

Facial Expression Recognition

Raymond Ptucha

Rochester Institute of Technology, USA

Andreas Savakis

Rochester Institute of Technology, USA

INTRODUCTION

Facial expression recognition is the task of autonomously analyzing the human face to estimate a person's emotional state, mood or other form of facial communication. The ability to automatically extract facial semantic information has widespread implications on a number of industries including security, entertainment, and human computer interfaces. State-of-the-art techniques have become adept at recognizing posed expressions in laboratory conditions and have migrated to recognizing spontaneous expressions in uncontrolled settings.

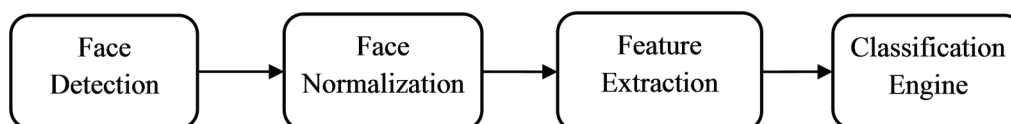
The importance of facial expressions was recognized in Charles Darwin's 1872 book "The Expression of the Emotions in Man and Animals." In the late 1960's, psychologist Paul Ekman and his collaborator Wallace Friesen empirically showed all cultures exhibit the six universal expressions of fear, sadness, happiness, anger, disgust, and surprise. Ekman, Friesen, and their colleagues then created a taxonomy of facial expressions and documented forty-three facial movements constituting over ten thousand facial expressions. They discovered facial expressions consisted of both voluntary and involuntary muscle contractions, noted differences between genuine and posed expressions, and documented quick bursts of involuntary facial expressions called microexpressions.

Affective computing, or computing that deliberately senses and influences emotion have spawned a new era in human computer interfaces. Facial expression recognition techniques initially mastered posed datasets and in the past decade researchers began introducing variability in the expression data. Today, techniques are focusing on recognizing facial expression in unconstrained conditions which include variations of facial pose, facial occlusions, illumination, image fidelity, and background clutter.

BACKGROUND

The study of the six universal expressions, i.e. fear, sadness, happiness, anger, disgust, and surprise, has made great strides in recent years from constrained frontal posed faces to unconstrained faces in natural conditions (Maja Pantic & Rothkrantz, 2000; Shuai-Shi, Yan-Tao, & Dong, 2009b; Zhihong, Pantic, Roisman, & Huang, 2009). Figure 1 shows the basic steps necessary for a facial expression recognition system. Face detection is often accomplished with the Viola-Jones approach because of its low computational requirements and high detection rates (Viola & Jones, 2001). Recent face detection methods improve the accuracy, efficiency, or robustness (Zhu & Ramanan, 2012). Following face detection, faces are normalized to a

Figure 1. Flowchart of fundamental operations used for facial expression recognition



DOI: 10.4018/978-1-4666-5888-2.ch051

reference shape and size. Typically, the eye and mouth corners are localized, and an affine warp to canonical frontal face is defined.

Facial expression methods can be broadly categorized as geometric or appearance-based (Fasel & Luettin, 2003; Shuai-Shi, Yan-Tao, & Dong, 2009a). Geometric methods localize facial landmarks such as the outline of eyes, lips, nose, etc. (Martin, Werner, & Gross, 2008; Yeongjae & Daijin, 2009). Appearance-based methods work holistically with facial pixels enabling the capture of facial muscle subtleties such as nose wrinkles or dimple formation (Shan, Gong, & McOwan, 2009).

Geometric methods require computing size, shape, and location of key facial features such as the eyes, mouth, and eyebrows. Active Shape Model (ASM) or Active Appearance Model (AAM) are two of the most popular facial landmark localization methods (Cootes, Edwards, & Taylor, 2001). Given enough training data and accurate facial landmark localization, shape models perform very well for expression classification (Martin et al., 2008; Yeongjae & Daijin, 2009).

Appearance based methods often compute intermediate representations of images using features such as Gabor wavelets (Buciu, Kotropoulos, & Pitas, 2003; Zhengyou, Lyons, Schuster, & Akamatsu, 1998) or Local Binary Patterns (LBP) (Feng, Pietikainen, & Hadid, 2007; Shan et al., 2009). Gabor wavelets compute directional band pass filters based on the human visual system, but are slow and memory intensive. LBP features capture various texture primitives and are quite tolerant to illumination changes.

The classification engine for facial processing has been studied extensively (Sebe et al., 2007). With proper feature extraction, common methods such as k-Nearest Neighbor (k-NN), Support Vector Machines (SVMs), Logistic Regression, Adaboost, regression trees, and artificial neural networks all yield acceptable results. Sparse Representations (SRs) have proven to be effective at facial recognition, and recently have been adopted for facial expression classification (Ptucha, Tsagkatakis, & Savakis, 2011; Liu, Han, Tong, 2013). Deep learning methods learn weights of multi-layer neural networks, most layers being learned in an unsupervised fashion (Liu, Li, Shan, Chen, 2013).

Sign-based methods decipher facial motion into action classes such as Facial Action Coding System (FACS) (Ekman & Friesen, 1978), whereby groupings of muscles in the face form Action Units (AUs), the motions and combinations of which enable final classification (Donato, Bartlett, Hager, Ekman, & Sejnowski, 1999). FACS captures all the possible atomic facial signals which can then be used as features into a reasoning engine. For example, human subjects exhibiting AU six (contraction of orbicularis oculi and pars orbitalis, or the cheek raiser muscles) in combination with AU twelve (pulling up of the zygomatic major, or the corners of the lips) are generally experiencing happiness. Interestingly, Ekman discovered that if someone is asked to act as if they are happy, they will perform only AU twelve. He found it almost impossible for subjects to exercise the orbicularis oculi and pars orbitalis properly on command. Equally intriguing, it was just as difficult for humans to stop those muscles from contracting when they were genuinely happy.

Temporal evidence has been shown to significantly aid emotion recognition (Curio, Bulthoff, & Giese, 2010; M. Pantic et al.; Ambadar, Schooler, & Cohn, 2005). Recent works demonstrate that temporal dynamics can improve AU detection considerably (Koelstra, Pantic, & Patras, 2010). The timing and duration of onset, apex, offset, and neutral stages are critical to the interpretation of the observed behavior. Temporal dynamics have been employed in discerning between genuine and acted pain (Patrick Lucey, Cohn, Matthews, Prkachin, & Solomon, 2011), telling the truth vs. lying (Bhaskaran, Nwogu, Frank, & Govindaraju, 2011), detecting depression, and many other applications.

DATASETS

There are numerous publically available facial expression datasets. The two most commonly referenced are the datasets by Cohn-Kanade (CK) (Kanade, Cohn, & Yingli, 2000) and Japanese Female Facial Expression (JAFEE) (Lyons, Akamatus, Kamachi, & Gyoba, 1998). The expressions in these datasets are frontal, acted and often exaggerated. Datasets containing spontaneous facial behavior are difficult because triggering the

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/facial-expression-recognition/112366

Related Content

Ergonomic Design of a Driver Training Simulator for Rural India

Prabir Mukhopadhyay and Vipul Vinzuda (2018). *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 1260-1276).

www.irma-international.org/chapter/ergonomic-design-of-a-driver-training-simulator-for-rural-india/183840

A RNN-LSTM-Based Predictive Modelling Framework for Stock Market Prediction Using Technical Indicators

Shruti Mittal and Anubhav Chauhan (2021). *International Journal of Rough Sets and Data Analysis* (pp. 1-13).

www.irma-international.org/article/a-rnn-lstm-based-predictive-modelling-framework-for-stock-market-prediction-using-technical-indicators/288521

Simulation Research on Smart Identification Model of English MT Based on Artificial Neural Network

Difei Wang and Sihan Yang (2025). *International Journal of Information Technologies and Systems Approach* (pp. 1-16).

www.irma-international.org/article/simulation-research-on-smart-identification-model-of-english-mt-based-on-artificial-neural-network/393669

Regional Health Information Organizations in the US

Jonathan Becker, Neelam Dwivedi and Sandeep Puro (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 3496-3505).

www.irma-international.org/chapter/regional-health-information-organizations-in-the-us/112781

A Study of Precipitation Inversion in Xinjiang Region Based on FY-4A and Machine Learning Models

Zhengyang Zhang, Jiang Li, Aqiao Li and Andrew Li (2024). *International Journal of Information Technologies and Systems Approach* (pp. 1-16).

www.irma-international.org/article/a-study-of-precipitation-inversion-in-xinjiang-region-based-on-fy-4a-and-machine-learning-models/351218