Pantic and Bartlett, 2007).



**Fig. 4** Facial expressions labeled with *Valence* and *Arousal* (Cowie, Douglas-Cowie, Savvidou, McMahon, Sawey, and Schroder, 2000).

crucial for successful interpretation of facial expressions. The differences between facial expressions are often conveyed more powerfully by dynamic transitions between different stages of expressions rather than any single state represented by a still image. This is especially true for spontaneous facial expressions without any deliberate exaggerated posing (Ambadar, Schooler, and Cohn, 2005). Therefore, capturing and analyzing temporal dynamics of facial expressions in videos or image sequences is important for facial expression analysis. Recently many approaches have been introduced to model dynamic facial expressions in videos or image se-

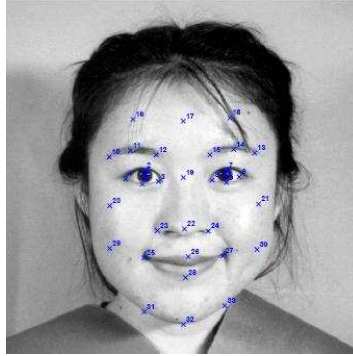
quences (Oliver, Pentland, and Berard, 2000; Zhang and Ji, 2005; Pantic and Patras, 2006). For example, Cohen et al (2003) presented a multi-level Hidden Markov Model (HMM) classifier to model temporal behaviors exhibited by facial expressions, which not only performs expression classification in a video segment, but also automatically segments an arbitrary long video sequence to different expression segments. Facial expression dynamics can also be captured in low dimensional manifolds embedded in the input image space (Chang, Hu, and Turk, 2004). Shan et al (2005a, 2006b) proposed to model facial expression dynamics by discovering the underlying low-dimensional manifold. An example of expression manifold learning is shown in Fig. 5. Lee and Elgammal (2005) recently introduced a framework to learn decomposable generative models for dynamic appearance of facial expressions where facial motion is constrained to one dimensional closed manifolds.



**Fig. 5** (Best viewed in color) Six image sequences of basic expressions of a person are mapped into the 3D embedding space (Shan, Gong, and McOwan, 2005a). Representative faces are shown next to circled points in different parts of the space. Different expressions are color coded as: Anger (red), Disgust (yellow), Fear (blue), Joy (magenta), Sadness (cyan), and Surprise (green).

### 2.3 Facial Feature Extraction and Representation

A vital step for successful facial expression analysis is deriving an effective facial representation from original face images. Two types of features (Tian, Kanade, and Cohn, 2005), geometric features and appearance features, are usually considered for facial representation. Geometric features deal with the shape and locations of facial components (such as mouth, eyes, brows, and nose), which are extracted to represent the face geometry. Appearance features present the appearance changes (skin texture) of the face (such as wrinkles, bulges and furrows), which are extracted by applying image filters to either the whole face or specific facial regions. The geometric features based approaches commonly require accurate and reliable facial feature detection and tracking, which is difficult to accommodate in real-world unconstrained scenarios, e.g., under head pose variation. In contrast, appearance features suffer less from issues of initialization and tracking errors, and can encode changes in skin texture that are critical for modeling. However, most of the existing appearance-based facial representations still require face registration based on facial feature detection, e.g., eye detection.



**Fig. 6** Geometric features (Zhang, Lyons, Schuster, and Akamatsu, 1998): 34 fiducial points for representing the facial geometry.

Fiducial facial feature points have been widely used in facial representation. As shown in Fig. 6, Zhang et al (1998) utilized the geometric positions of 34 fiducial points as facial features. A shape model defined by 58 facial landmarks was adopted in (Chang, Hu, and Turk, 2003, 2004). Pantic and her colleagues (Pantic and Rothkrantz, 2004; Pantic and Patras, 2006; Valstar and Pantic, 2006) used a set of facial characteristic points to describe facial actions (as shown in Fig. 7). Tian et al (2003) considered both location features and shape features for facial representation. Specifically, as shown in Fig. 8, six location features (eye centers, eyebrow inner endpoints, and corners of the mouth) are extracted and transformed into 5 parameters, and the mouth shape features are computed from zonal shape histograms of the edges in the mouth region. In image sequences, facial movements can be mea-

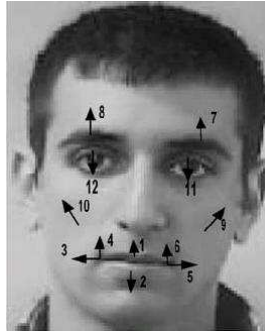
sured by the geometrical displacement of facial feature points between the current frame and the initial frame (Pantic and Rothkrantz, 2000b; Kaliouby and Robinson, 2004). Tian et al (2001) developed multi-state facial component models to track and extract the geometric facial features. Cohen et al (2003) referred the tracked motions of facial features (such as the eyebrows, eyelids, and mouth) at each frame as Motion-Units (MUs) (as shown in Fig. 9). The MUs represent not only the activation of a facial region, but also the direction and intensity of the motion. Optical flow analysis has also been used to model muscles activities or estimate the displacements of feature points (Yacoob and Davis, 1996; Essa and Pentland, 1997; Yeasin, Bulot, and Sharma, 2004). Although it is effective to obtain facial motion information by computing dense flow between successive image frames, flow estimation has its disadvantages such as easily disturbed by lighting variation and non-rigid motion, and sensitive to the inaccuracy of image registration and motion discontinuities (Zhang and Ji, 2005).



**Fig. 7** The facial points used to describe facial actions (Pantic and Bartlett, 2007).



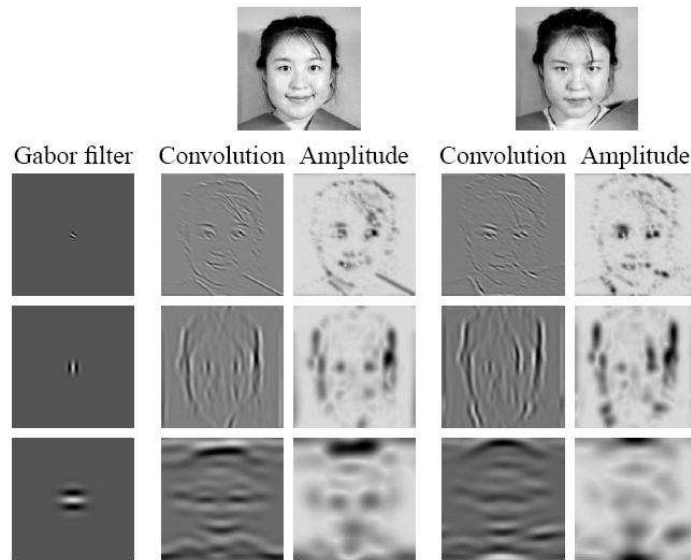
**Fig. 8** Geometric features (Tian, Brown, Hampapur, Pankanti, Senior, and Bolle, 2003): (Left) location features; (Right) normalized face and zones of the edge map of the normalized face.



**Fig. 9** The Motion-Units introduced in (Cohen, Sebe, Garg, Chen, and Huang, 2003).

Most of the existing appearance-based methods have adopted Gabor-wavelet features (Lyons, Budynek, and Akamatsu, 1999; Guo and Dyer, 2003; Littlewort, Bartlett, Fasel, Susskind, and Movellan, 2006a; Tong, Liao, and Ji, 2006). Gabor filters are obtained by modulating a 2D sine wave with a Gaussian envelope. Representations based on the outputs of Gabor filters at multiple spatial scales, orientations, and locations have proven successful for facial image analysis. Zhang et al (1998) computed a set of Gabor-filters coefficients at fiducial points (as shown in Fig. 10). Lyons et al (1999) represented each face using a set of Gabor filters at the facial feature points sampled from a sparse grid covering the face. Tian (2004) compared geometric features and Gabor-wavelet features with different image resolutions, and her experiments show that Gabor-wavelet features work better for low-resolution face images. Although widely adopted, it is computationally expensive to convolve face images with a set of Gabor filters. It is inefficient in both time and memory due to the high redundancy of Gabor-wavelet features. For example, in (Bartlett, Littlewort, Frank, Lainscsek, Fasel, and Movellan, 2005), the Gabor-wavelet representation derived from each  $48 \times 48$  face image has the high dimensionality of  $O(10^5)$ . Another kind of widely used appearance features is statistical subspace analysis, which will be discussed in Section 2.4.

The appearance-based approaches are believed to contain more information than those based on the relative positions of a finite set of facial features (Bartlett, Littlewort, Frank, Lainscsek, Fasel, and Movellan, 2006), so providing superior performance. For example, the Gabor-wavelet representation outperforms the performance upper-bound computed based on manual feature tracking (Littlewort, Bartlett, Fasel, Susskind, and Movellan, 2006a). However, some recent studies indicate that this claim does not always hold. For example, the AU detection method (Valstar and Pantic, 2006) based on tracked facial points achieved similar or higher detection rates than other methods. It seems that using both geometric and appearance features might be the best choice (Pantic and Bartlett, 2007). Recently feature selection methods have been exploited to select the most effective facial features. Guo and Dyer (2003) introduced a linear programming technique that jointly performs feature selection and classifier training so that a subset of features is optimally



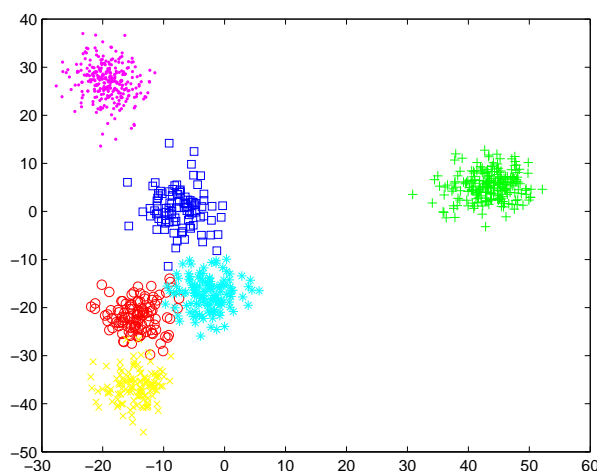
**Fig. 10** Gabor-wavelet representation (Zhang, Lyons, Schuster, and Akamatsu, 1998): two examples with three Gabor kernels.

selected together with the classifier. Bartlett et al (2005) selected a subset of Gabor filters using Adaboost.

## 2.4 Expression Subspace Learning

Appearance-based statistical subspace learning has been shown to be an effective approach for facial representation. This is because that despite a facial image space being commonly of a very high dimension, the underlying facial expression space is a sub-manifold of much lower dimensionality embedded in the ambient space. Subspace learning is a natural approach to resolve this problem. Traditionally, linear subspace methods including Principal Component Analysis (PCA) (Turk and Pentland, 1991), Linear Discriminant Analysis (LDA) (Belhumeur, Hespanha, and Kriegman, 1997), and Independent Component Analysis (ICA) (Bartlett, Movellan, and Sejnowski, 2002) have been used to discover both facial identity and expression manifold structures. For example, Lyons et al (1999) adopted PCA based LDA with the Gabor-wavelet features to classify facial images. Donato et al (1999) explored PCA, LDA, ICA, and Gabor-wavelet representation for facial action classification. Best performances were obtained using the Gabor-wavelet features and ICA. Recently, a number of graph-based linear subspace techniques have been proposed in literature, e.g., Locality Preserving Projections (LPP) (He and Niyogi, 2003). Shan et al (2006a) comprehensively investigate and evaluate these linear subspace

methods for facial expression analysis. An example of 2D embedding subspace of emotional facial expressions is shown in Fig. 11, which was derived using the supervised LPP. As facial muscles are contracted in unison to display facial expressions, different facial parts have strong correlations. Shan et al (2007b) further employed Canonical Correlation Analysis, a statistical technique that is well suited for relating two sets of signals, to model correlations among facial parts.



**Fig. 11** (Best viewed in color) The 2D embedding subspace of emotional facial expressions (Shan, Gong, and McOwan, 2006a). Different expressions are color coded as: Anger (red), Disgust (yellow), Fear (blue), Joy (magenta), Sadness (cyan), and Surprise (green).

## 2.5 Facial Expression Classification

Different techniques have been proposed for facial expression recognition, such as Neural Networks (Tian, Kanade, and Cohn, 2001), Support Vector Machine (SVM) (Bartlett, Littlewort, Frank, Lainscsek, Fasel, and Movellan, 2005), Bayesian Networks (Cohen, Sebe, Garg, Chen, and Huang, 2003), HMM (Yeasin, Bullot, and Sharma, 2004), Dynamic Bayesian Network (Zhang and Ji, 2005), and rule-based classifiers (Pantic and Rothkrantz, 2000b). Facial expression recognition can be divided into image-based or sequence-based. The image-based approaches use features extracted from a single image to recognize the expression of that image, while the sequence-based methods aim to capture the temporal pattern in a sequence to recognize the expression for one or more images.

Yacoob and Davis (1996) used local motions of facial features to construct a mid-level description of facial motions, which was classified into one of six facial expressions using a set of heuristic rules. Using the dual-view (front and profile) geometric features, Pantic and Rothkrantz (2000b) performed facial expression recognition by comparing the AU-coded description of an observed expression against rule descriptors of basic emotions. Recently they further adopted the rule-based reasoning to recognize action units and their combination (Pantic and Rothkrantz, 2004). Tian et al (2001) used a three-layer Neural Network with one hidden layer to recognize AUs by a standard back-propagation method. Lyons et al (1999) adopted a nearest neighbor classifier to recognize facial images using discriminant Gabor-wavelet features. As a powerful discriminative machine learning technique, SVM has been widely adopted for facial expression recognition. More recently Bartlett et al (2005) performed comparison of Adaboost, SVM, and LDA, and best results were obtained by selecting a subset of Gabor filters using Adaboost and then training SVM on the outputs of the selected filters. This strategy is also adopted in (Tong, Liao, and Ji, 2006; Valstar and Pantic, 2006). For example, Valstar and Pantic (2006) recognized AU temporal segments using a subset of most informative spatio-temporal features selected by Adaboost.

As one of the basic probabilistic tools used for time series modeling, HMM has been exploited to capture temporal behaviors exhibited by facial expressions. For example, Oliver et al (2000) associated each of the mouth-based expressions, e.g. sad and smile, with an HMM trained using the mouth features, and facial expressions were identified by computing the maximum likelihood of the input sequence with respect to all trained HMMs. Yeasin et al (2004) presented a two-stage approach to classify six basic emotions, and derive the level of interest using psychological evidences. First, a bank of linear classifiers were applied at frame level and the output was coalesced to produce the temporal signature for each observation. Second, temporal signatures computed from the training data set were used to train discrete HMMs to learn the underlying models for each expression. Dynamic Bayesian Networks (DBNs) are graphical probabilistic models which encode dependencies among sets of random variables evolving in time. DBNs are capable of accounting for uncertainty in facial expression recognition, representing probabilistic relationships among different actions and modeling the dynamics in facial action development. Zhang and Ji (2005) explored the multisensory information fusion technique with DBNs for modeling and understanding the temporal behaviors of facial expressions in image sequences. Kaliouby and Robinson (2004) proposed a system for inferring complex mental states from videos of facial expressions and head gestures in real-time. Their system was built on a multi-level DBN classifier which models complex mental states as a number of interacting facial and head displays. Recently Tong et al (2006) proposed to exploit the relationships among different AUs and their temporal dynamics using DBN.

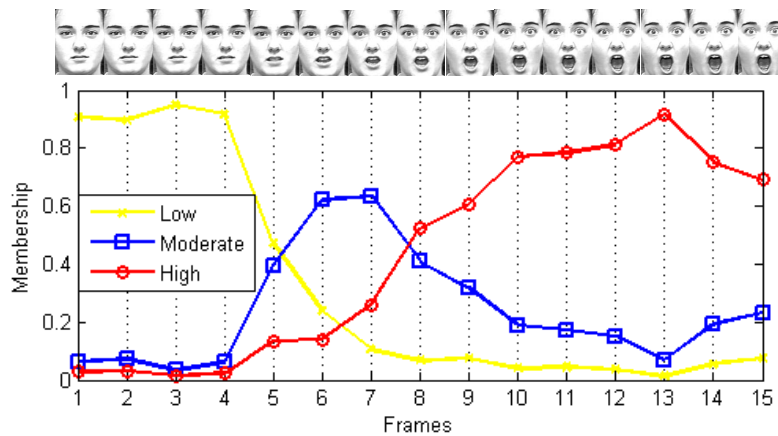
## 2.6 *Spontaneous Versus Posed Expression*

Most of the existing works have been carried out on expression data that were collected by asking subjects to deliberately pose facial expressions. However, the exaggerated facial expressions occur rarely in real-life situations. Spontaneous facial expressions induced in natural environments are more subtle and fleeting, such as tightening of the lips in anger or lowering the lip corners in sadness (Tian, Kanade, and Cohn, 2005). Recently research attention has started to shift to spontaneous facial expression analysis. Sebe et al (2004) collected a database containing spontaneous facial expressions, and compared a wide range of classifiers, including Bayesian Networks, the Decision Trees, SVM, k-Nearest-Neighbor, Bagging, and Boosting, for expression recognition. Surprisingly, the best classification results were obtained with the k-NN classifier. It seems that all the models tried were not able to entirely capture the complex decision boundary that separates different spontaneous expressions. Cohn et al (2004) developed an automated system to recognize brow actions in spontaneous facial behaviors captured in interview situations. Their recognition accuracy was relatively worse than that for the posed facial behaviors. Zeng et al (2006) treated the problem of emotional expression detection in a realistic conversation setting as an one-class classification problem, and adopted Support Vector Data Description to distinguish emotional expressions from non-emotional ones. Bartlett et al (2006) recently presented preliminary results on facial action detection in spontaneous facial expressions by adopting their AU recognition approach (Littlewort, Bartlett, Fasel, Susskind, and Movellan, 2006a).

Spontaneous facial expressions differ from posed expressions both in terms of which muscles move and how they move dynamically. Cohn and Schmidt (2004) observed that posed smiles were of larger amplitude and has a less consistent relationship between amplitude and duration than spontaneous smiles. A psychological study (Ekman and Rosenberg, 1997) has indicated that posed expressions may differ in appearance and timing from spontaneous ones. For example, spontaneous expressions have fast and smooth onsets, with distinct facial action peaking simultaneously, while posed expressions tend to have slow and jerky onsets, and the actions typically do not peak simultaneously (Bartlett, Littlewort, Frank, Lainscsek, Fasel, and Movellan, 2006). So facial dynamics is a key parameter in differentiating posed expressions from spontaneous ones (Valstar, Pantic, Ambadar, and Cohn, 2006). Recently Valstar et al (2006) experimentally showed that temporal dynamics of spontaneous and posed brow actions are different from each other. They built a system to automatically discern spontaneous brow actions from deliberately posed ones, based on the parameters of temporal dynamics such as speed, duration, intensity, and the occurrence order. They achieved the classification rate of 90.7% on 189 samples taken from three different databases.

## 2.7 Expression Intensity

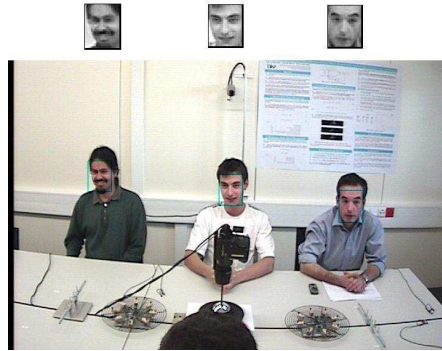
Facial expressions vary in intensity. For example, manual FACS coding uses a 3- or more recently 5-point intensity scale to describe intensity variation of action units (Tian, Kanade, and Cohn, 2005). A distinct emotional state of an expresser can not be correctly perceived unless the expression intensity exceeds a certain level. Methods that work for intense expressions may generalize poorly to subtle expressions with low intensity. It has been experimentally shown that the expression decoding accuracy and the perceived intensity of the underlying affective state vary linearly with the physical intensity of a facial display (Hess, Blairy, and Kleck, 1997). Explicit analysis of expression intensity variation is also essential for distinguishing between spontaneous and posed facial behaviors (Pantic and Bartlett, 2007). So expression intensity estimation is necessary and helpful for accurate assessment of facial expressions. It is desirable to decide the expression intensity from the face data without human labeling (Amin, Afzulpurkar, Dailey, Esichaikul, and Batanov, 2005). Some methods have been presented to automatically quantify expression intensity variation in emotional expressions (Kimura and Yachida, 1997) and action units (Lien, Kanade, Cohn, and Li, 1998). Essa and Pentland (1997) represented intensity variation in smiling using optical flow. Tian et al (2000) compared manual and automatic coding of intensity variation; using Gabor wavelets and a Neural Network, their approach discriminated intensity variation in eye closure as reliably as did human coders. By adopting a 3-grade intensity scale (*Low*, *Moderate*, and *High*), Shan (2007) applied Fuzzy K-Means to derive fuzzy clusters in the embedded space. An example of expression intensity estimation is shown in Fig. 12, where the cluster memberships were mapped to expression degrees.



**Fig. 12** (Best viewed in color) Expression intensity estimation on an example image sequences (Shan, 2007).

## 2.8 *Controlled Versus Uncontrolled Data Acquisition*

Most of the existing facial expression recognition systems attempt to recognize facial expressions from data collected in a highly controlled environment with very high resolution frontal faces (e.g.  $200 \times 200$  pixels). However, in real-life uncontrolled environments such as smart meeting and visual surveillance, only low-resolution compressed video input is available. Moreover, head pose variation is always present due to head movements. Fig. 13 shows a real-world image recorded in a smart meeting scenario. There have been few studies that address facial expression recognition from data collected in real-world unconstrained scenarios. Tian et al (2003) made the first attempt to recognize facial expressions with lower resolution (e.g.  $50 \times 70$  pixels). In their real-time system, to handle the full range of head motion, not face but head was first detected, and the head pose was then estimated. For faces of frontal and near frontal views, the geometric features were extracted for facial expression recognition using a Neural Network. Tian (2004) further explored the effects of different image resolutions for each step of facial expression analysis. Similarly, Shan et al (2005c) studied recognizing facial expressions at low resolution. In the existing research, the recognition performance is usually assessed on faces normalized based on the manually labeled eyes position, i.e., without considering face registration errors. However, in uncontrolled environments, face registration errors are always present and difficult to avoid due to face/eye detection errors. Therefore, it is necessary to investigate facial expression analysis in presence of face registration errors. In a more recent work (Gritti, Shan, Jeanne, and Braspenning, 2008), local features based facial expression recognition approaches were studied with face registration errors, where Gaussian noise is added to the manually annotated eye positions to simulate registration errors.



**Fig. 13** An example of low-resolution facial expressions recorded in real-world environments. (from PETS 2003 data set).

## 2.9 Correlation with Bodily Expression

The human face is usually perceived not as an isolated object but as an integrated part of the whole body. The human body configuration and movement also reveal and enhance emotions (Burgoon, Buller, and Woodall, 1996). For example, an angry face is more menacing when accompanied by a fist. When we see a bodily expression of emotion, we immediately know what specific action is associated with a particular emotion, as is the case for facial expressions (de Gelder, 2006). See Fig. 14 for examples of affective body gestures. Psychological studies (Ambady and Rosenthal, 1992; Meeren, Heijnsbergen, and Gelder, 2005) suggest that the visual channel combining facial and bodily expressions is most informative, and the recognition of facial expression is strongly influenced by the concurrently presented emotional bodily expression. Therefore, analyzing bodily expressions could provide helpful complementary cues for facial expression analysis. However, in affective computing, a great deal of attention has been focused on how emotions are communicated through facial expressions and voice. Little attention has been placed on the perception of emotion from bodily expressions.



**Fig. 14** Examples of affective body gestures, from the FABO database (Gunes and Piccardi, 2006b). From *top* to *bottom*: Fear, Joy, Uncertainty, and Surprise.

Affective bodily expression analysis is an unresolved area in psychology and nonverbal communication. Coulson (2004) presented experiments on attributing six universal emotions to static body postures using computer-generated mannequin figures, and his experiments suggest that recognition from body posture is comparable to recognition from voice, and some postures are recognized as well as facial expressions. Observers tend to be accurate in decoding some negative emotions like anger and sadness from static body postures and the gestures like head inclination

and face touching often accompany affective states like shame and embarrassment (Costa, Dinsbach, Manstead, and Bitti, 2001). Neagle et al (2003) reported a qualitative analysis on affective motion features of virtual ballet dancers, and their results show that human observers are highly accurate in assigning an emotion label to each dance exercise. Recently Ravindra De Silva et al (2006) presented an affective gesture recognition system that recognize child's emotion with intensity through body gesture in the context of a game. Nayak and Turk (2005) developed a system that allows virtual agents to express their mental state through body language (see Fig. 15 for some examples). In computer vision, the existing studies on gesture recognition primarily deal with non-affective gestures such as sign language (Pavlovic, Sharma, and Huang, 1997). There has been few investigations into affective body posture and gesture analysis. This is probably due to the high variability of the possible posture and gesture that can be displayed. Recently some tentative attempts have been made (Balomenos, Raouzaïou, Ioannou, Drosopoulos, Karpouzis, and Kollias, 2005; Gunes and Piccardi, 2006a). Shan et al (2007a) investigated affective body gesture analysis on a large dataset by exploiting spatial-temporal features.



**Fig. 15** Examples of body language displayed by the virtual agent in (Nayak and Turk, 2005). From *left to right*: anger, defensiveness, and headache.

Human emotional and interpersonal states are not conveyed by a single indicator rather by a set of cues. It is the combination of movements of the face, arms, legs and other body parts, as well as voice and touching behaviors, that yields an overall display (Argyle, 1988). Meeren et al (2005) showed that the recognition of facial expressions is strongly influenced by the concurrently presented emotional body language, and that the affective information from the face and the body start to interact rapidly, and the integration is a mandatory automatic process occurring early in the processing stream. To more accurately simulate the human ability to assess affect, an automatic affect recognition system should make use of multimodal data, which potentially can accomplish more effective affect analysis. However, most of the existing works have relied on a single modality. Existing work combining different modalities investigated mainly the combination of facial and vocal signals (Pantic and Rothkrantz, 2003), and there is little effort on human affect analysis by combining face and body gestures (Pantic, Sebe, Cohn, and Huang, 2005a). Kapoor

and Picard (2005) presented a multi-sensor affect recognition system for classifying interest (or disinterest) in children trying to solve a puzzle on the computer. The multimodal information from face expressions and postures are sensed through a camera and a chair respectively, which are combined with the state of the puzzle. The multimodal method is shown to outperform classification using the individual modalities. Balomenos et al (2005) and Gunes and Piccardi (2006a) made tentative attempts to analyze emotions from facial expressions and body gestures. Recently Shan et al (2007a) exploited statistical techniques to fuse the two modalities at the feature level by deriving a semantic joint feature space.

## **2.10 Databases**

To carry out research on facial expression analysis, dealing with different dimensions of the problem space as discussed above, large data sets are needed for evaluation and benchmark. We summarize the main databases of facial expressions in Table 1. The JAFFE database (Lyons, Budynek, and Akamatsu, 1999) is one of widely used databases, which consists of 213 images of 10 Japanese females displaying posed expressions of six basic emotions and neutral faces. The Cohn-Kanade database (Kanade, Cohn, and Tian, 2000) is the most widely used database, which contains image sequences of 100 subjects posing a set of 23 facial displays. The MMI database (Pantic, Valstar, Rademaker, and Maat, 2005b) is another comprehensive database, which contains both deliberately displayed and spontaneous facial expressions of AUs and emotions. Recently a 3D facial expression database has been built (Yin, Wei, Sun, Wang, and Rosato, 2006), which contains 3D range data for emotional expressions at a variety of intensities. Although several databases containing spontaneous facial expressions have been reported recently (Sebe, Lew, Cohen, Sun, Gevers, and Huang, 2004; Bartlett, Littlewort, Frank, Lainscsek, Fasel, and Movellan, 2006; Skelley, Fischer, Sarma, and Heisele, 2006), most of them currently are not available to the public, due to ethical and copyright issues. In addition, manual labeling of spontaneous expressions is very time consuming and error prone due to subjectivity. One of the available databases containing spontaneous facial expression was collected at UT Dallas (O'Toole, Harms, Snow, Hurst, Pappas, Ayyad, and Abdi, 2005), which contains videos of more than 200 subjects. Videos of spontaneous expressions (including happiness, sadness, disgust, puzzlement, laughter, surprise, and boredom) were captured when the subject watching videos that intend to elicit different emotions. Some other databases (Douglas-Cowie, Cowie, and Schroder, 2000) were recorded from talk TV shows which contain speech-related spontaneous facial expressions.

	Subjects	Expressions	Type	Labeled
JAFFE Database (Lyons et al, 1999)	10	7 classes	Posed	Yes
Cohn-Kanade Database (Kanade et al, 2000)	100	wide range	Posed	Yes
MMI Database (Pantic et al, 2005b)	53	wide range	Posed/Spontaneous	Yes
FGNet Database (Wallhoff, 2006)	18	7 classes	Spontaneous	Yes
Authentic Expression Database (Sebe et al, 2004)	28	4 classes	Spontaneous	Yes
UTDallas-HIT (O'Toole et al, 2005)	284	11 classes	Spontaneous	No
RU-FACS (Bartlett et al, 2006)	100	wide range	Spontaneous	Yes
MIT-CBCL (Skelley et al, 2006)	12	9 classes	Spontaneous	Yes
BU-3DFE Database (Yin et al, 2006)	100	7 classes	Posed	Yes
FABO Database (Gunes and Piccardi, 2006b)	23	wide range	Posed	Yes

**Table 1** Summary of the existing databases of facial expressions.

### 3 Facial Expression Recognition with Discriminative Local Features

In this section, we discuss our recent contribution to the field. We investigate discriminative local statistical features for facial expression recognition.

#### 3.1 Local Binary Patterns

Local Binary Patterns (LBP), an efficient non-parametric method summarizing the local structure of an image, has been introduced for facial representation recently (Ahonen, Hadid, and Pietikäinen, 2004; Shan, Gong, and McOwan, 2005d). The most important properties of LBP features are their tolerance against monotonic illumination changes and their computational simplicity. The original LBP operator (Ojala, Pietikäinen, and Harwood, 1996) labels the pixels of an image by thresholding a  $3 \times 3$  neighborhood of each pixel with the center value and considering the results as a binary number. Formally, given a pixel at  $(x_c, y_c)$ , the resulting LBP can be expressed in the decimal form as

$$LBP(x_c, y_c) = \sum_{n=0}^7 s(i_n - i_c) 2^n \quad (1)$$

where  $n$  runs over the 8 neighbors of the central pixel,  $i_c$  and  $i_n$  are the gray-level values of the central pixel and the surrounding pixel, and  $s(x)$  is 1 if  $x \geq 0$  and 0 otherwise.

Ojala et al (2002) later made two extensions of the original operator. Firstly, the operator was extended to use neighborhood of different sizes, to capture dominant features at different scales. Using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood. The notation  $(P, R)$  denotes a neighborhood of  $P$  equally spaced sampling points on a

circle of radius of  $R$ . Secondly, they proposed to use a small subset of the  $2^P$  patterns, produced by the operator  $LBP(P, R)$ , to describe the texture of images. These patterns, called *uniform patterns*, contain at most two bitwise transitions from 0 to 1 or vice versa when considered as a circular binary string. For example, 001110000 and 11100001 are uniform patterns. The uniform patterns represent local primitives such as edges and corners. It was observed that most of the texture information was contained in the uniform patterns. Labeling the patterns which have more than 2 transitions with a single label yields an LBP operator, denoted  $LBP(P, R, u2)$ , which produces much less patterns without losing too much information.

After labeling an image with a LBP operator, a histogram of the labeled image  $f_l(x, y)$  can be defined as

$$H_i = \sum_{x,y} I(f_l(x, y) = i), \quad i = 0, \dots, L - 1 \quad (2)$$

where  $L$  is the number of different labels produced by the LBP operator and  $I(A)$  is 1 if  $A$  is true and 0 otherwise. The LBP histogram contains information about the distribution of local micro-patterns over the whole image, so can be used to statistically describe image texture characteristics. LBP features have been exploited in many applications (see a comprehensive bibliography related to LBP methodology online<sup>1</sup>).

### 3.2 Learning Discriminative LBP-Histogram Bins

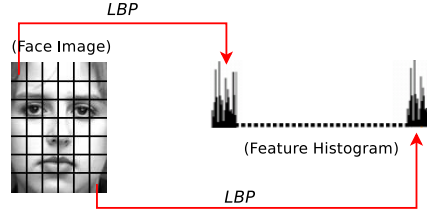
Each face image can be seen as a composition of micro-patterns that can be effectively detected by the LBP operator. Ahonen et al (2004) introduced LBP for face recognition. To consider the shape information of faces, face images are divided into  $M$  small non-overlapping regions  $R_0, R_1, \dots, R_M$  (as shown in Fig. 16). The LBP histograms extracted from each sub-region are then concatenated into a single, spatially enhanced feature histogram defined as:

$$H_{i,j} = \sum_{x,y} I(f_l(x, y) = i) I((x, y) \in R_j) \quad (3)$$

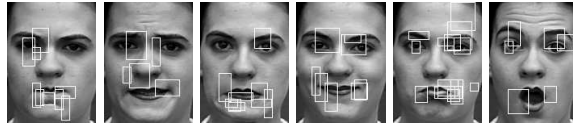
where  $i = 0, \dots, L - 1, j = 0, \dots, M - 1$ . The extracted feature histogram describes the local texture and global shape of face images. This face representation has also been proved effective for facial expression recognition (Shan, Gong, and McOwan, 2005d, 2008). Possible criticisms of this method are that dividing the face into a grid of sub-regions is somewhat arbitrary, as sub-regions are not necessary well aligned with facial features, and that the resulting facial representation suffers from fixed size and position of sub-regions. To address these limitations, the optimal sub-regions (in term of LBP histogram) were selected from a large pool of sub-regions generated by shifting and scaling a sub-window over face images (Zhang, Huang,

<sup>1</sup> [http://www.ee.oulu.fi/mvg/page/lbp\\_bibliography](http://www.ee.oulu.fi/mvg/page/lbp_bibliography)

Li, Wang, and Wu, 2004; Shan, Gong, and McOwan, 2005b). Some examples of selected sub-regions for facial expressions are shown in Fig. 17.



**Fig. 16** A face image is divided into sub-regions from which LBP histograms are extracted and concatenated into a single, spatially enhanced feature histogram.



**Fig. 17** The sub-regions selected by Adaboost for each facial expression. From left to right: Anger, Disgust, Fear, Joy, Sadness, and Surprise.

In the above methods, LBP histograms are extracted from local facial regions as the region-level description, where the  $n$ -bin histogram is utilized as a whole. However, not all bins in the LBP histogram are necessary to contain useful information for facial expression recognition. It is helpful and interesting to have a closer look at the local LBP histogram at the bin level, to identify the discriminative LBP-Histogram (LBPH) bins for better facial representation (Shan and Gritti, 2008). We adopt Adaboost to learn the discriminative LBPH bins. Adaboost (Schapire and Singer, 1999) learns a small number of weak classifiers whose performance can be just better than random guessing, and boosts them iteratively into a strong classifier of higher accuracy. The process of Adaboost maintains a distribution on the training samples. At each iteration, a weak classifier which minimizes the weighted error rate is selected, and the distribution is updated to increase the weights of the misclassified samples and reduce the importance of the others.

As the traditional Adaboost works on two-class problems, the multi-class problem here is accomplished by using the one-against-rest technique, which trains Adaboost between one expression with all others. For each Adaboost learner, the images of one expression were positive samples, while the images of all other expressions were negative samples. The weak classifier  $h_j(x)$  is designed to select the single LBPH bin that best separates the positive and negative examples, which consists of a feature  $f_j$  (corresponding to each LBPH bin), a threshold  $\theta_j$  and a parity  $p_j$  indicating the direction of the inequality sign:

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) \leq p_j \theta_j \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

### 3.3 Experiments

We carried out experiments on the Cohn-Kanade database. Fig. 18 shows some sample images from the database. For our experiments, we selected 320 image sequences of basic emotions, which come from 96 subjects, with 1 to 6 emotions per subject. For each sequence, the neutral face and three frames of expressions at apex were used for prototypic emotional expression recognition, resulting in 1,280 images (108 Anger, 120 Disgust, 99 Fear, 282 Joy, 126 Sadness, 225 Surprise, and 320 Neutral). To test the algorithms' generalization performance, we adopt 10-fold cross-validation scheme in our experiments. Following the existing works (Tian, 2004), we scaled the faces to a fixed distance between the two eyes. Facial images of  $110 \times 150$  pixels were then cropped from original frames based on the two eyes location.



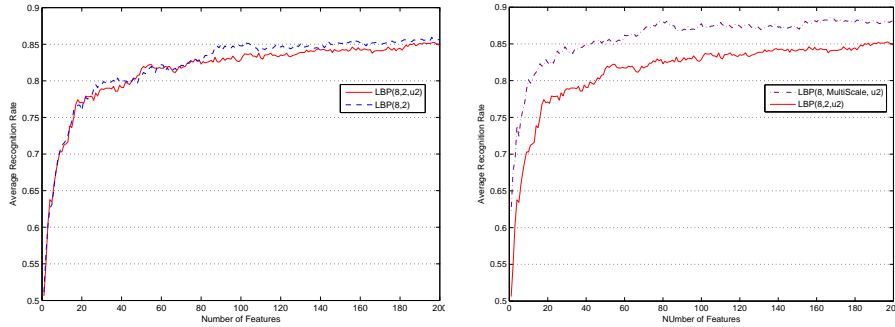
Fig. 18 The sample face expression images from the Cohn-Kanade database.

**Experiment: Limited Sub-regions** — As shown in Fig. 16, we divided face images into 42 sub-regions and applied the 59-label  $LBP(8, 2, u2)$  operator, resulting in a LBP histogram of 2,478 ( $42 \times 59$ ) bins for each face image. These parameter settings were suggested in (Ahonen, Hadid, and Pietikäinen, 2004). We adopted Adaboost to learn discriminative LBPH bins and boost a strong classifier. We plot in the left side of Fig. 19 the recognition performance of the boosted strong classifier as a function of the number of features selected. With the 200 selected LBPH bins, the boosted strong classifier achieves recognition rate of 85.3%.

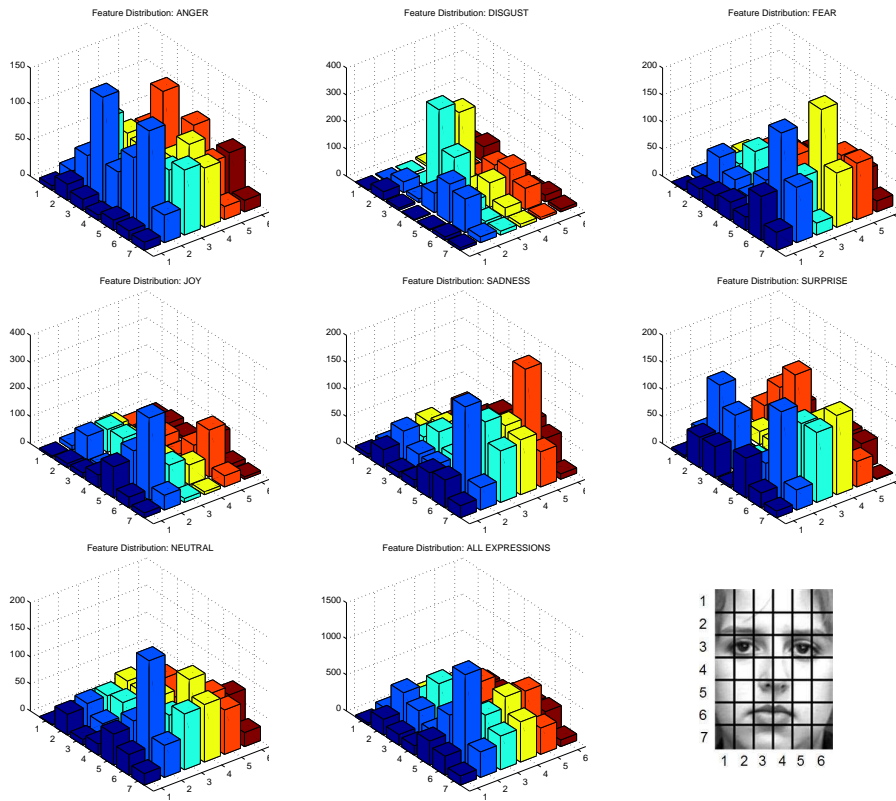
- Uniform patterns are usually adopted to reduce the length of LBP histograms. Here we verify the validity of uniform patterns for facial representation from a point view of machine learning. By using the  $LBP(8, 2)$  operator, each face image

was represented by a LBP histogram of 10,752 ( $42 \times 256$ ) bins. We plot in the left side of Fig. 19 the recognition performance of the boosted strong classifier. We can see that the boosted strong classifier of  $LBP(8,2)$  performs similarly with that of  $LBP(8,2,u2)$ , which illustrates that the non-uniform patterns do not provide more discriminative information for facial expression recognition. We took a closer look at the learned LBPH bins of  $LBP(8,2)$ , and found that 91.1% of them are uniform patterns. Therefore, most of discriminative information for facial expression recognition was contained in the uniform patterns.

- Fig. 20 shows the spatial distribution of the top 200 features selected in the 10-fold cross-validation experiments. As can be observed, different facial expressions have different distribution patterns. For example, for “Disgust”, most discriminative features locate in the eye inner corners, while most discriminative features for “Joy” are distributed in the mouth corners. Overall, discriminative features for facial expression classification mostly distribute in eyes and mouth regions.
- By varying the sampling radius  $R$ , LBP of different resolutions can be obtained. Here we also investigate multiscale LBP for facial expression recognition. We applied the  $LBP(8,R,u2)(R = 1, \dots, 8)$  to extract multiscale LBP features, resulting a LBP histogram of 19,824 ( $42 \times 59 \times 8$ ) bins for each face image. We then run Adaboost to learn discriminative LBPH bins from the multiscale feature pool. We plot in the right side of Fig. 19 the recognition performance of the boosted strong classifier. As can be observed, the boosted strong classifier of multiscale  $LBP(8,R,u2)(R = 1, \dots, 8)$  produces consistently better performance than that of single scale  $LBP(8,2,u2)$ , providing recognition rate of 88.6% with the 200 selected LBPH bins. Thus the multiscale LBP brings more discriminative information, and should be considered for facial expression recognition.



**Fig. 19** Average recognition rate of the boosted strong classifiers, as a function of the number of feature selected. *Left:*  $LBP(8,2,u2)$  vs  $LBP(8,2)$ ; *Right:*  $LBP(8,2,u2)$  vs  $LBP(8, Multiscale, u2)$ .

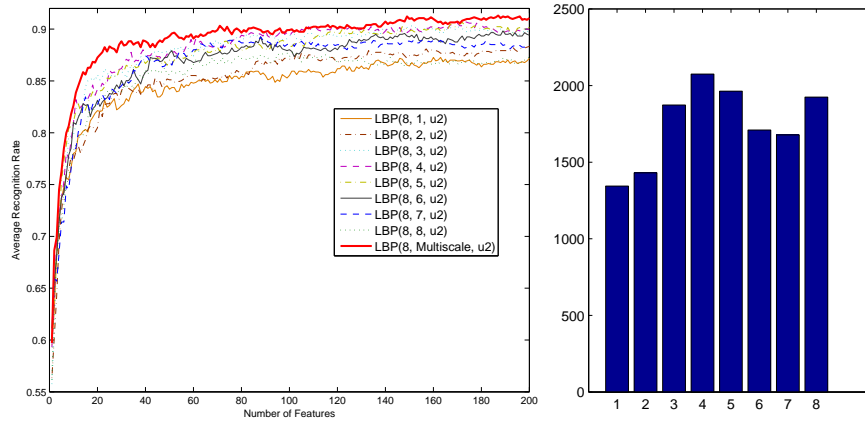


**Fig. 20** Spatial distribution of the selected LBPH bins (an example face image divided in sub-regions is included in the bottom right corner for illustration).

**Experiment: More Sub-regions** — By shifting and scaling a sub-window over face images, we can get many more sub-regions, which potentially contain more complete and discriminative information about face images. We shifted the sub-window with the shifting step of 14 pixels vertically and 12 pixels horizontally. The sub-window was scaled as 14, 21, or 28 pixels (height) and 12, 18, or 24 pixels (width) respectively. In total 725 sub-regions were obtained. By using multiscale  $LBP(8, R, u2)$  ( $R = 1, \dots, 8$ ), a histogram of 342,200 ( $725 \times 59 \times 8$ ) bins was extracted from each face image. To improve computation efficiency, we adopted a coarse to fine feature selection scheme: We first run Adaboost to select LBPH bins from each single scale  $LBP(8, R, u2)$ , then applied Adaboost to the selected LBPH bins at different scales to obtain final feature selection results. We plot in the left side of Fig. 21 the recognition performance of the boosted strong classifiers as a function of the number of features selected. We can see that the final boosted strong classifier of multiscale LBP provides better performance than that of each single scale. Among

strong classifiers of single scales, it seems that scales ( $R = 3, 4, 5, 6$ ) perform better, while the performance of scales ( $R = 1, 8$ ) is poor.

- The scale distribution of final selected multiscale LBPH bins is shown in the right side of Fig. 21. We can observe that most discriminative LBPH bins come from scales ( $R = 3, 4, 5, 8$ ).
- Finally we adopted SVM to recognize facial expressions using the selected LBPH bins. SVM using Gabor features selected by Adaboost (AdaSVM) achieves the best performance (93.3%) reported so far on the Cohn-Kanade database (Littlewort, Bartlett, Fasel, Susskind, and Movellan, 2006a). We compare our LBP-based methods with the Gabor-based methods in Table 2. We used the SVM implementation in the library SPIDER<sup>1</sup>, and the multi-class classification was accomplished by using the one-against-rest technique. It can be observed that the boosted LBPH bins produce comparable results to the boosted Gabor features.



**Fig. 21** *Left*: Average recognition rate of boosted strong classifiers, as a function of the number of feature selected. *Right*: Scale distribution of the selected LBPH bins

Features	Recognition Rates		
	Adaboost	AdaSVM(Linear)	AdaSVM(RBF)
LBP	91.3%	93.0%	93.1%
Gabor(Littlewort et al, 2006a)	90.1%	93.3%	93.3%

**Table 2** Comparison between the boosted LBPH bins and Gabor wavelet features.

<sup>1</sup> Public available at <http://www.kyb.tuebingen.mpg.de/bs/people/spider/index.html>

## 4 Discussion

Automating the recognition of facial expressions is a key step to realize natural and effective human-computer interaction, and to boost applications in many fields such as medicine, security and education. The chapter describes the problem domain and state of the art of automatic facial expression analysis, and summarizes our recent work on learning discriminative local features for recognizing expressions.

In summary, although human cognitive process appears to detect and interpret facial expressions with little or no effort, designing and developing an automated system that accomplishes this task is rather difficult. In last two decades, computer recognition of facial expressions from video or image has been widely studied, and much progress has been made. However, it is still far from the stage of realizing real-life applications in natural environments. Most of the existing systems attempt to recognize a small set of prototypic emotional facial expressions deliberately posed in controlled environments. But recently we have seen some progress made beyond that, for example, spontaneous facial action analysis. For future research, many problems remain open, for which answers must be found. Some major challenges are considered here:

- **How to make use of the temporal information?**  
Psychological studies indicate the facial dynamics are crucial for successful interpretation of facial expressions. This is especially true for spontaneous facial expressions without any deliberate exaggerated posing. Spontaneous facial expressions also differ from posed expressions in terms of which muscles move and how they move dynamically.
- **How to recognize spontaneous facial expressions in real life?**  
Real-life facial expression recognition is much more difficult than recognizing the posed expressions captured in controlled environments. Spontaneous facial expressions induced in natural environments are more subtle and fleeting. Head pose variations and low-resolution input also make it more complex.
- **How to combine facial expressions with other modalities (e.g, body)?**  
Facial expression is one of modes for non-verbal communication. Human emotional and interpersonal states are conveyed not by a single indicator but by a set of cues, which include the movements of the face, arms, legs and other body parts, as well as voice and touching behaviors. Therefore, integrating multiple modalities could potentially accomplish better performance.

## Acknowledgments

We sincerely thank Prof. Jeffery Cohn for granting access to the Cohn-Kanade database.

## References

- Ahonen T, Hadid A, Pietikäinen M (2004) Face recognition with local binary patterns. In: European Conference on Computer Vision (ECCV), pp 469–481
- Ambadar Z, Schooler JW, Cohn JF (2005) Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science* 16(5):403–410
- Amaby N, Rosenthal R (1992) Thin slices of expressive behaviour as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin* 111(2):256–274
- Amin MA, Afzulpurkar NV, Dailey MN, Esichaikul VE, Batanov DN (2005) Fuzzy-c-mean determines the principle component pairs to estimate the degree of emotion from facial expressions. In: International Conference on Natural Computation and International Conference on Fuzzy Systems and Knowledge Discovery, pp 484–493
- Argyle M (1988) *Bodily Communication* (2nd ed.). Methuen & Co. Ltd., New York
- Balomenos T, Raouzaïou A, Ioannou S, Drosopoulos A, Karpouzis K, Kollias S (2005) Emotion analysis in man-machine interaction systems. In: *Machine Learning for Multimodal Interaction*, LNCS 3361, pp 318–328
- Bartlett M, Littlewort G, Frank M, Lainscsek C, Fasel I, Movellan J (2006) Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia* 1(6):22–35
- Bartlett MS, Movellan JR, Sejnowski TJ (2002) Face recognition by independent component analysis. *IEEE Transactions on Neural Networks* 13(6):1450–1464
- Bartlett MS, Littlewort G, Frank M, Lainscsek C, Fasel I, Movellan J (2005) Recognizing facial expression: Machine learning and application to spontaneous behavior. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 568–573
- Bassili JN (1979) Emotion recognition: The role of facial movement and the relative importance of upper and lower area of the face. *Journal of Personality and Social Psychology* 37(11):2049–2058
- Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7):711–720
- Braathen B, Bartlett M, Littlewort G, Smith E, Movellan JR (2002) An approach to automatic recognition of spontaneous facial actions. In: *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp 231–235
- Burgoon JK, Buller DB, Woodall WG (1996) *Nonverbal Communication: The Unspoken Dialogue*. McGraw-Hill, New York
- Chang Y, Hu C, Turk M (2003) Manifold of facial expression. In: *IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG)*, pp 28–35
- Chang Y, Hu C, Turk M (2004) Probabilistic expression analysis on manifolds. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 520–527

- Cohen I, Sebe N, Garg A, Chen L, Huang TS (2003) Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision and Image Understanding* 91:160–187
- Cohn JF, Schmidt KL (2004) The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing* 2:1–12
- Cohn JF, Reed LI, Ambadar Z, Xiao J, Moriyama T (2004) Automatic analysis and recognition of brow actions in spontaneous facial behavior. In: *IEEE International Conference on Systems, Man, and Cybernetics*, p 610
- Costa M, Dinsbach W, Manstead ASR, Bitti PER (2001) Social presence, embarrassment, and nonverbal behavior. *Journal of Nonverbal Behavior* 25(4):225–240
- Coulson M (2004) Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior* 28(2):117–139
- Cowie R, Douglas-Cowie E, Savvidou S, McMahon E, Sawey M, Schroder M (2000) 'feeltrace': An instrument for recording perceived emotion in real time. In: *Proceeding of the ISCA Workshop on Speech and Emotion*, pp 19–24
- Darwin C (1872) *The Expression of the Emotions in Man and Animals*. John Murray, London
- Donato G, Bartlett M, Hager J, Ekman P, Sejnowski T (1999) Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(10):974–989
- Douglas-Cowie E, Cowie R, Schroder M (2000) A new emotion database: Considerations, sources and scope. In: *Proceeding of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, pp 39–44
- Ekman P, Friesen W (1971) Constants across cultures in the face and emotion. *J Personality Social Psychol* 17(2):124–129
- Ekman P, Rosenberg E (1997) *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS)*. New York: Oxford Univ. Press
- Ekman P, Friesen WV, Hager JC (2002) *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*. San Francisco: Consulting Psychologist
- Essa I, Pentland A (1997) Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7):757–763
- Fasel B, Luetttin J (2003) Automatic facial expression analysis: a survey. *Pattern Recognition* 36:259–275
- de Gelder B (2006) Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience* 7:242–249
- Gritti T, Shan C, Jeanne V, Braspenning R (2008) Local features based facial expression recognition with face registration errors. In: *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, Amsterdam, The Netherlands
- Gu H, Ji Q (2005) Information extraction from image sequences of real-world facial expressions. *Machine Vision and Applications* 16:105–115

- Gunes H, Piccardi M (2006a) Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications* (In press)
- Gunes H, Piccardi M (2006b) A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In: *International Conference on Pattern Recognition (ICPR)*, vol 1, pp 1148–1153
- Guo G, Dyer CR (2003) Simultaneous feature selection and classifier training via linear programming: A case study for face expression recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 346–352
- He X, Niyogi P (2003) Locality preserving projections. In: *Advances in Neural Information Processing Systems (NIPS)*
- Hess U, Blairy S, Kleck RE (1997) The intensity of emotional facial expression and decoding accuracy. *Journal of Nonverbal Behavior* 21(4):241–257
- Kaliouby RE, Robinson P (2004) Real-time inference of complex mental states from facial expressions and head gestures. In: *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pp 154–154
- Kanade T, Cohn J, Tian Y (2000) Comprehensive database for facial expression analysis. In: *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp 46–53
- Kapoor A, Picard RW (2005) Multimodal affect recognition in learning environments. In: *ACM International Conference on Multimedia*, pp 677–682
- Kimura S, Yachida M (1997) Facial expression recognition and its degree estimation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 295–300
- Lee CS, Elgammal A (2005) Facial expression analysis using nonlinear decomposable generative models. In: *IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG)*, pp 17–31
- Lien JJ, Kanade T, Cohn JF, Li C (1998) Subtly different facial expression recognition and expression intensity estimation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 853–859
- Littlewort G, Bartlett M, Fasel I, Susskind J, Movellan J (2006a) Dynamics of facial expression extracted automatically from video. *Image and Vision Computing* 24(6):615–625
- Littlewort G, Bartlett MS, Lee K (2006b) Faces of pain: Automated measurement of spontaneous facial expressions of genuine and posed pain. In: *Joint Symposium on Neural Computation*, pp 1–1
- Lyons MJ, Budynek J, Akamatsu S (1999) Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(12):1357–1362
- Meeren H, Heijnsbergen C, Gelder B (2005) Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of USA* 102(45):16,518–16,523
- Mehrabian A (1968) Communication without words. *Psychology Today* 2(4):53–56
- Nayak V, Turk M (2005) Emotional expression in virtual agents through body language. In: *International Symposium on Visual Computing*, pp 313–320

- Neagle RJ, Ng K, Ruddle RA (2003) Studying the fidelity requirements for a virtual ballet dancer. In: *Vision, Video and Graphics*, pp 181–188
- Ojala T, Pietikäinen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distribution. *Pattern Recognition* 29(1):51–59
- Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(7):971–987
- Oliver N, Pentland A, Berard F (2000) Lafter: a real-time face and lips tracker with facial expression recognition. *Pattern Recognition* 33:1369–1382
- O’Toole AJ, Harms J, Snow SL, Hurst DR, Pappas MR, Ayyad JH, Abdi H (2005) A video database of moving faces and people. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(5):812–816
- Pantic M, Bartlett MS (2007) Machine analysis of facial expressions. In: Kurihara K (ed) *Face Recognition, Advanced Robotics Systems*, Vienna, Austria, pp 377–416
- Pantic M, Patras I (2006) Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics* 36(2):433–449
- Pantic M, Rothkrantz L (2000a) Automatic analysis of facial expressions: the state of art. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(12):1424–1445
- Pantic M, Rothkrantz L (2000b) Expert system for automatic analysis of facial expression. *Image and Vision Computing* 18(11):881–905
- Pantic M, Rothkrantz L (2003) Toward an affect-sensitive multimodal human-computer interaction. In: *Proceeding of the IEEE*, vol 91, pp 1370–1390
- Pantic M, Rothkrantz LJM (2004) Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics* 34(3):1449–1461
- Pantic M, Sebe N, Cohn J, Huang T (2005a) Affective multimodal human-computer interaction. In: *ACM International Conference on Multimedia*, pp 669–676
- Pantic M, Valstar M, Rademaker R, Maat L (2005b) Web-based database for facial expression analysis. In: *IEEE International Conference on Multimedia and Expo (ICME)*, pp 317–321
- Pavlovic VI, Sharma R, Huang TS (1997) Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7):677–695
- Ravindra De Silva P, Osano M, Marasinghe A (2006) Towards recognizing emotion with affective dimensions through body gestures. In: *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp 269–274
- Russell JA (1994) Is there universal recognition of emotion from facial expression. *Psychological Bulletin* 115(1):102–141
- Schapire RE, Singer Y (1999) Improved boosting algorithms using confidence-rated predictions. *Machine Learning* 37(3):297–336

- Sebe N, Lew MS, Cohen I, Sun Y, Gevers T, Huang TS (2004) Authentic facial expression analysis. In: IEEE International Conference on Automatic Face & Gesture Recognition (FG), pp 517–522
- Shan C (2007) Inferring facial and body language. PhD thesis, Queen Mary, University of London
- Shan C, Gritti T (2008) Learning discriminative lbp-histogram bins for facial expression recognition. In: British Machine Vision Conference (BMVC), Leeds, UK
- Shan C, Gong S, McOwan PW (2005a) Appearance manifold of facial expression. In: Sebe N, Lew MS, Huang TS (eds) *Computer Vision in Human-Computer Interaction*, Lecture Notes in Computer Science, vol 3723, Springer, pp 221–230
- Shan C, Gong S, McOwan PW (2005b) Conditional mutual information based boosting for facial expression recognition. In: British Machine Vision Conference (BMVC), Oxford, UK, vol 1, pp 399–408
- Shan C, Gong S, McOwan PW (2005c) Recognizing facial expressions at low resolution. In: IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), Como, Italy, pp 330–335
- Shan C, Gong S, McOwan PW (2005d) Robust facial expression recognition using local binary patterns. In: IEEE International Conference on Image Processing (ICIP), Genoa, Italy, vol 2, pp 370–373
- Shan C, Gong S, McOwan PW (2006a) A comprehensive empirical study on linear subspace methods for facial expression analysis. In: IEEE Conference on Computer Vision and Pattern Recognition Workshop, New York, USA, pp 153–158
- Shan C, Gong S, McOwan PW (2006b) Dynamic facial expression recognition using a bayesian temporal manifold model. In: British Machine Vision Conference (BMVC), Edinburgh, UK, vol 1, pp 297–306
- Shan C, Gong S, McOwan PW (2007a) Beyond facial expressions: Learning human emotion from body gestures. In: British Machine Vision Conference (BMVC), Warwick, UK
- Shan C, Gong S, McOwan PW (2007b) Capturing correlations among facial parts for facial expression analysis. In: British Machine Vision Conference (BMVC), Warwick, UK
- Shan C, Gong S, McOwan PW (2008) Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*
- Skelley J, Fischer R, Sarma A, Heisele B (2006) Recognizing expressions in a new database containing played and natural expressions. In: International Conference on Pattern Recognition (ICPR), pp 1220–1225
- Tian Y (2004) Evaluation of face resolution for expression analysis. In: International Workshop on Face Processing in Video, pp 82–82
- Tian Y, Kanade T, Cohn J (2000) Eye-state action unit detection by gabor wavelets. In: International Conference on Multimodal Interfaces (ICMI), pp 143–150
- Tian Y, Kanade T, Cohn J (2001) Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(2):97–115

- Tian Y, Brown L, Hampapur A, Pankanti S, Senior A, Bolle R (2003) Real world real-time automatic recognition of facial expression. In: IEEE workshop on performance evaluation of tracking and surveillance (PETS), Australia
- Tian Y, Kanade T, Cohn J (2005) Handbook of Face Recognition, Springer, chap 11. Facial Expression Analysis
- Tong Y, Liao W, Ji Q (2006) Inferring facial action units with causal relations. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1623–1630
- Turk M, Pentland AP (1991) Face recognition using eigenfaces. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
- Valstar M, Pantic M (2006) Fully automatic facial action unit detection and temporal analysis. In: IEEE Conference on Computer Vision and Pattern Recognition Workshop, p 149
- Valstar M, Pantic M, Ambadar Z, Cohn JF (2006) Spontaneous vs. posed facial behavior: Automatic analysis of brow actions. In: International Conference on Multimodal Interfaces (ICMI), pp 162–170
- Wallhoff F (2006) Facial expressions and emotion database. <http://www.mmk.ei.tum.de/waf/fgnet/feedtum.html>
- Yacoob Y, Davis LS (1996) Recognizing human facial expression from long image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(6):636–642
- Yeasin M, Bullot B, Sharma R (2004) From facial expression to level of interests: A spatio-temporal approach. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 922–927
- Yin L, Wei X, Sun Y, Wang J, Rosato M (2006) A 3d facial expression database for facial behavior research. In: IEEE International Conference on Automatic Face & Gesture Recognition (FG)
- Zeng Z, Fu Y, Roisman GI, Wen Z, Hu Y, Huang TS (2006) Spontaneous emotional facial expression detection. *Journal of Multimedia (JMM)* 1(5):1–8
- Zhang G, Huang X, Li SZ, Wang Y, Wu X (2004) Boosting local binary pattern (lbp)-based face recognition. In: Chinese Conference on Biometric Recognition, pp 179–186
- Zhang Y, Ji Q (2005) Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(5):1–16
- Zhang Z, Lyons MJ, Schuster M, Akamatsu S (1998) Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In: IEEE International Conference on Automatic Face & Gesture Recognition (FG), pp 454–461