Boxing the face: A comparison of dynamic facial databases used in facial analysis and animation

Pasquale Dente*
Jacobs University Bremen

Dennis Küster†
Jacobs University Bremen

Eva Krumhuber‡
University College London

Keywords: Facial animation, Facial analysis, Databases, Facebox

Facial animation is a difficult task that is based on an approximation of subtle facial movements [Trutoiu et al. 2014], and that needs to be well grounded in real life dynamic facial behaviour to be convincing. Yet while the endpoints of expressions in still images can be defined relatively precisely using the Facial Action Coding System FACS [Ekman et al. 2002], the design of facial dynamics requires additional high-resolution data (e.g., [Trutoiu et al. 2014]). This is particularly the case for the creation of more naturalistic expressions, i.e., everyday patterns of "partial" and "mixed" movements that depart from simplistic assumptions of omnipresent stereotypic expressions of "basic" patterns of Action Units (AUs). However, for animation designers who do not have the resources to elicit, record, and validate such expressions, the question arises which of the extant and freely available dynamic facial databases might best serve this purpose. One of the most important steps in this decision process is the selection of technically adequate databases that offer sufficient resolution of the face and expressions to allow adequate modelling.

Study Overview: Databases and "Facebox" comparisons: In this work, we examine the technical quality of 16 databases of dynamic facial expressions, including 7 databases claiming to show spontaneous expressions. Dynamic facial databases typically consist of organized sets of video files in various formats and resolutions, although a few instead or additionally include sequences of still images. Of particular interest for the modelling of facial animations would be naturalistic databases with high-quality 3D models constructed on the basis of multiple camera views. However, among the presently available free-access databases compared in this study, only three databases provide such pre-generated 3D models – and none of these deal with spontaneously elicited expressions. In this study, we systematically compared the technical quality of the available video files by estimating the effective size of the visible facial area containing the expressions that we call "FaceBox" (FB). To estimate the effective face size, we wrote a custom software application using OpenCV [Bradski, 2000], and a Haar classifier [Viola, and Jones, 2001]. The software extracted one frame for each video clip at a fixed position (sec 2), then the classifier identified the human face in the image, and returned width and height of the face bounding box.

Results and Discussion: The relative proportion of the visible facial area (%(A)) as estimated by the FB-algorithm, showed a lot of variation, ranging from about 6% (HUMAINE) to up to 57% (STOIC) of the image. More important than these proportional values, however, is arguably the effective amount of pixels² upon which facial animation modelling can be based. Table 1 shows the comparison of the most important technical parameters of the 15 included databases.

Database	Format	FACS	Type	vid. Res.	FB (SD)	% (A)
ADFES	V, S	Y	P	720 x 576	358² (20)	31
BINED	V, A	N	S	720 x 576	173² (45)†	7
BNED	V, A	N	S	352 x 288	124² (39)	15
CK	S	Y	P	640 x 490	288² (25)	26
CK+	S	Y	P	640 x 490	290² (26)	27
D3D-FACS	S, D	Y	P	1280 x 1024	664² (44)	34
DaFEx	V, A	N	P	360 x 288	140² (12)	19
DISFA	V	Y	S	1024 x 768	346² (24)	15
DynEmo	V, A	N	S	768 x 576	308² (6)	21
FG-NET FEED	V	N	S	640 x 480	248² (14)	20
GEMEP (Core)	V, A	Y	P	720 x 576	205² (22)	10
HUMAINE	V, A	N	S	384 x 288	82² (19)†	6
MMI	V, S, A	Y	P,S	1200 x 1600	983² (184)†	50
MPI	V, A, D	N	P	384 x 288	133² (10)	16
MPI Bio	V, D	N	P	384 x 288	174² (9)	27
STOIC	V, A	N	P	256 x 256	193² (16)	57

Table 1: *Comparison of main technical parameters, including the resolution of the "FaceBox" (FB) in relation to the video resolution (vid. Res.). Format: Video (V), audio (A), slides (S), 3D (D). Type: Posed (P), Spontaneous (S). The FB entry (expressed in square-pixels) is followed by the SD in parentheses. †Box estimate is based only on the most suitable subsets.*

For some otherwise promising databases, the estimated visible facial area was only slightly above 100 pixels². In combination with sometimes high compression rates, this suggests that some of these databases may not be suitable for animation modelling despite their conceptual relevance. Notably, a few databases are downloadable at a considerably higher resolution. We suggest that databases should provide the possibility to enlarge the face (i.e. via zoom-in), as well as to provide profile views or 3D models for facial animation design. Overall, spontaneous databases appeared to be somewhat less advanced in respect to these technical parameters.

Summary: Our findings suggest that the creation of naturalistic facial databases deserves more attention. Often, a higher resolution of the visible facial area would be desirable for the dynamic modelling of expressions. Our overview provides at-a-glance suggestions to choose appropriate databases.

References

BRADSKI, G. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

EKMAN, P., FRIESEN, W., AND HAGER, J. 2002. *Facial Action Coding System: Second Edition*. Salt Lake City: Research Nexus eBook.

TRUTOIU, L. C., CARTER, E. J., POLLARD, N., COHN, J. F., AND HODGINS, J. K. 2014. Spatial and temporal linearities in posed and spontaneous smiles. *ACM Transactions on Applied Perception (TAP)*, *11*, 12.

VIOLA, P., AND JONES, M. 2001. Rapid Object Detection using a Boosted Cascade of Simple Features. In *Computer Vision and Pattern Recognition (CVPR 2001). Proceedings of the 2001 IEEE Computer Society Conference on*, vol.1, IEEE, I-511.