Affectiva-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected In-the-Wild

Daniel McDuff†‡, Rana El Kaliouby†‡, Thibaud Senechal†, May Amr†, Jeffrey F. Cohn§, Rosalind Picard†‡
†‡ Affectiva Inc., Waltham, MA, USA
† MIT Media Lab, Cambridge, MA, USA
§ Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA
djmcduff@media.mit.edu, {kaliouby, thibaud.senechal, may.amr}@affectiva.com,
jeffcohn@cs.cmu.edu, picard@media.mit.edu

Abstract

Computer classification of facial expressions requires large amounts of data and this data needs to reflect the diversity of conditions seen in real applications. Public datasets help accelerate the progress of research by providing researchers with a benchmark resource. We present a comprehensively labeled dataset of ecologically valid spontaneous facial responses recorded in natural settings over the Internet. To collect the data, online viewers watched one of three intentionally amusing Super Bowl commercials and were simultaneously filmed using their webcam. They answered three self-report questions about their experience. A subset of viewers additionally gave consent for their data to be shared publicly with other researchers. This subset consists of 242 facial videos (168,359 frames) recorded in real world conditions. The dataset is comprehensively labeled for the following: 1) frame-by-frame labels for the presence of 10 symmetrical FACS action units, 4 asymmetric (unilateral) FACS action units, 2 head movements, smile, general expressiveness, feature tracker fails and gender; 2) the location of 22 automatically detected landmark points; 3) self-report responses of familiarity with, liking of, and desire to watch again for the stimuli videos and 4) baseline performance of detection algorithms on this dataset. This data is available for distribution to researchers online, the EULA can be found at: http://www.affectiva.com/facial-expression-dataset-am-fed/.

1. Introduction

The automatic detection of naturalistic and spontaneous facial expressions has many applications, ranging from medical applications such as pain detection [1], or monitoring of depression [4] and helping individuals on the autism spectrum [10] to commercial uses cases such as advertising research and media testing [14] to understanding non-verbal communication [19]. With the ubiquity of cameras on computers and mobile devices, there is growing interest in bringing these applications to the real-world. To do so, spontaneous data collected from real-world environments is needed. Public datasets truly help accelerate research in an area, not just because they provide a benchmark, or a common language, through which researchers can communicate and compare their different algorithms in an objective manner, but also because compiling such a corpus and getting it reliably labeled, is tedious work - requiring a lot of effort which many researchers may not have the resources to do.

There are a number of publicly available labeled databases for automated facial analysis, which have helped accelerate research in automated facial analysis tremendously. Databases commonly used for facial action unit and expression recognition include; Cohn-Kanade (in its extended edition know as CK+) [11], MMI [23], RU-FACS [2], Genki-4K [24] and UNBC-McMaster Pain archive [12]. These datasets are reviewed in Section 2. However, all (except the Genki-4K and UNBC-McMaster Pain archives) were captured in controlled environments which do not reflect the the type of conditions seen in real-life applications. Computer-based machine learning and pattern analysis depends hugely on the number of training examples [22]. To date much of the work automating the analysis of facial expressions and gestures has had to make do with limited datasets for training and testing. As a result this often leads to over-fitting.

Inspired by other researchers who made an effort to share their data publicly with researchers in the field, we present a database of spontaneous facial expressions that was collected in naturalistic settings as viewers watched video content online. Many viewers watched from the comfort of their homes, which meant that the facial videos contained a range of challenging situations, from nonuniform lighting