Face for Interface

Maja Pantic
Delft University of Technology, 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; it regulates conversation by gazing or nodding and interprets 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 also can be seen from someone's face. Thus, the face is a multi-signal sender/receiver capable of tremendous flexibility and specificity. In general, the face conveys information via four kinds of signals listed in Table 1.

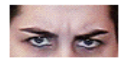
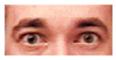
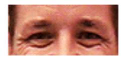
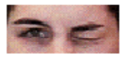
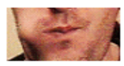
Automating the analysis of facial signals, especially rapid facial signals, would be highly beneficial for fields as diverse as security, behavioral science, medicine, communication, and education. In security contexts, facial expressions play a crucial role in establishing or detracting from credibility. 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 interfaces between humans and computers (i.e., 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 seems to have a natural place in various vision subsystems, 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

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*.

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

using synthesized speech and facial expressions are compulsory. The human ability to read emotions from someone's facial expressions is the basis of facial affect processing that can lead to expanding interfaces with emotional communication and, in turn, obtain a more flexible, adaptable, and natural interaction between humans and machines.

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, indicating a specific change in the appearance of a particular facial feature (e.g., smile, smirk, frown, sneer), or for 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 et al. (1978, 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 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,

Related Content

Security, Trust, and Privacy on Mobile Devices and Multimedia Applications
Edgar R. Weippl (2009). *Handbook of Research on Mobile Multimedia, Second Edition (pp. 115-132).*
www.irma-international.org/chapter/security-trust-privacy-mobile-devices/20999

A Comparative Study of Graph Kernels and Clustering Algorithms
Riju Bhattacharya, Naresh Kumar Nagwaniand Sarsij Tripathi (2021). *International Journal of Multimedia Data Engineering and Management (pp. 33-48).*
www.irma-international.org/article/a-comparative-study-of-graph-kernels-and-clustering-algorithms/271432

Adaptive Multi-Agent Control Strategy in Heterogeneous Countermeasure Environments
Wei Wang, Hui Liuand Wangqun Lin (2021). *International Journal of Multimedia Data Engineering and Management (pp. 31-56).*
www.irma-international.org/article/adaptive-multi-agent-control-strategy-in-heterogeneous-countermeasure-environments/276399

Requirements to a Search Engine for Semantic Multimedia Content
Lydia Weiland, Felix Hanserand Ansgar Scherp (2014). *International Journal of Multimedia Data Engineering and Management (pp. 53-65).*
www.irma-international.org/article/requirements-to-a-search-engine-for-semantic-multimedia-content/120126

Deep-Learning-Based Classification and Diagnosis of Alzheimer's Disease
Rekh Ram Janghel (2018). *Feature Dimension Reduction for Content-Based Image Identification (pp. 193-217).*
www.irma-international.org/chapter/deep-learning-based-classification-and-diagnosis-of-alzheimers-disease/207235