A High-Resolution 3D Dynamic Facial Expression Database

Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale
Department of Computer Science
State University of New York at Binghamton

Abstract

Face information processing relies on the quality of data resource. From the data modality point of view, a face database can be 2D or 3D, and static or dynamic. From the task point of view, the data can be used for research of computer based automatic face recognition, face expression recognition, face detection, or cognitive and psychological investigation. With the advancement of 3D imaging technologies, 3D dynamic facial sequences (called 4D data) have been used for face information analysis. In this paper, we focus on the modality of 3D dynamic data for the task of facial expression recognition. We present a newly created high-resolution 3D dynamic facial expression database, which is made available to the scientific research community. The database contains 606 3D facial expression sequences captured from 101 subjects of various ethnic backgrounds. The database has been validated through our facial expression recognition experiment using an HMM based 3D spatio-temporal facial descriptor. It is expected that such a database shall be used to facilitate the facial expression analysis from a static 3D space to a dynamic 3D space, with a goal of scrutinizing facial behavior at a higher level of detail in a real 3D spatio-temporal domain.

1. Introduction

Research on facial information analysis has been intensified recently, driven mainly by its important applications: face recognition (FR) [10, 13, and 43] and facial expression recognition (FER) [5, 7, 18, 20, 21, and 42]. Most research for face analysis utilized conventional 2D static images or 2D dynamic videos [1, 4, 13, 25, 30, 31, 39, and 43]. In recent years, 3D range data has been extensively used for face analysis due to its explicit representation of geometric information and its inherent capability of handling facial pose and illumination variations [2, 12, and 22]. Similar to the 2D modality, 3D face data can also be represented in a static space and a dynamic space.

(1) 3D static: Most existing work utilizes the static 3D range data acquired through laser-scanning, stereo photogrammetry, or active light projection [2, 3, 8, 12, 22, 38, and 40]. Many successful works have been reported recently for 3D face recognition [2, 12, 14, 22, and 37]. Most recently, some work has also been reported for 3D facial expression recognition [35, 36, and 41]. For example, Yin et al. have investigated the 3D facial expression recognition [35] using 3D surface primitive features based on the 3D static facial expression database [40]. Wang et al. [36] used the in-house 3D static face models to study the facial expressions in 3D space and 2D space, and achieved encouraging results in identifying facial expression abnormality in schizophrenia. Note that all the data that have been used are still based on static range models.

(2) 3D dynamic: Facial expression is by nature a dynamic facial behavior. The 3D dynamic face representation is believed to be the best reflection of this nature. Psychological research shows that facial dynamics provide important cues that can be interpreted in order to represent an individual's characteristics [27]. The recent findings indicate that the dynamic cues from expressive and talking movements of human faces provide information about individuals' facial structure, and therefore play a great role in facilitating the subsequent recognition [27]. A human face is a bumpy and mobile surface. Neither 2D dynamic data nor 3D static data may be sufficient to depict such a property. 3D static face models lacking a temporal context may be a profound handicap to recognizing facial expressions as well as identifying faces with varied expressions.

Recent technological advances in 3D imaging systems allow a high quality 3D shape to be acquired in real time [3, 8, and 38]. Such 3D dynamic data (or so-called *4D data*) captures the dynamics of time-varying 3D facial surfaces, making it possible to analyze the dynamic facial behavior in a 3D spatio-temporal domain. It is conceivable that more information concerning the individual's characteristics or expressive traits can be derived from 3D dynamic sequences.

There were a few works reported using 4D data for facial expression analysis. For example, Wang et al. [38] successfully developed a hierarchical framework for tracking high-density 3D facial expression sequences captured from a structure-lighting imaging system. Recent work reported by Chang and Turk et al. in [3] utilized 3D model sequences for expression analysis and editing. The work is accomplished through usage of a probabilistic model on the generalized expression manifold of the standard model. Notice that the existing reported works

978-1-4244-2154-1/08/$25.00 ©2008 IE

were all based on in-house 3D dynamic data sets. There is no 3D dynamic facial expression database publicly available to the research community. It is therefore highly demanded to create an accessible benchmark 4D facial expression database in order to facilitate the new algorithm design, assessment, and comparison for facial expression recognition.

1.1 Existing Face Databases

Essentially, a common testing resource is crucial to fostering research on face information processing. Generally, two aspects are considered for constructing a face database depending on the data modality: *2D* vs. *3D* and *static* vs. *dynamic*. To date, there are a number of standard face databases available to public (as listed in Table 1). A few commonly used databases can also be found in the site [10]. However, the data formats presented are either 2D static, 2D dynamic, or 3D static. To the best of our knowledge, no 3D dynamic face database is readily accessible to public. In this paper, we present a new high-resolution 3D dynamic face database for the research community. Motivated by our existing 3D static facial expression database [40], we continue to focus on creating a 3D dynamic facial expression database for facial behavior research, which extends the work from a static 3D space to a dynamic 3D space.

The database contains 3D videos from 101 subjects with six universal facial expressions, with a total of 606 3D model sequences and approximately 60,600 model frames. The database has been validated through our 3D facial expression classification experiment. In the following, we will introduce the database from its creation to its validation in order to demonstrate the usefulness of such data for facial expression recognition.

Databases	STATIC	DYNAMIC
2D	FERET [23], CMU-PIE [26], JAFFE [15], AR [16], Many others [10], ...	Cohn-Kanade [11], MMI [19], RU-FACS-1 [24], xm2vtsdb [17], UT-Dallas [32], Newcastle [9], Many others [10], [28], [34] ...
3D	3D FRGC [22], DARPA-HumanID [33], xm2vtsdb [12, 17], BU-3DFE [40], …	**None**

Table1. List of some public face databases

2. High-Resolution Data Acquisition

2.1. System setup

We used the Di3D (Dimensional Imaging [8]) dynamic face capturing system to generate our 3D facial expression sequences, which include both 3D model sequences and 2D texture videos. The system consists of two stereo cameras and one texture video camera. The three cameras are placed on a tripod with two lighting lamps separated in the two sides. A blue board and a calibrating board are used for calibration and background segmentation. With two master machines (PCs) running in parallel, the system captures the 3D videos at a speed of 25 frames per second. Figure 1 illustrates the system at work. Each pair of stereo images is processed using a passive stereo photogrammetry approach to produce its own range map. The range maps are then combined to produce a temporally varying sequence of high-resolution 3D images with an RMS accuracy of 0.2 mm. As the 3D range model sequence is captured from the top and bottom cameras, a corresponding 2D texture video captured from the middle camera is also recoded.

Figure 1: 3D dynamic imaging system setup

2.2. Data capture and processing

Each participant was instructed to sit in front of the 3D face capture system at about one and a half meters distance from the cameras. The subjects were instructed to perform expressions dynamically with a certain range of movement that would not go beyond the range of three cameras. With the guidance of a psychologist from our institute, the subject was requested to perform six universal expressions, i.e., *angry, disgust, fear, happiness, sadness, and surprise*. Each expression sequence contains neutral expressions in the beginning and the end. In other words, each expression was performed gradually from neutral appearance, low intensity, high intensity, and back to low intensity and neutral, with each such sequence being approximately 4 seconds in length. For each expression sequence, three videos are captured as shown

in Figure 2: two grey-scale videos (upper and lower) for range data creation, and one color-video for texture generation. As a result, for each subject, six 3D expression model sequences and six corresponding 2D video sequences were created respectively.

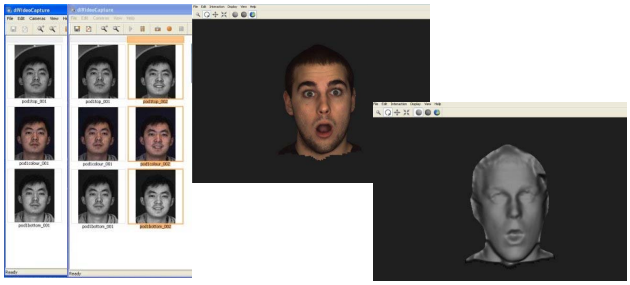

Figure 2: 3D dynamic face data capture and processing. Left: three video data are captured over time from three cameras (upper row, middle row, and lower row). Right: generated 3D videos for illustration (with / without textures).

There were 101 subjects who participated in the face scans, including undergraduates, graduates and faculty from our institute's departments of Psychology, Computer Science, and Engineering, with an age range of 18 – 45 years old. A psychologist helped explain the capture procedure and the performance of the six expressions. The majority of participants were undergraduates from the Psychology Department. All of the participants signed the consent form allowing the distribution of data for public scientific research. The resulting database consists of 58 female and 43 male subjects, with a variety of ethnic/racial ancestries, including Asian (28), Black (8), Hispanic/Latino (3) and White (62). Table 2 is the summary of the database.

# of Subjects	# of Expressions	# of 3D model sequences	# of 2D texture videos	Approximate # of 3D Models
101	6	606	606	60,600

Table 2: Summary of 3D dynamic face expression database

3. Data Organization and Visualization

The database is structured by subjects. Each subject has six prototypic expressions. Each expression has three sequences: a 3D model sequence, a 2D texture sequence, and an AVI video for visualization. Figure 3 illustrates an example of the individual 3D frame models using our developed visualization tool, which displays individual models by geometric shape, texture, and point clouds. As illustrated in Figure 4, the database is searchable by a subject's gender, type of expression, geometric data, texture data, and video data. Each 3D model of a 3D video sequence has the resolution of approximately 35,000 vertices. The texture video has a resolution of about *1040×1329* pixels per frame. The resulting database is in the size of approximate 500Gbytes. Figure 5 shows several samples of the 3D dynamic facial expression videos.

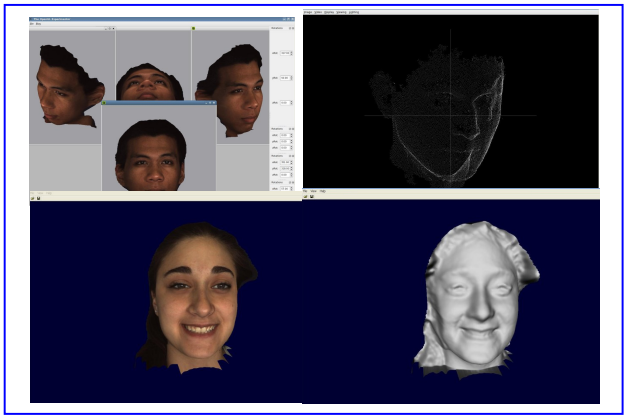

Figure 3: Snapshot of our visualization tool to display frame models with arbitrary views with/without textures or point clouds.

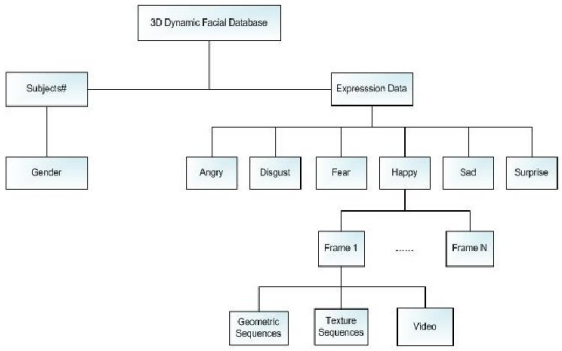

Figure 4: Data organization of the 3D dynamic facial expression database.

4. Validation and Evaluation

The range system produces high quality 3D dynamic facial expression data. However, it is not trivial to use such data. The main challenges are: (1) the data obtained by the range system are in a raw format with a large amount of points, which is "blind" without the information of facial structures or vertex correspondences; (2) each instant capture generates different number of vertices, which increases the difficulty of finding the vertex correspondence across expression sequences. Thus it increases difficulty of tracking and analyzing the motion of facial surfaces over time.

We proposed a 3D spatio-temporal analysis approach to track the range model vertices based on a tracking model

(e.g., a generic model). Here we give a brief description of the approach (For details, please refer to [29]).

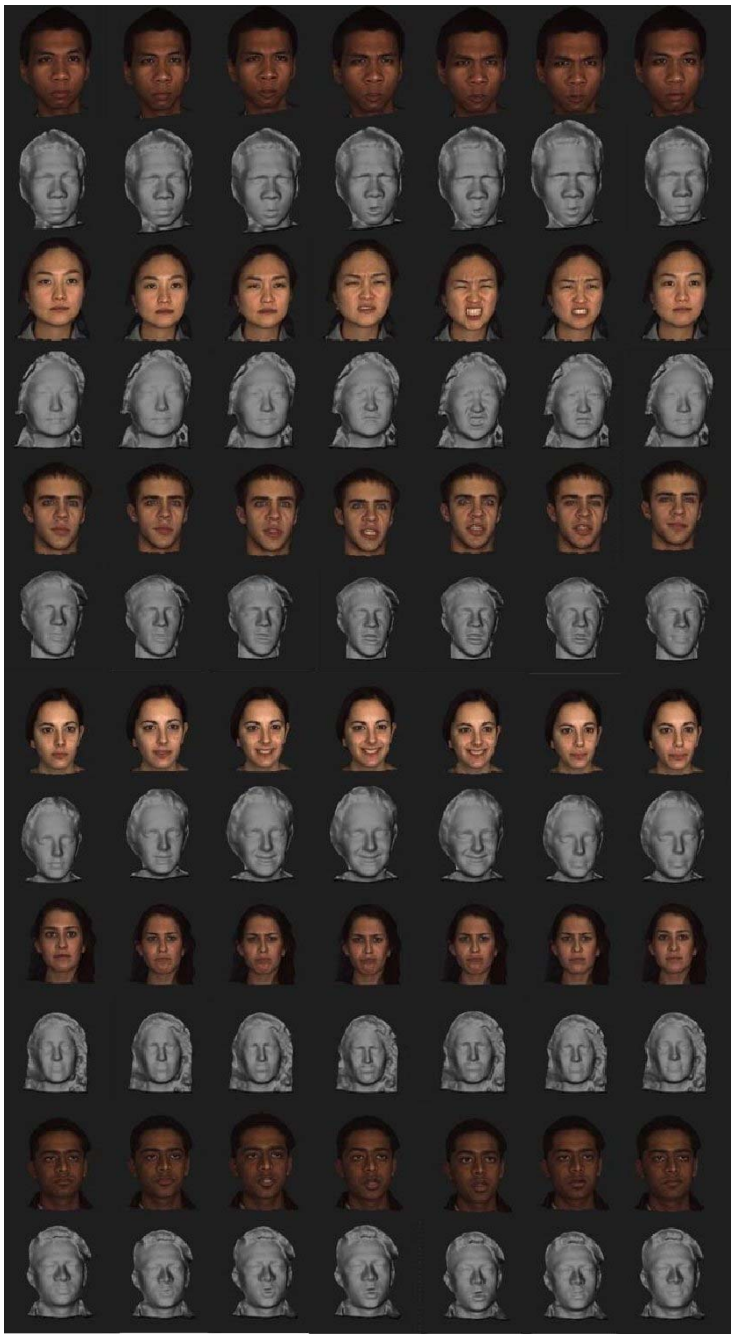

Figure 5: 3D-video samples displayed in the format of textured models and shaded models. Subjects from top to bottom exhibit expressions *angry* (male, black), *disgust* (female, East-Asia), *fear* (male, White), *happy* (female, White), *sad* (female, Latino), and *surprise* (male, India), respectively.

Our validation method is composed of two major components: 1) vertex-level correspondence establishment based on a 2D active appearance model and generic model matching; 2) expressive model labeling based on primitive surface feature classification and an HMM-based classifier [29].

4.1. Vertex level correspondence and tracking

We defined 83 feature points around the facial areas of eyes, nose, mouth, eyebrows, and chin contour at the initial frame of a video sequence (Figure 6 shows an example). Then we applied the active appearance model [6] to track the 83 points in the texture sequence. Since the range model sequence is aligned with the corresponding 2D video textures from the output of the 3D imaging system, the features defined on 2D textures can be exactly found or mapped to the 3D range models. Given the established alignment between 2D-textures and 3D-models from the Di3D dynamic system [8], the tracked points can be projected to the corresponding 3D range models and generated a set of 3D feature vertices. Then we applied a generic model to adapt to the range model based on the 83 control points using a radius-based model interpolation approach.

The adapted generic model represents each frame of range model. More importantly, the adapted models have the same number of vertices across the corresponding 3D video sequence. Thereby, the vertex correspondence across the range model sequence is established. The vertices on the adapted model can be easily tracked by finding the displacement of tracked vertices of two neighboring frames. Figure 6 illustrates an example showing a 3D range sequence with tracked feature points and the adapted 3D models.

Figure 6: A sample 3D sequence with tracked 83 points (upper row) and the adapted 3D models (lower row).

4.2. Expression analysis and recognition

To validate the usefulness of the data, we also applied a Hidden Markov Model to learn and classify the facial expressions across 3D dynamic model sequences. We propose to use a two-dimensional HMM (2D-HMM) [29] to learn the temporal dynamics and spatial relationships of facial regions. We labeled facial surface using a surface curvature classification approach (similar to the primitive feature classification approach in [35]), and classified each vertex to one of the eight surface labels. The label map of a face region forms the face descriptor, which is a vector

composed of the vertices' labels on the face region of a model. The 6-state 2D-HMM is then used to learn the face temporal variation across each set of 6 frames. We conducted a person-independent experiment on 60 subjects. Following a 10-fold cross-validation procedure, we randomly partitioned the 60 subjects into two subsets: one with 54 subjects for training and the other with 6 subjects for test. The experimental paradigm guarantees that any subject used for testing does not appear in the training set. The tests were executed 10 times with different partitions, and achieved an average correct recognition rate of 90.44% for distinguishing 6 prototypic expressions. To further evaluate our approach, we conducted a comparison study by implementing the 3D static model based approach using geometric primitive features [35] and the 2D texture based approach using Gabor-wavelet features [15]. The average recognition rates for the two compared approaches are 53.24% and 63.72% respectively.

4.3. Future development

It is worth noting that our current database was designed for the task of facial expression recognition. There are still some limitations on the database in terms of data variety, data quality, data quantity, data usability, and its applicability for other applications. We will address the following aspects for the future development.

1) *Variety:* Limited by the storage capacity and the processing capability, the current system can only generate short 3D videos for every capture, which makes it hard to capture spontaneous expressions using any elicitation method. We are currently expanding the hardware setup using multiple processors and larger storage space in an attempt to capture a greater variety of non-deliberate facial expressions for a longer period of time. As an option, for example, each subject can choose to perform mixed expressions freely so that various types of expressions can be included in a longer video sequence. In addition, we will collect data with more types of subtle expressions in order to evaluate the sensitivity of 3D motions versus 2D motions for action unit (AU) detection.

2) *Quality*: Limited by the range of capture with the current imaging system, the dramatic facial pose change could cause the partial facial surface missing or occlusion. To allow for the performance evaluation of algorithms for 3D tracking and expression classification with respect to the partial face occlusion or partial missing data, we plan to systematically capture 3D dynamic data under multiple specified poses. As such the collected ground-true data can be used for assessment of any new algorithms. To improve the data quality, we will seek to expand the current system with more distributed cameras for a wider range of views.

3) *Quantity and Applicability*: Although the current database was designed for facial expression recognition task, it is applicable to testing face recognition algorithms. However, the size of the database (101 subjects) is still small with respect to the requirement from the face recognition task. We plan to expand the database to a larger scale in order to meet the requirement of real world applications.

4) *Usability*: One of the big challenges facing the 3D dynamic database is the storage of a huge amount of data, including both sequential geometric data and sequential texture data. Our current database has the size of about 500 gigabytes. With the future expansion, the data size will be on the order of terabytes. The larger amount of the data increases the difficulty of data organization, annotation, distribution, processing, and evaluation. We will address this issue as a future work in order to make it easier to manage and search data. An automatic 3D model registration and annotation approach will be developed so that the temporal segments of facial expression sequences could be parsed and archived. Another issue to be investigated is the data representation or compression for both geometric data and texture data.

5. Conclusion

We have developed a 3D dynamic facial expression database, which is made available to the research community. Such a database can be a valuable resource in the research and development of applications in security, HCI, telecommunication, entertainment, cognition and psychology research, and biomedical applications. The limitations and issues discussed in the previous section give rise to a number of new tasks for our future work so that our facial expression research could be moved towards a more realistic scenario.

Acknowledgement

This material is based upon the work supported in part by the National Science Foundation under grants IIS-0541044, IIS-0414029, and the NYSTAR's James D. Watson Investigator Program. We would like thank Gina Shroff and Peter Gerhardstein of the Department of Psychology, Jeff Rasmussen, Tadeusz Jordan, Leah Haas, David Finley, Janette Wamber, and Titi Gu of the Computer Science Department for the help during the process of creating the database.

References

[1] M. Bartlett, G. Littlewort, I. Fasel, J. Chenu, and J. Movellan, Fully automatic facial action recognition in spontaneous behavior. In *FGR06*, p223-228, 2006.

[2] K. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *CVIU*, 101(1):1–15, 2006.

[3] Y. Chang, M. Vieira, M. Turk, and L. Velho. Automatic 3D facial expression analysis in videos. In *ICCV05 Workshop on Analysis and Modeling of Faces and Gestures*.

[4] I. Cohen, N. Sebe, A. Garg, L. Chen, and T. Huang. Facial expression recognition from video sequences: temporal and static modeling. *CVIU*, 91(1), p160-187, 2003.

[5] J. Cohn, Foundations of human computing: facial expression and emotion, *International Conference on Multimodal Interfaces*, 233-238, 2006.

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

[7] R.Cowie, E. Douglas-Cowie, et al. Emotion recognition in human computer interaction. *IEEE Signal Processing Magazine* 18 (1). 2001.

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

[9] E. Douglas-Cowie, R. Cowie and M. Schroder. A new emotion database: considerations, sources and scope. Proc. of the ISCA ITRW on Speech and Emotion, Newcastle, 2000, pp. 39-44.

[10] M. Grgic and K. Delac, Face Recognition Homepage, http://www.face-rec.org/

[11] T. Kanade, J. Cohn, and Y. Tian, Comprehensive database for facial expression analysis, *International Conference on Automatic Face and Gesture Recognition*, 46-53, 2000.

[12] J. Kittler, A. Hilton, et al, 3D assisted face recognition: a survey of 3D imaging, modeling and recognition approaches, in *CVPR 2005 Workshops on A3DISS*.

[13] S. Li and A. Jain. *The Handbook of Face Recognition*, 2004.

[14] X. Lu and A. Jain, Deformation modeling for robust 3D face matching, *IEEE Trans. on PAMI*, 30(8):1346-1356, 2008.

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

[16] A. Martinez and R. Benavente, The AR face database, http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html.

[17] K. Messer, J Matas, J Kittler, et al. Xm2vtsdb: The extended m2vts database. In *International Conference of AVBPA*, March 1999.

[18] M. Pantic and M. Bartlett, Machine analysis of facial expressions. *Face Recognition*, Book edited by: Kresimir Delac and Mislav Grgic, ISBN 978-3-902613-03-5, pp.558, I-Tech, Vienna, Austria, June 2007

[19] M. Pantic, Man machine interaction group. Imperial College London. http://www.mmifacedb.com/

[20] M. Pantic, *et al*. Automatic analysis of facial expressions: the state of the art. *IEEE Trans. on PAMI*, 22(12), 2000.

[21] M. Pantic, N. Sebe, J. Cohn, and T. Huang. Affective multimodal human computer interaction. In *Proc. ACM Intl. Conf. Multimedia*, pages 669–676, 2005.

[22] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *In IEEE CVPR, 2005*.

[23] P. Phillips, H. Moon, P. Rauss, et al., The FERET evaluation methodology for face recognition algorithm. *IEEE Trans. PAMI*, 22 (10), 2000.

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

[25] N. Sebe, M. Lew, I. Cohen, Y Sun, T. Gevers, and T. Huang. Authentic facial expression analysis. *Image Vision Computing*, 12(25), 1856-1863, 2007.

[26] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination and expression database. *IEEE Trans. PAMI*, 25(12), 2003.

[27] P. Sinha, B. Balas, Y. Ostrovsky, R. Russell, Face Recognition by Humans: 19 Results All Computer Vision Researchers Should Know About, *Proceedings of the IEEE*, 94(11):1948-1962, 2006

[28] J. Skelley and et al. Recognizing expressions in a new database containing played and natural expressions. *IEEE ICPR* 2006.

[29] Y. Sun and L. Yin, Facial expression recognition based on 3D dynamic range model sequences, *The 10th European Conference on Computer Vision, (ECCV'08)*. Marseille, France.

[30] Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *IEEE Trans. on PAMI*, 23(2), 2001.

[31] Y. Tong, W. Liao, and Q. Ji. Facial action unit recognition by exploiting their dynamic and semantic relationships. *IEEE PAMI*, 10(29), 1683-1699, 2007.

[32] A. O'Toole, J. Harms, et al. A video database of moving faces and people. *IEEE Trans. PAMI*, 27(5), 2005.

[33] USF DARPA Human ID 3D face database. Courtesy of Prof. Sudeep Sarkar, University of South Florida, Tampa, FL.

[34] F. Wallhoff B. Schuller, et al., Efficient recognition of authentic dynamic facial expressions on the Feedtum database, *IEEE ICME 2006*. p493-496.

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

[36] P. Wang and R. Verma *et al*. Quantifying facial expression abnormality in schizophrenia by combining 2d and 3d features. In *CVPR07*, 2007.

[37] S. Wang, Y. Wang, X. Gu, and D. Samaras, *3D surface matching and recognition using conformal geometry*. IEEE Conf. on CVPR, 2006. p2453-2460.

[38] Y. Wang, D. Samaras, D. Metaxas, A. Elgammal, *et al*. High resolution acquisition, learning, transfer of dynamic 3D face expression. In *Eurographics* 2004.

[39] P. Yang, Q. Liu, and D. Metaxas. Boosting coded dynamic features for facial action units and facial expression recognition. In *CVPR07*, 2007.

[40] L. Yin, X. Wei, J. Wang, Y. Sun, and M. Rosato, A 3D facial expression database for facial behavior research. *IEEE International Conference on Automatic Face and Gesture Recognition.* 2006.

[41] L. Zalewski and S. Gong. Synthesis and recognition of facial expressions in virtual 3d views. In *FGR 2004*.

[42] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions. *International Conference on Multimodal Interfaces*, 126-133. 2007

[43] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4), Dec. 2003.