Facial Expression Recognition: A Brief Tutorial Overview

Claude C. Chibelushi, Fabrice Bourel

***Abstract*—Facial expressions convey non-verbal cues, which play an important role in interpersonal relations. Automatic recognition of facial expressions can be an important component of natural human-machine interfaces; it may also be used in behavioural science and in clinical practice. Although humans recognise facial expressions virtually without effort or delay, reliable expression recognition by machine is still a challenge. This paper presents a high-level overview of automatic expression recognition; it highlights the main system components and some research challenges.**

I. Introduction

A facial expression is a visible manifestation of the affective state, cognitive activity, intention, personality, and psychopathology of a person [6]; it plays a communicative role in interpersonal relations. Facial expressions, and other gestures, convey non-verbal communication cues in face-to-face interactions. These cues may also complement speech by helping the listener to elicit the intended meaning of spoken words. As cited in [14] (p. 1424), Mehrabian reported that facial expressions have a considerable effect on a listening interlocutor; the facial expression of a speaker accounts for about 55 percent of the effect, 38 percent of the latter is conveyed by voice intonation and 7 percent by the spoken words.

As a consequence of the information that they carry, facial expressions can play an important role wherever humans interact with machines. Automatic recognition of facial expressions may act as a component of natural human-machine interfaces [20] (some variants of which are called perceptual interfaces [16] or conversational [21] interfaces). Such interfaces would enable the automated provision of services that require a good appreciation of the emotional state of the service user, as would be the case in transactions that involve negotiation, for example. Some robots can also benefit from the ability to recognise expressions [3]. Automated analysis of facial expressions for behavioural science or medicine is another possible application domain [6] [9].

From the viewpoint of automatic recognition, a facial expression can be considered to consist of deformations of facial components and their spatial relations, or changes in the pigmentation of the face. Research into automatic recognition of facial expressions addresses the problems surrounding the representation and categorisation of static or dynamic characteristics of these deformations or face pigmentation. Further details on the problem space for facial expression analysis are given in [11].

This paper is a high-level tutorial overview of automatic facial expression recognition. Due to length restrictions, only a small sample of recognition techniques is explicitly referred to. Further details can be found in the cited references. The next section gives an overview of facial expression recognition systems. Thereafter, some outstanding research problems are pointed out, and a summary of the overview is given.

II. Main Architectural Components

A. Generic Architecture

Despite the task duality that exists between facial expression recognition and face recognition, it can be observed in the literature [4] [6] [14] [18] that similar architectures and processing techniques are often used for both recognition tasks. The duality arises from the following considerations. In addition to conveying expressions, faces also carry other information such as the identity of a person. By definition, the expression of a face is the focal element in facial expression recognition. Hence, personal identity information conveyed by a face is an unwanted source of variability in expression recognition. Conversely, variability arising from facial expression is unwanted in face recognition, where the uniqueness of a face is the central recognition criterion.

Automatic systems for facial expression recognition usually take the form of a sequential configuration of processing blocks, which adheres to a classical pattern recognition model (see Figure 1) [10] [14] [18]. The main blocks are: image acquisition, pre-processing, feature extraction, classification, and post-processing. Table I presents a small sample of approaches to facial expression recognition.

B. General Description

With regard to the interconnection between the blocks shown in Figure 1, feedback paths between blocks are absent from most expression recognition systems, although feedback could be beneficial for improving recognition accuracy.

C.C. Chibelushi is with the School of Computing, Staffordshire University, Beaconside, Stafford ST18 0DG, UK. E-mail: C.C.Chibelushi@staffs.ac.uk

F. Bourel is with ORSYP, 101 quartier Boieldieu, La Défense 8, F-92042 Paris La Défense Cedex, France. E-mail: Fabrice.Bourel@orsyp.com

© 2002

IPR2023-01014
Nielsen Company EX2009 Page 1

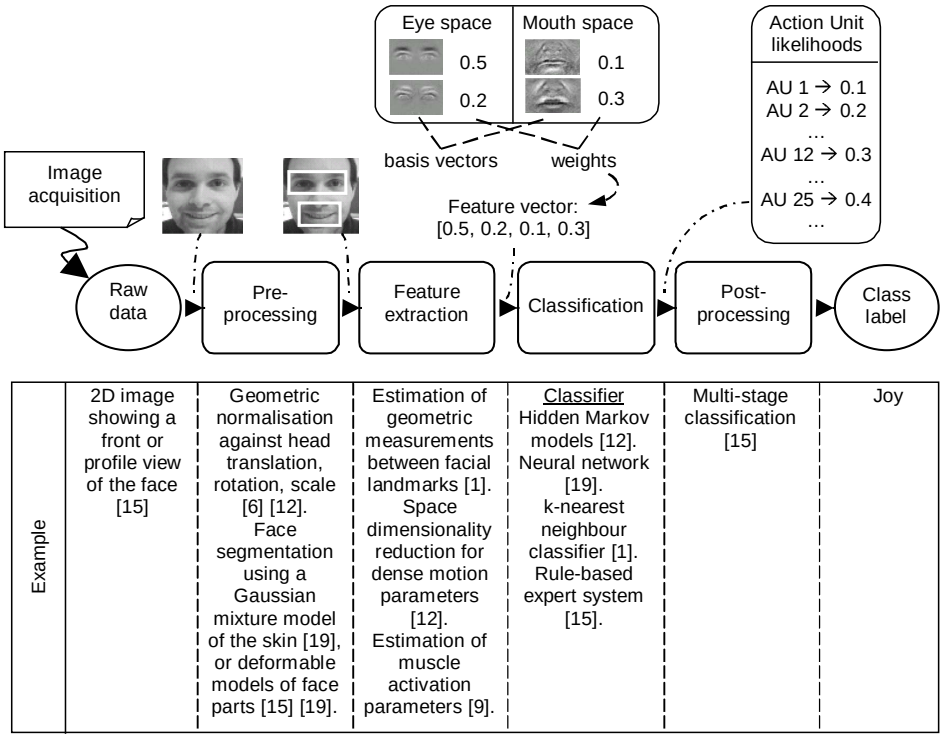

Figure 1: Some techniques for facial expression recognition, shown in the context of the classical pattern recognition model.

Based on the spatial extent of the face parts to which feature extraction and subsequent classification are applied, facial expression recognition can be classified as piecemeal or holistic recognition. Piecemeal recognition typically involves an ensemble of feature extractors or classifiers, together with a combination unit. In holistic recognition, the whole face provides a single input to the recognition system.

The main approaches embedded in the components of an automatic expression recognition system are reviewed below. Some unsolved problems are discussed in the next section.

1) Image Acquisition: Images used for facial expression recognition are static images or image sequences. An image sequence contains potentially more information than a still image, because the former also depicts the temporal characteristics of an expression. With respect to the spatial, chromatic, and temporal dimensionality of input images, 2-D monochrome (grey-scale) facial image sequences are the most popular type of pictures used for automatic expression recognition. However, colour images could become prevalent in future, owing to the increasing availability of low-cost colour image acquisition equipment, and the ability of colour images to convey emotional cues such as blushing.

2) Pre-processing: Image pre-processing often takes the form of signal conditioning (such as noise removal, and normalisation against the variation of pixel position or brightness), together with segmentation, location, or tracking of the face or its parts. Expression representation can be sensitive to translation, scaling, and rotation of the head in an image. To combat the effect of these unwanted transformations, the facial image may be geometrically standardised prior to classification. This normalisation is usually based on references provided by the eyes or nostrils.

Segmentation is concerned with the demarcation of image portions conveying relevant facial information. Face segmentation is often anchored on the shape, motion, colour, texture, and spatial configuration of the face or its components [13]. The face location process yields the position and spatial extent of faces in an image; it is typically based on segmentation results. A variety of face detection techniques have been developed [13]. However, robust detection of faces or their constituents is difficult to attain in many real-world settings. Tracking is often implemented as location, of the face or its parts, within an image sequence, whereby previously determined location is typically used for estimating location in subsequent image frames.

3) Feature Extraction: Feature extraction converts pixel data into a higher-level representation — of shape, motion, colour, texture, and spatial configuration of the face or its components. The extracted representation is used for subsequent expression categorisation. Feature extraction generally reduces the dimensionality of the input space. The reduction procedure should (ideally) retain essential information possessing high discrimination power and high stability. Such dimensionality reduction may mitigate the 'curse of dimensionality' [10]. Geometric, kinetic, and statistical- or spectral-transform-based features are often used as alternative representation of the facial expression prior to classification [14].

TABLE I: ILLUSTRATIVE SAMPLE OF FACIAL EXPRESSION RECOGNITION APPROACHES

Reference	Pre-processing	Feature extraction	Classification	Post-processing	Recognition performance (%)
Bourel et al. [1]	Tracking of facial landmarks using a point tracker specialised to work on facial features.	Scalar quantisation of facial dynamics (which are represented as temporal evolutions of scale-normalised geometric measurements between facial landmarks).	*Classifier*: rank-weighted k-nearest neighbour classifier. *Classes*: 6 basic expression prototypes (anger, disgust, fear, joy, sadness and surprise).	—	*Data set*: 300 image sequences, front face view, many subjects (database: CMU-Pittsburgh AU-Coded Face Expression Image Database [11]). *Recognition accuracy*: relatively little degradation in recognition under partial face occlusion or tracker noise.
Pantic and Rothkrantz [15]	Location / tracking of the face and its parts (front and profile contours of the head, eyebrow region, eye region, mouth region) using multiple detectors (e.g. snakes, neural networks, …).	Extraction of static geometric measurements from landmarks on the contours of eyebrows, eyes, nostrils, mouth, and chin.	*Classifier*: rule-based expert system for each stage of a two-stage classification hierarchy: encoding of face geometry into AUs, followed by classification into basic expression prototypes. *Classes*: 28 AUs and 6 basic expression prototypes (happiness, sadness, surprise, anger, fear, and disgust).	Two-stage classification architecture	*Data set*: 265 still images (two views each, front and profile), more than 8 subjects. *Recognition accuracy*: 91% recognition of basic expression prototypes.
Essa and Pentland [9]	Location of the face by fitting a 3D mesh model of face geometry to a 2D face image. The model-fitting process is based on View-based and Modular Eigenspace methods [17], coupled to image warping. Face tracking by the mesh model is tied to optical flow computation.	Alternative features: (i) peak activation of each muscle. Muscle activation is extracted using a physics-based model of facial deformation. The deformation is estimated from the optical flow. (ii) 2D motion-energy computed from motion estimates, which have been corrected by the physics-based model.	*Classifier*: alternative similarity measures — (i) maximum correlation with muscle activation template, (ii) minimum distance to motion energy template. *Classes*: 5 facial expressions (smile, surprise, anger, disgust, raised brows).	—	*Data set*: 52 image sequences, front face view, 8 subjects. *Recognition accuracy*: 98% recognition of facial expressions.
Tian et al. [19]	Segmentation and tracking of permanent face parts and transient features based on: Gaussian mixture model (for lip colour), deformable template (for lip shape, and eyes), Lucas-Kanade tracking algorithm (cited in [19], p. 102) (for brows and cheeks), Canny edge detector (for wrinkles and furrows).	Continuous and discrete (state-based) representation of magnitude and direction for motion of face parts (lips, eyes, brows, cheeks), as well as state-based (present / absent) representation of transient features (furrows and wrinkles). The continuous representation is normalised against image scale and in-plane head motion.	*Classifiers*: 1 multi-layer perceptron for upper-face AUs, 1 multi-layer perceptron for lower-face Aus. *Classes*: 6 upper-face AUs, 10 lower-face AUs, and neutral expression. (Classified into a single AU class or a combination of AU classes).	—	(i) *Database*: CMU-Pittsburgh AU-Coded Face Expression Image Database [11]. *Recognition accuracy*: 96.4% recognition of upper-face AUs, and neutral expression. Data set: 50 image sequences, front face view, 14 subjects. 96.7% recognition of lower-face AUs, and neutral expression. Data set: 63 image sequences, front face view, 32 subjects. (ii) *Database*: Ekman-Hager Facial Action Exemplars Database (cited in [19], p. 105). *Recognition accuracy*: 93.3% recognition of upper-face AUs, lower-face AUs, and neutral expression. Data set: 122 image sequences, front face view, 21 subjects.
Donato et al. [6]	Geometric normalisation against rigid translation, rotation, and scaling of the head. Segmentation: cropping to region of interest (containing the upper or lower face). Brightness normalisation: partial histogram equalisation. Motion estimation: dense optical flow, or difference image between each frame and the first (neutral) frame.	Alternative representations generated by: (i) holistic spatial analysis. Technique applied to original image sequence: optical flow. Techniques applied to sequence of difference images: principal component analysis (PCA), local feature analysis (LFA), independent component analysis (ICA), or Fisher's linear discriminant (FLD). (ii) local spatial analysis techniques applied to sequence of difference images: local PCA, Gabor wavelet decomposition, or local PCA jets.	*Classifier*: nearest neighbour classifier or template matching. *Classes*: 6 upper-face AUs, 6 lower-face AUs.	—	*Data set*: 111 image sequences, front face view, 20 subjects. *Recognition accuracy*: best performance resulted from using a Gabor wavelet representation (95.5% AU recognition) or independent component representation (95.5% AU recognition). Worst performance resulted from using a smoothed optical flow representation (53.1% AU recognition).
Lien et al. [12]	Alternative segmentation or motion estimation processes: facial feature point tracking, dense optical flow, or edge detection. Geometric normalisation against rigid translation, rotation, and scaling of the head. Segmentation: cropping to region of interest (upper or lower face).	Alternative representations: (i) displacement (of 6 points the on upper boundary of brows) relative to the first frame (ii) PCA of dense optical flow (iii) block-based density and variance of pixels, which show significant change in edge magnitude relative to the first frame.	*Classifier*: vector quantiser followed by hidden Markov models. *Classes*: 3 upper-face AUs associated with brow movement, and neutral expression. (Classified into a single AU class or a combination of AU classes).	—	*Data set*: 260 image sequences, front face view, 60 subjects. *Recognition accuracy*: 85% AU recognition (feature-point displacement, or edge density and variance). 93% AU recognition (PCA features of dense optical flow).

Nielsen Company EX2009 Page 3 IPR2023-01014

4) Classification: Expression categorisation is performed by a classifier, which often consists of models of pattern distribution, coupled to a decision procedure. A wide range of classifiers, covering parametric as well as non-parametric techniques, has been applied to the automatic expression recognition problem [14]. The two main types of classes used in facial expression recognition are action units (AUs) [6], and the prototypic facial expressions defined by Ekman [8].

The 6 prototypic expressions relate to the emotional states of happiness, sadness, surprise, anger, fear, and disgust [8]. However, it has been noted that the variation in complexity and meaning of expressions covers far more than these six expression categories [12]. Moreover, although many experimental expression recognition systems use prototypic expressions as output categories, such expressions occur infrequently, and fine changes in one or a few discrete face parts communicate emotions and intentions [6] [19]. An AU is one of 46 atomic elements of visible facial movement or its associated deformation; an expression typically results from the agglomeration of several AUs [6] [8]. AUs are described in the Facial Action Coding System (FACS) [7].

Sometimes, AU and prototypic expression classes are both used in a hierarchical recognition system — for example, categorisation into AUs can be used as a low-level of expression classification, followed by a high-level classification of AU combinations into basic expression prototypes [15].

5) Post-processing: Post-processing aims to improve recognition accuracy, by exploiting domain knowledge to correct classification errors, or by coupling together several levels of a classification hierarchy, for example.

III. Some Challenges

Although humans recognise facial expressions virtually without effort or delay, reliable expression recognition by machine is still a challenge. The problems that have haunted the pattern recognition community at large (see [10]) still require attention. A key challenge is achieving optimal pre-processing, feature extraction or selection, and classification, particularly under conditions of input data variability. To attain successful recognition performance, most current expression recognition approaches require some control over the imaging conditions. The controlled imaging conditions typically cover the following aspects.

(i) View or pose of the head. Although constraints are often imposed on the position and orientation of the head relative to the camera, and the setting of camera zoom, it should be noted that some processing techniques have been developed, which have good insensitivity to translation, scaling, and in-plane rotation of the head. The effect of out-of-plane rotation is more difficult to mitigate, as it can result in wide variability of image views. Further research is needed into transformation-invariant expression recognition.

(ii) Environment clutter and illumination. Complex image background pattern, occlusion, and uncontrolled lighting have a potentially negative effect on recognition. These factors would typically make image segmentation more difficult to perform reliably. Hence, they may potentially cause the contamination of feature extraction by information not related to facial expression. Consequently, many researchers use uncluttered backgrounds and controlled illumination, although such conditions do not match the operational environment of some potential applications of expression recognition.

(iii) Miscellaneous sources of facial variability. Facial characteristics display a high degree of variability due to a number of factors, such as: differences across people (arising from age, illness, gender, or race, for example), growth or shaving of beards or facial hair, make-up, blending of several expressions, and superposition of speech-related (articulatory) facial deformation onto affective deformation.

The controlling of imaging conditions is detrimental to the widespread deployment of expression recognition systems, because many real-world applications require operational flexibility. However, few studies have systematically investigated the robustness of automatic expression recognition under adverse conditions [2]. Further research into techniques that are robust against variability of the primary input is needed. In particular, research into automatic expression recognition systems capable of adapting their knowledge periodically or continuously has not received much attention. The authors are of the opinion that robustness of expression recognition, against the variability of facial characteristics, would be difficult to achieve without incorporating adaptation in the recognition framework.

Emotions also have acoustic characteristics. Although the combination of acoustic and visual characteristics promises improved recognition accuracy, the development of effective combination techniques is a challenge, which has not been addressed by many. Published research on audio-visual speech or speaker recognition has reported some potentially useful approaches [5].

IV. Summary

This paper has briefly overviewed automatic expression recognition. Similar architectures and processing techniques are often used for facial expression recognition and face recognition, despite the duality that exists between these recognition tasks. 2-D monochrome facial image sequences are the most popular type of pictures used for automatic expression recognition. Although a variety of face detection techniques have been developed, robust detection and location of faces or their constituents is difficult to attain in many cases. Features for automatic expression recognition aim to capture static or dynamic facial information specific to individual expressions. Geometric, kinetic, and statistical- or spectral-transform-based features are often used as alternative representation of the facial expression prior to classification.

IPR2023-01014
Nielsen Company EX2009 Page 4

A wide range of classifiers, covering parametric as well as non-parametric techniques, has been applied to automatic expression recognition.

Generally speaking, automatic expression recognition is a difficult task, which is afflicted by the usual difficulties faced in pattern recognition and computer vision research circles, coupled with face specific problems. As such, research into automatic expression recognition has been characterised by partial successes, achieved at the expense of constraining the imaging conditions, in many cases. Unresolved research issues are encapsulated in the challenge of achieving optimal pre-processing, feature extraction or selection, and classification, under conditions of data variability. Sensitivity of automatic expression recognition to data variability is one of the key factors that have curtailed the spread of expression recognisers in the real world. However, few studies have systematically investigated robustness of automatic expression recognition under adverse conditions.

REFERENCES

[1] F. Bourel, C.C. Chibelushi, A.A. Low, "Robust Facial Expression Recognition Using a State-Based Model of Spatially-Localised Facial Dynamics", *Proc. Fifth IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 106-111, 2002

[2] F. Bourel, *Models of Spatially-Localised Facial Dynamics for Robust Expression Recognition*, Ph.D. Thesis, Staffordshire University, 2002

[3] V. Bruce, "What the Human Face Tells the Human Mind: Some Challenges for the Robot-Human Interface", *Proc. IEEE Int. Workshop Robot and Human Communication*, pp. 44-51, 1992

[4] R. Chellappa, C.L. Wilson, S. Sirohey, "Human and Machine Recognition of Faces: a Survey", *Proc. IEEE*, Vol. 83, No. 5, pp. 705-741, 1995

[5] C.C. Chibelushi, F. Deravi, J.S.D. Mason, "A Review of Speech-Based Bimodal Recognition", *IEEE Trans. Multimedia*, Vol. 4, No. 1, pages 23-37, 2002

[6] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, T.J. Sejnowski, "Classifying Facial Actions", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 21, No. 10, pp. 974-989, 1999

[7] P. Ekman, W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press, 1978

[8] P. Ekman, *Emotion in the Human Face*, Cambridge University Press, 1982

[9] I.A. Essa, A.P. Pentland, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 757-763, 1997

[10] A.K. Jain, R.P.W. Duin, J. Mao, "Statistical Pattern Recognition: A Review", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 1, pp. 4-37, 2000

[11] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis", *Proc. 4th IEEE Int.Conf. on Automatic Face and Gesture Recognition* , pp. 46–53, 2000

[12] J.J. Lien, T. Kanade, J.F. Cohn, C-C. Li, "Automated Facial Expression Recognition Based on FACS Action Units", *Proc. Third IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 390-395, 1998

[13] Yang Ming-Hsuan, D.J. Kriegman, N. Ahuja, "Detecting Faces in Images: a Survey", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, pp. 34-58, 2002

[14] M. Pantic, L.J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: the State of the Art", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp. 1424-1445, 2000

[15] M. Pantic, L.J.M. Rothkrantz, "An Expert System for Multiple Emotional Classification of Facial Expressions", *Proc. 11th IEEE Int. Conf. on Tools with Artificial Intelligence*, pp. 113-120, 1999

[16] A. Pentland, "Looking at People: Sensing for Ubiquitous and Wearable Computing", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 1, pp. 107-119, 2000

[17] A. Pentland, B. Moghaddam, T. Starner, "View-based and Modular Eigenspaces for Face Recognition", *Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 84-91, 1994

[18] A. Samal, P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey", *Pattern Recognition*, Vol. 25, No. 1, pp. 65-77, 1992

[19] Y-L. Tian, T. Kanade, J.F. Cohn, "Recognizing Action Units for Facial Expression Analysis", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp. 97-115, 2001

[20] A. van Dam, "Beyond WIMP", *IEEE Computer Graphics and Applications*, Vol. 20, No. 1, pp. 50-51, 2000

[21] V.W. Zue, J.R. Glass, "Conversational Interfaces: Advances and Challenges", *Proc. IEEE*, Vol. 88, No. 8, pp. 1166-1180, 2000