DISFA: A Spontaneous Facial Action Intensity Database

S.Mohammad Mavadati, *Student Member, IEEE*, Mohammad H. Mahoor, *Member, IEEE*, Kevin Bartlett, Philip Trinh, and Jeffrey F. Cohn

Abstract

Access to well-labeled recordings of facial expression is critical to progress in automated facial expression recognition. With few exceptions [1], publicly available databases are limited to posed facial behavior that can differ markedly in conformation, intensity, and timing from what occurs spontaneously. To meet the need for publicly available corpora of well-labeled video, we collected, ground-truthed, and prepared for distribution the Denver Intensity of Spontaneous Facial Action (DISFA) database. Twenty-seven young adults were video-recorded by a stereo camera while they viewed video clips intended to elicit spontaneous emotion expression. Each video frame was manually coded for presence, absence, and intensity of facial action units according to the Facial Action Unit Coding System [2]. Action units are the smallest visibly discriminable changes in facial action; they may occur individually and in combinations to comprise more molar facial expressions. To provide a baseline for use in future research, protocols and benchmarks for automated action unit intensity measurement are reported. Details are given for accessing the database for research in computer vision, machine learning, and affective and behavioral science.

Index Terms

FACS, action units, intensity, spontaneous facial behavior, facial expression, video corpus

1 Introduction

Perception and production of facial expression are central to nonverbal communication. These capacities emerge early in life and become more varied and nuanced with development [3], [4]. Motivated by basic science and a wide range of applications, researchers in computer vision and pattern recognition have become increasingly interested in developing algorithms for automatic recognition and synthesis of facial expressions in still images and videos [5], [6]. Potential applications of these efforts include human-computer interaction, social robots, consumer photography, automated pain detection, marketing, biometrics, and behavioral and neuroscience. Initial work in some of these application domains is emerging [7]–[10].

Critical to the success of automated facial expression recognition and synthesis is access to well-labeled video of facial behavior. With few exceptions [1], publicly available databases are limited to posed facial expressions. Posed expressions, however, can differ markedly in configuration, intensity, and timing from those that occur spontaneously. Because of the differences in underlying neural circuits, for instance, some facial actions occur rarely in posed behavior but are common in spontaneous facial behavior [11].

Spontaneous facial behavior is less intense, more difficult to annotate reliably, and has different complexity and timing from posed or deliberate facial behavior [12], [13]. In their authoritative review, [5] emphasized the need for publicly available databases of spontaneous facial behavior. Recent efforts to address this need include GEMEP [14] and the Belfast induced emotions database [16]. GEMEP consists of emotion portrayals by trained actors, which may be thought of as intermediate between posed and spontaneous, or naturally occurring. Facial behaviour was FACS coded and served as the basis for a recent grand challenge in automated facial expression analysis [15]. A few naturalistic databases with different modalities (e.g. physiological signal, such as electroencephalogram (EEG) and face video recording) have been presented for emotion recognition [1], [16], [17], and affective states analysis [18]. Spontaneous facial expression in the Belfast database is coded holistically (i.e. valence) rather than with anatomically based descriptors (i.e. FACS action units). Efforts in the direction of spontaneous facial expression recognition are just beginning [5].

Ekman and colleagues described two approaches to facial measurement: message-based and sign-based [12]. With message-based measurement, the implied meaning of the expression is labeled. Labels for ”basic” emotions are frequently used [19]; they include joy, surprise, anger, fear, disgust, and sadness, and more recently embarrassment

- *S. M. Mavadati, M. H. Mahoor, K. Bartlett, and P. Trinh are with the Department of Electrical and Computer Engineering, University of Denver, Co.*
- *J. F. Cohn is with the Department of Psychology, University of Pittsburgh and Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.*
  E-mails: (smavadat, mmahoor, ptrinh)@du.edu; kbartlett3@msn.com; jeffcohn@pitt.edu.

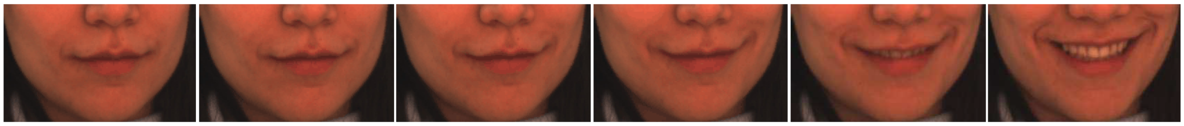

Fig. 1: Sample facial images with AU intensity variations (AU12: lip corner puller, AU25: lips part).

and contempt. Because these expressions occur relatively infrequently in their prototypic forms and may mistakenly imply emotions or cognitive-emotional states that are absent [20], more descriptive approaches have been proposed. These are what Ekman refers to as sign-based approaches in which the configuration or timing of facial actions is described without reference to the action's presumed meaning. The most well-known and widely used sign-based approach is the Facial Action Coding System (FACS) [2], which was initially developed in the 1970s [21] and was informed by earlier research by [22]–[24]. FACS action units are anatomically based facial actions. Each is based on one or at most a few facial muscles and may occur individually or in combinations. Action units (AU) provide a terminology for describing nearly all possible facial actions. For example, AU12 (lip corner puller) codes contractions of the zygomatic major muscle and AU25 contractions of the Orbicularis Oris muscle (lips part) [21]. FACS is widely used in behavioral science to investigate emotion, pain, psychopathology, and related topics [25]. By using FACS, expert coders can determine onset (start or beginning), peak, and offset (stop or end), as well as graded changes in intensity of AUs [2].

FACS assigns the letters A through E to represent intensity variation from barely detectable or trace (A) to maximum (E). In our usage, we substitute the descriptors 1-5 to represent the corresponding range of variation (A-E) and 0 when the AU fails to occur. Figure 1 shows examples of intensity variation in two action units, AU12 and AU25 [26].

To annotate facial action units, a human expert manually labels every video frame, which is laborious and time consuming. An hour or more may be required to label the onset and offset of each AU for every minute of video [25]. To label graded changes in intensity, additional time would be required. Although manually annotating facial action units is standard, it is not ideal, especially when too few manual FACS coders are available for the amount of video to be coded or when instantaneous FACS coding is needed. Automated measurement is essential to the feasibility of real-time applications, such as tutoring systems [27], [28].

To meet the need for automated measurement, a number of approaches have been proposed [4], [24], [28]. Several are sufficiently advanced that they have been used offline either alone or in combination with manual FACS coding in behavioural research. Examples include distinguishing between posed and spontaneous facial expressions (i.e. deliberately posed and spontaneously occurring smiles) [29], and categorizing pain-related facial expressions [30], [31]. We refer readers to [5], [20], [25], [27], [32], [33] for more details.

Almost all research into automated measurement has been directed at binary AU detection. That is, whether actions are present or absent. Graded changes in intensity have rarely been considered. Yet, action units can and do vary continuously in intensity, and the temporal envelope of intensity variation is critical to their meaning and communicative functions. Interpersonal synchrony, for instance, is manifest in the coordination of cycles of increasing and decreasing positive and negative affect over time [34]. Measurement of graded changes in intensity has so far received relatively little attention. Exceptions include [35], [37], [38], [39] .

Automatic spontaneous facial expression analysis faces several challenges [40]. Among these is the lack of access to well-annotated, publicly available databases of spontaneous facial expression. Most available databases include only posed facial expressions and message-based descriptors (i.e. emotion labels rather than more informative sign-based labels). Even when spontaneous facial expressions have been included, they are a small subset of the databases. See, for instance, Cohn-Kanade [41], MMI [42], etc.

With few exceptions (e.g. the UNBC shoulder-expression archive [43]), no publicly available database includes full intensity FACS coding. At most, only peak AU intensity is coded. To train and test classifiers to detect AU intensity variation, ample train and test video with full intensity coding are essential. To meet the need for well-annotated video of spontaneous facial expression, we created the **Denver Intensity of Spontaneous Facial Action (DISFA)** database[1]. DISFA contains approximately 130,000 annotated frames from 27 adult participants. For every video frame, the intensity of 12 action units was manually annotated on a six-point ordinal scale (0 and five increasing levels of intensity) using the continuous measurement system (CMS) [44]. The AUs chosen were among those most common in emotion expression and social interaction and that have been studied previously in computer vision and machine learning [45]. In addition to introducing DISFA, we report benchmark metrics for automatic AU intensity measurement in this new database.

1. Instructions for obtaining DISFA can be found in [26].

TABLE 1: Description of facial expression databases.

Database	Database information	Expression description	Posed vs Spontaneous
Cohn-Kanade [41]	- 100 Subjects (multi-ethnicity) - 69% female, 31% male (age 18-50 yrs) - Frontal and 30 degree imaging	- AU-coded face database - 23 series of facial display - Single and combinations of AUs - Available for *non-commercial use*	- Posed
Extended Cohn-Kanade(CK+) [48]	- Extension of Cohn-Kanade - 123 Subjects (multi-ethnicity)	- Onset to peak coded - Spontaneous smiles (66 subjects)	- Posed - Spontaneous
MMI [42]	- 25 subjects (multi-ethnicity) - 12 female , 13 male (age 20-32 yrs)	- Single and combinations of AUs - Temporal analysis (e.g. onset,apex,offset) - Available for *scientific applications*	- Posed & Spontaneous - Spontaneous, six basic - Expressions added later
JAFFE [49]	- 10 female Japanese models - Grayscale images	- Neutral+6 Basic expressions - 2 to 4 samples per expression - Available for *non-commercial research*	- Posed
Bosphorus [51]	- 105 subjects - 44 female , 61 male	- AU-coded (2D/3D data) - Pose and illumination variations - Available for *scientific applications*	- Posed
BU-3DFE [52]	- 100 subjects (multi-ethnicity) - 56 female , 44 male (18 to 70 yrs) - 2,500 3D facial expression models	- Neutral+ 6 basic expressions (4 intensity) - Available for *Research applications*	- Posed
RU-FACS [53]	- 100 subjects	- AU-FACS coded (33 AUs) - *Private database*	- Posed & Spontaneous
NVIE [54]	- 215 Student (age 17-31 yrs) - Visible and infrared imaging	- Basic facial expressions - Temporal analysis for posed data - Available for *scientific applications*	- Posed & Spontaneous
UNBC-McMaster Pain Archive [43]	- 129 Participants (63 male,66 female) - 200 video sequence	- Pain related AUs coded - Available for *non-commercial use*	- Spontaneous
Belfast Induced Emotion [16]	- Lab-based emotion induction tasks - Three sets of tasks	- Emotion and Intensity (self-report) - Intensity and Valence (trace-style rating)	- Natural emotion
DISFA [26]	- 27 Participants (15 male,12 female) - 130,000 video frames	- Intensity of 12 AUs coded - Available for *non-commercial use*	- Spontaneous

The remainder of this paper is organized as follows: Section 2 reviews facial expression coding systems and existing face databases. Section 3 describes observational procedures, imaging, and other operational details of DISFA. Section 4 presents our approach to automatic AU intensity estimation and relevant performance metrics. Section 5 presents results for automated AU intensity measurement. Section 6 discusses challenges and future directions in facial expression analysis and considers potential applications of the DISFA as a testbed for evaluating alternative algorithms. Section 7 concludes the paper.

2 Background

2.1 Facial Coding System

Following seminal efforts by Darwin [23], Duchenne [24] and Hjortsjo [22], Ekman and Friesen [21] developed the Facial Action Coding System (FACS) to describe nearly all possible facial actions. With the exception of visual speech (i.e. visemes), relatively few facial expressions cannot be decomposed into one or more anatomically based FACS action units. No other system has such descriptive power [25]. In part for this reason, FACS has become the standard for anatomically-based description of facial expression and muscle-based facial animation (e.g. MPEG-4 facial animation parameters [46]).

The 1978 edition of FACS described facial expressions in terms of 44 anatomically based action units and additional action descriptors [21]. Action descriptors are actions for which the anatomical basis is either unknown or not specified. A 2002 revision reduced the number of AU to slightly more than 30, revised the scoring criteria, and improved usability. The coding of intensity was expanded to more of the AU and the ordinal scaling of intensity was increased from 3 levels (X, Y, and Z) to 5 (A through E) [2]. Cohn et al. [47] describe the changes made in the 2002 revision of FACS.

2.2 Previous Databases

For automatic action unit detection, a number of FACS-coded databases have become available for research. The most widely used are Cohn-Kanade [41] and MMI [42]. Cohn-Kanade contains images of 100 subjects that performed 23 facial displays. The image sequences consist of posed facial expressions that include single AU and combinations of AUs in which peak intensity is labeled for each AU. A recent extension of Cohn-Kanade [48] adds additional subjects and spontaneous facial actions. MMI contains over 1500 samples of both static images and image sequences (in frontal and profile views). MMI includes labels for onset (start frame), apex (maximum intensity), and offset

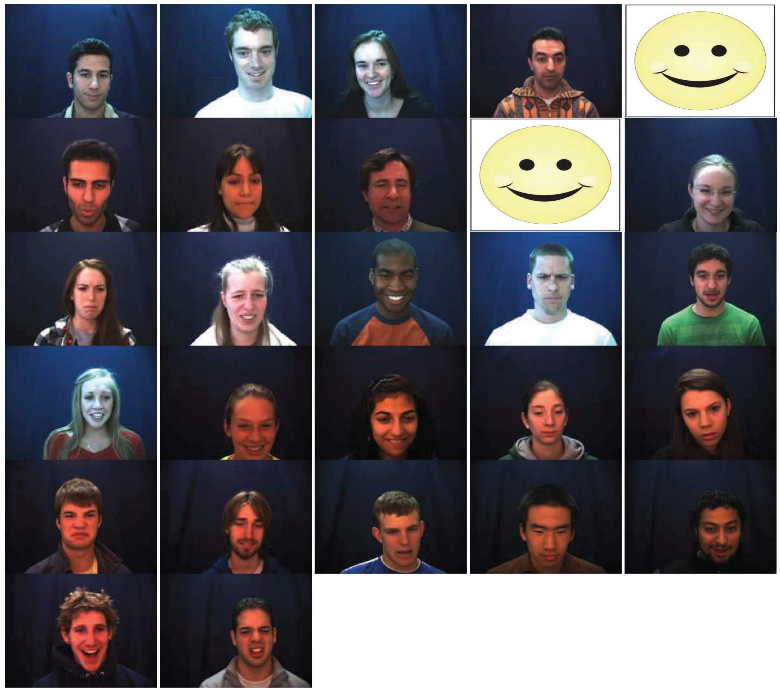

Fig. 2: Facial images of 25 of 27 participants. Two participants did not consent to use of their images in publications.

(end frame) of each AU but not intensity labels for each frame [42]. MMI also includes some spontaneous facial expressions.

Among other FACS-coded databases are RU-FACS (aka M3) [53] and the UNBC Pain Archive [43]. RU-FACS, which has restricted availability, contains video of spontaneous behavior of 100 subjects that were observed during 2.5 minute interviews [53]. The Bosphorus dataset contains 2D and 3D images of 105 subjects which is coded using FACS and can be used for posed facial expression recognition and facial action unit detection [51]. In [35], Fasel and Luettin utilized a posed database of 9 single AUs and 7 combinations of AUs. The database was FACS coded by Kaiser and Wehrle [36]. The UNBC-McMaster Shoulder Pain Expression Archive [43] provides facial expressions associated with pain. This database includes both FACS and pain intensity coding. To provide a common ground for detecting action unit intensity of spontaneous facial expression, we created DISFA [26]. Table 1 summarizes FACS-coded facial expression databases. Also included are ones that use holistic descriptors, such as emotion-specified expressions and valence.

3 DISFA

This section describes the acquisition setup for eliciting and imaging spontaneous facial behavior for DISFA, a description of the video, and metadata. Metadata are FACS AU and facial landmarks obtained using AAM.

3.1 Participants

Participants were 27 adults (12 women and 15 men) that vary in age from 18 to 50 years. Three were Asian, 21 Euro-American, two Hispanic, and one African-American. All participants gave informed consent for the distribution and use of their video images for non-commercial research. Twenty-five of the 27 gave permission for use of their images in publications (Figure 2).

3.2 Stimulus and Recording Procedure

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.

Fig. 3: DISFA imaging setup.

3.3 Manual FACS Annotation

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) (see Table 2 for more details).

TABLE 2: The number of AU events (EV) and total number of frames for each intensity level.

	AU Info		# Frames for each AU Intensity					
	AU#	**#EV**	**0**	**1**	**2**	**3**	**4**	**5**
Upper-Face	1	161	122036	2272	1749	2809	1393	555
	2	111	123450	1720	934	3505	836	369
	4	238	106220	4661	7636	6586	4328	1383
	5	100	128085	1579	719	293	104	34
	6	170	111330	9157	5986	3599	601	141
Lower-Face	9	73	123682	1659	2035	3045	316	77
	12	250	100020	13943	6869	7233	2577	172
	15	97	122952	5180	1618	1017	47	0
	17	271	117884	6342	4184	2281	112	11
	20	99	126282	1591	1608	1305	28	0
	25	296	84762	9805	13935	15693	5580	1039
	26	321	105838	13443	7473	3529	314	217

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. Inter-observer 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].

TABLE 3: AU description and inter-observer reliability.

AU	**Description**	**Reliability (ICC)**
1	Inner Brow Raiser	0.83
2	Outer Brow Raiser	0.86
4	Brow Lowerer	0.94
5	Upper Lid Raiser	0.81
6	Cheek Raiser	0.80
9	Nose Wrinkler	0.85
12	Lip Corner Puller	0.90
15	Lip Corner Depressor	0.82
17	Chin Raiser	0.81
20	Lip Stretcher	0.83
25	Lips Part	0.92
26	Jaw Drop	0.93

3.4 AAM Facial Landmark Points

Most techniques for facial expressions recognition and AU detection require a set of landmark points marked in facial images. These points may be used for facial image registration, head pose estimation and compensation, and high-level feature extraction. To detect and extract landmark points, we used active appearance models (AAM) [56]. We extracted 66 landmark points in each image. After finding the 66 landmarks, a similarity transformation was used to map facial landmark points to a canonical orientation. Using the transformed landmark points, a binary mask then was fitted to the image to remove non-face background (see Figure 4). The cropped facial regions were resized into 108×128 pixels.

(a) (b)

Fig. 4: (a) 66 Landmark points, (b) Normalized masked image.

4 Automatic Action Unit Analysis System

This section describes a benchmark for measuring the intensity of action units on a 0-5 scale.

4.1 Feature Extraction

Three sets of appearance features were extracted from the cropped and aligned images. These are local binary pattern histogram, histogram of oriented gradient, and localized Gabor features, as described below.

4.1.1 Local Binary Pattern Histogram

The local binary pattern (LBP) codifies local texture information of an image. It is a well-known technique for object representation and classification and is widely used for biometrics [58] and facial expression recognition [59]. An important property is that it is robust to variation in illumination [57]. LBP histogram (LBPH) is the histogram of LBP labels computed over a region as a texture descriptor. With LBPH, there is an inherent trade-off between amount of retained shape information and the dimensionality of the data.

To calculate the LBPH, we defined a rectangular region (i.e. 18×16 pixels) around every landmark point and then used a uniform LBPH feature with 59 bins [59]. By concatenating the LBPH features for each region, we obtained a feature vector of length 3894.

4.1.2 Histogram of Oriented Gradients

The histogram of oriented gradient (HOG) descriptor counts the occurrences of gradient orientations in a localized portion of an image. This technique was introduced by Dalal and Triggs for the application of human detection [60], and since then has attracted attention in the face analysis field [61]. HOG represents both appearance and shape information, which makes it suitable for representing different facial expressions. To represent spatial information using this technique, images are divided into small cells; for each cell, the histogram of gradients is calculated [60].

We defined a cell of size 18×16 pixels around every landmark point, for a total of 66 cells for each image. We also applied the horizontal gradient filter $[-1\ 0\ 1]$ with 59 orientation bins [60]. In order to form the HOG feature vector, the HOG representation for all the cells were stacked together. This resulted in a HOG feature vector of size 3894.

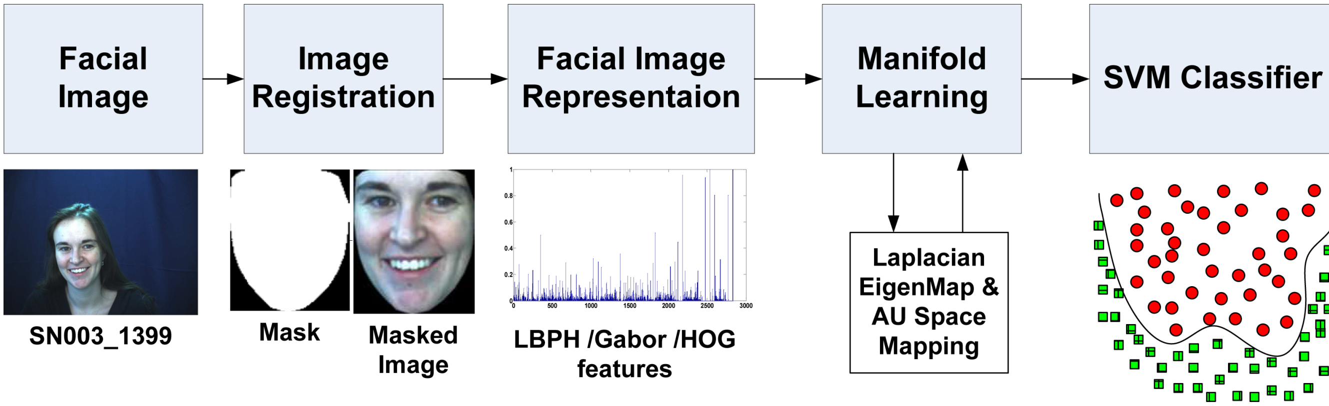

Fig. 5: Benchmark system for automatic AU intensity measurement.

4.1.3 Localized Gabor filters

Gabor filters are another widely-used technique for representing facial textures in face and expression recognition [62], [63]. To extract a global appearance of a face, Gabor filters are applied to the whole face. We extracted local features by applying Gabor filters to specific regions of the face (e.g. around facial landmark points). We applied 40 Gabor filters (5 scales and 8 orientations) around each of 66 landmark points, which resulted in 2640 Gabor features.

4.2 Dimensionality Reduction

To reduce the high dimensionality of the features, we used nonlinear *manifold learning*. This approach assumes that the features lie in a low-dimensional manifold embedded into a high dimensional space. Mathematically speaking, given a set of points $x_1, ..., x_n \in R^D$ find a set of points $y_1, ..., y_n \in R^d$ ($d << D$), such that y_i represents x_i efficiently [64].

The *Laplacian Eigenmap* [65] followed by spectral regression (SR) technique was used to extract low-dimensional features of facial images. We randomly selected a small number of images from all subjects to learn a separate manifold for each action unit. Similar to our previous work [38], we applied the spectral regression algorithm to calculate the projection function for each action unit, and then used it to transform high-dimensional LBPH, HOG, and Gabor features into separate low dimension spaces for building a more efficient classifier.

4.3 Classification

The support vector machine is a classification approach that has gained popularity for pattern classification in the last decade. Relative to some other approaches (e.g. discriminant analysis and logistic regression), it can be applied to both linear and non-linear classification and is computationally relatively efficient. The SVM classifier applies the kernel trick, which uses dot product, to keep computational loads reasonable. The kernel functions enable the SVM algorithm to fit a hyperplane with a maximum margin into the transformed high dimensional feature space. Several kernel functions have been proposed. Some of the most popular kernels are *Linear* kernel, *polynomial* kernel, *Gaussian radial basis function* (RBF) kernel, and *Sigmoid* kernel. SVM is originally designed for binary classification however it can be applied for multi-class SVM too. To classify C classes of data, two alternative approaches are used. One is *one-against-rest* strategy in which $C-1$ different classifiers are defined to determine whether a data point belongs to class C_i or to other remaining classes $C_{j\neq i}$. Another strategy is *one-against-one* in which $C(C-1)/2$ binary discriminant functions are defined, one for every possible pair of classes. More details about SVM can be found in [66].

We used a multi-class SVM classifier for AU intensity measurement. We examined three different kernels (i.e. Linear, Polynomial, and Gaussian RBF). The Gaussian RBF kernel outperformed the other two kernels. To train SVM classifiers, the one-against-one strategy was utilized (i.e. 15 classifiers were trained for six-intensity levels of each AU). Test samples were assigned to the class with the highest probability. The schematic framework for AU measurement is shown in Figure 5.

4.4 Concurrent Validity of the System

Any measurement is associated with some error. Because error can be variously defined, several reliability indices have been proposed. Widely used in automated facial expression detection and behavioural science are F1-score

TABLE 4: AU intensity measurement results; (a)LBPH, (b)HOG, (c)Gabor features.

(a) LBPH features

		AU1	AU2	AU4	AU5	AU6	AU9	AU12	AU15	AU17	AU20	AU25	AU26	AVG.
ICC		0.75	0.74	0.83	0.50	0.73	0.65	0.83	0.60	0.53	0.49	0.91	0.72	**0.69**
Accuracy(%)		83.22	85.34	78.31	90.79	78.84	89.67	72.94	88.97	74.04	88.98	72.73	74.67	**81.54**
6-Level Intensity Accuracy (%)	’0’	85.82	87.31	84.33	91.80	83.98	92.16	77.83	91.69	77.40	90.56	78.41	80.45	**85.14**
	’1’	33.08	27.73	39.15	34.78	43.83	18.93	53.95	39.62	42.94	33.02	53.99	46.71	**38.98**
	’2’	58.50	67.03	47.27	54.19	53.14	53.18	63.95	58.40	51.45	70.16	59.28	55.46	**57.67**
	’3’	48.67	55.71	53.45	54.30	65.72	58.83	52.11	68.13	37.07	28.75	64.22	52.77	**53.31**
	’4’	52.26	47.34	70.35	74.75	53.24	63.84	79.42	80.00	1.79	32.14	74.79	44.40	**56.11**
	’5’	49.73	90.81	56.74	91.18	0.00	39.66	12.79	-	0.00	-	87.23	62.79	**49.09**

(b) HOG features

		AU1	AU2	AU4	AU5	AU6	AU9	AU12	AU15	AU17	AU20	AU25	AU26	AVG.
ICC		0.75	0.74	0.84	0.58	0.77	0.74	0.75	0.65	0.57	0.46	0.79	0.69	**0.70**
Accuracy(%)		80.24	83.92	75.31	92.85	80.70	91.97	65.55	87.58	72.56	84.91	60.98	73.28	**79.14**
6-Level Intensity Accuracy (%)	’0’	82.23	85.36	79.65	93.78	86.29	93.86	71.91	89.48	75.06	85.99	66.89	77.41	**82.33**
	’1’	34.63	26.25	37.67	39.43	43.99	21.54	40.93	57.98	40.12	43.16	38.80	50.57	**39.59**
	’2’	63.28	42.56	54.49	71.68	45.01	58.62	40.76	64.52	56.72	60.74	50.93	63.31	**56.05**
	’3’	50.96	83.07	56.00	54.64	76.12	73.80	52.72	52.79	69.46	60.88	58.75	58.82	**62.33**
	’4’	59.38	64.52	71.82	47.47	51.75	64.17	65.80	22.22	6.25	46.43	44.08	61.19	**50.42**
	’5’	83.24	36.77	81.81	82.35	0.00	27.59	0.00	-	0.00	-	51.67	73.26	**43.67**

(c) Gabor features

		AU1	AU2	AU4	AU5	AU6	AU9	AU12	AU15	AU17	AU20	AU25	AU26	AVG.
ICC		0.80	0.83	0.87	0.58	0.81	0.80	0.84	0.71	0.69	0.54	0.94	0.79	**0.77**
Accuracy(%)		86.24	91.28	80.94	94.46	83.98	92.32	79.67	91.35	80.66	88.12	79.85	79.61	**85.71**
6-Level Intensity Accuracy (%)	’0’	88.06	92.45	84.36	95.49	87.69	93.61	84.55	93.12	82.71	89.02	82.11	82.52	**87.98**
	’1’	52.55	61.83	57.35	36.20	58.16	75.06	63.65	62.81	56.63	74.95	70.11	63.11	**61.03**
	’2’	77.74	66.26	69.10	66.03	65.69	64.62	68.97	59.83	69.93	55.53	76.25	68.03	**67.33**
	’3’	61.92	86.20	70.93	54.64	77.83	74.39	64.09	82.07	65.71	59.57	73.81	79.08	**70.85**
	’4’	64.49	56.98	64.82	53.54	59.57	60.26	59.72	0.00	46.43	7.14	85.97	85.07	**53.65**
	’5’	31.17	28.13	59.58	55.88	0.00	15.52	0.00	-	0.00	-	96.20	93.60	**38.01**

[67], Cohen’s-Kappa [68], and intraclass correlation (ICC) [69]. F1 and Kappa are appropriate for nominal data (e.g. presence/absence of AU) and ICC for ordinal or interval-level data (e.g. AU intensity).

ICC is used to calculate the reliability of judgments in multi-class recognition applications. It thus may be used to quantify reliability for AU intensity, which is measured on a 0 to 5 scale. ICC has several definitions that may yield different results for the same data. Selecting the best ICC technique is not possible without knowing the judgment model of the problem. For instance, to model n instances by k judges, two possible models are: 1) each target is rated by k judges, each of which is randomly selected from a larger set of judges, and 2) each target is rated by the same k judges of interest. [69] discusses these and other measurement models.

In our case, the second model applies (equivalent to ICC(3, 1) in [69]). There were k judges and n targets, and by defining the within-target sum square (WSS), between raters sum squares (RSS), between-target sum squares (BSS) and residual sum of squares ($ESS = WSS - RSS$) the ICC can be calculated by:

$$ICC = \frac{BMS - EMS}{BMS + (k-1)EMS} \tag{1}$$

where $BMS = \frac{BSS}{n-1}$ is between-targets mean square and $EMS = \frac{ESS}{(k-1)\times(n-1)}$ is the residual mean squares [69]. ICC is a powerful reliability measure for multi-class classification problems because it takes into account the difference between large and small errors between judges (e.g. predicting label ’5’ for a true label ’4’ has higher ICC than prediction ’1’ for the true label ’4’).

5 Experimental Results

We detected the intensity of 12 AUs on a 0-5 scale using three facial representations, LBPH, HOG, and Gabor. We used a leave-one-subject-out (LOSO) cross validation technique in which a subset of the videos of 26 subjects was used for SVM training (i.e. about 3000 randomly selected video frames). Testing was performed on the left-out subject. This scenario was repeated for every subject and finally the accuracy and ICC are reported. Accuracy was defined as an average of correctly detected intensity levels for each AU. In our experiment we utilized LIBSVM library with parameters $[-s\ 0, -t\ 2, -c\ 32, -g\ 0.1]$ (refer to [70] for definition of LIBSVM parameters).

As shown in Table 4, LBPH and HOG features led to accuracy less that 82% and ICC value about 0.70. Overall, the Gabor feature had 86% accuracy and 0.77 ICC value, which yielded the best result among the three facial representations. Table 4 shows the accuracy for each of the 6 intensity levels.

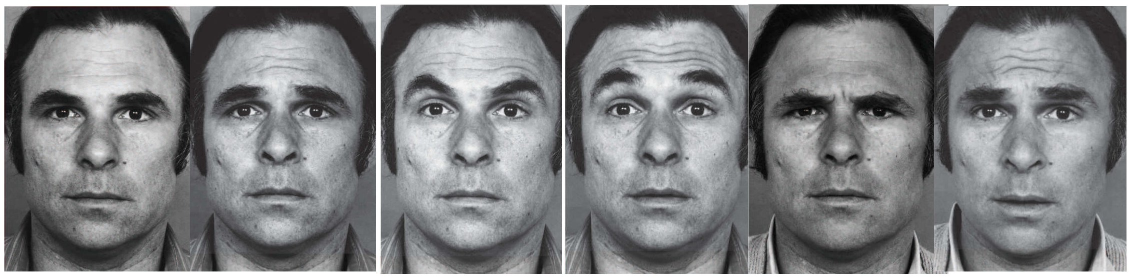

Fig. 6: AU based facial appearance variations [2].

Comparing Tables 2 and 4 suggests the importance of having sufficient samples for each intensity level. Detection metrics are lowest for those intensity levels that occur least frequently. For instance, AUs, such as AU12 and AU25, have large number of samples for all six levels of intensity, and performance metrics are generally high (e.g. $ICC > 0.84$). AU20, on the other hand, has relatively sparse rates of occurrence for some intensity levels. Performance metrics for AU20 are lower than those for AU12 and AU25. Because all three involve similar amounts of change in appearance, it is likely that these differences are due to the uneven representation of AU intensity for AU20 (i.e. $ICC \approx 0.55$).

6 Research Problems and Outlook

Spontaneous facial expression recognition is an important research area that has recently grasped the attention of the research community. As mentioned earlier, the lack of manually coded spontaneous facial behavior database has hampered research in automated facial expression recognition. This paper introduces the spontaneous facial expression database, DISFA, which can be applied to study the dynamics, intensity, and spatiotemporal structure of spontaneous FACS action units.

Besides the data collection, measurement and coding of spontaneous facial expressions and FACS action units are challenging compared to posed expressions. The challenges can generally be divided into two categories: (a) *Extrinsic challenges* that are caused by external sources, such as image modality and acquisition setup, viewing geometry, and illumination condition. (b) *Intrinsic challenges* that are independent of any external sources and are due to the physical nature of spontaneous facial expressions. They include the co-occurrence of multiple AUs, eye movement and blinking, head movement, subtle expressions, speaking, appearance variations across subjects and AUs and social dynamics. These challenges apply to both manual and automated measurement of action unit intensity. In the following, we discuss some of the intrinsic challenges we encountered in our study along with suggestions for future research.

Co-occurring action units may or may not present a challenge. Eye movement and blinking make the face annotation harder when they co-occur with the AUs that describe the eyes actions. For instance, eye movement can influence the detection and intensity measurement of AU5. In other words, in AU5 with maximum intensity (i.e. AU5(E)), the upper eyelid is pulled up maximally, revealing a maximum amount of sclera above the iris. In this case an eye movement can influence the upper lids muscles and AU5.

When AUs are in different regions or otherwise do not influence each others appearance, then measurement is straightforward [50]. An example of a non-additive combination is shown in Figure 6. The AU do not influence each others appearance (or the influence is the same direction), and thus detection of one is not complicated by the occurrence of the other. Alternatively, co-occurrence can create a distinctively new appearance, which is referred to as a non-additive combination. This is one of the most difficult situations for detecting and measuring spontaneous facial expressions. As illustrated in Figure 6, in the AU1+2+4 combination, the lifting of the eyebrows due to the AU1+2 is opposed by AU 4, which pulls the brows down and toward each other medially [2]. With DISFA, one could evaluate AU detection as a function of types of co-occurrence and train new classifiers accordingly. For instance, it might be advantageous to train separate classifiers for non-additive combinations.

Another issue arises from the fact that different types of features may be more or less revealing of different AU. Some AU create bulges or pouches (e.g. AU14, AU15, AU24); others create lines, wrinkles, or furrows (e.g. AU12, AU9). And others reveal new features (e.g., teeth in AU 25) or occlude existing features (e.g., closed eyes in AU 43 or 45). Some AU may produce multiple changes. Obviously, a single facial feature representation cannot describe all these appearance changes appropriately. In our work, we utilized facial image representations, like LBPH, HOG and Gabor features. Discovering a set of features that has both powerful representing and discriminating capabilities can be another interesting research topic in the field of spontaneous facial expression recognition. With DISFA, one could identify optimal features for each AU, which could increase efficiency of AU measurement.

Facial representation techniques typically entail high-dimensional feature vectors. It is well-known in machine learning and pattern recognition that the number of samples required for training increases exponentially with the number of dimensions (i.e. the curse of dimensionality problem). In practice, linear (e.g. PCA and LDA) and non-linear techniques, such as manifold learning have been developed for reducing the dimensionality of features. We utilized the Laplacian Eigenmap manifold learning technique combined with the spectral regression to map the high dimensional facial features into a lower dimensional space. However, finding the best number of features and the most efficient and robust dimensionality reduction algorithm is a subject of future research that DISFA would enable.

In addition to feature extraction and dimensionality reduction algorithms, building a robust classifier for detecting the action units and measuring their intensity levels is another important area of research. We utilized the support vector machines with Gaussian RBF kernel among others for AU intensity measurement. Although, our results are quite promising, better and more reliable classifiers need to be developed.

Traditionally 2D images are used for facial expression recognition. Common problems such as head pose variation and self-occlusion can be handled using 3D images (i.e. depth information). Currently there are a few databases available for 3D facial expression recognition (e.g. BU-3DFE face [52] and Bosphorus database [51]). However, these databases are not intensity FACS coded. In recording the DISFA videos, we used stereo imaging system (Bumblebee2). Both left and right video recordings are released and can be used for multi-view and 3D facial expression recognition. DISFA does not provide depth information and we would encourage interested readers to extract depth map of these facial images and provide them to the community for 3D spontaneous AU measurement.

In DISFA pilot coding 12 AUs were coded. Other researchers may harvest additional AUs. We would encourage that effort and ask that they vet their coding and make it available to the community. The frame rate in DISFA was set to 20 fps. The frame rate was chosen to accommodate the hardware synchronized acquisition with resolution of 1024×768 pixel for both left and right images. While this necessitated a slower frame rate than is usual, motion artefact or discontinuity was absent. This frame rate may be considers as a not sufficient frame rate for analyzing the very subtle facial expression analysis such as micro expression analysis and one might need to obtain videos with higher frame rates for those types of applications. Last but not least, we realized that some action units occurred less frequently than others (see Table 2). Hence, if there are not enough number of samples to train a classifier then we will have a less reliable system. In this regard, designing a system that can be trained with a small number of training samples is desirable.

7 Conclusion

Data drives research. Without adequate data for training and testing classifiers, progress is difficult at best. Most previous databases for automated facial expression analysis are limited to posed facial expressions, holistic labels (e.g. emotion labels), or continuous ratings along a molar dimension (e.g. valence). We present a new, publicly available database of spontaneous facial behaviour in response to emotion eliciting videos. Facial behaviour is exhaustively annotated using anatomically based descriptions, FACS action units. The intensity of each action unit is annotated on a 6-point intensity scale. DISFA thus provides continuous annotation of graded changes in spontaneous facial expression of emotion. For investigators that want a continuous representation along a single dimension, the action unit intensities may be combined into scaled scores for affective valence (e.g. Messinger et al., [34]).

To promote research in automated facial expression analysis, DISFA includes 66-facial landmarks for each video image and three appearance representations, LBPH, HOG, and Gabor coefficients. For each representation, we report benchmarks performance for SVM classifiers for AU intensity on an ordinal scale. We believe that this new facial database will be a great asset for the science community and will help researchers to develop and evaluate novel methods for spontaneous facial expression recognition.

Appendix

DISFA Video Stimulus

In pilot testing, we identified 9 YouTube videos that elicited AUs related to one or more emotions and facial action units of interest. Facial actions of interest were those common in emotion and social communication and that have been a focus of on-going research. Mindful of time constraints, we selected the most potent segment from each video for inclusion. The 9 segments were concatenated to create the video stimulus clip, which included a one-second pause before each segment. Table 5 presents the URL for each clip, its begin and end time, the target emotions, and the percentage of time that each emotion have been occurred. Table 6 specifies the emotion description based on [2].

TABLE 5: Description of 9 segments of DISFA stimulus.

Seg. #	Link	Start	Stop	Target Emotion	% of time target emotion occurred
Seg. 1	http://www.youtube.com/watch?v=qXo3NFqkaRM	00:09	00:20	- Happy	67.3%
Seg. 2	http://www.youtube.com/watch?v=wwAbtizFCzo	01:00	01:22	- Happy	52.8 %
Seg. 3	http://www.youtube.com/watch?v=1eH9YeZjGIQ&feature=youtu.be	00:19	00:35	- Surprise	9.2%
Seg. 4	http://www.youtube.com/watch?v=AYsq4fqbMG8	00:48	00:55	- Fear	6.3%
Seg. 5	http://www.youtube.com/watch?v=QuB3kr3ckYE	00:18 02:09	00:21 03:00	- Disgust	23.6%
Seg. 6	http://www.youtube.com/watch?v=2NEzyvuzs1I&feature=youtu.be	00:08	00:20	- Disgust	5.5%
Seg. 7	http://www.youtube.com/watch?v=DtiSRaB9Sk8&feature=related	01:25	01:59	- Sadness	5.1%
Seg. 8	http://www.youtube.com/watch?v=6PLS7HXo8Bc&feature=plcp	00:00	00:33	- Sadness	1.2%
Seg. 9	http://www.youtube.com/watch?v=rW29Ex1k2Yc	00:00 01:08	00:15 1:34	- Surprise	11.7%

TABLE 6: Emotion expression in terms of facial action units [2].

Emotion	Criteria
Joy	- Must present AU 12 or AU 6+12 or AU 7+12
Surprise	- Must present AU 1+2+5(low) or AU 1+2+26. (low: Intensity level A or B)
Disgust	- Must present AU 9 or only AU 10
Sad	- Must present AU 1 or AU 1+4
Fear	- Must present AU 1+2+4 or AU 20

Acknowledgments

This research was partially supported by the grant IIS-0957983 from the National Science Foundation. The authors would like to thank all subjects who participated in this study. We also would like to thank Mr. Mu Zhou and Ms. Saba Bakhshi for their help for video collection.

References

[1] M. Soleymani, J. Lichtenauer, T. Pun, M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging, " *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42-55, Jan.-March 2012,

[2] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial Action Coding System". Salt Lake City, UT: A Human Face, 2002.

[3] J. J. Campos, K. C. Barrett, M. E. Lamb, H. H. Goldsmith, and C. Stenberg, "Socioemotional development," in Handbook of child psychology. vol. II, M. M. Haith and J. J. Campos, Eds., 4th ed New York: Wiley, 1983, pp. 783-916.

[4] C. Z. Malatesta, C. Culver, J. R. Tesman, and B. Shephard, "The development of emotion expression during the first two years of life," *Monographs of the Society for Research in Child Development*, vol. 54, 1989.

[5] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *Pattern Analysis and Machine Intelligence*, vol. 31, pp. 31-58, 2009.

[6] F. De la Torre and J. F. Cohn, "Visual analysis of humans: Facial expression analysis," in Looking at people, T. B. Moeslund, A. Hilton, A. U. Volker Krger, and L. Sigal, Eds., ed: Springer-Verlag, 2011, pp. 377-410.

[7] J. Whitehill, G. Littlewort, I. Fasel, M. S. Bartlett, and J. Movellan, "Towards practical smile detection," IEEE *Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 2106 - 2111, 2009.

[8] P. Lucey, J. F. Cohn, J. Howlett, S. Lucey, and S. Sridharan, "Recognizing emotion with head pose variation: Identifying pain segments in video," *Systems, Man, and Cybernetics - Part B*, vol. 41, pp. 664-674, 2011.

[9] D. S. Messinger, W. I. Mattson, M. H. Mohammad, and J. F. Cohn, "The eyes have it: Making positive expressions more positive and negative expressions more negative," Emotion, In press.

[10] J. F. Cohn, "Advances in behavioral science using automated facial image analysis and synthesis", *IEEE Social Signal Processing Magazine*, vol. 128, pp. 128-133, 2010.

[11] W. E. Rinn, "The neuropsychology of facial expression,", *Psychological Bulletin*, vol. 95, pp. 52-77, 1984.

[12] J. F. Cohn, Z. Ambadar, and P. Ekman, "Observer-based measurement of facial expression with the Facial Action Coding System," in The handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science, J. A. Coan and J. J. B. Allen, Eds., ed New York, NY: Oxford University, 2007, pp. 203-221.

[13] K. L. Schmidt, Y. Lui, and J. F. Cohn, "The role of structural facial asymmetry in asymmetry of peak facial expressions," *Laterality*, vol. 11, pp. 540-561, 2006.

[14] T. Bnziger, and K.R. Scherer, "Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus". *Blueprint for affective computing: A sourcebook* pp. 271-294. Oxford, England: Oxford university Press, 2010

[15] *http://sspnet.eu/fera2011/*

[16] I. Sneddon, M. McRorie, G. McKeown, J. Hanratty, "The Belfast Induced Natural Emotion Database", *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 32-41, Jan.-March 2012,

[17] G. McKeown, M. Valstar, R. Cowie, M. Pantic, M. Schrder, "The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent", *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 5-17, Jan.-March 2012,

[18] Sander Koelstra, Christian Mhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, Ioannis (Yiannis) Patras, “DEAP: A Database for Emotion Analysis ;Using Physiological Signals,” *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18-31, Jan.-March 2012

[19] Keltner, D., Ekman, P. “Facial expression of emotion”. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions (2nd ed., pp. 236-249)*. New York: Guilford, 2000.

[20] L. F. Barrett, ”Was Darwin wrong about emotional expressions?,” *Current Directions in Psychological Science*, vol. 20, pp. 400-406, 2011.

[21] P. Ekman and W.V. Friesen, “Facial Action Coding System (FACS),” *Manual. Palo Alto: Consulting Psychologists Press*, 1978.

[22] C. H. Hjortsjo, “Man’s face and mimic language“.*Lund: Studentlitterature*, 1970.

[23] C. Darwin, The expression of the emotions in man and animals (3rd edition). New York: Oxford University, 1872/1998.

[24] B. Duchenne, Mechanisme de la physionomie humaine; ou, Analyse electrophysiologique de l’expression des passions. Paris: Bailliere, 1862.

[25] P. Ekman, E.L. Rosenberg,” What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS),” 2nd Edition. New York: Oxford University Press, 2005

[26] *http://www.engr.du.edu/mmahoor/DISFA.htm*

[27] S. D’Mello, R. Picard, and A. Graesser, ”Towards an affect-sensitive AutoTutor,” *IEEE Intelligent Systems*, 2007.

[28] J. Whitehill, M. S. Bartlett, and J. Movellan, ”Automatic facial expression recognition for intelligent tutoring systems,”, *IEEE Workshop for Computer Vision and Pattern Recognition Conference*, Anchorage, Alaska, 2008.

[29] K. L. Schmidt, Z. Ambadar, J. F. Cohn, L. I. Reed, “Movement Differences Between Deliberate and Spontaneous Facial Expressions: Zygomaticus Major Action In Smiling”, *Journal of Nonverbal Behavior*. vol. 30(1), pp. 37-52, 2006

[30] A. B. Ashraf, S. Lucey, J. F. Cohn, T. Chen, Z. Ambadar, K. Prkachin, P. Solomon, B. J. Theobald, ”The painful face: Pain expression recognition using active appearance models”; *International Conference on Multimodal Interfaces*,pp. 9-14, 2007

[31] *http://www.emo-pain.ac.uk/*

[32] M. H. Mahoor, M. Zhou, K. L. Veon, S. M. Mavadati and J. F. Cohn, “Facial action recognition with sparse representation,” *In the proceedings of the 9th IEEE International Conference on Automatic Face and Gesture Recognition (FG011)*, Santa Barbra CA, March 2011.

[33] Z. Yunfeng; F. De la Torre; J.F. Cohn; Z. Yu-Jin, ”Dynamic Cascades with Bidirectional Bootstrapping for Action Unit Detection in Spontaneous Facial Behavior”, *Affective Computing, IEEE Transactions on* , vol.2, no.2, pp.79-91, April-June 2011

[34] D. S. Messinger,M. Mahoor, S. Chow, and J.F. Cohn. “Automated Measurement of Facial Expression in Infant-Mother Interaction: A Pilot Study”. *Infancy*, vol. 14, pp. 285-305, 2009

[35] B. Fasel, J. Luettin, ”Recognition of Asymmetric Facial Action Unit Activities and Intensities”, Pattern Recognition, International Conference on, *15th International Conference on Pattern Recognition (ICPR’00)*, Vol. 1, pp.1100-1103, 2000

[36] S. Kaiser and T. Wehrle, ”Automated coding of facial behavior in human-computer interactions with facs”. Journal of Nonverbal Behavior, 16(2):6783, 1992.

[37] A. Savran, B. Sankur, M. Taha Bilge, ”Regression-based intensity estimation of facial action units” , *Image and Vision Computing*, Vol 30, issue 10, pp. 774-784., 2012

[38] M. H. Mahoor, S. Cadavid, D. S. Messinger, and J. F. Cohn, ”A Framework for Automated Measurement of the Intensity of Non-Posed Facial Action Units,” *2nd IEEE Workshop on CVPR for Human communicative Behavior analysis (CVPR4HB)*, Miami Beach, June 25, 2009

[39] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, ”Automatic Recognition of Facial Actions in Spontaneous Expressions,” *Journal of Multimedia*, vol. 1, pp. 22-35, 2006.

[40] M. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, “The First Facial Expression Recognition and Analysis Challenge,” *in Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, Workshop on Facial Expression Recognition and Analysis Challenge*, 2011.

[41] T. Kanade, J. Cohn, and Y. Tian. “Comprehensive database for facial expression analysis,” *In Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 46-53, 2000.

[42] M. Pantic, M. Valstar, R. Rademaker and L. Maat, “Web-based database for facial expression analysis”, *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on* , vol., no., pp. 5 pp., 6-8 July 2005

[43] P. Lucy, J.F. Cohn, K.M. Prkachin, P. Solomon, I. Matthrews . Painful data: The UNBC-McMaster Shoulder Pain Expression Archive Database. *IEEE International Conference on Automatic Face and Gesture Recognition (FG2011),Santa Barbra CA, March 2011*

[44] *http://measurement.psy.miami.edu/cms.phtml*

[45] Y.-I. Tian, T. Kanade, J.F. Cohn , “Recognizing action units for facial expression analysis,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.23, no.2, pp.97-115, Feb 2001

[46] Y. Zhang, Q. Ji, Z. Zhu, and B. Yi. “Dynamic facial expression analysis and synthesis with MPEG-4 facial animation parameters?. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(10), pp. 1383-1396), 2008.

[47] J. F. Cohn, Z. Ambadar, and P. Ekman, “Observer-based measurement of facial expression with the Facial Action Coding System“. In J. A. Coan and J. J. B. Allen (Eds.),*The handbook of emotion elicitation and assessment. Oxford University Press Series*,2007

[48] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews,“The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression,” *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on* , pp.94-101, 13-18 June 2010

[49] M. J. 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*, Nara Japan, IEEE Computer Society, pp. 200-205, April 14-16 1998

[50] J. F. Cohn, and T. Kanade, ”Automated facial image analysis for measurement of emotion expression”. In J. A. Coan and J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science*, pp.222-238. New York: Oxford.

[51] N. Alyz, B. Gkberk, H. Dibeklioglu, A. Savran, A. A. Salah, L. Akarun and B. Sankur, “3D Face Recognition Benchmarks on the Bosphorus Database with Focus on Facial Expressions,” *The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008)*, Roskilde University, Denmark, May 2008

[52] L. Yin, X. Wei, Y. Sun, J. Wang and M. J. Rosato, “A 3D Facial Expression Database For Facial Behavior Research , *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pp: 211 - 216, 2006

[53] M. Frank,J. Movellan, M. Bartlett,G. Littleworth: ”RU-FACS-1 database”, *Machine Perception Laboratory*, U.C. San Diego

[54] W. Shangfei, L. Zhilei, L. Siliang, L. Yanpeng, W. Guobing, P. Peng, C. Fei and W. Xufa, “A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference”, *IEEE Transactions on Multimedia* , vol.12, no.7, pp.682-691, Nov. 2010

[55] . J. Cohen, “Statistical power analysis for the social sciences.” *Hillsdale, NJ: Lawrence Erlbaum Associates*, 1988

[56] T.F. Cootes, G.J Edwards, C.J. Taylor, “Active appearance models,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.23, no.6, pp.681-685, Jun 2001

[57] T. Ojala, M. Pietikainen, and D. Harwood, “A comparative study of texture measures with classification based on feature distributions,” *Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.

[58] T. Ahonen, A. Hadid, and M. Pietikainen “Face Recognition with Local Binary Patterns“,*ECCV*,pp. 469-481, 2004

[59] C. Shan, S. Gong,P.W McOwan, “Facial expression recognition based on Local Binary Patterns:A comprehensive study,” *Image and Vision Computing*, 27 (6), p.803-816, May 2009

[60] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *Computer Vision and Pattern Recognition 2005*, vol.1, no., pp.886-893 ,June 2005
[61] C. Shu, X. Ding and C. Fang; “Histogram of the Oriented Gradient for Face Recognition,” *Tsinghua Science Technology, Tsinghua University Press*. 16(2), 216-224. 2011
[62] M. H. Mahoor and M. Abdel-Mottaleb, ”A Multi-modal Approach for Face Recognition Based on Ridge Images and Attributed Relational Graph”, *IEEE Transactions on Information Forensics and Security*, Volume 3, Issue 3, pp.431- 440, Sept. 2008
[63] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, “Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior,” *Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition (CVPR ’05)*, pp. 568-573, 2005.
[64] L. Cayton,“Algorithms for manifold learning”,*University of California, San Diego, Tech. Rep. CS2008-0923*,2005
[65] M. Belkin and P. Niyogi, “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation”, *Neural Computation*, pp. 1373-1396 , vol.15, 2003
[66] V.N. Vapnik; , “An overview of statistical learning theory,” *Neural Networks, IEEE Transactions on* , vol.10, no.5, pp.988-999, Sep 1999
[67] M. Sokolova, N. Japkowicz, S. Szpakowicz, “Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation”. *In: Australian conference on artificial intelligence*, vol 4304. LNCS, pp 1015-1021, 2006
[68] J. Cohen, “A Coefficient of Agreement for Nominal Scales”, *Educational and Psychological Measurement*, vol. 20, 1: pp. 37-46, April 1960.
[69] P.E. Shrout and J.L. Fleiss, “Intraclass correlations: Uses in assessing rater reliability”, *Psychological Bulletin*, Vol 86(2),pp.420-428, Mar 1979
[70] C.-C. Chang and C.-J. Lin. ”LIBSVM : a library for support vector machines”. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.