F

Facial Expression Recognition

MAJA PANTIC
Imperial College London, London, UK; University of Twente, AE Enschede, The Netherlands

Synonyms

Facial Expression Analysis; Facial Action Coding

Definition

Facial expression recognition is a process performed by humans or computers, which consists of:

1. Locating faces in the scene (e.g., in an image; this step is also referred to as *face detection*),
2. Extracting facial features from the detected face region (e.g., detecting the shape of facial components or describing the texture of the skin in a facial area; this step is referred to as *facial feature extraction*),
3. Analyzing the motion of facial features and/or the changes in the appearance of facial features and classifying this information into some facial-expression-interpretative categories such as facial muscle activations like smile or frown, emotion (affect) categories like happiness or anger, attitude categories like (dis)liking or ambivalence, etc. (this step is also referred to as *facial expression interpretation*).

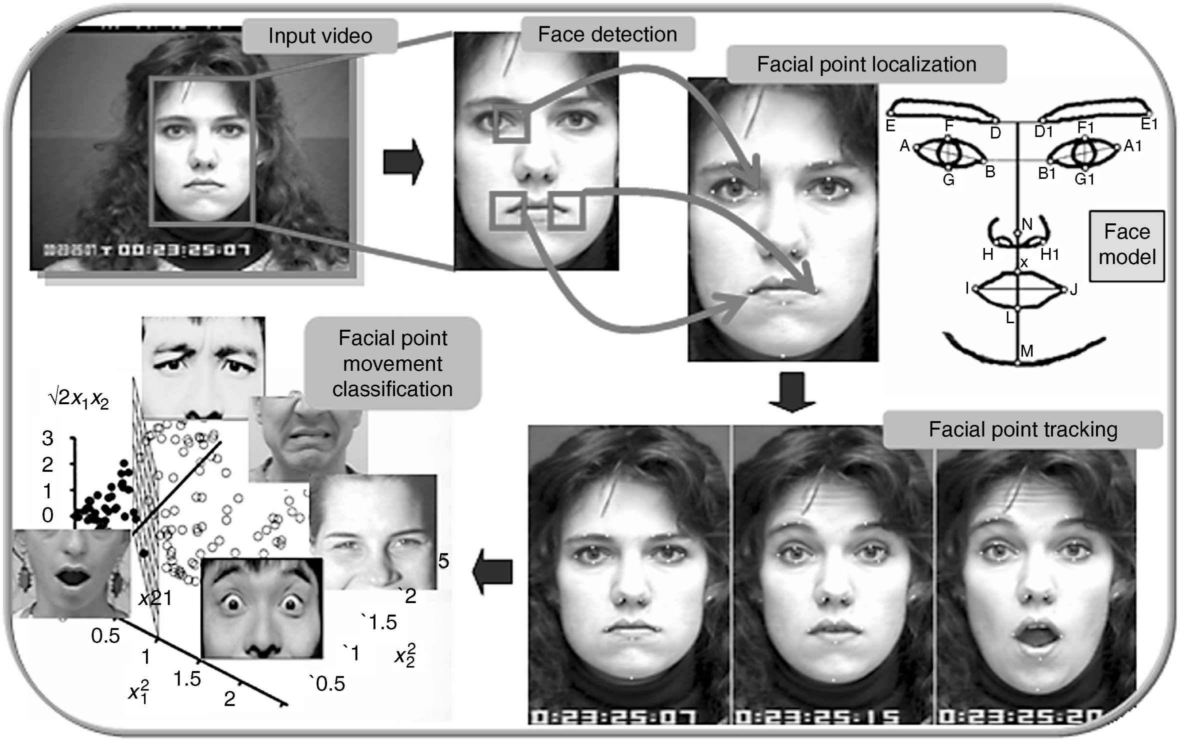
Introduction

A widely accepted prediction is that computing will move to the background, weaving itself into the fabric of our everyday living and projecting the human user into the foreground. To realize this goal, next-generation computing (a.k.a. pervasive computing, ambient intelligence, and ▶ human computing) will need to develop human-centered ▶ user interfaces that respond readily to naturally occurring, multimodal, human communication [1]. These interfaces will need the capacity to perceive and understand intentions and emotions as communicated by social and affective signals. Motivated by this vision of the future, automated analysis of nonverbal behavior, and especially of facial behavior, has attracted increasing attention in computer vision, pattern recognition, and human-computer interaction [2–5]. To wit, facial expression is one of the most cogent, naturally preeminent means for human beings to communicate emotions, to clarify and stress what is said, to signal comprehension, disagreement, and intentions, in brief, to regulate interactions with the environment and other persons in the vicinity [6, 7]. Automatic analysis of facial expressions forms, therefore, the essence of numerous next-generation-computing tools including ▶ affective computing technologies (proactive and affective user interfaces), learner-adaptive tutoring systems, patient-profiled personal wellness technologies, etc.

The Process of Automatic Facial Expression Recognition

The problem of machine recognition of human facial expression includes three subproblem areas (Fig. 1): (1) finding faces in the scene, (2) extracting facial features from the detected face region, (3) analyzing the motion of facial features and/or the changes in the appearance of facial features, and classifying this information into some facial-expression-interpretative categories (e.g., emotions, facial muscle actions, etc.).

The problem of *finding faces* can be viewed as a segmentation problem (in machine vision) or as a detection problem (in pattern recognition). It refers to identification of all regions in the scene that contain a human face. The problem of finding faces (*face localization*, *face detection*) should be solved regardless of



Facial Expression Recognition. Figure 1 Outline of an automated, geometric-features-based system for facial expression recognition (for details of this system, see [4]).

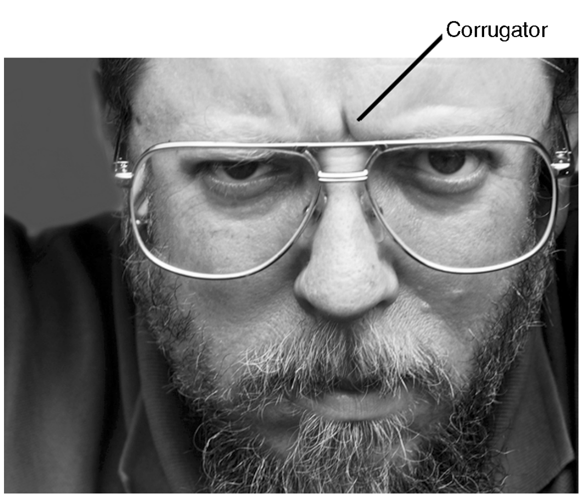
clutter, occlusions, and variations in head pose and lighting conditions. The presence of non-rigid movements due to facial expression and a high degree of variability in facial size, color and texture make this problem even more difficult. Numerous techniques have been developed for face detection in still images [8, 9], (see ► Face Localization). However, most of them can detect only upright faces in frontal or near-frontal view. Arguably the most commonly employed face detector in automatic facial expression analysis is the real-time face detector proposed by Viola and Jones [10].

The problem of feature extraction can be viewed as a dimensionality reduction problem (in machine vision and pattern recognition). It refers to transforming the input data into a reduced representation set of features which encode the relevant information from the input data. The problem of *facial feature extraction* from input images may be divided into at least three dimensions [2, 4]: (1) Are the features holistic (spanning the whole face) or analytic (spanning subparts of the face)?; (2) Is temporal information used?; (3) Are the features view- or volume based (2-D/3-D)?. Given this glossary, most of the proposed approaches to facial expression recognition are directed toward static, analytic, 2-D facial feature extraction [3, 4]. The usually extracted facial features are either *geometric features* such as the shapes of the facial components (eyes, mouth, etc.) and the locations of facial fiducial points (corners of the eyes, mouth, etc.), or *appearance features* representing the texture of the facial skin in specific facial areas including wrinkles, bulges, and furrows. Appearance-based features include learned image filters from Independent Component Analysis (ICA), Principal Component Analysis (PCA), Local Feature Analysis (LFA), Gabor filters, integral image filters (also known as box-filters and Haar-like filters), features based on edge-oriented histograms, etc, (see ► Face Features, ► Skin Texture, and ► Feature Extraction). Several efforts have also been reported which use both geometric and appearance features (e.g., [3]). These approaches to automatic facial expression analysis are referred to as *hybrid* methods. Although it has been reported that methods based on geometric features are often outperformed by those based on appearance features using, e.g., Gabor wavelets or

Comp. by: SunselvakumarGalleys0000870950 Date:27/11/08 Time:14:34:44 Stage:First Proof
File Path://ppdys1108/Womat3/Production/PRODENV/0000000005/0000010929/0000000016/
0000870950.3D Proof by: QC by:

eigenfaces, recent studies show that in some cases geometric features can outperform the appearance-based ones [4, 11]. Yet, it seems that using both geometric and appearance features might be the best choice in the case of certain facial expressions [11].

Contractions of facial muscles, which produce facial expressions, induce movements of the facial skin and changes in the location and/or appearance of facial features (e.g., contraction of the Corrugator muscle induces a frown and causes the eyebrows to move towards each other, usually producing wrinkles between the eyebrows; Fig. 2). Such *changes can be detected* by analyzing optical flow, facial-point- or facial-component-contour-tracking results, or by using an ensemble of classifiers trained to make decisions about the presence of certain changes (e.g., whether the nasolabial furrow is deepened or not) based on the passed appearance features. The optical flow approach to describing face motion has the advantage of not requiring a facial feature extraction stage of processing. Dense flow information is available throughout the entire facial area, regardless of the existence of facial components, even in the areas of smooth texture such as the cheeks and the forehead. Because optical flow is the visible result of movement and is expressed in terms of velocity, it can be used to represent directly the facial expressions. Many researchers adopted this approach [2, 3]. Until recently, standard optical flow techniques were, arguably, most commonly used for tracking facial characteristic points and contours as well [4]. In order to address the limitations inherent in optical flow techniques such as the accumulation of error and the sensitivity to noise, occlusion, clutter, and changes in illumination, recent efforts in automatic facial expression recognition use sequential state estimation techniques (such as Kalman filter and Particle filter) to track facial feature points in image sequences (e.g., [4, 11]).



Facial Expression Recognition. Figure 2 Facial appearance of the Corrugator muscle contraction (coded as in the FACS system, [14]).

Eventually, dense flow information, tracked movements of facial characteristic points, tracked changes in contours of facial components, and/or extracted appearance features are translated into a description of the displayed facial expression. This description (*facial expression interpretation*) is usually given either in terms of shown affective states (emotions) or in terms of activated facial muscles underlying the displayed facial expression. This stems directly from two major approaches to facial expression measurement in psychological research [12]: message and sign judgment. The aim of message judgment is to infer what underlies a displayed facial expression, such as affect or personality, while the aim of sign judgment is to describe the "surface" of the shown behavior, such as facial movement or facial component shape. Thus, a brow frown can be judged as "anger" in a message-judgment and as a facial movement that lowers and pulls the eyebrows closer together in a sign-judgment approach. While message judgment is all about interpretation, sign judgment attempts to be objective, leaving inference about the conveyed message to higher order decision making. Most commonly used facial expression descriptors in message judgment approaches are the six basic emotions (fear, sadness, happiness, anger, disgust, surprise; see Fig. 3) proposed by Ekman and discrete emotion theorists [13], who suggest that these emotions are universally displayed and recognized from facial expressions. Most commonly used facial action descriptors in sign judgment approaches are the Action Units (AUs) defined in the Facial Action Coding System (FACS; [14]). Most facial expressions analyzers developed, so far, target human facial affect analysis and attempt to recognize a small set of prototypic emotional facial expressions like happiness and anger [2, 5]. However, several promising prototype systems were reported that can recognize deliberately produced AUs in face images and even few attempts towards recognition of spontaneously displayed AUs have been recently reported as well [3–5]. While the older methods employ simple approaches including

Facial Expression Recognition. Figure 3 Prototypic facial expressions of six basic emotions (left-to-right from top row): disgust, happiness, sadness, anger, fear, and surprise.

expert rules and machine learning methods such as neural networks to classify the relevant information from the input data into some facial-expression-interpretative categories, the more recent (and often more advanced) methods employ probabilistic, statistical, and ensemble learning techniques, which seem to be particularly suitable for automatic facial expression recognition from face image sequences [3, 5].

Evaluating Performance of an Automated System for Facial Expression Recognition

The two crucial aspects of evaluating performance of a designed automatic facial expression recognizer are the utilized training/test dataset and the adopted evaluation strategy.

Having enough labeled data of the target human facial behavior is a prerequisite in designing robust automatic facial expression recognizers. Explorations of this issue showed that, given accurate 3-D alignment of the face (see ► Face Alignment), at least 50 training examples are needed for moderate performance (in the 80% accuracy range) of a machine-learning approach to recognition of a specific facial expression [4]. Recordings of spontaneous facial behavior are difficult to collect because they are difficult to elicit, short lived, and filled with subtle context-based changes. In addition, manual labeling of spontaneous facial behavior for ground truth is very time consuming, error prone, and expensive. Due to these difficulties, most of the existing studies on automatic facial expression recognition are based on the "artificial" material of deliberately displayed facial behavior, elicited by asking the subjects to perform a series of facial expressions in front of a camera. Most commonly used, publicly available, annotated datasets of posed facial expressions include the Cohn-Kanade facial expression database, JAFFE database, and MMI facial expression database [4, 15]. Yet, increasing evidence suggests that deliberate (posed) behavior differs in appearance and timing from that which occurs in daily life. For example, posed smiles have larger amplitude, more brief duration, and faster onset and offset velocity than many types of naturally occurring smiles. It is not surprising, therefore, that approaches that have been trained on deliberate and often exaggerated behaviors usually fail to generalize to the complexity of expressive behavior found in real-world settings. To address the general lack of a reference set of (audio and/or) visual recordings of human spontaneous behavior, several efforts aimed at development of such datasets have been recently reported. Most commonly used, publicly available, annotated datasets of spontaneous human behavior recordings include SAL dataset, UT Dallas database, and MMI-Part2 database [4, 5].

In pattern recognition and machine learning, a common evaluation strategy is to consider correct classification rate (*classification accuracy*) or its complement error rate. However, this assumes that the natural distribution (prior probabilities) of each class are known and balanced. In an imbalanced setting, where the prior probability of the positive class is significantly less than the negative class (the ratio of these being defined as the *skew*), accuracy is inadequate as a performance measure since it becomes biased towards the majority class. That is, as the skew increases, accuracy tends towards majority class performance, effectively ignoring the recognition capability with respect to the minority class. This is a very common (if not the default) situation in facial expression recognition setting, where the prior probability of each target class (a certain facial expression) is significantly less than the negative class (all other facial expressions). Thus, when evaluating performance of an automatic facial expression recognizer, other performance measures such as *precision* (this indicates the probability of correctly detecting a positive test sample and it is independent of class priors), *recall* (this indicates the fraction of the positives detected that are actually correct and, as it combines results from both positive and negative samples, it is class prior dependent), *F1-measure* (this is calculated as *2*recall*precision/(recall + precision)*), and ROC (this is calculated as $P(x|positive)/P(x|negative)$, where $P(x|C)$ denotes the conditional probability that a data entry has the class label C, and where a ROC curve plots the classification results from the most positive to the most negative classification) are more appropriate. However, as a confusion matrix shows all of the information about a classifier's performance, it should be used whenever possible for presenting the performance of the evaluated facial expression recognizer.

Applications

The potential benefits from efforts to automate the analysis of facial expressions are varied and numerous and span fields as diverse as cognitive sciences, medicine, communication, education, and security [16].

When it comes to computer science and computing technologies, facial expressions provide a way to communicate basic information about needs and demands to the machine. Where the user is looking (i.e., gaze tracking) can be effectively used to free computer users from the classic keyboard and mouse. Also, certain facial signals (e.g., a wink) can be associated with certain commands (e.g., a mouse click) offering an alternative to traditional keyboard and mouse commands. The human capability to "hear" in noisy environments by means of lip reading is the basis for bimodal (audiovisual) speech processing (see ▶ Lip-Movement Recognition), which can lead to the realization of robust speech-driven *user interfaces*. To make a believable *talking head* (avatar) representing a real person, recognizing the person's facial signals and making the avatar respond to those using synthesized speech and facial expressions is important. Combining facial expression spotting with facial expression interpretation in terms of labels like "did not understand", "disagree", "inattentive", and "approves" could be employed as a tool for monitoring human reactions during videoconferences, web-based lectures, and automated tutoring sessions. The focus of the relatively, recently initiated research area of *affective computing* lies on sensing, detecting and interpreting human affective states (such as pleased, irritated, confused, etc.) and devising appropriate means for handling this affective information in order to enhance current ▶ HCI designs. The tacit assumption is that in many situations human-machine interaction could be improved by the introduction of machines that can adapt to their users and how they feel. As facial expressions are our direct, naturally preeminent means of communicating emotions, machine analysis of facial expressions forms an indispensable part of affective HCI designs.

Monitoring and interpreting facial expressions can also provide important information to lawyers, police, security, and intelligence agents regarding *person's identity* (research in psychology suggests that facial expression recognition is much easier in familiar persons because it seems that people display the same, "typical" patterns of facial behaviour in the same situations), *deception* (relevant studies in psychology suggest that visual features of facial expression function as cues to deception), and *attitude* (research in psychology indicates that social signals including accord and mirroring – mimicry of facial expressions, postures, etc., of one's interaction partner – are typical, usually unconscious gestures of wanting to get along with and be liked by the interaction partner). Automated facial reaction monitoring could form a valuable tool in law enforcement, as now only informal interpretations are typically used. Systems that can recognize friendly faces or, more importantly, recognize unfriendly or aggressive faces and inform the appropriate authorities represent another application of facial measurement technology.

Concluding Remark

Faces are tangible projector panels of the mechanisms which govern our emotional and social behaviors. The automation of the entire process of facial expression recognition is, therefore, a highly intriguing problem, the solution to which would be enormously beneficial for fields as diverse as medicine, law, communication, education, and computing. Although the research in the field has seen a lot of progress in the past few years, several issues remain unresolved. Arguably the most important unattended aspect of the problem is how the grammar of facial behavior can be learned (in a human-centered, context-profiled manner) and how this information can be properly represented and used to handle ambiguities in the observation data. This aspect of machine analysis of facial expressions forms the main focus of the current and future research in the field.

Related Entries

▶ Face Alignment
▶ Face Features
▶ Face Localization
▶ Feature Extraction
▶ Skin Texture

References

1. Pantic, M., Pentland, A., Nijholt, A., Huang, T.S.: Human computing and machine understanding of human behavior: A Survey. Lect. Notes Artif. Intell. **4451**, 47–71 (2007)
2. Pantic, M., Rothkrantz, L.J.M.: Toward an affect-sensitive multimodal HCI. Proceedings of the IEEE **91**(9), 1370–1390 (2003)
3. Tian, Y.L., Kanade, T., Cohn, J.F.: Facial expression analysis. In: Li, S.Z., Jain, A.K. (eds.) Handbook of Face Recognition, pp. 247–276. Springer, New York (2005)
4. Pantic, M., Bartlett, M.S.: Machine analysis of facial expressions. In: Delac, K., Grgic, M. (eds.) Face Recognition, pp. 377–416. I-Tech Education and Publishing, Vienna, Austria (2007)
5. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Trans. Pattern Anal. Mach. Intell. **30**, (2008) Au1
6. Ambady, N., Rosenthal, R.: Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. Psychol. Bull. **111**(2), 256–274 (1992)
7. Ekman, P., Rosenberg, E.L. (eds.): What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system. Oxford University Press, Oxford, UK (2005)
8. Yang, M.H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: A survey. IEEE Trans. Pattern Anal. Mach. Intell. **24**(1), 34–58 (2002)
9. Li, S.Z., Jain, A.K. (eds.): Handbook of face recognition. Springer, New York (2005)
10. Viola, P., Jones, M.: Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004)
11. Pantic, M., Patras, I.: Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. IEEE Trans. Syst. Man Cybern. B Cybern. **36**(2), 433–449 (2006)
12. Cohn, J.F., Ekman, P.: Measuring facial actions. In: Harrigan, J.A., Rosenthal, R., Scherer, K. (eds.) The New Handbook of Methods in Nonverbal Behavior Research, pp. 9–64. Oxford University Press, New York (2005)
13. Keltner, D., Ekman, P.: Facial expression of emotion. In: Lewis, M., Haviland-Jones, J.M. (eds.) Handbook of Emotions, pp. 236–249. Guilford Press, New York (2000)
14. Ekman, P., Friesen, W.V., Hager, J.C.: Facial action coding system. A Human Face, Salt Lake City, USA (2002)
15. Pantic, M., Valstar, M.F, Rademaker, R., Maat, L. Web-based database for facial expression analysis. Proc. IEEE Int'l Conf. Multimedia & Expo (ICME) 317–321 (2005)
16. Ekman, P., Huang, T.S., Sejnowski, T.J., Hager, J.C. (eds.): NSF Understanding the Face. A Human Face eStore, Salt Lake City, USA, (see Library) (1992)

Comp. by: SunselvakumarGalleys0000870950 Date:27/11/08 Time:14:34:48 Stage:First Proof
File Path://ppdys1108/Womat3/Production/PRODENV/0000000005/0000010929/0000000016/
0000870950.3D Proof by: QC by:

Author Query Form

Encyclopedia of Biometrics
Alpha: F

Query Refs.	Details Required	Author's response
AU1	Kindly provide the page range for the article.	