Emotion expression from different angles: a video database for facial expressions of actors shot by a camera array

Seuss, Dominik (ed.); Dieckmann, Anja (ed.); Hassan, Teena (ed.); Garbas, Jens-Uwe (ed.); Ellgring, Johann Heinrich (ed.); Mortillaro, Marcello (ed.); Scherer, Klaus R. (ed.)

Emotion Expression from Different Angles:
A Video Database for Facial Expressions of Actors Shot by a Camera Array

Dominik Seuss
Electronic Imaging Department
Fraunhofer IIS
Erlangen, Germany
dominik.seuss@iis.fraunhofer.de

Anja Dieckmann
Behavioral Science
Nuremberg Institute for Market Decisions
Nürnberg, Germany
anja.dieckmann@nim.org

Teena Hassan & Jens-Uwe Garbas
Electronic Imaging Department
Fraunhofer IIS
Erlangen, Germany
{teena.hassan;
jens.garbas}@iis.fraunhofer.de

Johann Heinrich Ellgring
Department of Psychology
University of Würzburg
Würzburg, Germany
ellgring@uni-wuerzburg.de

Marcello Mortillaro
Campus Biotech
University of Geneva
Geneva, Switzerland
marcello.mortillaro@unige.ch

Klaus Scherer
Department of Psychology
University of Geneva
Geneva, Switzerland
klaus.scherer@unige.ch

Abstract*—Over the last few decades, there has been an increasing call in the field of computer vision to use machine-learning techniques for the detection, categorization, and indexing of facial behaviors, as well as for the recognition of emotion phenomena. Automated Facial Expression Analysis has become a highly attractive field of competition for academic laboratories, startups and large technology corporations. This paper introduces the new Actor Study Database to address the resulting need for reliable benchmark datasets. The focus of the database is to provide real multi-view data, that is not synthesized through perspective distortion. The database contains 68-minutes of high-quality videos of facial expressions performed by 21 actors. The videos are synchronously recorded from five different angles. The actors' tasks ranged from displaying specific Action Units and their combinations at different intensities to enactment of a variety of emotion scenarios. Over 1.5 million frames have been annotated and validated with the Facial Action Coding System by certified FACS coders. These attributes make the Actor Study Database particularly applicable in machine recognition studies as well as in psychological research into affective phenomena—whether prototypical basic emotions or subtle emotional responses. Two state-of-the-art systems were used to produce benchmark results for all five different views that this new database encompasses. The database is publicly available for non-commercial research.

***Keywords*—affective database, facial expression analyses**

I. INTRODUCTION

Recent years have seen immense acceleration in the technological sector of computer vision to use machine-learning techniques for Automated Facial Expression Analysis (AFEA), and for the recognition of emotion phenomena through the detection, categorization, and indexing of facial behaviors. There remains, however, a significant need for benchmark datasets, with multiple facial views and validated face annotation, for method optimization and performance evaluation.

This paper introduces the new Actor Study Database, which contains high-speed and high-resolution video recordings of facial actions and facial expressions of emotion. These are posed by actors in response to scripted tasks and emotion as well as appraisal scenarios. The dataset consists of penta-view (or five-angle) recordings of these expressive displays in frame-aligned videos. The frames are annotated with the Action Units (AUs) of the Facial Action Coding System (FACS) developed by Ekman and Friesen [1].

These attributes locate the Actor Study Database on the cutting-edge of currently available benchmark datasets. It extends beyond other non-commercial research databases by encompassing not only full-frontal pictures of the face, but also views from multiple angles. It includes not only the basic and high-intensity facial expressions of emotions, but also enactments of emotional appraisal scenarios for more subtle responses in facial expressivity.

The above-mentioned innovative features of the Actor Study Database make it particularly applicable for machine-learning recognition studies into facial expressions of emotions. It can be used for the development, optimization and evaluation of AFEA algorithms. The dataset is available for non-commercial research, see section ActorStudy under http://www.iis.fraunhofer.de/shore. This paper details the theoretical background and the data collection procedure. It includes a comparison of the dataset with existing AU-annotated datasets. Finally, it also provides baseline benchmark results for this new dataset using two state-of-the-art AFEA systems. Benchmark results are provided for all five views—top, bottom, center, left, right.

II. THEORETICAL BACKGROUND

Much work in AFEA is based on Basic Emotions Theory (BET) that, for several decades, has also dominated empirical research in emotion psychology. According to BET, there are prototypical facial behaviors for particular emotion phenomena, and these patterns of muscular contraction mechanisms and skin appearance changes are biologically specified as well as culturally universal [2].

In line with the assumption that there are prototypical expression configurations for the basic emotions, much of the automated recognition work in computer vision is guided by an implicit notion of template matching—that is, the comparison of the facial elements of a query face in an image record with those in a matched class from a face database. The relative success achieved with these machine recognition studies confirms that there are indeed some typical expression patterns for a small number of basic emotions (joy, anger,

978-1-7281-3888-6/19/$31.00 ©2019 IEEE

disgust, fear, sadness, and to some extent surprise), see [3]. However, even for the basic emotions there are a large variety of quite different expression patterns. This is even more true for the large variety of relatively subtle emotions such as doubt, impatience, interest, or relief. Consequently, automated recognition that is based on template matching has so far attained only limited validity.

Extending beyond template-matching techniques, a further approach in affective computing is to recognize AUs, or functional groups of facial muscles that generally respond in unison. AUs are the most fine-grained visually differentiable indicators of emotions and are defined in FACS [1]. Using still photography and moving video annotated in accordance with FACS, machine-learning algorithms can be trained to detect AUs [4]. In a subsequent step, AU-data can be used rather flexibly to infer basic emotions, via rule-based methods according to the Emotion Facial Action Coding System (EMFACS) [5] or via data-driven machine-learning techniques [3, 6]. The AU-data could also be used for predictions based on emotion-antecedent appraisals (as proposed by Scherer and collaborators) [7].

Besides offering higher flexibility, an AFEA system that is based on AUs also yields greater transparency. Whereas direct inference from pixel-level facial features to basic emotions is frequently considered a "black box" [8, 9], inferences from a system based on AU detection as the first step can still be evaluated in terms of correspondence with theoretical predictions (that is, whether the inference is based on the theorized AUs that have been found to usually accompany certain emotions).

It may be due to these advantages in terms of both flexibility and transparency that the two most recent FERA (Facial Expression and Analysis) Challenges, in which Affective Computing research groups from around the world compete against each other based on an objective benchmark, focused exclusively on AU detection [10, 11], while the first challenge still included the direct recognition of discrete emotions as a sub-challenge [12].

To develop an AU-recognition system, face recordings with AU ground truth are required as training material. However, especially when the goal for the automated system is to also recognize subtle intensities in facial expressions of emotions, these machine-learning algorithms are "data hungry." This means that, for the algorithm to perform both accurately and robustly, a large dataset is necessary to train it. Whereas for recognition of discrete emotions, pictures of faces that are annotated relatively quickly by laypeople can be used as training material, AU annotations in accordance with FACS are significantly more time-consuming (the coding-time to real-time ratio can often be 100:1) and requires trained experts [13]. Thus, unsurprisingly, only a few AU-annotated image and video databases are publicly available, such as the Extended Cohn-Kanade (CK+) database [14, 15], MMI database [16], DISFA database [17], Bosphorus [18], UNBC-McMaster Pain Archive [19], BP4D [20], BP4D+ [21] and GFT [22].

Moreover, most AU-annotated databases exclusively provide full-frontal pictures of the face. But if a detection system is supposed to detect facial expressions of emotions "in the wild"—that is, in "real world" natural conditions—it is highly desirable that emotions can be recognized when the face is viewed from different angles.

Additionally, there are some picture sets that show facial expressions of the prototypical basic emotions (e.g., Pictures of Facial Affect [23], JAFFE [24]). However, due to typicality and high intensity of the facial expressions, an automated system trained on such pictures may not prove sensitive enough to detect more subtle emotion expressions.

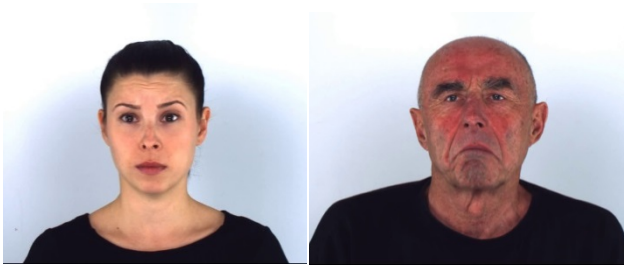

Fig. 1: AU performance from the Actor Study Database. Left: AU01, AU02 and AU04 in combination. Right: AU15.

Therefore, we decided to create a database in which actors posed facial expressions of single AUs, AU combinations, and enacted facial expressions according to different emotion and appraisal scenarios. There are many AU combinations, which are important in facial appraisal and often difficult to detect. We therefore provide individual AUs and their relevant combinations. All video recordings were annotated in terms of AU onset and offset by certified FACS coders. Extending beyond the functionality of existing databases, we recorded faces from different angles simultaneously, and included recordings of appraisal scenarios intended to elicit relatively subtle emotion expressions. The dataset thus contains real, multi-view data, which is captured using frame-synchronized cameras, in contrast to e.g. BP4D [20] dataset, which is synthesized through perspective distortion.

III. Data Collection Procedure

This section describes the recording settings, the facial expression elicitation methods and the video annotation procedure that were used to produce the Actor Study Database

A. Recording specifications

The database was created for applications where the target person's responses are directed towards a frontal stimulus or observer (e.g. interactions with a robot, responses to advertisement). In such cases, facial angles in a range of -30 to 30 degrees are expected. Moreover, AUs do not necessarily occur symmetrically. Therefore, we recorded videos from 30° left and 30° right views.

We used a total of seven cameras to record the videos. This includes five JAI CB-200 GE industrial cameras (24 frames per second; 1624x1236 px) and two high-speed Optronics CL300/2m cameras (125 frames per second; 1280x1024 px). The five low-speed cameras were positioned at five different angles, namely center, 30° left, 30° right, top and bottom. The high-speed cameras were positioned at 30° right and center positions. The average radial distance to the low-speed cameras was 1.6m and that to the high-speed cameras was 1.8m. The cameras were synchronized to achieve perfect frame-alignment for videos recorded from all seven cameras. To achieve uniform lighting over the actors' faces and for best visibility of AUs, six high power MultiLED softbox lights were used.

The total length of the recordings is 68 minutes (1,503,495 frames). Of these, 1002330 frames were from the high-speed cameras, and 501165 frames from low-speed cameras. The frames of the low-speed frontal camera were

annotated by FACS-coders, and the annotations were interpolated for the high-speed camera frames.

B. Expression elicitation methods

Professional actors, 21 in total (10 males, 11 females, all Caucasian) and with an average age of 42 years (ranging from 26 to 68 years), performed the facial actions and facial expressions of emotion. We recruited these actors from the Munich Artists Employment Agency, and each received an honorarium of 500€ for their work.

In preparation for the recording session, the actors received information material regarding the purpose of the study. This material contained an overview of the tasks that the actors would be asked to complete, and included a list of all single AUs and AU combinations with pictures of these facial expressions that they would be asked to perform (see Table I and Table II). The actors were asked to practice the AUs at home in front of a mirror before the experimental session took place.

The actors were then invited to individual recording sessions of about 2.5 hours duration. They were asked to perform a series of four tasks while being seated in front of the camera array.

- Task 1: Display of 32 single AUs and AU combinations, 18 of which were performed at two intensities (medium and high, corresponding to *c* and *d-e* in FACS, respectively), with two video recordings made for each display.
- Task 2: Display of AU combinations that correspond to 5 basic emotions (happiness, anger, disgust, fear, and sadness), with two video recordings made for each display.
- Task 3: Response to 8 emotional appraisal scenarios. The actors received scripts consisting of three parts in which the protagonist made three different appraisals of a fictitious event. The appraisals included novelty, pleasantness and coping potential. The actors were instructed to show their facial expression in response to the scenarios.
- Task 4: Enactment of 13 emotion scenarios, using the standard emotion portrayal procedure [25], wherein actors were asked to imagine their emotional response to the different described scenarios, and produce the facial expression they considered appropriate.

Actors were instructed to move facial muscles in a specific way in tasks 1 and 2. In Tasks 3 and 4, actors were asked to express the required emotional appraisal (task 3), or enact an emotion (task 4) but the choice of facial expressions was at their own discretion. Actors were given written appraisal and emotion scenarios that they should vividly imagine, that is, using a Stanislavski-like enactment method to induce an appropriate emotional state (see [25]) and then show expressions they would likely show in that situation. Therefore, subtle displays of AUs can also be observed here.

Each recording session involved two "experimenters." One of them was a certified and experienced FACS expert who served as "face experimenter." The other was a "technical experimenter", who operated the camera array and made the recordings. The "face experimenter" gave instructions to the actors, and for tasks 1 and 2, confirmed to the "technical experimenter" when an AU was correctly performed.

In tasks 1 and 2, first the "face experimenter" verbally described and displayed (with his own face) each facial expression, before the actor rehearsed them in a "dry run." As soon as the actor performed a facial action with sufficient accuracy (in accordance to FACS), the "face experimenter" signaled the "technical experimenter" to start the video recording. The "face experimenter" then requested the actor to "freeze" or hold the facial expression at its medium or maximum intensity. In case of insufficient accuracy (i.e. deviating from those described by FACS), repetitions were attempted. In cases where no sufficient movement was made after some repetitions, the "technical experimenter" took a corresponding note, and the actor went on to the next display. Fig. 1 shows examples of AUs performed by two actors in task 1. In task 2, that is, the display of AU configurations corresponding to five basic emotions, the "face experimenter" highlighted to the actors that they would now be asked to display more complex combinations of movements by activating groups of muscles. In this task, the intensity of expression was not varied. Apart from that, the procedure was the same as for task 1.

In task 3, the actor was asked to specifically attend to the three different appraisal components in each scenario. The performing actor and "face experimenter" together read aloud the scenario, before the actor gave an "ok" when ready to facially express her or his emotional response. Then, the "technical experimenter" gave the signal for the actor to look at the camera. The actor expressed facially the course of her or his appraisal response to the fictional situation, while the "face experimenter" read again the main appraisal-related elements of the scenario (novelty, pleasantness, coping); for example, "The task is not as expected - you don't like it at all - you think it is too difficult for you").

In task 4, as in task 3, the actor and "face experimenter" together read the scenario. Then, the "face experimenter" asked the actor to intensively imagine this situation, before giving an "ok" when ready. At that point, the "face experimenter" read for a second time the last sentence of the scenario (e.g., "I have to remove the vomit of a guest," for the basic emotion of disgust), and the actor facially expressed her or his emotional response while simultaneously making a vocalization commonly associated with this response (e.g., "ɔ", "aah" like in "bar"). For tasks 3 and 4, performance rehearsals—or dry runs—ensured that the actors understood the display sequencing for each scenario. The exact wording of the scenarios used in Tasks 3 and 4 is available on request.

C. FACS annotation

One of the main contributions of this work is the provision of high quality frame-wise FACS annotations for the multi-view video data. Much effort has been spent to generate reliable annotations for all portrayal tasks as described in the following subsections.

Annotation of single AUs and AU-configurations. We recruited sixteen certified FACS coders to annotate AUs in the video recordings for tasks 1 and 2. To evaluate their performance, they were at first given only a subset of the recordings. Performance evaluation was based on coding speed and inter-coder reliability.

Only full-frontal recordings were annotated. For tasks 1 and 2, only the second, and if applicable, the high intensity recording was annotated. The intensity annotation was, however, based on the actual portrayal and was coded on a

TABLE I. AUs AND THEIR COMBINATIONS PERFORMED BY THE ACTORS WITH INTENSITIES MEDIUM (M) AND HIGH (H), CORRESPONDING TO C AND D-E IN FACS

TASK 1		
Single AUs		
AU	**FACS label**	**Intensity**
AU1	Inner Brow Raiser	M, H
AU2	Outer Brow Raiser	M, H
AU4	Brow Lowerer	M, H
AU5	Upper Lid Raiser	M, H
AU6	Cheek Raiser	H
AU7	Lid Tightener	M, H
AU9	Nose Wrinkler	M, H
AU10	Upper Lip Raiser	M, H
AU11	Nasolabial Furrow Deepener	H
AU12	Lip Corner Puller	M, H
AU13	Cheek Puffer	H
AU14	Dimpler	M, H
AU15	Lip corner Depressor	M, H
AU16	Lower Lip Depressor	M, H
AU17	Chin Raiser	M, H
AU18	Lip Puckerer	H
AU20	Lip Stretcher	M, H
AU22	Lip Funneler	H
AU23	Lip Tightener	H
AU24	Lip Pressor	M, H
AU25	Lips Part	H
AU26	Jaw Drop	H
AU27	Mouth Stretch	M, H
AU38	Nostril Dilator	H
AU43	Eyes Closed	H
AU45	Squint	H
AU46	Wink	H
AU Combinations		
AUs		**Intensity**
AU1+2		M, H
AU1+4		M, H
AU1+2+4		H
AU6+12		M, H
AU12+25		H

three-level scale. Annotations from four coders were excluded because they did not finish in the allotted time period. Excluding the four coders that were removed from our coder sample because they were too slow, average duration to code one minute of video was 2.61 hours per video minute. Two more coders dropped out for private reasons. Of the remaining ten coders, we selected the top six coders in terms of inter-coder reliability computed as Cronbach's Alpha. Their reliabilities ranged from 0.64 to 0.72 (average = 0.69). After this evaluation of coder performance, the remaining videos were distributed among these selected six FACS coders. Each of these videos was annotated by only one coder. For the videos that had been coded during the performance evaluation phase, the AUs provided by the coder with the highest inter-coder reliability for the respective video were chosen.

Coders received a basic payment of 15.00€ per coding hour plus an hourly bonus contingent on their coding experience, inter-coder reliability and speed (i.e. the number of hours it took them to complete the coding). On average, this amounted to an hourly payment of 17.44€.

Annotation of AUs in emotion appraisal scenarios. To annotate the recordings from tasks 3 and 4, we recruited fifteen certified FACS coders: five new coders, and ten from the previous coding task. We expected this video material to be more challenging to annotate than the material from tasks 1 and 2. In tasks 3 and 4, the actors responded to and enacted complex scenarios, rather than an instructed display of specific AUs. This would result in more AUs displayed and more elaborate combinations. Therefore, we conducted another performance evaluation based on a subset of Task 3 recordings. For that purpose, a subset of 40 recordings was evenly distributed among five groups of three coders. Performance evaluation was again based on coding speed and inter-coder agreement (Cronbach's Alpha).

All FACS coders finished the annotation within a reasonable time. It took the coders on average 4.00 hours to annotate one minute of video. We excluded three coders because their inter-coder reliabilities were below 0.60. One more coder dropped out for private reasons. The reliabilities of the remaining eleven coders ranged from 0.65 to 0.87 (average=0.75).

The remaining recordings from task 3, and all videos from task 4, were distributed among these eleven FACS coders. Now, each video was annotated by only one coder. For the videos that had been coded during the performance evaluation phase, we chose the AU coding provided by the coder with the highest inter-coder reliability for the respective video.

Coders received a basic payment of 15.00€ per coding hour plus a bonus contingent on their coding experience and inter-coder reliability. On average, this amounted to an hourly payment of 18.00€.

Coding instructions followed the FACS manual [1]. All 44 AUs were coded in a dynamic fashion. Coders were instructed to code the onset, apex and offset phase of each AU. The onset phase was defined as starting with the frame where the first appearance change associated with the AU is observed. The apex phase was defined as starting at the frame where all appearance changes have reached a plateau or peak where no further increase is noticed. The beginning of the offset phase was defined at the frame where the first evidence of a decrease in intensity is observed. The offset phase continues until disappearance of the AU or a new onset.

Additionally, the intensities of AUs were scored at apex. To increase reliability between coders, three levels of intensity were used instead of five [26]. These levels are A (small action, corresponding to *a* and *b* in FACS manual), B (moderate to strong action, corresponding to *c* in FACS

TABLE II. AU CONFIGURATIONS CORRESPONDING TO BASIC EMOTIONS PERFORMED BY THE ACTORS

TASK 2	
Prototypical AU Configurations Corresponding to Basic Emotions	
AUs	**Emotion Label**
AU4+5+23 or 24, if possible with AU7+17	Anger
AU9+10+16+19+26 or 25	Disgust
AU1+2+4, if possible with AU5+7+20+26	Fear
AU1+4+6, if possible with AU7+15+64	Sadness
AU6+12	Joy

manual), and C (estimated maximum action, corresponding to *d* and *e* in FACS manual).

IV. Comparison to other databases and performance benchmarks

Table III summarizes the commonly used FACS-coded facial expression databases that are available for research use. The last row of the table describes our new Actor Study Database. The Actor Study Database contains frames recorded synchronously from five different views. They are frame-aligned and annotated according to FACS. A total of 1.5 million FACS-annotated frames containing single AUs as well as AU combinations (see Table I) performed by 21 different actors are available. Most of the databases provide AU intensity annotations for only very few selected AUs (e.g. BP4D provides intensities of only two AUs). The Actor Study provides intensity values for 40 AUs on a three-level coding style (small, moderate and maximum action), which makes it the database with the most comprehensive FACS intensity annotation, to the best of our knowledge. The Actor Study is the first large scale corpus (completely annotated by certified facial action unit coders) of facial appraisal and emotion-related expressions (portrayals by professional actors) that has been constructed and validated entirely on the basis of theoretical predictions by the Component Process Model of Emotion (CPM) [27].

In order to provide researchers with a benchmark for all the five views, we used two different state-of-the-art systems. None of these systems has been developed by the authors nor trained or tuned with the Actor Study Database. Therefore, the following results give an indication of how they perform on previously unseen data. Our focus was not a direct comparison of the approaches.

A. Description of the benchmarking system

In this subsection, we describe the two state-of-the-art automatic AU recognition systems that were used to produce benchmark results on the new Actor Study Database.

OpenFace. The first system that we used for benchmarking is a facial behavior analysis toolkit named OpenFace [28, 29]. It is an open source software capable of facial landmark detection, head pose estimation, AU recognition (for 18 AUs), and eye-gaze estimation. It is not limited to frontal faces. For benchmarking, only the AU recognition output was taken into account.

OpenFace uses a combination of appearance and geometric features to detect AUs in single images or entire video sequences. In the first step, it detects the face in the provided image or current video frame. Then, it extracts geometric features and performs face alignment and masking. In the next step, appearance features are computed by using Histograms of Oriented Gradients (HOGs). For benchmarking, we provided the video files as inputs to OpenFace. It extracts frames from the input video and performs a calibration by estimating the person's expression at rest, before providing the AU detection results.

AU detection system from ISIR. The second system that we used for benchmarking is a state-of-the-art system for expression recognition and AU detection [30] from the Institute for Intelligent Systems and Robotics (ISIR http://www.isir.upmc.fr/) in Paris, France. This system was not yet public as of writing this paper. The executable is, however, available upon request from the authors of the journal article [30]. This system uses a random forest consisting of trees, each of which uses features extracted from

TABLE III. Summary of FACS-annotated facial expression databases

Database	Database Statistics	Database Description	Size	Posed / Spontaneous
Cohn-Kanade [14]	- 100 subjects - 69% female, 31% male (age 18-50 years) - Frontal and 30 degree imaging	- AU-coded - Single and combinations of AUs		Posed
Extended Cohn-Kanade (CK+) [15]	- 123 subjects - Extension to Cohn-Kanade	- AU-coded (Onset to peak) - Spontaneous smiles (66 subjects)		Posed and Spontaneous
MMI[16]	- 25 subjects - 48% female, 52% male (age 20-32)	- AU-coded (onset, apex, offset) - Single and combinations of AUs		Posed and Spontaneous
DISFA [17]	- 27 subjects - 44% female, 56% male	- AU-coded (12 AUs) - Intensity annotation	130,000 video frames	Spontaneous
Bosphorus [18]	- 105 subjects - 42% female, 58% male	AU-coded		Posed
UNBC-McMaster Pain Archive [19]	- 129 subjects - 51% female, 49% male	- AU-coded (only pain relevant AUs) - Intensity annotation	200 video sequences	Spontaneous
BP4D [20]	- 41 subjects - 56% female, 44% male	- AU-coded (27 AUs) - Intensity annotation (2 AUs)	368,036 video frames	Spontaneous
BP4D+ [21]	- 140 subjects - 59% female, 41% male	- AU-coded (34 AUs) - Intensity annotation (5 AUs)	1,400,000 videos frames	Spontaneous
GFT [22]	- 96 subjects - 42% female, 58% male	- AU-coded (20 AUs) - Intensity annotation (5 AUs)	172,800 video frames	Spontaneous
Actor Study (Current)	- 21 actors - 52% female, 48% male (age 26-68, mean: 42) - 5 views, 7 cameras (Top, bottom, center, 30° left, 30° right) including two high speed cameras	- AU-coded (40 AUs) - Intensity annotation - Single and combinations of AUs	- 777 video sequencies - 1,505,495 frames	Posed

TABLE IV. BENCHMARK RESULTS FOR CENTER VIEW

System		OpenFace	ISIR
Analyzed Frames		100233	98943
Area Under Curve	AU01	0.65	0.76
	AU02	0.60	0.83
	AU04	0.80	0.73
	AU05	0.58	0.83
	AU06	0.93	0.81
	AU07	0.85	-
	AU09	0.62	0.83
	AU10	0.88	-
	AU12	0.94	0.83
	AU14	0.78	-
	AU15	0.63	0.75
	AU17	0.70	0.70
	AU20	0.56	0.69
	AU23	0.53	-
	AU25	0.76	0.89
	AU26	0.73	0.88
	AU45	0.81	-

a randomly selected local facial region to predict prototypical expressions. The local predictions from the trees are then aggregated to obtain the global prototypical expression probability. The local predictions are used as inputs to another random forest to predict AUs. To achieve robustness against partial facial occlusion, an autoencoder network is used to estimate confidence measures, which are then used to weight the emotion predictions. The output provided by the system includes probability and confidence scores for 6 basic facial expressions, neutral expression, and 12 AUs. For the purpose of benchmarking, we used only the probability scores for the 12 AUs. Since the system was trained on frontal faces, we evaluated it only on the center view videos in the Actor Study Database.

B. Benchmark results

Benchmarking was done on videos recorded using the low speed cameras (24 frames per second), since it matches the commonly used frame rate in facial videos available in existing databases on which the benchmarking systems were trained. Additionally, the low speed camera videos cover all five views. Although the high speed camera recordings were excluded from benchmarking, they could be used for deeper research into facial dynamics and for developing systems that are sensitive to subtle facial motion.

Both systems used for benchmarking provide scores for the detected AUs. For comparison, we computed the Receiver Operating Characteristic (ROC) curve for each AU based on the output scores. The Area Under the ROC Curve (AUC) for each AU was used as the performance metric.

The benchmarking results for the center view videos are given in Table IV. The difference in the number of analyzed frames is mainly caused by a loss of the face during tracking in some videos. OpenFace was able to analyze 1290 more frames than the system from ISIR.

The mean AUC value for ISIR's system (0.79) is slightly higher than that achieved by OpenFace (0.73). One possible explanation could be that OpenFace recognizes more number of AUs, which increases the chances for inter-AU confusion. Noteworthy is the fact that in this evaluation the systems analyzed frames from an unseen database. So the benchmarking results indicate the generalization performance of the systems.

TABLE V. BENCHMARK RESULTS FOR THE FIVE VIEWS OBTAINED USING OPENFACE, LISTED IN THE ORDER: BOTTOM; CENTER; TOP; RIGHT; LEFT

AU	F1 Score	Area Under Curve
AU01	0.34;0.33;0.31;0.25;0.27	0.64;0.65;0.64;0.61;0.62
AU02	0.36;0.37;0.35;0.29;0.28	0.60;0.60;0.59;0.57;0.57
AU04	0.46;0.42;0.37;0.25;0.35	0.79;0.80;0.82;0.72;0.75
AU05	0.12;0.14;0.19;0.13;0.15	0.57;0.58;0.57;0.55;0.57
AU06	0.45;0.47;0.38;0.37;0.32	0.93;0.93;0.89;0.82;0.85
AU07	0.29;0.33;0.31;0.40;0.32	0.82;0.85;0.79;0.69;0.71
AU09	0.16;0.21;0.20;0.16;0.14	0.62;0.62;0.62;0.60;0.62
AU10	0.17;0.19;0.20;0.11;0.07	0.82;0.88;0.85;0.77;0.77
AU12	0.54;0.54;0.50;0.54;0.27	0.94;0.94;0.90;0.89;0.89
AU14	0.16;0.12;0.09;0.16;0.14	0.79;0.78;0.74;0.68;0.77
AU15	0.08;0.06;0.07;0.06;0.08	0.61;0.63;0.62;0.59;0.58
AU17	0.19;0.21;0.17;0.16;0.14	0.64;0.70;0.65;0.61;0.57
AU20	0.06;0.07;0.03;0.04;0.07	0.56;0.56;0.56;0.56;0.55
AU23	0.05;0.04;0.03;0.04;0.04	0.52;0.53;0.53;0.54;0.53
AU25	0.46;0.44;0.38;0.31;0.40	0.75;0.76;0.73;0.70;0.74
AU26	0.36;0.37;0.27;0.28;0.30	0.72;0.73;0.69;0.66;0.68
AU28	0.04;0.03;0.04;0.00;0.02	-
AU45	0.21;0.21;0.23;0.20;0.19	0.80;0.81;0.80;0.76;0.77

Since OpenFace detects AUs also in non-frontal faces, we computed AUC scores on the other four views. OpenFace also outputs a binary decision whether an AU occurs or not. Based on these binary predictions, we computed F1 scores. Table V lists the F1 and AUC scores obtained using OpenFace on all five views.

The results of both systems seem promising. However, the fact that automatic AU detection sometimes fails even on this high quality data, indicates that there are still challenges in AU detection. The authors hope that the new Actor Study Database will complement existing datasets to support improvements in AU detection. A few samples of the penta-view video recordings and corresponding AU annotations as well as a more detailed illustration of the camera positions and the recording setting are provided as supplementary material to this paper.

V. CONCLUSION

In this paper, we introduced our new Actor Study Database that contains high-quality, frame-synchronized, penta-view recordings of facial expressions performed by 21 actors. The recordings include expressions of 27 single AUs, five different combinations of AUs, AU configurations corresponding to five of the basic emotions, and responses to 21 different appraisal and emotional scenarios. The five views were recorded at 24 frames per second with cameras mounted at the center, top, bottom, 30° left, and 30° right positions. In addition, high-speed recordings of two views (center, 30° right) at 125 frames per second are also available. The frames have been annotated for 40 AUs by certified FACS coders, resulting in a total of over 1.5 million annotated and validated frames. Two state-of-the-art AU recognition systems were used for producing benchmark results on the five views. The recordings as well as the annotations are available for non-commercial research, especially for benchmarking new as well as existing approaches for automatic AU recognition, and for research into emotional phenomena and facial expression dynamics.

REFERENCES

[1] P. Ekman and W. V. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA, 1978.

[2] P. Ekman and D. Cordaro. What is meant by calling emotions basic. *Emotion Review*, 3(4):364–370, 2011.

[3] M. F. Valstar, M. Mehu, B. Jiang, M. Pantic, and K. Scherer. Meta-analysis of the first facial expression recognition challenge. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(4):966–979, 2012.

[4] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 223–230, 2006.

[5] P. Ekman, W. Friesen, and J. C. Hager. Emotion predictions (Table 10-1), p174. In *Facial Action Coding System - Investigator's Guide*. Salt Lake City: Research Nexus, 2002.

[6] S. Velusamy, H. Kannan, B. Anand, A. Sharma, and B. Navathe. A method to infer emotions from facial action units. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2028–2031, 2011.

[7] K. R. Scherer, M. Mortillaro, and M. Mehu. Facial expression is driven by appraisal and generates appraisal inference. In J.-M. Fernández-Dols and J. A. Russell, editors, *The science of facial expression*, pages 353–373. Oxford University Press, New York, NY, 2017.

[8] M. Mortillaro, B. Meuleman, and K. R. Scherer. Advocating a componential appraisal model to guide emotion recognition. *International Journal of Synthetic Emotions*, 3(1):18–32, 2012.

[9] N.-S. Pai and S.-P. Chang. An embedded system for real-time facial expression recognition based on the extension theory. *Computers & Mathematics with Applications*, 61(8):2101–2106, 2011. Advances in Nonlinear Dynamics.

[10] M. F. Valstar, T. Almaev, J. M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic, and J. F. Cohn. Fera 2015 – second facial expression recognition and analysis challenge. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 06, pages 1–8, 2015.

[11] M. F. Valstar, E. Sánchez-Lozano, J. F. Cohn, L. A. Jeni, J. M. Girard, Z. Zhang, L. Yin, and M. Pantic. Fera 2017 - addressing head pose in the third facial expression recognition and analysis challenge. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pages 839–847, 2017.

[12] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer. The first facial expression recognition and analysis challenge. In *Face and Gesture 2011*, pages 921–926, 2011.

[13] K. M. Prkachin. Assessing pain by facial expression: Facial expression as nexus. *Pain Research and Management*, 14(1):53–58, 2009.

[14] T. Kanade, J. F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, pages 46–53, 2000.

[15] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.

[16] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-Based Database for Facial Expression Analysis," *IEEE International Conference on Multimedia and Expo*, 2005.

[17] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "DISFA: A Spontaneous Facial Action Intensity Database," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 151–160, 2013.

[18] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus Database for 3D Face Analysis," *Lecture Notes in Computer Science Biometrics and Identity Management*, pp. 47–56, 2008.

[19] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews, "Painful data: The UNBC-McMaster shoulder pain expression archive database," *Face and Gesture 2011*, 2011.

[20] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, and P. Liu. A high-resolution spontaneous 3D dynamic facial expression database. *FG*, 2013.

[21] Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, J. F. Cohn, Q. Ji, and L. Yin. Multimodal spontaneous emotion corpus for human behavior analysis. In *CVPR*, pages 3438–3446, 2016.

[22] J. M. Girard, W. Chu, L. A. Jeni and J. F. Cohn, "Sayette Group Formation Task (GFT) Spontaneous Facial Expression Database," *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, Washington, DC, 2017, pp. 581-588.

[23] P. Ekman and W. V. Friesen. *Pictures of Facial Affect*. Consulting Psychologists Press, Palo Alto, CA, 1976.

[24] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998.

[25] K. R. Scherer and T. Bänziger. On the use of actor portrayals in research on emotional expression. In K. R. Scherer, T. Bänziger, and E. B. Roesch, editors, *Blueprint for affective computing: A sourcebook*, pages 166–178. Oxford University Press, Oxford, 2010.

[26] M. A. Sayette, J. F. Cohn, J. M. Wertz, M. A. Perrott, and D. J. Parrott. A psychometric evaluation of the facial action coding system for assessing spontaneous expression. *Journal of Nonverbal Behavior*, 25(3):167–185, 2001.

[27] Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York: Oxford University Press.

[28] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "OpenFace: An open source facial behavior analysis toolkit," *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.

[29] T. Baltrušaitis. (2017, Nov. 15). *OpenFace: an open source facial behavior analysis toolkit* [Online]. Available: https://github.com/TadasBaltrusaitis/OpenFace

[30] A. Dapogny, K. Bailly, and S. Dubuisson, "Confidence-Weighted Local Expression Predictions for Occlusion Handling in Expression Recognition and Action Unit Detection," *International Journal of Computer Vision*, Aug. 2017.