WEB-BASED DATABASE FOR FACIAL EXPRESSION ANALYSIS

*Maja Pantic, Michel Valstar, Ron Rademaker and Ludo Maat**
Delft University of Technology
EEMCS / Man-Machine Interaction Group
Delft, the Netherlands
mpantic@ieee.org, M.F.Valstar@ewi.tudelft.nl, ron@ttvavanti.nl, ludomaat@zonnet.nl

ABSTRACT

In the last decade, the research topic of automatic analysis of facial expressions has become a central topic in machine vision research. Nonetheless, there is a glaring lack of a comprehensive, readily accessible reference set of face images that could be used as a basis for benchmarks for efforts in the field. This lack of easily accessible, suitable, common testing resource forms the major impediment to comparing and extending the issues concerned with automatic facial expression analysis. In this paper, we discuss a number of issues that make the problem of creating a benchmark facial expression database difficult. We then present the MMI Facial Expression Database, which includes more than 1500 samples of both static images and image sequences of faces in frontal and in profile view displaying various expressions of emotion, single and multiple facial muscle activation. It has been built as a web-based direct-manipulation application, allowing easy access and easy search of the available images. This database represents the most comprehensive reference set of images for studies on facial expression analysis to date.

1. INTRODUCTION

Facial expression is one of the most cogent, naturally preeminent means for human beings to communicate emotions, to clarify and stress what is said, to signal comprehension, disagreement, and intentions, in brief, to regulate interactions with the environment and other persons in the vicinity [1, 2]. Automatic analysis of facial expression attracted the interest of many AI researchers since such systems will have numerous applications in behavioral science, medicine, security, and human-computer interaction.

To develop and evaluate such applications, large collections of training and test data are needed [3, 4]. While motion records are necessary for studying temporal dynamics of facial expressions, static images are important for obtaining information on the configuration of facial expressions which is essential, in turn, for inferring the related meaning (e.g., in terms of emotions). Therefore both static face images and face videos are needed.

While the researchers of machine analysis of facial affect are interested in facial expressions of emotions such as the prototypic expressions of happiness, sadness, anger, disgust, surprise, and fear, the researchers of machine analysis of atomic facial signals are interested in facial expressions produced by activating a single facial muscle or by activating a combination of facial muscles. Therefore both kinds of training and test material are needed. For general relevance, the reference images should be scored in terms of AUs defined in the Facial Action Coding System (FACS) [5, 6]. FACS is a system designed for human observers to describe changes in facial expression in terms of observable facial muscle actions (i.e., facial action units, AUs). FACS provides the rules for visual detection of 44 different AUs and their temporal segments (onset, apex, offset) in a video of an observed face. Using these rules, a human coder decomposes a shown facial expression into the specific AUs that produced the expression.

In a frontal-view face image, facial actions such as tongue pushed under the upper lip (AU36t) or pushing the jaw forwards (AU29) represent out-of-image-plane non-rigid facial movements that are difficult to detect. Such facial actions are clearly observable in a profile view of the face. Because the usage of face-profile view promises a qualitative enhancement of AU detection performed, efforts have been made to automate facial expression analysis from face-profile images [7, 8]. Hence, both frontal and profile facial views are of interest for the research in the field.

In spite of repeated calls for the need of a comprehensive, readily accessible reference set of face images that could provide a basis for benchmarks for all different efforts in the research on machine analysis of facial expressions, no such database has been yet created that is shared by all diverse facial-expression-research communities [4]. In general, only isolated pieces of such a facial database exist. An example is the unpublished database of Ekman-Hager Facial Action Exemplars [9]. It has been used by several research groups (e.g., [10, 11]) to train and test their methods for AU detection from frontal-view facial expression sequences. An overview of the databases of face images that have been made

Table 1: Overview of the existing Face Databases

	[3]	[12]	[13]	[14]	[15]	MMI
static images	✗	1581	219	4000	40000	740
videos	2105	✗	✗	✗	✗	848
emotion exp.	✓	smile	✓	4	4	✓
single AU exp.	some	✗	✗	✗	✗	✓
multiple AUs exp.	✓	smile	✓	4	4	✓
AU-coded	✓	✗	✗	✗	✗	✓
no. subjects	210	?	10	126	68	19
subjects' age	18-50	?	?	?	?	19-62
no. ethnicities	3	?	1	?	?	3
gender	♀♂	♀♂	♀	♀♂	♀♂	♀♂
lighting	uni	?	?	var	var	uni
facial hair, glasses	✗	✗	✗	✓	✓	✓
profile view	✗	✓	✗	✗	✓	✓
downloadable	✗	✓	✓	✗	✗	✓
searchable	✗	✓	✗	✗	✗	✓

*The work of M. Pantic is supported by the Netherlands Organization for Scientific Research Grant EW-639.021.202. The work of M.F. Valstar is supported by the Netherlands BSIK-MultimediaN-N2 Interaction project.

0-7803-9332-5/05/$20.00 ©2005 IEEE

publicly available is provided in Table 1. From these, the Cohn-Kanade facial expression database [3] is most comprehensive and the most commonly used database in research on automated facial expression analysis. Two main drawbacks of this facial expression data set are as follows. First, each recording ends at the apex of the shown expression. This makes research of facial expression temporal activation patterns (onset → apex → offset) impossible to conduct using this data set. Second, many recordings contain the date/time stamp recorded over the chin of the subject. This makes changes in the appearance of the chin less visible and motions of the chin obscured by changes of the time/date stamp.

Overall, none of these existing databases contains images of all possible single-AU activations, none contains face images in profile view, and none contains both static images and videos of faces. Also, the metadata (labels) associated with each database object do not identify the temporal segments (onset, apex, offset) of shown AU and emotion facial displays. Finally, except of the PICS database [12], none of the existing databases is either easily accessible or easily searchable. Once permission for usage has been issued, large, unstructured files of material are sent. This lack of easily accessible, suitable, and common training and testing material forms the major impediment to comparing, resolving, and extending the issues concerned with automated facial expression analysis from face images. It is this critical issue that we tried to address by building a novel face image database, which we call the MMI Facial Expression Database[**].

We discuss a number of issues that make the establishment of a readily accessible, benchmark database of facial expression images difficult. Then, we describe the MMI Facial Expression Database and discuss its properties in terms of these issues. A discussion of possible database extensions concludes the paper.

2. FACIAL EXPRESSION DATA BASE

The establishment of an easily accessible, comprehensive, benchmark database of facial expression exemplars has become an acute problem that needs to be resolved if fruitful avenues for new research in automatic facial expression analysis are to be opened. A number of issues make this problem complex.

What kind of samples should be included into the database so that it meets multiple needs of scientists working in the field? As noted above, the benchmark database should include both static and motion images of faces showing prototypic expressions of emotion and various expressions of single AU and multiple AU activation. Images of non-occluded and partially occluded faces, with facial hair and glasses, in various poses (at least in the frontal and the profile view), acquired under various lighting conditions should be included as well.

Another important issue is the distinction between deliberate actions performed on request vs. spontaneous actions not under volitional control. Posed expressions may differ in appearance and timing from spontaneously occurring expressions [17]. Also, while few people are able to perform certain facial actions voluntarily (e.g., AU2), many are able to perform these actions spontaneously [3]. Examples of both categories should be included in the database in order to study the essential question of the difference between these. However, eliciting spontaneous facial behavior represents a research challenge on its own right.

In addition, virtually all the existing facial expression analyzers assume that the input expressions are isolated or pre-segmented, showing a single temporal activation pattern (onset → apex → offset) of either a single AU or an AU combination that begins and ends with a neutral expression. In reality, such segmentation is not available; facial expressions are more complex and transitions from an action or combination of actions to another does not have to involve intermediate neutral state. Examples of both the neutral-expressive-neutral and the variable-expressive-variable behavior should be included in the database in order to study the essential question of how to achieve parsing the stream of behavior.

A crucial aspect of the benchmark database is the metadata that is to be associated with each database object and to be used as the ground truth in validating automated facial expression analyzers. As already noted above, for general relevance, the images should be scored in terms of the AUs and their temporal segments. The interpretations of displayed facial expressions in terms of affective state(s) should be associated with each sample. Note that it takes more than one hour to manually score 100 still images or a minute of videotape in terms of AUs and their temporal segments [5]. Hence, obtaining large collections of AU-coded facial-expression data is extremely tedious and, in turn, difficult to achieve.

How should the samples be collected? First, several technical considerations for the database should be resolved including field of sensing, spatial resolution, frame rate, data formats, and compression methods. The choice should enable sharing the data set between different research communities all over the world.

Further, the database objects should represent a number of demographic variables including ethnic background, gender, and age, and should provide a basis for generality of research findings.

For the acquisition of deliberate actions performed on request, the subjects should be either experts in production of expressions (e.g., trained FACS coders) or individuals being instructed by such experts on how to perform the required facial expressions.

Given the large number of expressions that should be included into the database, provision should be made for individual researchers to append their own research material to the database. However, a secure handling of such additions has to be facilitated. At least, an automatic control of whether the addition matches the specified technical and other formats defined for the database objects should be realized.

How does one facilitate efficient, fast, and secure retrieval? The benchmark database of facial expression exemplars could be valuable to hundreds of researchers in various scientific areas if it would be easy to access and use. A relaxed level of security, which allows any user a quick, web-based access to the database and frees administrators of time-consuming identity checks, can attain such an easy access. However, in this case, nonscientists such as journalists and hackers would be able to access the database. If the database is likely to contain images that can be made available only to certain authorized users (e.g., images of psychiatric patients), then a more comprehensive security strategy should be used. For example, a Mandatory Multilevel Access Control model could be used in which the users can get rights to use database objects at various security levels (e.g., confidential, for internal use only, no security).

[**] The acronym MMI comes from M&M Initiative where the M's are the initials of the two main authors. Although other colleagues joined the development efforts of the main authors, the acronym remained in use.

Fig 1: Examples of static frontal-view images of facial expressions in the MMI Facial Expression Database

3. MMI FACIAL EXPRESSION DATABASE

The MMI Face Database has been developed to address most (if not all) the issues mentioned above. It contains more than 1500 samples of both static images and image sequences of faces in frontal and in profile view displaying various facial expressions of emotion, single AU activation, and multiple AU activation. It has been developed as a web-based direct-manipulation application, allowing easy access and easy search of the available images [16]. The properties of the database can be summarized in the following way (see also the last column of Table 1).

Sensing: The static facial-expression images are all true color (24-bit) images which, when digitized, measure 720×576 pixels. There are approximately 600 frontal and 140 dual-view images (Fig. 1). Dual-view images combine frontal and profile view of the face, recorded using a mirror. All video sequences have been recorded at a rate of 24 frames per second using a standard PAL camera. There are approximately 30 profile-view and 750 dual-view facial-expression video sequences (Fig. 2, Fig. 3). The sequences are of variable length, lasting between 40 and 520 frames, picturing one or more neutral-expressive-neutral facial behavior patterns.

Subjects: Our database includes 19 different faces of students and research staff members of both sexes (44% female), ranging in age from 19 to 62, having either a European, Asian, or South American ethnic background.

Samples: The subjects were asked to display 79 series of expressions that included either a single AU (e.g., AU2) or a combination of a minimal number of AUs (e.g., AU8 cannot be displayed without AU25) or a prototypic combination of AUs (such as in expressions of emotion). They were instructed by an expert (a FACS coder) on how to display the required facial expressions, and they were asked to include a short neutral state at the beginning and at the end of each expression. The subjects were asked to display the required expressions while minimizing out-of-plane head motions. For approximately one fourth of samples, natural lighting and variable backgrounds were used (e.g., Fig 2). For the other three fourths, blue screen background and two high-intensity lamps with reflective umbrellas were used (e.g., Fig. 3).

Metadata: Two experts (FACS coders) were asked to depict the AUs displayed in the images constituting the MMI Face Database. To date, approximately two thirds of the samples have been FACS coded for target actions. In the case of facial-expression video sequences, the coders were also asked to depict the temporal segments of displayed AUs. To date, 169 image sequences have been FACS coded per frame for temporal segments of target actions. Since human observers may sometimes disagree in their judgments and may make mistakes occasionally, the inter-observer agreement is usually measured in the case of data sets used in psychological studies. However, inter-observer agreement was not measured in the case of MMI Facial Expression Database. When a disagreement was observed, we asked the FACS coders to make a final decision on the relevant metadata by consensus.

Retrieval and Inclusion of Samples: In order to allow a quick and easy web-based access to the database and yet to shield parts of the database that can be made available only to certain authorized users, we used a Mandatory Multilevel Access Control model. In this model, the users can get rights to use database objects at various security levels (i.e., administrators, confidential, users).

The database allows easy search according to a number of criteria such as data format, facial view, shown AU, shown emotion, gender, age, etc. The provision of a preview per sample makes the search even easier. The user can inspect the previews before deciding on the desired material to be downloaded.

The current implementation of the database allows easy addition of either new samples or entire databases. However, only the database administrators can perform such additions; an automatic control of whether an addition matches the specified formats defined for the database objects has not been realized yet.

The database has been developed as a user-friendly, direct-manipulation application. The Graphical User Interface of the web-application is easy to understand and to use. On-demand online help and a flash (animation) demonstrating the path that the user

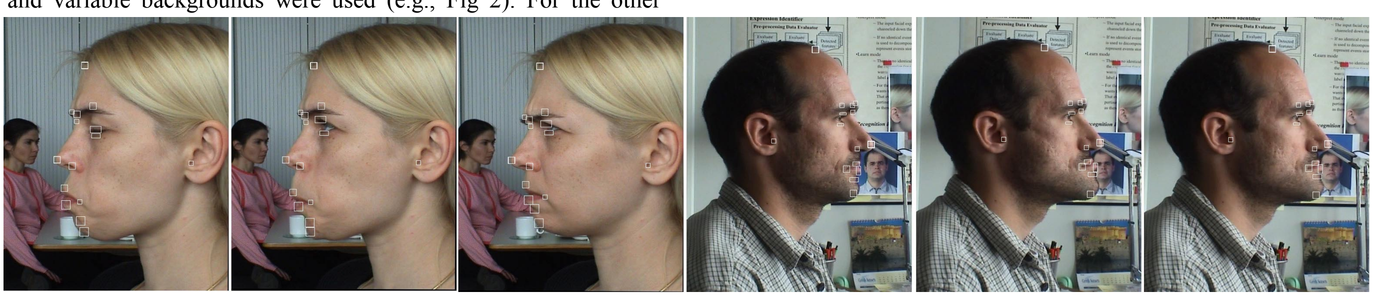

Fig 2: Automated facial fiducial points tracking in profile-face image sequences contained in the MMI Facial Expression Database

Fig 3: Examples of apex frames of dual-view image sequences in MMI Facial Expression Database

should follow to download the required facial expression exemplars form the other aspects of user friendliness.

4. CONCLUSIONS

With more than 1500 samples of both static images and image sequences from 19 male and female subjects of varying ethnic background, in frontal and in profile view, displaying various facial expressions of emotion, single AU and multiple AU activation, the MMI Facial Expression Database provides a large test-bed for research on automated facial expression analysis. The level of facial expression description supports analyses of single AU, AUs combination, and prototypic emotion displays. In contrast to other existing facial expression databases (Table 1) and in the case of 169 samples up to now, analysis of AU temporal activation patterns (onset → apex → offset) is supported as well. Most importantly, the database is easily accessible and easily searchable. In turn, the MMI Facial Expression Database is the most comprehensive, easy to access and to search, reference set of images for studies on facial expression analysis to date. Up to now, it has been successfully used for validation of the systems proposed in [7] and [8], which could not be validated using other existing facial expression databases due to their lack of both face-profile view test samples and test samples showing full AU temporal activation patterns (onset → apex → offset).

Nonetheless, the database has several limitations in its present form. Except for few videos recorded while subjects were preparing to pose, the database contains exemplars of neither spontaneous facial expressions nor variable-expressive-variable behavior (all the recordings picture one or more neutral-expressive-neutral patterns of facial behavior). Yet these are essential for studies on human natural behavior and, in turn, for research on natural human-computer interaction. Also, except for two or three videos, the database does not contain exemplars of partially occluded faces (e.g., by a hand or by a long beard). Out-of-plane head motions captured in the recordings were small; all the database objects are either in nearly-frontal or in nearly-profile facial view. For their relevance in evaluating the achievements of automated methods for facial expression analysis in tackling the relevant problems, images of (temporarily) partially occluded faces, with long beards and moustache, and in various poses (e.g., 30-degree to the side views) should be included into the database. Finally, scoring the temporal segments of AUs is incomplete (it has been conducted only for 169 image sequences); for most sequences only the target (apex) frames rather than the entire sequence have been coded. The coding of the temporal segments of AUs for the rest of the image sequences is currently conducted.

Except addressing the issues listed above, our current and future work will also consider other issues that should be resolved if a common benchmark facial expression database is to be established. These consider the following questions. How could the performance of a tested automated system be included into the database? How should the relationship between the performance and the database objects used in the evaluation be defined?

REFERENCES

[1] J. Russell and J. Fernandez-Dols, *The psychology of facial expression*, Cambridge University Press, 1997.
[2] D. Keltner and P. Ekman, "Facial expression of emotion", *Handbook of Emotions*, Guilford Press, pp. 236-249, 2000.
[3] T. Kanade, J.F. Cohn and Y. Tian, "Comprehensive database for facial expression analysis", *Proc. IEEE FGR*, pp. 46-53, 2000.
[4] M. Pantic and L.J.M. Rothkrantz, "Toward an affect-sensitive multimodal HCI", *Proceedings of the IEEE*, vol. 91, no. 9, pp. 1370-1390, 2003.
[5] P. Ekman and W.V. Friesen, *Facial Action Coding System*, Palo Alto: Consulting Psychologist Press, 1978.
[6] P. Ekman, W.V. Friesen and J.C. Hager, *Facial Action Coding System*, Salt Lake City: A Human Face, 2002.
[7] M. Pantic and L.J.M. Rothkrantz, "Facial action recognition for facial expression analysis from static face images", *IEEE TSMC-B*, vol. 34, no. 3, pp. 1449-1461, 2004.
[8] M. Pantic and I. Patras, "Temporal modeling of facial actions from face profile image sequences", *Proc. IEEE ICME*, vol. 1, pp. 49-52, 2004.
[9] P. Ekman, J.C. Hager, C.H. Methvin and W. Irwin, "Ekman-Hager Facial Action Exemplars", unpublished. San Francisco: Human Interaction Laboratory, University of California.
[10] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, T.J. Sejnowski, "Classifying Facial Actions", *IEEE TPAMI*, vol. 21, no. 10, pp. 974-989, 1999.
[11] Y. Tian, T. Kanade and J.F. Cohn, "Recognizing action units for facial expression analysis", *IEEE TPAMI*, vol. 23, no. 2, pp. 97-115, 2001.
[12] PICS: http://pics.psych.stir.ac.uk
[13] JAFFE: www.mic.atr.co.jp/~mlyons/**jaffe**.html
[14] AR: http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html
[15] PIE: http://www.ri.cmu.edu/projects/project_418.html
[16] The MMI Facial Expression Database: www.mmifacedb.com
[17] K.L. Schmidt and J.F. Cohn, "Dynamics of facial expression: Normative characteristics and individual differences", *Proc. IEEE ICME*, pp. 547-550, 2001.