

Statistical Analysis of Facial Expressions

Luigi Ippoliti¹ - Pasquale Valentini

Department of Quantitative Methods and Economic Theory
University G. d'Annunzio, Chieti-Pescara, Italy

Antonio Gattone

Department SEFeMeQ
University of Tor Vergata, Rome

Abstract: Over the last decade, the statistical analysis of facial expressions has become an active research topic that finds potential applications in many areas. Some examples of interest are human-computer interfaces, image retrieval and human emotion analysis. Facial expressions reflect not only emotions but also other mental activities, social interaction and physiological signals. Humans detect and interpret faces and facial expressions in a scene with little effort; however, the development of an automated system that accomplishes this task is rather difficult. There are, in fact, several related problems. This paper, provides an overview of the different stages which can be distinguished in the organization of an automatic facial expression analysis system.

Keywords: Facial action encoding, facial expression emotional, Face detection.

1. Introduction

The analysis of facial expression started into the nineteenth century. In 1872 in fact, Darwin demonstrated the universality of facial expressions and their continuity in man and animals; among other things, he claimed that exists specific inborn emotions which originates in serviceable associated habits. In 1971, Ekman and Friesen also postulated six primary emotions each having a distinctive content together with a unique facial expression. These prototypic emotional displays are also referred to as *basic emotions*. They seem to be universal across humans ethnicities and cultures and comprise happiness, sadness, fear, disgust, surprise and anger.

In the past, facial expression analysis (FEA) was primarily a research subject to psychologists, but already in 1978, Suwa *et al.*, presented a preliminary investigation on automatic facial expression analysis from an image sequence. However, in the nineties, automatic FEA gained much inertia starting with the pioneering work of Mase and Pentland (1991). The reasons for this renewed interest in FEA are multiple, but mainly due to advancements accomplished in related research areas such as face detection, face tracking and face recognition. Also, the recent availability of relatively cheap computational power constitutes another good reason.

Facial expression recognition should not be confused with human emotion recognition as is often done in the computer vision community. In fact, while *facial expression recognition* deals with the classification of facial motion and facial feature deformation into

¹Address of correspondence: DMQTE, Viale Pindaro 42, 65127 Pescara, e-mail: ippoliti@unich.it

abstract classes that are purely based on visual information, *human emotions* are a result of many different factors and their state may or may not be revealed through a number of signals such as emotional voice, pose, gesture, gaze direction and facial expressions. Thus, emotions are not the only source of facial expressions.

Facial expressions are generated by contractions of facial muscles, which results in temporally deformed facial features such as eye lids, eye brows, nose, lips and skin texture, often revealed by wrinkles and bulges. To provide facial expression measurements, of importance is the *location* of facial actions, their *intensity*, as well as their *dynamics*. Thus, facial expression intensities may be measured by determining either geometric deformations of facial features or the density of wrinkles appearing in certain face regions. However, since there are inter-personal variations with regard to the amplitudes of facial actions, it appears difficult to determine *absolute* facial expression intensities, without referring to the neutral face of a given subject.

There are three main temporal parameters which can be used to describe the facial expressions: *onset* (attack), *apex* (sustain), and *offset* (relaxation). These three characteristics can be measured by a *sign vehicle*-based approach, where facial motion and deformation are coded into visual classes. Facial actions are abstracted and described by their location and intensity; hence, a complete description framework would ideally contain all possible perceptible changes that may occur on a face. This is the goal of FACS (Facial Action Coding System), which was developed by Ekman and Friesen (1971) and has been considered as a foundation for describing facial expressions.

Automatic facial expression analysis is a complex task as physiognomies of faces vary from one individual to another quite considerably due to different age, ethnicity, gender, cosmetic products and occluding objects such as glasses and hair. Furthermore, faces appear disparate because of pose and lighting changes. Variations such as these have to be addressed at different stages of an automatic facial expression analysis system and, in general, three main steps can be distinguished in tackling the problem.

2. Automatic Facial Expression Analysis

2.1 Face Detection

Before a facial expression can be analyzed, the face must be detected in a scene. Ideally, a face acquisition stage features an automatic face detector that allows to allocate faces in complex scenes with cluttered backgrounds. In general, independently of the type of input images (facial images or arbitrary images) detection of the exact face position in an observed image or image sequence has been approached in two ways. In the *holistic approach*, the face is determined as a whole unit. In the second, *analytic approach*, the face is detected through some important facial features first (eg. the irises and the nostrils). The correspondence of the features with each other, then determines the overall location of the face.

For example, to represent the face in facial images, Huang and Huang (1997) apply a point distribution model (PDM). In order to achieve a correct placement of an initial PDM in an input image, Huang and Huang (1997) utilize a Canny edge detector to obtain a rough estimate of the face location in the image. The valley in pixel intensity that lies between the lips and the two symmetrical vertical edges representing the outer vertical boundaries of the face, generate a rough estimate of the face location.

To detect faces in arbitrary images, Essa and Pentland (1997) instead, use the eigenspace method of Pentland *et al.* (1994). The method employs eigenfaces approximated by using Principal Component Analysis (PCA).

Face analysis however, is also complicated by face *appearance changes* caused by pose variations and illumination changes. It may therefore be a good idea to normalize the acquired faces prior to their analysis.

2.2 Facial Feature Extraction

After the presence of a face is detected in a scene, the next step is to extract the information about the shown facial expression. In general, feature extraction methods can be categorized according to whether they focus on motion or deformation of faces and facial features, and whether they act holistically (where the face is processed as a whole) or locally (by focusing on local facial features and areas that are prone to change with facial expression).

Deformation of facial features are characterized by shape and texture changes and lead to high spatial gradients that are good indicators for facial actions. *Active Appearance Models* - AAM - (Lanitis *et al.*, 1997), for example, allow to simultaneously determine the shape, the scale and the pose by fitting an appropriate Point Distribution Model (PDM) to the object of interest. Huang and Huang (1997) used a PDM to represent the shape of a face, where shape parameters were estimated by employing a gradient-based method. However, a drawback of appearance-based models is the manual labor necessary for the construction of the shape models. The latter are based on landmarks that need to be precisely placed around intransient facial features during the training of the models.

Another type of holistic face models also constitute the so called *labeled graphs*, where each node of the graph, called *Gabor jet*, consists of an array. Specifically, each component of a jet is a filter response of a specific Gabor wavelet extracted at a given image point (Hong *et al.*, 1998). A labeled graph is matched to a test face by varying its scale and position. The obtained graph is then compared to reference graphs in order to determine the facial expression.

Instead, as an example of local approach to face representation, Kobayashi and Hara (1992) propose a geometric face model of 30 frontal *Facial Characteristic Points*. This model was successively extended by Pantic and Rothkrantz (2000) by using a side-view with 10 profile points.

Among the motion extraction methods that have been used for the task of FEA we find *dense optical flow* (Mase and Pentland, 1991), *feature point tracking* (Otsuka and Ohya, 1998) and *difference images* (Donato *et al.*, 1999).

2.3 Facial Expression Classification

Feature classification is performed in the last stage of an automatic facial expression analysis system. The mechanism of classification is either a template-based or a "rule"-based classification method. If template-based classification methods are used, the encountered facial expression is compared to the templates defined for each expression category. The best match decides the category of the shown expression. In general, it is difficult to achieve a template-based quantified recognition of a non prototypic facial expression. There are in fact, a lot of combinations of different facial actions and their intensities that should be modeled with a finite set of templates. The problem becomes even more diffi-

cult due to the fact that everybody has his own maximal intensity of displaying a certain facial action.

On the other hand, the rule-based classification methods classify the examined facial expression into basic emotion categories based on previously encoded facial actions. The prototypic expressions, which characterize the emotion categories, are first described in terms of facial actions. Then, the shown expression, described in terms of facial actions, is compared to the prototypic expressions defined for each of the emotion categories and classified in the optimal fitting category.

In general, commonly used classifier approaches are based on Hidden Markov Models, Neural Networks, Principal Component Analysis and Independent Component Analysis.

References

- Darwin C. (1872) *The Expression of the Emotions in Man and Animals*, J. Murray, London.
- Donato G., Bartlett S., Hager C., Ekman P. and Sejnowski J. (1999) Classifying facial actions, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 21, 974–989.
- Ekman P. and Friesen W. (1971) Constants across cultures in the face and emotions, *Journal of Personality and Social Psychology*, 17, 124–129.
- Essa I. and Pentland A. (1997) Coding, analysis, interpretation and recognition of facial expressions, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19, 757–763.
- Hong H., Neven H. and von der Malsburg C. (1998) On line facial expression recognition based on personalized galleries, *Proceedings International Conference Automatic Face and Gesture Recognition*, 354–359.
- Huang C. and Huang Y. (1997) Facial expression recognition using model-based feature extraction and action parameters classification, *Journal of Visual Communications and Image Representation*, 8, 278–290.
- Kobayashi H. and Hara F. (1992) Recognition of six basic facial expressions and their strength by neural network, *Proceedings International Workshop Robot and Human Communication*, 381–386.
- Lanitis A., Taylor C. and T.F. C. (1997) Automatic interpretation and coding of face images using flexible models, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19, 743–756.
- Mase K. and Pentland A. (1991) Recognition of facial expression from optical flow, *IEICE Transaction E*, 74, 3474–3483.
- Otsuka T. and Ohya J. (1998) Extracting facial motion parameters by tracking feature points, *Proceedings of the 1st International Conference on Advanced Multimedia Content Processing*, 442–453.
- Pantic M. and Rothkrantz L. (2000) Expert system for automatic analysis of facial expressions, *Image and Vision Computing Journal*, 18, 881–905.
- Pentland A., Moghaddam B. and Starner T. (1994) View-based and modular eigenspaces for face recognition, *Proceedings Computer Vision and Pattern Recognition*, 84–91.
- Suwa M., Sugie N. and Fujimora K. (1978) A preliminary note on pattern recognition of human emotional expression, *Proceedings of the 4th International Joint Conference on Pattern Recognition*, 408–410.