DATASETS(FACS)

Authors: Rahul. k, *IIITDM Kancheepuram.*

1 INTRODUCTION

Preparing datasets for FACS involves a meticulous process of capturing, annotating, and preprocessing facial images or video frames to ensure accurate detection and analysis of facial action units (AUs).

The preparation of a FACS dataset typically begins with the acquisition of high-quality facial images or videos, which should cover a diverse range of expressions, demographics, and lighting conditions to enhance model robustness. Once the data is collected, manual or automated annotation is performed to label the facial expressions according to FACS. This process involves identifying and categorizing subtle facial movements associated with specific AUs, which are the fundamental building blocks of facial expressions.

To further refine the dataset, preprocessing techniques are applied to standardize image quality and enhance facial features. Common preprocessing steps include facial alignment, normalization, resizing, and contrast enhancement to ensure consistent input for machine learning models. Data augmentation techniques, such as rotation, flipping, and color jittering, are often used to artificially increase dataset size and improve model generalization by introducing variability.

A well-prepared FACS dataset is crucial for training robust and accurate models for facial action unit detection and expression analysis. By meticulously capturing, annotating, and preprocessing the data, researchers can ensure that their models perform effectively across a wide range of conditions and facial expressions, contributing to advancements in emotion recognition and human-computer interaction technologies.

2 Datasets

2.1 DISFA(Denver Intensity of Spontaneous Facial Action)

' The **Denver Intensity of Spontaneous Facial Action Database** [3]is a **non-posed** facial expression database for those who are interested in developing computer algorithms for automatic action unit detection and their intensities described by FACS. This database contains stereo videos of 27 adult subjects (12 females and 15 males) with different ethnicities. The images were acquired using PtGrey stereo imaging system at high resolution (1024×768). The intensity of AU's (0-5 scale) for all video frames were manually scored by two human FACS experts. The database also includes 66 facial landmark points of each image in the database. Twenty-seven young adults were video-recorded by a stereo camera while they viewed video clips intended to elicit spontaneous emotion expression.

Participants viewed a 4-minute video clip (242 seconds in length) intended to elicit spontaneous AUs in response to videos intended to elicit a range of facial expressions of emotion. The clip consisted of 9 segments taken mostly from YouTube. Further information about the video clip is provided in the Appendix. While viewing the video, participants sat in a comfortable chair positioned in front of a video display and stereo cameras. They were alone with no one else present. Their facial behavior was imaged using a high-resolution (1024 × 768 pixels) BumbleBee Point Grey stereo-vision system at 20 fps under uniform illumination. For each participant, 4845 video frames were recorded.The imaging system is depicted in Figure 3.

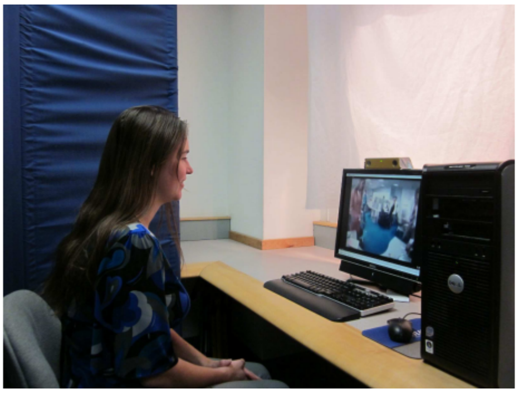


Figure 1: DISFA imaging setup.

AU intensity was coded for each video frame on a 0 (not present) to 5 (maximum intensity) ordinal scale. For each AU, we report the number of events and the number of frames for each intensity level. Event refers to the continuous occurrence of an AU from its onset (start frame) to its offset (end frame) .

FACS coding was performed by a single FACS coder. To evaluate inter-observer reliability, video from 10 randomly selected participants was annotated by a second FACS coder. Both coders were certified in use of FACS. Interobserver reliability, as quantified by intra-class correlation coefficient (ICC), ranged from 0.80 to 0.94 (as seen in Table 3). ICC value of 0.80 and higher is considered as high reliability [55].

2.2 DISFA+(Extended Denver Intensity of Spontaneous Facial Action Database)

2.2.1 Recording procedure for posed expressions

To acquire the posed expressions of individuals we designed a software, which instructed and guided users to imitate a set of 42 facial actions during the capturing session (some of these facial actions are shown in Figure 4). (The full list of these facial actions is provided in Table 1. Every subject watched a 3-minute demo to learn how to use the software and record her posed facial expressions for multiple trails. In each trial, the user was asked to mimic a full dynamic of facial actions[1] (i.e. begin with a neutral face, proceed to the maximum intensity of expression and finally end with a neutral face). The user could see her face on an LCD monitor as she mimicked the expression. Figure 5 illustrates a screen-shot of our software while one of the users mimiking 'Surprise' face. This software has been written in C++ and the OpenCV library [2] was used to record and time-stamp every frame of the video.

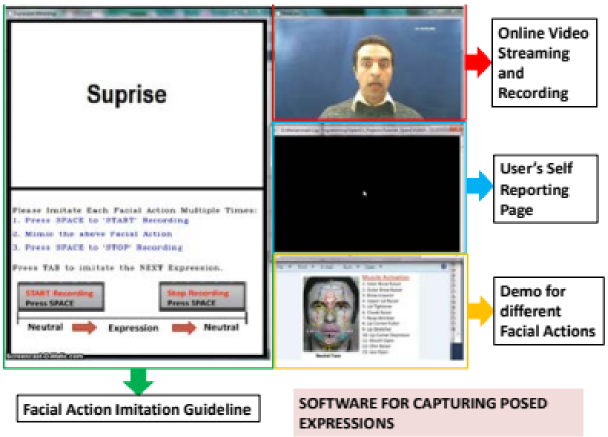
Each user was asked to imitate 30 facial actions (i.e. single AU or combinations of AUs) and 12 facial expressions corresponding to the emotional expression (e.g. Surprise, anger, etc.). A list of these facial expressions is provided in the Appendix. Meanwhile an HD camera recorded the facial responses of participants with 1280 x 720 pixel resolution in 20 frames per second. This frame rate and imFacial Action Imitation Guideline Online Video Streaming and Recording User's Self Reporting Page Demo for different Facial Actions SOFTWARE FOR CAPTURING POSED EXPRESSIONS Figure 2: A demo of the designed software for capturing the posed facial expressions of subjects in DISFA+ [2]database.

age resolution was selected to match with the DISFA video acquisition setting and make data comparison and analysis easy. Each individual was instructed to practice each of the facial action first and then begun to record a few trials for each facial action. Users imitated each facial action few times and after finishing each trial, participants rated (in the range of [0-10]) each mimicked facial expression by answering two questions:

- How difficult was it for you to make each facial expression?
- How accurate could you mimic/pose the facial expression?

We selected the best trials of each individual with highest score. These trials were then FACS coded and the intensity of each frame was labeled 1.

To annotate the ground-truth labels of DISFA+, a certified FACS coder annotated the intensity of 12 AUs



Please Imitate Each Facial Action Multiple Times:
1. Press SPACE to 'START' Recording
2. Mimic the above Facial Action
3. Press SPACE to 'STOP' Recording

Press TAB to imitate the NEXT Expression.

Suprise
START Recording
Press SPACE
Stop Recording
Press SPACE
Neutral
Expression
Neutral
Online Video Streaming and Recording
User's Self Reporting Page
Demo for different Facial Actions
Facial Action Imitation Guideline
SOFTWARE FOR CAPTURING POSED EXPRESSIONS

Figure 2: A demo of the designed software for capturing the posed facial expressions of subjects in DISFA+ database.

2.3 IFED(Indian Facial Expressions Dataset)

Background: Face is the most presented part of the human body in all interactions of our life. Face helps in identification of a person in expressing our interest verbally and non-verbally. Facial expression even though considered universal, there are substantial differences among different population across the globe. Databases on basic facial expressions are very limited among Indians.
Material and methods: Study was reviewed and ethically approved by Institutional Ethical Committee. Participants were selected by stratified random sampling from different zones of India. 112 participants were thus selected; informed consent was collected. Participants expressions were evoked by showing validated emotionally valent videos. Then responses were recorded and then classified, analyzed and tested statistically.
Results: There was significant difference in the expression of fear, anger, disgust, contempt, sadness and surprise among participants. Happiness was universally similarly expressed and welcomed expression among the basic expressions. Finally, a database for facial expression is compiled for facial expressions among Indians [IFED].

2.3.1 Materials and Methods

Observational, Qualitative cross sectional Study with sample size of 112 obtained after statistical calculation from previous studies. Participants were Indian subjects aged between 18-40 years who were apparently in good health and capable of expressing facial expressions. Participants were selected by stratified random sampling method. Participants were invited by announcements about the project in various classrooms of the University and college campus of the medical schools. 60% of participants belong to South zone of India, rest 40% belong to east zone, west zone and north zone of India.

2.3.2 Method

The intended database was of expressions of evoked type. The participants were shown pre-selected validated emotionally valent videos one after another for each emotion like anger, contempt, disgust, Happiness, surprise, fear and sadness. Each video was of time duration ranging between 17 seconds to 7 minutes dependent on the expression to be elicited. Participants were all given information sheet and requested to read through the sheet and any queries related to the project were answered in the language they are comfortable. Informed consent was then collected. Preliminary data was collected and instructions were given to each participant separately for the experiment. They were instructed to watch the video one after another and express the emotion they felt as they watch the video.
Meanwhile, as they watch the video, their face was recorded with the webcam Logitech C920HD with 1080p resolution. The participant was made to sit in-front of the computer which contained validated experimental videos. Experimental set up was well lit with photo studio lights and back-ground was kept darker to decrease reflection of light, capture the expressions better. The room had only the Participant and the investigator during the period of the experiment. The investigator was behind the shadow with another computer recording the expression. The participant was kept at ease and comfort. The investigator did not interrupt in any manner with the participant once the recording/experiment began. After each video was watched, the participant wrote a self-report on the same using a format about the emotion they felt and the valence of the same. The experiment took nearly 30 minutes for each participant to complete the watching and self-reports for each.

2.3.3 Experimental setup

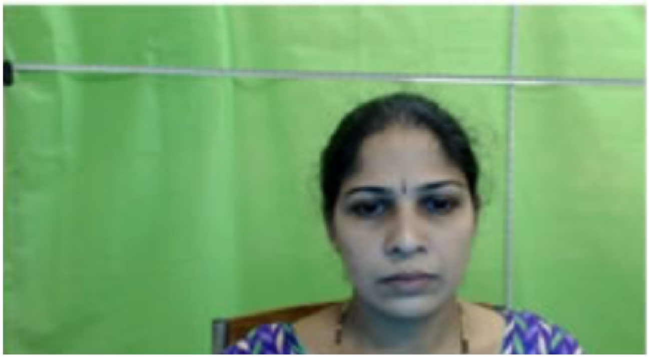


Figure 3: Experimental set up: [Representational; figures reproduced with consent

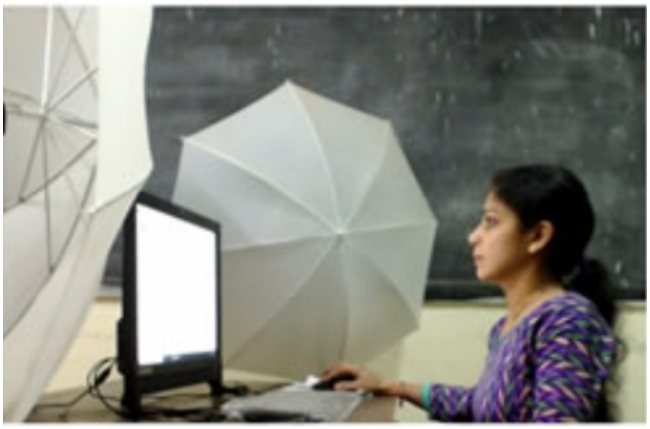


Figure 4: Lateral view of representational for showing the experimental set up. [figures reproduced with consent]

The obstacles like spectacles, beard, mustache, hair were not restricted for the participants as they would become more conscious of being recorded. The participants were watching these videos for first time and were not aware of the contents of the video.

These responses of the participants were cross-checked with their self-report, coder report and the iMotions software for analysis. The responses were compared statistically.

References

[1] Littlewort, G., Fasel, I., and Movellan, J. (2003). Real time face detection and facial expression recognition: Development and applications to human computer interaction. volume 5, pages 53–53.

[2] Mavadati, M., Sanger, P., and Mahoor, M. H. (2016). Extended disfa dataset: Investigating posed and spontaneous facial expressions. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1452–1459.

[3] Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., and Cohn, J. F. (2013). Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, 4(2):151–160.