A VIDEO DATABASE FOR FACIAL BEHAVIOR UNDERSTANDING

D. Freire-Obregón and M. Castrillón-Santana.

SIANI, Universidad de Las Palmas de Gran Canaria, Spain
dfreire@iusiani.ulpgc.es, mcastrillon@iusiani.ulpgc.es

Abstract. The use of new technologies in order to step up the interaction between humans and machines is the main proof that faces are important in videos. Therefore we suggest a novel Face Video Database for development, testing and verification of algorithms related to face-based applications and to facial recognition applications. In addition of facial expression videos, the database includes body videos. The videos are taken by three different cameras, working in real time, without varying illumination conditions.

Keywords: database, facial expression, multimedia, micro expressions, face tracking, digital video.

1 Introduction

Facial expression recognition has been studied for over 30 years. Especially in recent years, it has become one of the most active research areas in pattern recognition, computer vision, and psychology due to the extensive public expectation of its wide potential applications in public security, financial security, entertainment, intelligent humancomputer interaction, soft biometrics, etc. In addition, much progress has been made in the past few years. However, the facial expression recognition remains a research area far from maturity, and its applications are still limited. Therefore, it is becoming more and more significant to discover the bottleneck and the valuable future research topics by evaluating and comparing the potential facial expression recognition technologies exhaustively and objectively.

People look very different depending on a number of factors. Perhaps the three most significant factors are: (1) the pose; i.e. the angle at which you look at them, (2) the illumination conditions at the time, and (3) their facial expression; i.e. whether or not they are smiling, etc. Although several other face collections exist with a large number of subjects [2], and with significant pose and illumination variation [1], we felt that there was still a need for a collection consisting of a fairly large number of subjects, from different types of cameras, from several different poses, with a different frame rate, under significant illumination variation, and with a variety of facial expressions. This work responds to a highly demand of the researchers on more datasets that provide them the

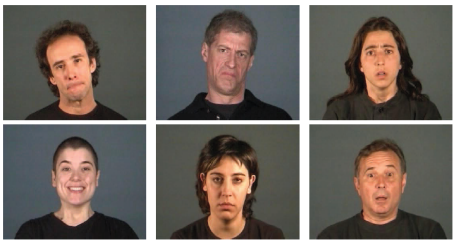

Fig. 1. Six basic emotions presented by six different individuals; extracted from the DaFEx database [4]. The displayed emotions are, from left to right and top to bottom: angry, disgust, fear, happiness, sadness, and surprise.

ability to benchmark they algorithms as faster and publicable as possible. In fact, the first task the researches face when they are going to develop a new work is the distributed searching. It is composed of three fundamental activities: choosing the specific databases to search, searching the chosen databases, and merging the results into a cohesive response.

On the other hand researchers are not just demanding a huge number of collections but also a higher quality of the datasets. They need databases that allow them to test their work under the most realistic enviroment possible. For example, portable electronic devices such as mobile phones and PDAs are becoming important means to provide wireless access to the Internet and other telecommunication networks, algorithms for facial authentication are required, then it is necessary a collection of facial images taked under controlled or uncontrolled condition by a mobile phone [3]. Same happens with webcams. What if we want to identify ourselves or communicate with computers but just our facial expression, none of us usually have next to our computer a high resolution camera or controlled conditions to communicate with it. For use in the development, training, and testing of facial identity or expression classifiers, appropriate extensive facial databases are required. Most of these databases are non-trivial to create, as they need to be sufficiently rich in both facial expression variety and representative samples of each expression. Moreover, the creators of the database need to make sure that the human models form their true facial expressions when posing.

1.1 Collections for Facial Expression Analysis

According to the elicitation method of emotions, facial expression collections can be categorized into three major classes: induced, acted, and naturalistic. Naturalistic data seems the ideal way to collect data reliable for evaluating life-like affective systems, but the reality is not that straightforward. Having such data is challanging due to several aspects, such as problems of copyright and privacy, need of high developed tools to deal with it, and unreliable ground truth, are some obstacles challenging the employing of such a data to evaluate real-life emotion analysis systems. Between the naturalistic facial expression, mentioned

above, and the acted one lie various emotion induction techniques. On the one hand, about the acted one, most technological research on emotion continues to be based on recordings of actors, skilled [4] or not skilled [6] [7]. That is because of the difficulties of having naturalistic databases. On the other hand, there are also the emotion induction techniques; various established methods such as listening to emotive music, looking at emotive pictures or films, and playing specially designed games. Such data, however, is difficult to tag, as a specific induction methods "emotive image" could elicite disgust by some subjects, while it could trigger fear by others. The reason of this is because some specific emotional states are related to specific physiological patterns neglects the important facts that physiology will vary with action, and that actions associated with the same emotional state will also often vary [8]. That is, most, if not all, peripheral (and to some degree, central) indices of physiological activity will vary as a function of the amount and type of somatic involvement and the accompanying demand for metabolic support. Put yourself, running (or preparing to run) will produce a very different configuration of physiological activity than sitting and observing, with activity in one system (e.g., cardiovascular) dependent, to some degree, on activity in another system (e.g., somatic).

Collection	Year	Subjects	Facial Exp.	Availability	Website
JAFFE	1999	126	6 Basic Emotion & Neutral	Online	[12]
Cohn-Kanade	2000	100	6 Basic Emotion & Neutral	By request	[14]
DaFEx	2004	8	6 Basic Emotion & Neutral 3 Intensity Levels	By request	[15]
MMI	2005	61	6 Basic Emotion	By request	[16]
U. Texas	2005	284	++	By request	[13]

Table 1. Overview of the recording conditions for all databases discussed in this subsection. Cases where the exact number of conditions is not determined (either because the underlying measurement is continuous or the condition was not controlled for during recording) are marked with "++"

In recent years no collections with dynamic stimuli appear to have been available. At the moment, DaFEX (Database of Human Expressions), MMI ("M&M Initiative" Face Database [9]) and University of Texas Database are the most used dynamic basic expression collections by researchers (See Table 1). Most of the collections have been created by asking actors to pose certain FACS action unit configurations on their faces, either after short guidance or after more extensive training. In JAFFE (Japanese Female Facial Expression Database) collection, actors posed emotions freely [5]. Arrangements were such that the actors were able to monitor their faces and take the photographs themselves.

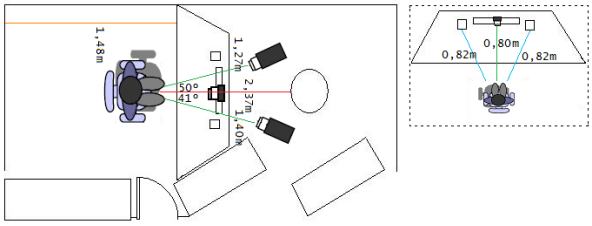

Fig. 2. Configuration for the facial expression experiment. Each subject was presented with 17 successively displayed video sequences. Each sequence length was approximately 20 seconds and the viewing distance to the monitor was 80 cm.

In DaFEX collection [4], professional actors were given short stories depicting certain emotions and asked to pose them by empathizing. The only collection containing spontaneous emotional expressions was the University of Texas collection [10], where subjects were videotaped while watching emotion-evoking films. As discussed earlier, some emotions are extremely difficult to evoke, possibly because of display rules, and it is even more difficult to obtain instances of pure basic emotions. The Cohn-Kanade AU-Coded Facial Expression Database [11] is publicly available from Carnegie Mellon University (CMU). It contains image sequences of facial expressions from men and women of varying ethnic backgrounds. The subjects perform a series of 23 facial displays that include single action units and combinations of action units. Ideally, facial expression collections should be evaluated both with FACS coding and by evaluation studies with subjects, although the latter could be claimed to be more important because they confirm that the collected material is actually perceived as intended. CMU and MMI collections have been only FACS coded. It appears that no evaluations have yet been conducted with the University of Texas collection.

2 Experimental Setup

In order to ensure spontaneity, the acquisition process consists in two basic and parallel steps:

- The subject was presented with multimedia content (short videos) on a screen in front of the three cameras. These short videos were expected to generate emotional states that mapped on the subject's face as a facial expression. For example, to have a subject assume a "smiling" expression, a funny content clip is shown to him/her.
- Data acquisition process by the three cameras. Subjects were asked to watch the multimedia content while they were recording by the three cameras simultaneously.

It is important to mention that in face database collection, faces are sampled in multiple dimensions, such as pose, illumination, expression, aging, etc.

In the case of recording videos, time is a dimension too. In our recording system, we sampled in the following dimensions: motion, pose, image resolution, frame rate, illumination and variations over time. Motion and pose were left participant-dependent. In order to vary image resolution and data rate, we used three different digital cameras from different recording angles. To sample variations over time, we conducted our data collection in two sessions, aiming for a week of separation between sessions. The total collection spanned 4 weeks, 12 sessions. Illumination was unvaried by our environment; a controlled, indoor environment. The indoor environment was fixed with a white background and a lamphouse with a bank diffuser.

To simulate real-world conditions, no glasses or hats were prepared in the room for used as accessories, they were left participant dependent. Subjects used their own accessories to further increase the diversity of the database. About the kind of accessories used by participants, there were mostly glasses and hats. The glasses consisted of sunglasses or eyeglasses (as they were left participant dependent, there was a great diversity of glasses; dark frame glasses, thin and white frame glasses, glasses without frame, etc). On the other hand hats were hoods in almost every case.

The variables that remained constant in this environment were the data rate for each camera and quality of the videos, camera angles, procedure, and the face of each participant. The poses for each sequence will have varied, as they are dependent on the participant.

About the procedure followed by each participant, there were two recording procedures because two different experiments were made for each session: (1) the main one was the facial expression collection, and (2) a second experiment recorded the upper part of the body while each subject raise and lower their arms for once. As it was pointed before, the reason for this second experiment was to take the chance to complete this database regarding other computer vision applications (p.e. clothes detection).

The body experiment took place after the facial expression experiment. Only a pose was recordered for this second experiment. Before starting this experiment, the subjects were asked to stand up, next to the wall, with is their back towards the wall and to lower their arms and place them on either sides of their bodies. Then, the procedure in this case was:

1. The subjects were asked to raise their arms, constantly and slowly, from their waist over their head.
2. After that, the subject must lower their arms back to the initial position next to their waist.

Bearing in mind the characteristics of an induction process, the experimental setup is a critical step of the data acquisition process. Indeed, environment should allow the maximum isolation for subjects to achieve a high concentration level, reporting an air of tranquility to feel confidence to express their emotions in the most natural way. The dimensions of the room is about 4.33m x 1.9m x 1.97m. To record facial expressions with different poses, expressions or accessories, special equipment was configured in the room including three digital cameras and a

Fig. 3. Different camera views with the subject looking right into the screen. It can be appreciated that recording settings are not the same for each camera according to the results this work is interested to achieve. The left pictures stand for the Sony Camera. This camera recorded 126 video sequences (two sessions per participant). The Webcam, the frontal one, also recorded 126 clips. The right pictures belong to the Casio. Due to storage and distribution constraints, only 93 out of 126 video sequences were recorded. Thus, three camera angles (frontal, left and right) are considered. The data could be used for evaluating the robustness of face recognition algorithms across pose and evaluating the pose estimation algorithms.

special spotlight. It is remarkable to point about the illumination configuration that the set is occluded with the aim to avoid distractions of participants during the acquisition process. For obtaining images of participants from three different poses, three different cameras recording simultaneously were used.

About the settings for each camera, it is a remarkable that, despite of the pose for each cam, each camera had different settings. The Sony Handycam HDR-XR520VE has 24-bit true color and data rate of 1920x1080 pixels at 30 frames per second. The Logitech Webcam C250 also has 24-bit true color and data rate of 640x480 pixels at 30 frames per second. The Casio Exilim EX-FH20 , has a maximum resolution of 480x360 pixels and 24-bit true color images (RGB) and capture 210 frames per second. Therefore, two different frame rates were used at the same time. Since we were aiming to capture facial expression, being able to capture and detect micro expressions is an essential step for the research in the field. The disadvantage of using such high rate is that the resulting video is quite heavy. For example, a video that is 07:36 minutes length on the Sony and the Webcam, it takes 53:53 length on the Casio.

The illumination approximated ambient lighting. Specifically, the recording setup consisted of one 1000 watt quartz halogen focusing open face floodlight, mounted with a height of 172 centimeters. A bank diffuser of 50 by 50 cm was used during the experiment. It is important to mention that illumination setup was a challenge because of the Casio's frame rate setting. The selected illumination avoids the flickering effect on this camera during the acquisition process. The light was 2,37 meters from the participant.

3 The Vedana Database

The Vedana facial expression database is divided into two different sets depending on the experiment. The main set contains 345 video sequences of 63 subjects

Fig. 4. Facial expressions responses from different subjects.

for the facial expression experiment and, the second set contains 126 video sequences of 63 subjects for the body experiment. In order to provide a frame for background subtraction, a set of video sequences with no subjects is also supplied (a video sequence per camera and experiment). The images on Figure 3 belong to the main set that is divided into three subsets depending on the camera used:

1. On the left view subset, all images are captured from the Sony Camera.
2. On the frontal view subset, all images are captured from the webcam, which is attached to the screen.
3. On the right view subset, all images are captured from the Casio Camera, which captured 210 frames per second.

The number of participants recorded in the Vedana database is 63, whom all of them came to both sessions. Of these participants, 29% are female and 71% are male. The youngest participant is 18 years old. The oldest participant is 57 years old. The mean age of the participants is 27,49 years old.

In FERET [2] and other tests, aging was always another important factor decreasing the recognition rates. In most face databases, images of one subject captured under different sessions are insufficient or absent because the subjects are hardly traced. In the Vedana database 63 subjects have been recorded in two sessions a week apart.

Vedana has three kinds of occlusion video sequences, involving hair, hoods and glasses. In addition, we recorded whether or not each participant wears glasses, hoods and/or has facial hair. This data, along with gender and age, is also documented. To simulate real-world conditions, no glasses or hats were prepared in the room for used as accessories, they were left participant dependent. Subjects used their own accessories to further increase the diversity of the database. About the kind of accessories used by participants, were mostly glasses and hoods. The glasses consisted of sunglasses or eyeglasses.

Fig. 5. Example sequence of the body experiment.

The dynamic facial expression clips capture emotions such as happiness, sadness or disgust (See Fig. 4). These are common non rigid movements of the face. We employed a simple method to capture dynamic, natural facial expressions. During filming, the subject watched a 8-minute video, which contained scenes from various internet sequences intended to elicit different emotions. The digital stream captured during the 8-minute filming session was scanned subsequently for instances of non rigid facial motions that corresponded to: happiness, sadness, fear, disgust, anger, puzzlement, laughter, surprise, boredom, or disbelief. It is important to note that the expression rating was not done yet. Indeed, without making additional assumptions about how to determine what constitutes a "smile" or "disgust" expression, there can be no ground truth for the expression videos. Thus, researchers are advised to carry out psychological expression-norming procedures prior to making claims about particular facial expressions found in the database. Due to this fact and even if several expressions were recorded for each individual in each session. To edit these sequences in order to have expression segments into small video clips was considered not the best choice. Also the reader must bear in mind that expressions may vary in length; some occurred over a few frames, others lasted many seconds.

For the second experiment, the upper part of the body was recorded while each subject raise and lower his/her arms once. The reason for this second experiment was to take the chance to complete this database regarding other computer vision applications (p.e. clothes detection) (See Fig. 5).

4 Conclusion

We have developed a 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 (Human-Computer interaction), telecommunication, entertainment, cognition and psychology research, and biomedical applications.

Some potential applications of the presented database have been presented throughout this work. To summarize, the following activities could benefit from using the Vedana Facial Expression Database:

1. Evaluation of the robustness of face recognition algorithms to the pose variations.
2. Evaluation of the performance of face/body pose estimation algorithms.
3. Evaluation of face/body recognition algorithms using video sequences as input.
4. Evaluation of face and facial features localization algorithms.
5. Development of either 2D statistical face/body shape models.

Automatic face expression recognition systems find applications in several interesting areas. With the recent advances in robotics, especially humanoid robots, the urgency in the requirement of a robust expression recognition system is evident. As robots begin to interact more and more with humans and start becoming a part of our living spaces and work spaces, they need to become more intelligent in terms of understanding the human's moods and emotions. Expression recognition systems will help in creating this intelligent visual interface between the man and the machine.

5 Dissemination

The database is available from the authors [17]. We maintain a searchable database that will be made available upon request. We will provide a brief key explaining the file naming conventions used with the various file types. This database is for noncommercial use only, as the consent forms signed by the subjects allow use only for research. A small number of subjects have additionally granted permission for their faces to appear in research publications. Requesters of the database will be required to sign a form agreeing to the terms of use and to respecting the limits of the subject's consent. The database is available by download.

References

1. A.S. Georghiades, P.N. Belhumeur and D.J. Kriegman:From few to many: Generative models for recognition under variable pose and illumination. Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (2000)
2. P.J. Philips, H. Moon, P. Rauss and S.A. Rizvi: The FERET evaluation methodology for face-recognition algorithms. Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (1997)
3. MOBIO Project: http://www.mobioproject.org/databases. [Online; accessed 31-March-2010]
4. A. Battocchi, F. Pianesi and D. Goren-Bar: A first evaluation study of a database of kinetic facial expressions (DAFEX). Proceedings of Int. Conf. on Multimodal Interfaces, pp. 214–221. (2004)
5. M. J. Lyons, J. Budynek and S. Akamatsu: Automatic Classification of Single Facial Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1357–1362. (1999)

6. M. J. Lyons, S. Akamastu, M. Kamachi and J. Gyoba: Coding facial expressions with Gabor wavelets. Proceedings of Third IEEE Conf. on Face and Gesture Recognition, pp. 200–205. (1998)
7. A. Savran et al.: Bosphorus database for 3D face analysis. Proceedings of First European Workshop on Biometrics and Identity Management Workshop, pp. 203–208. (2008)
8. M. M. Bradley and P. J. Lang: Motivation and emotion. Cambridge University Press. (2006)
9. M. Pantic and I. Patras: Dynamics of facial expression: Recognition of facial actions and their temporal segments form face profile image sequences. IEEE Transaction on Systems, Man, and Cybernetics, pp. 433–449. (2004)
10. A. O'Toole, J. Harms, S. Snow, D. R. Hurst, M. Pappas, and H. Abdi: A video database of moving faces and people. Vision Res., pp. 3145–3155. (2003)
11. T. Kanade, J. Cohn, and Y. Tian: Comprehensive database for facial expression analysis. Proceedings of the 4th IEEE Int Conf on Automatic Face and Gesture Recognition. (2000)
12. The Japanese Female Facial Expression (JAFFE) Database: http://kasrl.org/jaffe.html. [Online; accessed 11-April-2010]
13. University of Texas Video Database: http://bbs.utdallas.edu/facelab.htm. [Online; accessed 06-April-2010]
14. Cohn-Kanade AU-Coded Facial Expression Database: http://vasc.ri.cmu.edu/idb/html/face/facial_expression/index.html. [Online; accessed 11-April-2010]
15. Database of Human Facial Expressions: http://tcc.itc.it/research/dafex/index.html. [Online; accessed 12-April-2010]
16. MMI Facial expression database: http://www.mmifacedb.com/. [Online; accessed 10-April-2010]
17. Vedana database: http://new-mozart.dis.ulpgc.es/vedana/. [Online; accessed 05-December-2011]