Encyclopedia of Multimedia Technology and Networking

Second Edition

Margherita Pagani
Bocconi University, Italy

Volume II
Ex–N

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE
Hershey · New York

Director of Editorial Content:	Kristin Klinger
Senior Managing Editor:	Jennifer Neidig
Managing Editor:	Jamie Snavely
Assistant Managing Editor:	Carole Coulson
Cover Design:	Lisa Tosheff
Printed at:	Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com/reference

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanbookstore.com

Library of Congress Cataloging-in-Publication Data

Encyclopedia of multimedia technology and networking / Margherita Pagani, editor. -- 2nd ed.
p. cm.
Includes bibliographical references and index.
Summary: "This publication offers a research compendium of human knowledge related to the emerging multimedia digital metamarket"--Provided by publisher.
ISBN 978-1-60566-014-1 (hardcover) -- ISBN 978-1-60566-015-8 (ebook)
1. Multimedia communications--Encyclopedias. I. Pagani, Margherita, 1971-
TK5105.15.E46 2009
621.38203--dc22
2008030766

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Face for Interface

Maja Pantic
Imperial College London, UK
University of Twente, The Netherlands

INTRODUCTION: THE HUMAN FACE

The human face is involved in an impressive variety of different activities. It houses the majority of our sensory apparatus: eyes, ears, mouth, and nose, allowing the bearer to see, hear, taste, and smell. Apart from these biological functions, the human face provides a number of signals essential for interpersonal communication in our social life. The face houses the speech production apparatus and is used to identify other members of the species, to regulate the conversation by gazing or nodding, and to interpret what has been said by lip reading. It is our direct and naturally preeminent means of communicating and understanding somebody's affective state and intentions on the basis of the shown facial expression (Lewis & Haviland-Jones, 2000). Personality, attractiveness, age, and gender can also be seen from someone's face. Thus the face is a multisignal sender/receiver capable of tremendous flexibility and specificity. In general, the face conveys information via four kinds of signals listed in Table 1.

Automating the analysis of facial signals, especially rapid facial signals, would be highly beneficial for fields as diverse as security, behavioral science, medicine, communication, and education. In security contexts, facial expressions play a crucial role in establishing or detracting from credibility. In medicine, facial expressions are the direct means to identify when specific mental processes are occurring. In education, pupils' facial expressions inform the teacher of the need to adjust the instructional message.

As far as natural user interfaces between humans and computers (PCs/robots/machines) are concerned, facial expressions provide a way to communicate basic information about needs and demands to the machine. In fact, automatic analysis of rapid facial signals seem to have a natural place in various vision subsystems and vision-based interfaces (face-for-interface tools), including automated tools for gaze and focus of attention tracking, lip reading, bimodal speech processing, face/visual speech synthesis, face-based command issuing, and facial affect processing. 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 that can lead to the realization of robust speech-driven interfaces. To make a believable "talking head" (avatar) representing a real person, tracking the person's facial signals and making the avatar mimic those using synthesized speech and facial expressions is compulsory. The human ability to read emotions from someone's facial expressions is the basis of facial affect processing that can lead to expanding user interfaces with emotional communication and, in turn, to obtaining a more flexible, adaptable, and natural affective interfaces between humans and machines. More specifically, the information about when the existing interaction/processing should be adapted, the importance of such an adaptation, and how the interaction/ reasoning should be adapted, involves information about how the user feels (e.g., confused, irritated, tired, interested). Examples of affect-sensitive user interfaces are still rare, unfortunately, and include the systems of Lisetti and Nasoz (2002), Maat and Pantic (2006), and Kapoor, Burleson, and Picard (2007). It is this wide range of principle driving applications that has lent a special impetus to the research problem of automatic facial expression analysis and produced a surge of interest in this research topic.

BACKGROUND: FACIAL ACTION CODING

Rapid facial signals are movements of the facial muscles that pull the skin, causing a temporary distortion of the shape of the facial features and of the appearance of folds, furrows, and bulges of skin. The common terminology for describing rapid facial signals refers either to culturally dependent linguistic terms

Table 1. Four types of facial signals

- *Static facial signals* represent relatively permanent features of the face, such as the bony structure, the soft tissue, and the overall proportions of the face. These signals are usually exploited for person identification.
- *Slow facial signals* represent changes in the appearance of the face that occur gradually over time, such as development of permanent wrinkles and changes in skin texture. These signals can be used for assessing the age of an individual.
- *Artificial signals* are exogenous features of the face such as glasses and cosmetics. These signals provide additional information that can be used for gender recognition.
- *Rapid facial signals* represent temporal changes in neuromuscular activity that may lead to visually detectable changes in facial appearance, including blushing and tears. These (atomic facial) signals underlie facial expressions.

indicating a specific change in the appearance of a particular facial feature (e.g., smile, smirk, frown) or to the linguistic universals describing the activity of specific facial muscles that caused the observed facial appearance changes.

There are several methods for linguistically universal recognition of facial changes based on the facial muscular activity (Scherer & Ekman, 1982). From those, the facial action coding system (FACS) proposed by Ekman and Friesen (1978) and Ekman, Friesen, and Hager (2002) is the best known and most commonly used system. It is a system designed for human observers to describe changes in the facial expression in terms of visually observable activations of facial muscles. The changes in the facial expression (rapid facial signals) are described with FACS in terms of 44 different action units (AUs), each of which is anatomically related to the contraction of either a specific facial muscle or a set of facial muscles. Examples of different AUs are given in Table 2. Along with the definition of various AUs, FACS also provides the rules for visual detection of AUs and their temporal segments (onset, apex, offset) in a face image. Using these rules, a FACS coder (that is, a human expert having a formal training in using FACS) decomposes a shown facial expression into the AUs that produce the expression.

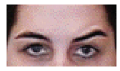
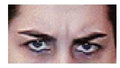
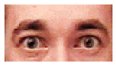
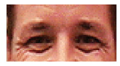
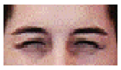
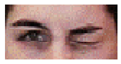
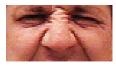
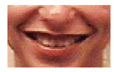
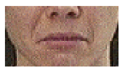
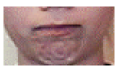
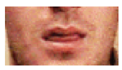
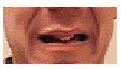
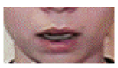
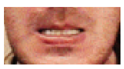
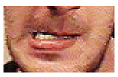
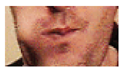
Although FACS provides a good foundation for AU-coding of face images by human observers, achieving AU recognition by an automated system for facial expression analysis is by no means a trivial task. A problematic issue is that AUs can occur in more than 7,000 different complex combinations (Scherer & Ekman, 1982), causing bulges (e.g., by the tongue pushed under one of the lips) and various in- and out-of-image-plane movements of permanent facial features (e.g., jetted jaw) that are difficult to detect in 2D face images.

AUTOMATED FACIAL ACTION CODING

Most approaches to automatic facial expression analysis attempt automatic facial affect analysis by recognizing a small set of prototypic emotional facial expressions, that is, fear, sadness, disgust, anger, surprise, and happiness, (For exhaustive surveys of the past work on this research topic, readers are referred to the work of Pantic & Rothkrantz, 2003, and Zeng, Pantic, Roisman, & Huang, 2007.) This practice may follow from the work of Darwin and more recently Ekman (Lewis & Haviland-Jones, 2000), who suggested that basic emotions have corresponding prototypic expressions. In everyday life, however, such prototypic expressions occur relatively rarely; emotions are displayed more often by subtle changes in one or few discrete facial features, such as raising of the eyebrows in surprise. To detect such subtlety of human emotions and, in general, to make the information conveyed by facial expressions available for usage in the various applications mentioned above, including user interfaces, automatic recognition of rapid facial signals (AUs) is needed.

A number of approaches have been reported up to date for automatic recognition of AUs in images of faces. For exhaustive surveys of the related work, readers are referred to the work of Tian, Kanade, and Cohn (2005), Pantic (2006), and Pantic and Bartlett (2007). Some researchers described patterns of facial motion that correspond to a few specific AUs, but did

Table 2. Examples of facial action units (AUs)

	AU1: Raised inner eyebrow		AU2: Raised outer eyebrow
	AU1 + AU2: Raised eyebrows		AU4: Lowered eyebrow Eyebrows drawn together
	AU5: Raised upper eyelid		AU6: Raised cheek Compressed eyelid
	AU7: Tightened eyelid		AU41: Drooped eyelid
	AU44: Squinted eyes		AU46: Wink
	AU9: Wrinkled nose		AU11: Deepened nasolabial furrow
	AU12: Lip corners pulled up		AU13: Lip corners pulled up sharply
	AU14: Dimpler – mouth corners pulled inwards		AU15: Lip corners depressed
	AU17: Chin raised		AU19: Tongue shown
	AU20: Mouth stretched horizontally		AU24: Lips pressed
	AU26: Jaw dropped		AU29: Jaw pushed forward
	AU30: Jaw sideways		AU36: Bulge produced by the tongue

not report on actual recognition of these AUs. Examples of such works are the studies of Mase (1991) and Essa and Pentland (1997). Historically, the first attempts to explicitly encode AUs in images of faces in an automatic way were reported by Bartlett, Cohn, Kanade, and Pantic (Pantic & Bartlett, 2007). These three research groups are still the forerunners in this research field. The focus of the research efforts in the field was first on automatic recognition of AUs in either static face images or face image sequences picturing facial expressions produced on command. Several promising prototype systems were reported that can recognize deliberately produced AUs in either (near-) frontal view face images (Bartlett, Littlewort, Frank, Lainscsek, Fasel, & Movellan, 1999; Pantic, 2006; Pantic & Rothkrantz, 2004; Tian, Kanade, & Cohn, 2001) or profile view face images (Pantic & Patras, 2006; Pantic & Rothkrantz, 2004). These systems addressed the problem of automatic AU recognition in face images/videos using both computer vision techniques like facial characteristic point tracking or analysis of optical flow, Gabor wavelets, and temporal templates, and machine learning techniques such as neural networks, support vector machines, and Hidden Markov Models (Pantic & Bartlett, 2007; Tian et al., 2005).

One of the main criticisms that these works received from both cognitive and computer scientists is that the methods are not applicable in real-life situations, where subtle changes in facial expression typify the displayed facial behavior rather than the exaggerated changes that typify posed expressions. Hence, the focus of the research in the field started to shift to automatic AU recognition in spontaneous facial expressions (produced in a reflex-like manner). Several works have recently emerged on machine analysis of AUs in spontaneous facial expression data (e.g., Bartlett et al., 2005; Cohn, Reed, Ambadar, Xiao, & Moriyama, 2004; Littlewort, Bartlett, & Lee, 2007; Valstar, Pantic, Ambadar, & Cohn, 2006; Valstar, Gunes, & Pantic, 2007). These methods employ probabilistic, statistical, and ensemble learning techniques, which seem to be particularly suitable for automatic AU recognition from face image sequences (Pantic & Bartlett, 2007; Tian et al., 2005).

CRITICAL ISSUES

Facial expression is an important variable for a large number of basic science studies (in behavioral science, psychology, psychophysiology, psychiatry) and computer science studies (in natural human–machine interaction, ambient intelligence, affective computing). However, the progress in these studies is slowed down by the difficulty of manually coding facial behavior (approximately 100 hours are needed to manually FACS code 1 hour of video recording) and the lack of non-invasive technologies like video monitoring capable of analyzing human spontaneous (as opposed to deliberately displayed) facial behavior. Although few works have been recently reported on machine analysis of facial expression in spontaneous data, the research on the topic is actually just beginning to be explored. Also, the only reported efforts to automatically discern spontaneous from deliberately displayed facial behavior are that of Valstar et al. (2006, 2007) and of Littlewort et al. (2007).

In a frontal-view face image (portrait), facial gestures such as showing the tongue (AU 19) or pushing the jaw forward (AU 29) represent out-of-image-plane nonrigid facial movements which are difficult to detect. Such facial gestures are clearly observable in a profile view of the face. Hence, the usage of face-profile view promises a qualitative enhancement of AU detection performed (by enabling detection of AUs that are difficult to encode in a frontal facial view). Furthermore, automatic analysis of expressions from face profile-view would facilitate deeper research on human emotion. Namely, it seems that negative emotions (where facial displays of AU2, AU4, AU9, etc., are often involved) are more easily perceivable from the left hemiface than from the right hemiface and that, in general, the left hemiface is perceived to display more emotion than the right hemiface (Mendolia & Kleck, 1991). However, only Pantic and Patras (2006) made an effort up to date in automating FACS coding from video of profile faces. Finally, it seems that facial actions involved in spontaneous emotional expressions are more symmetrical, involving both the left and the right side of the face, than deliberate actions displayed on request. Based upon these observations, Mitra and Liu (2004) have shown that facial asymmetry has sufficient discriminating power to significantly improve the performance of an automated genuine emotion classifier. In summary, the usage of both frontal and profile facial views and moving toward 3D analysis of facial expressions promises, therefore, a qualitative increase in facial behavior analysis that can be achieved.

There is now a growing body of psychological research that argues that temporal dynamics of facial behavior (i.e., the timing, the duration, and the intensity of facial activity) is a critical factor for the interpretation of the observed facial behavior (Ekman & Rosenberg, 2005). For example, Schmidt and Cohn (2001) have shown that spontaneous smiles, in contrast to posed smiles, are fast in onset, can have multiple AU12 apexes (i.e., multiple rises of the mouth corners), and are accompanied by other AUs that appear either simultaneously with AU12 or follow AU12 within one second. Similarly, it has been shown that the differences between spontaneous and deliberately displayed brow actions (AU1, AU2, AU4) is in the duration and the speed of onset and offset of the actions and in the order and the timing of actions occurrences (Valstar et al., 2006). Hence, it is obvious that automated tools for automatic analysis of temporal dynamic of facial behavior (i.e., for detection of FACS AUs and their temporal dynamics) would be highly beneficial. However, only three recent studies analyze explicitly the temporal dynamics of facial expressions in an automatic way. These studies explore automatic segmentation of AU activation into temporal segments (neutral, onset, apex, offset) in frontal- (Pantic & Patras, 2005; Valstar & Pantic, 2006) and profile-view (Pantic & Patras, 2006) face videos.

None of the existing systems for facial action coding in images of faces is capable of detecting all 44 AUs defined by the FACS system. Besides, truly robust facial expression analysis is yet to be achieved. In many instances, automated facial expression analyzers operate only under strong assumptions that make the problem more tractable (e.g., images contain faces with no facial hair or glasses, the illumination is constant, the subjects are young, of the same ethnicity, and they remain still while the recordings are made so that no head movements are present). Although methods have been proposed that can handle rigid head motions to a certain extent (e.g., Valstar & Pantic, 2006) and distractions like facial hair (beard, moustache) and glasses to a certain extent (e.g., Essa & Pentland, 1997; Valstar et al., 2004, 2006), truly robust facial expression analysis despite rigid head movements and facial occlusions is still not achieved (e.g., abrupt and fast head motions and faces covered by a large beard cannot be handled correctly by the existing methods). Also, none of the automated facial expression analyzers proposed in the literature up to date "fills in" missing parts of the observed face, that is, none "perceives" a whole face when a part of it is occluded (e.g., by a hand or some other object). In addition, no existing system for automatic facial expression analysis performs explicit coding of intensity of the observed expression (where intensity is the relative degree of change in facial expression as compared to a relaxed, expressionless face).

To develop and evaluate automatic facial expression analyzers capable of dealing with different dimensions of the problem space as defined above, large collections of training and test facial expression data are needed. However, there is no comprehensive reference set of face images that could provide a basis for all different efforts in the research on machine analysis of facial expressions (Pantic & Bartlett, 2007; Zeng et al., 2007). We can distinguish two main problems related to this issue. First, a large majority of the existing datasets of facial behavior recordings are yet to be made publicly available. Second, a large majority of the publicly available databases are not coded for ground truth (i.e., the recordings are not labeled in terms of AUs and affective states depicted in the recordings). The two exceptions for this overall state of the art are the Cohn–Kanade database and the MMI database. The Cohn–Kanade facial expression database (Kanade, Cohn, & Tian, 2000) is the most widely used database in research on automated facial expression analysis (Pantic & Bartlett, 2007; Tian et al., 2005). This database contains image sequences of approximately 100 subjects posing a set of 23 facial displays, and it contains FACS codes in addition to basic emotion labels. The release of this database to the research community enabled a large amount of research on facial expression recognition and feature tracking. Three main limitations of this facial expression dataset are as follows. First, each recording ends at the apex of the shown expression, which limits research of facial expression temporal activation patterns (onset → apex → offset). Second, many recordings contain the date/time stamp recorded over the chin of the subject. This makes changes in the appearance of the chin less visible and motions of the chin difficult to track. Third, the database does not contain recordings of spontaneous (as opposed to posed) facial behavior. To fill this gap, the MMI facial expression database was developed (Pantic, Valstar, Rademaker, & Maat, 2005). It has two parts: a part containing deliberately displayed facial expressions and a part containing spontaneous facial displays. The first part contains over 4,000 videos as well as over 600 static images depicting facial expressions of single AU activation, multiple AU activations, and six basic emotions. It has profile as well as frontal views, and was FACS coded by two certified coders. The second part of the MMI facial expression database contains currently 65 videos of spontaneous facial displays, that were coded in terms of displayed AUs and emotions by two certified coders. Subjects were 18 adults 21 to 45 years old and 11 children 9 to 13 years old: 48% female, 66% Caucasian, 30% Asian, and 4% African. The recordings of 11 children were obtained during the preparation of a Dutch TV program, when children were told jokes by a professional comedian or were told to mimic how they would laugh when something is not funny. The recordings contain mostly facial expressions of different kinds of laughter and were made in a TV studio, using a uniform background and constant lighting conditions. The recordings of 18 adults were made in subjects' usual environments (e.g., home), where they were shown segments from comedies, horror movies, and fear-factor series. The recordings contain mostly facial expressions of different kinds of laughter, surprise, and disgust expressions, which were accompanied by (often large) head motions, and were made under variable lighting conditions. Although the MMI facial expression database is the most comprehensive database for research on automated facial expression analysis, it still lacks metadata for the majority of recordings when it comes to frame-based AU coding. Also, although the MMI database is the only publicly available dataset containing recordings of spontaneous facial behavior at present, it still lacks metadata about the context in which these recordings were made such the utilized stimuli, the environment in which the recordings were made, the presence of other people, and so forth.

CONCLUSION

Faces are tangible projector panels of the mechanisms which govern our emotional and social behaviors. Analysis of facial expressions in terms of rapid facial signals (that is, in terms of the activity of the facial muscles causing the visible changes in facial expression) is, therefore, a highly intriguing problem. While the automation of the entire process of facial action coding from digitized images would be enormously beneficial for fields as diverse as medicine, law, communication, education, and computing, we should recognize the likelihood that such a goal still belongs to the future. The critical issues concern the establishment of basic understanding of how to achieve robust, (near) real-time, automatic spatiotemporal facial-gesture analysis from multiple views of the human face displaying spontaneous (as opposed to posed) facial behavior.

REFERENCES

Bartlett, M.S., & Littlewort, G., Frank, M.G., Lainscsek, C., Fasel, I., & Movellan, J. (2005). Recognizing facial expression: Machine learning and application to spontaneous behavior. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition* (pp. 568–573).

Cohn, J.F., Reed, L.I., Ambadar, Z., Xiao, J., & Moriyama, T. (2004). Automatic analysis and recognition of brow actions in spontaneous facial behavior. *Proceedings of the International Conference on Systems, Man, and Cybernetics, 1*, 610–616.

Ekman, P., & Friesen, W.V. (1978). *Facial action coding system.* Palo Alto, CA: Consulting Psychologist Press.

Ekman, P., Friesen, W.V., & Hager, J.C. (2002). *Facial action coding system.* Salt Lake City, UT: A Human Face.

Ekman, P., & Rosenberg, E.L. (Eds.). (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the FACS*. Oxford, UK: Oxford University Press.

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.

Kanade, T., Cohn, J., & Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Proceedings of the International Conference on Face and Gesture Recognition* (pp. 46–53).

Kapoor, A., Burleson, W., & Picard, R.W. (2007). Automatic prediction of frustration. *International Journal of Human–Computer Studies, 65*(8), 724–736.

Lewis, M., & Haviland-Jones, J.M. (Eds.). (2000). *Handbook of emotions.* New York: Guilford Press.

Lisetti, C.L., & Nasoz, F. (2002). MAUI: A multimodal affective user interface. In *Proceedings of the International Conference on Multimedia* (pp. 161–170).

Littlewort, G., Bartlett, M.S., & Lee, K. (2007). Faces of pain: Automated measurement of spontaneous facial expressions of genuine and posed pain. In *Proceedings of the International Conference on Multimodal Interfaces*.

Maat, L., & Pantic, M. (2006). Gaze-X: Adaptive, affective, multimodal interface for single-user office scenarios. In *Proceedings of the International Conference on Multimodal Interfaces* (pp. 171–178).

Mase, K. (1991). Recognition of facial expression from optical flow. *IEICE Transactions, E74*(10), 3474–3483.

Mendolia, M., & Kleck, R.E. (1991). Watching people talk about their emotions: Inferences in response to full-face vs. profile expressions. *Motivation and Emotion, 15*(4), 229–242.

Mitra, S., & Liu, Y. (2004). Local facial asymmetry for expression classification. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition* (pp. 889–894).

Pantic, M. (2006). Face for ambient interface. *Lecture Notes in Artificial Intelligence, 3864*, 35–66.

Pantic, M., & Bartlett, M.S. (2007). Machine analysis of facial expressions. In K. Delac & M. Grgic (Eds.), *Face recognition* (pp. 377–416). Vienna, Austria: I-Tech Education and Publishing.

Pantic, M., & Patras, I. (2005). Detecting facial actions and their temporal segments in nearly frontal-view face image sequences. In *Proceedings of the International Conference on Systems, Man and Cybernetics* (pp. 3358–3363).

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 (Part B), 36*(2), 433–449.

Pantic, M., & Rothkrantz, L.J.M. (2003). Toward an affect-sensitive multimodal human–computer interaction. *Proceedings of the IEEE, 91*(9), 1370–1390.

Pantic, M., & Rothkrantz, L.J.M. (2004). Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics (Part B), 34*(3), 1449–1461.

Pantic, M., Valstar, M.F., Rademaker, R., & Maat, L. (2005). Web-based database for facial expression analysis. In *Proceedings of the International Conference on Multimedia and Expo* (pp. 317–321). Retrieved May 18, 2008, from www.mmifacedb.com

Scherer, K.R., & Ekman, P. (Eds.). (1982). *Handbook of methods in non-verbal behavior research.* Cambridge: Cambridge University Press.

Schmidt, K.L., & Cohn, J.F. (2001). Dynamics of facial expression: Normative characteristics and individual differences. In *Proceedings of the International Conference on Multimedia and Expo* (pp. 547–550).

Tian, Y., Kanade, T., & Cohn, J.F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(2), 97–115.

Tian, Y., Kanade, T., & Cohn, J.F. (2005). Facial expression analysis. In S.Z. Li & A.K. Jain (Eds.), *Handbook of face recognition* (pp. 247–276). New York: Springer.

Valstar, M.F., Gunes, H., & Pantic, M. (2007). How to distinguish posed from spontaneous smiles using geometric features. In *Proceedings of the International Conference on Multimodal Interfaces* (pp. 38–45).

Valstar, M.F., & Pantic, M. (2004). Motion history for facial action detection in video. In *Proceedings of the International Conference on Systems, Man and cybernetics* (pp. 635-640).

Valstar, M.F., & Pantic, M. (2006). Fully automatic facial action unit detection and temporal analysis. *Proceedings of the International Conference on Computer Vision and Pattern Recognition, 3*, 149.

Valstar, M.F., Pantic, M., Ambadar, Z., & Cohn, J.F. (2006). Spontaneous vs. posed facial behavior: Automatic analysis of brow actions. In *Proceedings of the International Conference on Multimodal Interfaces* (pp. 162–170).

Zeng, Z., Pantic, M., Roisman, G.I., & Huang, T.S. (2007). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. In *Proceedings of the International Conference on Multimodal Interfaces* (pp. 126–133).

KEY TERMS

Ambient Intelligence: The merging of mobile communications and sensing technologies, with the aim of enabling a pervasive and unobtrusive intelligence in the surrounding environment supporting the activities and interactions of the users. Technologies like face-based interfaces and affective computing are inherent ambient-intelligence technologies.

Automatic Facial Expression Analysis: A process of locating the face in an input image, extracting facial features from the detected face region, and classifying these data into some facial-expression-interpretative categories such as facial muscle action categories, emotion (affect) categories, attitude categories, and so forth.

Face-Based Interface: Regulating (at least partially) the command flow that streams between the user and the computer by means of facial signals. This means associating certain commands (e.g., mouse pointing, mose clicking, etc.) with certain facial signals (e.g., gaze direction, winking, etc.). Face-based interface can be effectively used to free computer users from classic keyboard and mouse commands.

Face Synthesis: A process of creating a "talking head" which is able to speak, to display (appropriate) lip movements during speech, and to display expressive facial movements.

Lip Reading: The human ability to "hear" in noisy environments by analyzing visible speech signals, that is, by analyzing the movements of the lips and the surrounding facial region. Integrating both visual speech processing and acoustic speech processing results in a more robust bimodal (audiovisual) speech processing.

Machine Learning: A field of computer science concerned with the question of how to construct computer programs that automatically improve with experience. The key algorithms that form the core of machine learning include neural networks, genetic algorithms, support vector machines, Bayesian networks, and Markov models.

Machine Vision: A field of computer science concerned with the question of how to construct computer programs that automatically analyze images and produce descriptions of what is imaged.