A High-Resolution Spontaneous 3D Dynamic Facial Expression Database

Xing Zhang[1], Lijun Yin[1], Jeffrey F. Cohn[2], Shaun Canavan[1], Michael Reale[1], Andy Horowitz[1], and Peng Liu[1]
[1]State University of New York at Binghamton; [2]University of Pittsburgh

***Abstract*—Facial expression is central to human experience. Its efficient and valid measurement is a challenge that automated facial image analysis seeks to address. Most publically available databases are limited to 2D static images or video of posed facial behavior. Because posed and un-posed (aka "spontaneous") facial expressions differ along several dimensions including complexity and timing, well-annotated video of un-posed facial behavior is needed. Moreover, because the face is a three-dimensional deformable object, 2D video may be insufficient, and therefore 3D video archives are needed. We present a newly developed 3D video database of spontaneous facial expressions in a diverse group of young adults. Well-validated emotion inductions were used to elicit expressions of emotion and paralinguistic communication. Frame-level ground-truth for facial actions was obtained using the Facial Action Coding System. Facial features were tracked in both 2D and 3D domains using both person-specific and generic approaches. The work promotes the exploration of 3D spatiotemporal features in subtle facial expression, better understanding of the relation between pose and motion dynamics in facial action units, and deeper understanding of naturally occurring facial action.**

Keywords: 3D facial expression; FACS; spontaneous expression; dynamic facial expression database.

I. INTRODUCTION

Research on computer-based facial expression and affect analysis has intensified since the first FG conference in 1995. The resulting advances have made possible the emerging field of affective computing. The continued development of emotion-capable systems greatly depends on access to well-annotated, representative affective corpora [13]. A number of 2D facial expression databases have become available (e.g., [1][2][16][7][8]), as well as some with 3D imaging (e.g., [9][14][15][24][25][45]). Because the face is a 3D object and many communicative signals involve changes in depth and head rotation, inclusion of 3D images is an important addition. A major limitation of existing databases is that most have only posed or acted facial behavior, and thus are not representative of spontaneous affective expression, which may differ in timing, complexity, and intensity [22]. No currently available dataset contains dense, *dynamic, 3D* facial representations of *spontaneous* facial expression with *anatomically-based (FACS) annotation* [36].

Currently, most approaches to automatic facial expression analysis attempt to recognize a set of prototypic emotional expressions (e.g., anger, disgust, fear, happiness, sadness, and surprise) [3][5][13]. Many studies about emotion use "acting" or "emotion portrayals" in a restricted sense by recording subjects who are expressing emotions instructed via single labels of emotions, sometimes using scripts [6]. The resulting posed and exaggerated facial actions may occur only rarely in daily life [4].

Because posed and un-posed (aka "spontaneous") facial expression differ along several dimensions [32], including complexity (especially with respect to segmentation), well-annotated video of un-posed facial behavior is needed. Moreover, as noted above, because the face is a three-dimensional deformable object, a 3D video archive would be especially important. Two-dimensional databases, such as RU-FACS [23] or Cohn-Kanade [2], are insufficient. The CMU Multi-PIE database [34], 3D dynamic AU database [35], Bosphorus database [9], KDEF [33], BU 3D Facial Expression Databases [14][15], and ICT-3DRFE database [24] begin to address the need for 3D (or multi-view) data but are limited to posed facial behavior.

Recent efforts to collect, annotate, and analyze spontaneous facial expression for community use have begun [26][27][28]. All are limited to the 2D domain or thermal imaging.

To address the need for well-annotated, dynamic 3D video of spontaneous facial behavior in response to meaningful and varied emotion inductions, we developed a 3D database for the community of researchers in automated facial expression analysis. We used a series of effective tasks for authentic emotion induction. The tasks include social interviews between previously unacquainted people (one a naïve subject and the other a professional actor/director), pre-designed activities (e.g., games), viewing of film clips, a cold pressor test to elicit pain, social challenge to elicit anger followed by reparation, and olfactory stimulation to elicit disgust. Well-experienced, certified FACS coders annotated the video. Additionally, person-specific and generic face tracking was performed. The new 3D spontaneous dynamic facial expression database is intended for use by the research community.

II. HIGH RESOLUTION DATA ACQUISITION

A. System Setup

A Di3D dynamic face capturing system [12] captured and generated 3D facial expression sequences. Passive stereo photogrammetry was used to create 3D model sequences at the frame rate of 25 frames per second. The geometric face model contains 30,000 ~ 50,000 vertices. The 2D texture videos are 1040×1392 pixels/frame. Figure 1 shows an example of the imaging system at work.

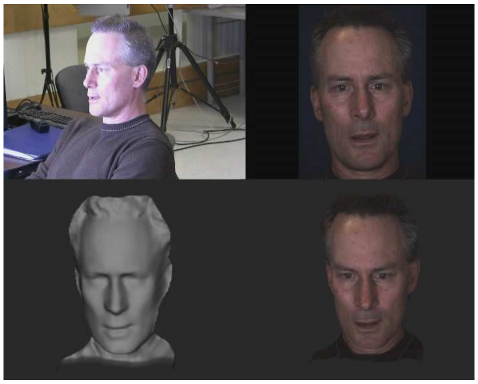

Figure 1: Upper-left: general view from a regular camera; Upper-right: 2D video; Lower-left: 3D dynamic geometric model; Lower-right: 3D dynamic geometric model with mapped texture.

B. *Data Capture*

1) Emotional expression elicitation

For recording spontaneous affective behavior, a good trade-off between acquisition of natural emotional expressions and data quality is needed. If the recording environment is too constrained, genuine emotion and social signaling become difficult to elicit. If the recording environment is unconstrained, much error may be introduced in the recordings. In the psychology literature, well-validated emotion techniques and guidelines have been proposed to meet this challenge [43].

To elicit target emotional expressions and conversational behavior, we used approaches adapted from other investigators plus techniques that proved promising in pilot testing. All sessions were conducted by a professional actor and director of performing arts. The tasks include *face-to-face interview*, *social games*, *documentary film watching*, *cold pressor task*, *social anger induction*, and *experience of smell*. Film clips and games [10][46] are well-validated approaches to elicit emotion; cold pressor is well studied to safely elicit pain expressions without risk of tissue injury [44]; olfactory stimuli can reliably elicit disgust; and interviews elicit a wide range of emotion expression and interpersonal behavior. These methods evoke a range of authentic emotions in a laboratory environment [11].

After participants gave informed consent to the procedures and permissible uses of their data, the experimenter explained the procedure and began the emotion inductions. Following usage in the psychology literature, each emotion induction is referred to as a “task.” The experimenter was a professional actor and director. Each participant experienced 8 tasks, as summarized in Table 1. Those tasks were seamlessly spaced with smooth transitions between them. Immediately after each task, participants completed self-report ratings of their feelings unless otherwise noted.

The protocol began with a conversation, which included joke telling, between the participant and the experimenter. The relaxed exchange and shared positive emotion were intended to build rapport and elicit expressions of amusement. After rating the first experience, the participant watched and listened to a documentary about a real emergency involving a child, followed by an interview that gave them opportunity to talk about their feelings in response to the task. Reactions of sadness were intended responses.

TABLE I. EIGHT TASKS FOR EMOTIONAL EXPRESSION ELICITATION

Task	Activity	Target Emotion
1	Talk to the experimenter and listen to a joke (Interview).	Happiness or Amusement
2	Watch and listen to a recorded documentary and discuss their reactions.	Sadness
3	Experience sudden, unexpected burst of sound.	Surprise or startle
4	Play a game in which they improvise a silly song.	Embarrassment
5	Anticipate and experience physical threat.	Fear or nervous
6	Submerge their hand in ice water for as long as possible.	Physical pain
7	Experience harsh insults from the experimenter.	Anger or upset
8	Experience an unpleasant smell.	Disgust

Next, the participant was asked to participate in several activities with the experimenter. These included startle triggered by a siren; embarrassment elicited by having to improvise a silly song; fear while playing a game that occasioned physical danger; and physical pain elicited by submerging their hand in ice water. Following this cold pressor task, the experimenter intentionally berated the participant to elicit anger followed by reparation.

Finally, the participant was asked to smell an unpleasant odor to evoke strong feelings and expressions of disgust. The tasks concluded with a debriefing by the experimenter. Each task lasted about 1 to 4 minutes and was recorded as described below in sub-section C.

The procedures elicited a range of emotions and facial expressions that include happiness/amusement, disgust, sadness, surprise/startle, embarrassment, nervous/fear, physical pain, and anger/upset.

2) Participants

Forty-one participants (23 women, 18 men) were recruited from the departments of psychology and computer science as well as from the School of Engineering. They were 18 – 29 years of age; 11 were Asian, 6 were African-American, 4 were Hispanic, and 20 were Euro-American (Table 2).

C. *Database Organization*

The database is structured by participants. Each participant is associated with 8 tasks. For each task, there is both 3D and 2D video. Although tasks varied in duration, to reduce storage demands and processing time, each video consists of the segment during which the participant was most expressive (about 1 min. on average). This reduced retention of frames in which little facial expression occurred. The video data are about 3 TB in size.

Metadata consists of manually annotated action units (FACS AU), automatically tracked head pose, and 2D/3D facial landmarks. Table 2 summarizes the 3D spontaneous dynamic facial expression database. Figure 2 shows the data structure of each task. Figure 3 shows several samples of 3D spontaneous dynamic facial expression sequences. The meta-data (e.g., AU codes, tracked features, head poses, etc.) will be detailed in the next section on data processing, annotation, and evaluation.

TABLE II. SUMMARY OF 3D SPONTANEOUS DYNAMIC FACIAL EXPRESSION DATABASE

# of participants	# of tasks	# of 3D+2D sequences	# of metadata sequences (i.e., annotated AUs, facial landmarks, and poses)
41	8	328	328

Note: Asian (11), African-American (6), Hispanic (4), and Euro-American (20).

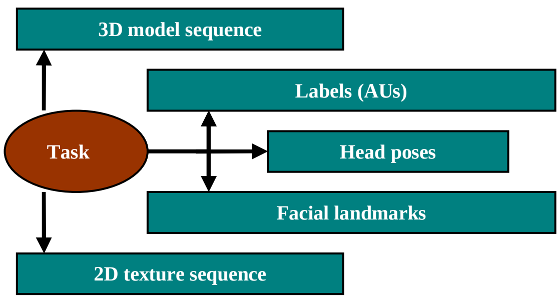

Figure 2: Database organization.

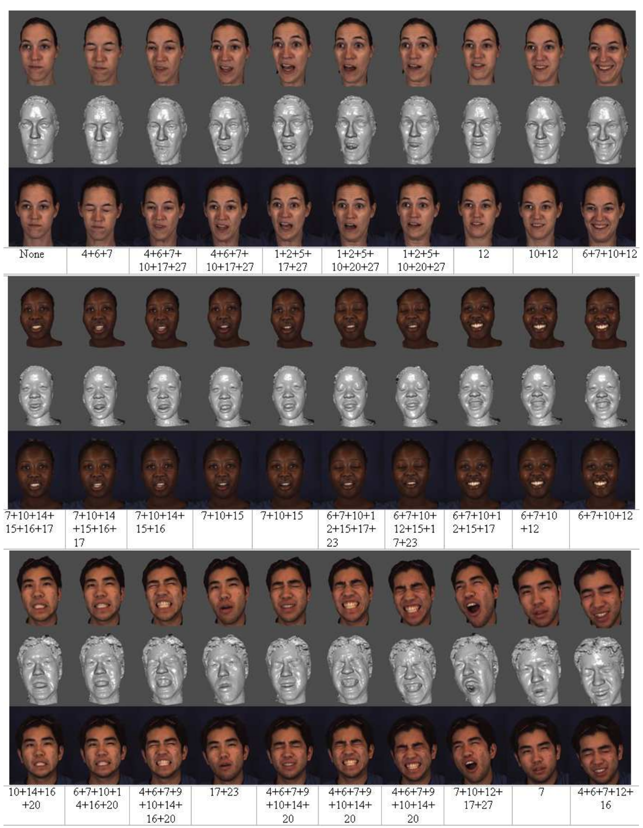

Figure 3: Samples of textured models, shaded models, original 2D videos, and the annotated Action Units (AUs).

III. DATA PROCESSING, ANNOTATION, AND EVALUATION

A. *FACS Coding*

Automatic detection of FACS action units is a major thrust of current research in automated facial image analysis [22]. To provide necessary ground truth in support of these efforts, we annotated facial expressions using the Facial Action Coding System (FACS) [17][18].

For each participant, we code action units associated with emotion and paralinguistic communication. Because FACS coding is time intensive, we prioritized coding to focus on 20-second segments that were most productive of facial expression.

For 8 conditions (tasks), FACS coders coded a 20-second segment that had the highest density of facial expression. Coders were free to code for longer than 20 seconds if expression was continuing. If a video was less than 20 seconds, it was coded in its entirety. Descriptive statistics are reported in Table 3.

For each condition, two experienced FACS-certified coders independently coded onsets and offsets of 27 action units per the 2002 edition of FACS [36] using Observer Video-Pro Software [21]. The Observer system makes it possible to manually code digital video in stop-frame and at variable speed and later synchronize codes according to digital time stamp. For AU 12 and AU 14, intensity was coded as well on a 0-5 ordinal scale using custom software.

Inter-observer exact (25f/s) agreement was quantified using coefficient kappa [37], which is the proportion of agreement above what would be expected to occur by chance, and *F1*, which is the geometric mean of precision and recall. For intensity coding, reliability was quantified using intra-class correlation coefficients (ICC). Table 4 reports the number of events (from onset to offset) and number of frames coded for each AU and kappa reliability.

TABLE III. DESCRIPTIVE STATISTICS FOR FACS- CODED VIDEOS (UNIT OF MEASURE IS SECONDS)

Task	Activity	Minimum	Maximum	Mean
1	Talk to the experimenter and listen to a joke (Interview).	13.00	29.70	19.60
2	Watch and listen to a recorded documentary and discuss their reactions.	12.12	25.00	20.21
3	Experience sudden, unexpected burst of sound.	8.56	16.76	12.24
4	Play a game in which they improvise a silly song.	16.14	24.12	19.73
5	Anticipate and experience physical threat.	18.52	31.00	20.04
6	Submerge their hand in ice water for as long as possible.	8.00	23.00	18.95
7	Experience harsh insults from the experimenter.	17.24	25.00	19.91
8	Experience an unpleasant smell.	3.60	21.40	11.49

Note. Unit of measure is seconds. Data are based on video from the first 30 participants.

In summary, the expression sequences were AU-coded by two experts. For each sequence, 27 AUs were considered for coding. For each of the target AUs, we have various numbers of coded events, where an event is defined as the contiguous frames from onset to offset.

TABLE IV. DESCRIPTIVE STATISTICS FOR EVENTS, FRAMES, AND KAPPA RELIABILITY.

Action Unit	Kappa	#Events	#Frames
1	0.894	411	27610
2	0.967	317	20898
4	0.953	351	25204
5	0.972	176	6418
6	0.905	428	51498
7	0.927	440	62001
9	0.902	89	5066
10	0.918	518	67086
11	0.999	7	1153
12	0.906	379	67586
13	n/a	2	138
14	0.927	477	48017
15	0.926	542	16892
16	0.609	158	3420
17	0.876	1010	40430
18	0.261	30	418
19	0.845	50	901
20	0.955	86	2718
22	0.951	39	623
23	0.777	616	18405
24	0.878	363	16039
27	0.946	55	1529
28	0.968	94	4797
30	0.952	17	631
32	0.984	22	1365
38	0.94	33	1208
39	n/a	7	232
Overall	0.931	n/a	n/a

Note: Data are based on video from the first 30 participants. Overall kappa is weighted average. An event is defined as a set of contiguous frames from onset frame to offset frame.

B. Head Pose

Head pose, which includes rigid head motion, is important for image registration and is itself of communicative value (e.g., downward head pitch when coordinated with smiling communicates embarrassment). Head pose was measured from the 2D videos using a cylindrical head tracker of [19]. This tracker is person-independent, robust, and has concurrent validity with person-specific 2D+3D AAM [20] and with magnetic motion capture device [19]. The head pose (yaw, roll, and pitch) were measured with respect to the frontal pose.

C. Statistics of Self-Reports

Participants used 5-point Likert-type scales to report their felt emotions for each task. The emotions, or affective states, listed were relaxed, happiness/amusement, disgust, nervous/fear, anger/upset, sadness, sympathy, surprise, startle, physical pain, and embarrassment. After each task, the participants were asked to read the items, choose the emotions (if any) that best described how they felt during the task and indicate the degree to which they experienced the emotion (i.e., from “very slightly” to “extremely”).

Participants could and did experience more than one emotion for each task. Figure 4 shows the highest rated emotion reported by the participants for each task. Except for task 7, the target emotion for each task (see Table 1) was the one most highly rated by the majority of participants. For instance, the highest bar of task 8 shows that the majority of subjects rated the “disgust” emotion as the main emotion for that task. The highest bar of task 6 shows the majority of subjects rated the “pain” feeling as the main emotion. Accordingly, almost all of the other tasks show this property as well. For task 7, one might note that there is no clear highest-ranking emotion if we only consider the emotion with the strongest scale. However, based on the self-reporting results for all scales, the majority of participants reported experiencing “anger/upset” from at least scale 2 (“a little”) to 5 (“extremely”) during this task. Thus, the task generally succeeded in evoking the target emotion.

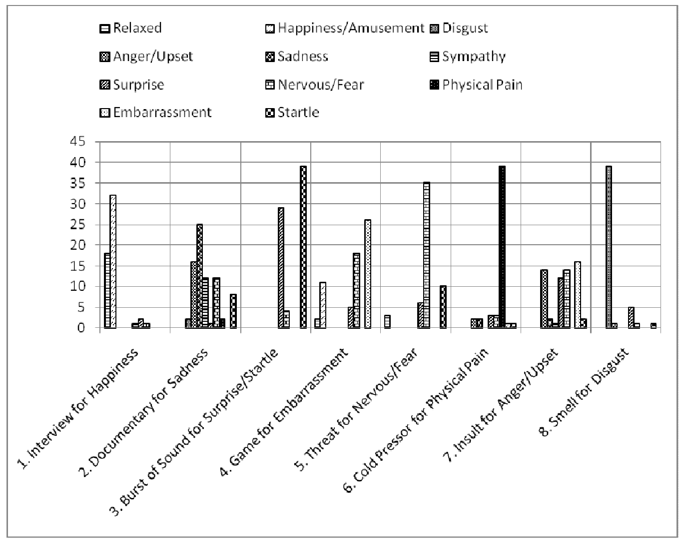

Figure 4: Statistics of self-report emotion distribution for task 1 to task 8 (from left to right); vertical axis is the number of votes.

D. Feature Points Tracking

1) 3D-TDSM based tracking

We defined 83 feature points around the 3D facial areas of eyes, nose, mouth, eyebrows, and chin contour at the initial frame of a video sequence. Extended from the active appearance model approach [30], we applied our newly developed 3D geometric surface based Temporal Deformable Shape Model [40] to track 83 points on the 3D dynamic surface directly. Our developed method involves fitting a new multi-frame constrained 3D temporal deformable shape model (TDSM) to range data sequences. We consider this a temporal based deformable model as we concatenate consecutive deformable shape models into a single model driven by the appearance of facial expressions. This allows us to simultaneously fit multiple models over a sequence of time with one TDSM.

To construct a temporal deformable shape model, we applied a representation of the point distribution model to describe the 3D shape, in which a parameterized model *S* was constructed by 83 landmark points on each model frame. Such a set of feature points (shape vector) was aligned by the Procrustes analysis method [30]. Principal component analysis (PCA) was then performed on the new aligned feature vector. This was done to estimate the different variations of all the training shape data. When approximating a new shape *S*, the point distribution model was constrained by both the variations in shape and the shapes of neighbor frames. Figure 5 (lower row) shows several sample frames of the tracked 83 feature points on a 3D model sequence. The detailed algorithm is described in [40].

2) 2D-CLM tracking

Two-dimensional facial expression sequences were automatically tracked using the constrained local model (CLM) approach of [38][39]. All CLM tracking was reviewed offline for tracking errors. Coded were: 1) Good tracking; 2) Multiple errors; 3) Jawline off; 4) Occlusion; and 5) Face out of frame. Figure 5 (upper row) shows several sample frames of the tracked points.

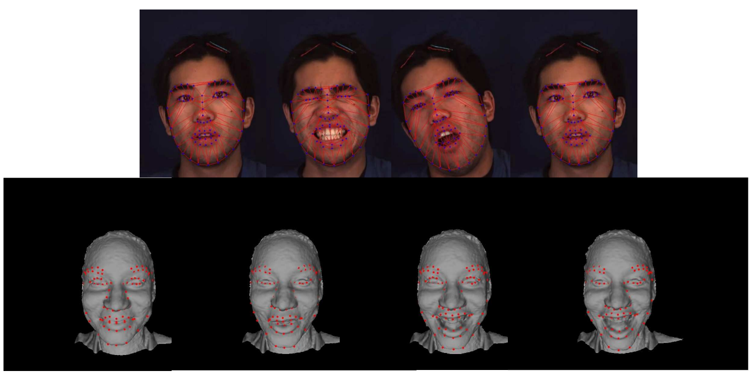

Figure 5: CLM-tracked feature points on a 2D sequence of a male subject (upper row); a sample 3D sequence with 3D-TDSM tracked feature points of a female subject (lower row).

E. Expression Analysis and Recognition

1) Spontaneous expression classification

To validate the data for prototypic emotion expression recognition, we applied the existing 3D dynamic facial expression descriptor [42] for expression classification. An HMM was used to learn the temporal dynamics and spatial relationships of facial regions. We conducted a person-independent experiment on 16 subjects. Following a 10-fold cross-validation procedure, we used 14 subjects for training and 2 subjects for testing, and achieved an average correct recognition rate of 70.2% for distinguishing six spontaneous emotion expressions. Note that spontaneous expression data are more difficult to classify than posed expression data. When the same approach was applied to the 3D posed dynamic facial expression database BU-4DFE [15], over 80% recognition rate was achieved for classifying six posed expressions. The performance degradation on classifying 3D spontaneous expressions is due to the complexity, mixture, and subtlety of the spontaneous expressions in the new database. To further evaluate our approach, we conducted a comparison study by implementing the 3D static model based approach using geometric primitive features [29] and the 2D texture based approach using Gabor-wavelet features [31]. The average recognition rates for the two approaches were 51.3% and 63.2%, respectively.

2) Action Unit recognition on spontaneous 4D data

We also performed experiments in AU recognition on the new spontaneous 3D dynamic database. We extended the idea of a 3D surface primitive feature into 4D space and developed a new feature representation: the so-called “Nebula” features [41]. Given a spatiotemporal volume, the data is voxelized and fit to a cubic polynomial $f(x; y; t) = z$. A label is assigned based on the principal curvature values; we use this label and the polar angles of the direction of least curvature to build a 3D histogram for each region of the face. The concatenated histograms from each of the regions give us our final feature vector. We selected 16 subjects and tested on 12 AUs using a support vector classifier. The average recognition AUC (Area Under Receiver Operating Characteristic Curve) was over 0.738. Details are described in [41].

IV. Conclusion and Future Work

In this paper, we reported our newly developed spontaneous 3D dynamic facial expression database, which will be made available to the research community. Such a database can be a valuable resource to facilitate the research and development of human behavior analysis in security, HCI, psychology and biomedical applications.

Limited by the working environment, data collection was conducted in a lab environment. The guided format using a professional actor and director as experimenter sought to simulate a more natural setting. In future work, other settings and image capture setups might be considered. Data quality could be improved by using a wider range imaging system with more robust illumination control. The database will also be expanded to include more subjects.

Moreover, our current database includes sequential geometric model data and texture data. In addition to the facial feature tracking algorithms, more powerful approaches need be investigated in order to make the data processing and visualization fast and accurate. Automatic data annotation, registration, and efficient data representation (or compression) for micro-expression analysis will also be our next research direction.

Acknowledgment

This material is based upon the work supported in part by the National Science Foundation under grants IIS-1051103 and IIS-1051169. We would like to thank Nicki Siverling, Dean Rosenwald, and Shaun Zuratovic for FACS coding.

REFERENCES

[1] http://www.mmifacedb.com/., Man machine interaction group. 2005.

[2] T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, France, 2000.

[3] M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions: the state of the art. *IEEE Trans. on PAMI*, 22(12), 2000.

[4] M. Pantic and L. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370-1390, 2003.

[5] Y. Zhang and Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Trans. on PAMI*, 27 (5):699-714, May 2005.

[6] H. Wallbott and K. Scherer, Cues and channels in emotion recognition. *Journal of Personality and Social Psychology 51*(4), 690–699, 1996.

[7] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. McRorie, J. C. Martin, L. Devillers, S. Abrilian, A. Batliner, N. Amir, and K. Karpouzis, "The humaine database: Addressing the collection and annotation of naturalistic and induced emotional data," in *Proc. 2nd Inter. Conf. on Affective Computing and Intelligent Interaction (ACII),* Lisbon, Portugal., 2007, pp. 488–500.

[8] T. Anziger and K. Scherer, "Using actor portrayals to systematically study multimodal emotion expression: The gemep corpus," in *ACII, 2007,* pp. 476–487.

[9] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *BIOID*, 2008, pp. 47–56.

[10] J. Gross and R. Levenson, "Emotion elicitation using films," *Cognition and Emotion*, vol. 9, no. 1, pp. 87–108, 1995.

[11] R. Cowie and R. Cornelius, "Describing the emotional states that are expressed in speech," *Speech Communication*, 40(1-2):5–32, 2003.

[12] Di3D Inc. http://www.di3d.com.

[13] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions," *IEEE Trans. on PAMI*, vol. 31, no. 1, pp. 39–58, 2009.

[14] L. Yin, X. Wei, Y. Sun, W. J, and M. Rosato, "A 3D facial expression database for facial behavior research," in *IEEE Inter. Conf. on Automatic Face and Gesture Recognition*, Southampton, UK, 2006.

[15] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3D dynamic facial expression database," in *IEEE Inter. Conf. on Automatic Face and Gesture Recognition*, Amsterdam, the Netherlands, Sept. 2008.

[16] P. Lucey, J.F. Cohn, J. Saragih, I. Matthews, Z. Ambadar "The Extended Cohn-Kanade Database: A complete facial expression database for both facial action units and emotion detection", *IEEE CVPR4HB*, 2010.

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

[18] J.F. Cohn and P. Ekman, Measuring facial action by manual coding, facial EMG, and automatic facial image analysis. In J. A. Harrigan, R. Rosenthal & K. Scherer (Eds.), *Handbook of nonverbal behavior research methods in the affective sciences* (pp. 9-64). New York: Oxford, 2005.

[19] J. Jang and T. Kanade, Robust 3D head tracking by online feature registration. *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, The Netherlands. 2008.

[20] I. Matthews, J. Xiao, and S. Baker, 2D vs. 3D deformable face models: Representational power, construction, and real-time fitting. *International Journal of Computer Vision, 75*(1), 93–113, 2007.

[21] L. Noldus, R. Trienes, A. Henriksen, H. Jansen, and R. Jansen, The Observer Video-Pro : New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments, and Computers, 32*, 197-206, 2000.

[22] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, A survey of affect recognition methods: Audio, visual and spontaneous expressions, *ACM International Conference on Multimodal Interfaces*, 2007.

[23] RU-FACS. http://mplab.ucsd.edu/databases/databases.html.

[24] G. Stratou, A. Ghosh, P. Debevec, and L. Morency, Exploring the effect of illumination on automatic expression recognition using the ICT-3DRET database, *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2011.

[25] G. Fanelli, J. Gall, H. Romsdorfer, T. Weise, and L. Van Gool, A 3D Audio-Visual Corpus of Affective Communication. *IEEE Trans. on Multimedia*, 12(6): 591-598, Oct. 2012.

[26] S. Wang, Z. Liu, S. Lu, X. Wang, et al., A natural visible and infrared facial expression database for expression recognition and emotion inference, *IEEE Trans. on Multimedia*, 12(7):682 -691, Nov. 2010.

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

[28] M. Mahoor, S. Cadavid, D. 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.

[29] J. Wang, L. Yin, X. Wei, and Y. Sun, 3D facial expression recognition based on primitive surface feature distribution, *IEEE CVPR* 2006.

[30] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Trans. on PAMI*, 23:681–685, 2001.

[31] M. Lyons et al., Automatic classification of single facial images. *IEEE Trans. on PAMI* (21):1357–1362, 1999.

[32] K. Schmidt, Z. Ambadar, J.F. Cohn, and L. Reed, Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling. *Journal of Nonverbal Behavior*, 30:37-52, 2006.

[33] E. Goeleven, R. De Raedt, L. Leyman, and B. Verschuere, The Karolinska Directed Emotional Faces: A validation study. *Cognition and Emotion*, 22(6), 1094-1118, 2008.

[34] R. Gross, I. Matthews, J.F. Cohn, T. Kanade, and S. Baker, Multi-PIE. *Image and Vision Computing*, 28(5), 807-813, 2010.

[35] D. Cosker, E. Krumhuber, and A. Hilton, A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling, in *IEEE Inter. Conf. on Computer Vision (ICCV)*, 2011.

[36] P. Ekman, W. Friesen, and J. Hager, *Facial action coding system*: Research Nexus, Network Research Information, Salt Lake City, 2002.

[37] J. Fleiss, *Statistical methods for rates and proportions*. New York: Wiley, 1981.

[38] J. Saragih, S. Lucey, and J.F. Cohn, Deformable Model Fitting with a Mixture of Local Experts. *IEEE Inter. Conf. on Computer Vision*, 2009.

[39] J. Aragih, S. Lucey, and J.F. Cohn, Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision, 91*(2), 200-215, 2011.

[40] S. Canavan and L. Yin, "Fitting and tracking 3D/4D facial data Using a temporal deformable shape model", *Technique Report*, Binghamton University, 2012.

[41] M. Reale, X. Zhang, and L. Yin, "Nebula feature: a space-time feature for posed and spontaneous 4D facial behavior analysis", *IEEE Inter. Conf. on Automatic Face and Gesture Recognition (FG),* 2013.

[42] Y. Sun and L. Yin, "Facial Expression Recognition Based on 3D Dynamic Range Model Sequences". *The 10th European Conference on Computer Vision (ECCV08)*, Oct. 2008, Marseille, France.

[43] J. Coan and J. Allen (Eds.), *Oxford handbook on emotion elicitation and assessment*. NY: Oxford, 2007.

[44] C. Von Baeyer, T. Piira, C. Chambers, M. Trapannotto, and L. Zeltzer, Guidelines for the cold pressor task as an experimental pain stimulus for use with children. *The Journal of Pain, 6*(4), 218-227, 2005.

[45] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin. Static and Dynamic 3D Facial Expression Recognition: A Comprehensive Survey. *Image and Vision Computing*. 30(10):683-697, Oct. 2012.

[46] D. Clark, "On the induction of depressed mood in the laboratory: Evaluation and comparison of the velten and musical procedures," *Advances in Behaviour Research and Therapy*, 5(1):27–49, 1983.