A Web Survey for Facial Expressions Evaluation

Matteo Sorci Gianluca Antonini
Jean-Philippe Thiran
Ecole Polytechnique Federale de Lausanne
Signal Processing Institute
Ecublens, 1015 Lausanne, Switzerland
{Matteo.Sorci,Gianluca.Antonini,JP.Thiran}@epfl.ch

Michel Bierlaire
Ecole Polytechnique Federale de Lausanne, Operation Research Group
Ecublens, 1015 Lausanne, Switzerland
Michel.Bierlaire@epfl.ch

June 9, 2008

Abstract

In the recent years, researchers have investigated several methods of facial expression analysis. Their interest has been to apply their algorithms to sets of images labelled by a restrained number of experts. In order to deal with the generalizability of the proposed methods, databases of numerous facial expressions images have been collected. Less attention has been given to the experts. In this paper we present a web-based survey aiming at collecting the judgment on different facial expressions of a heterogeneous number of experts. In order to provide the researchers with a common set of features for a fair comparison of algorithms, the collected data combine the participants choices and a set of measures computed on the images. The resulting database consists, at the moment, of 40704 annotations from 1785 experts.

1 Introduction

Facial expressions are probably the most visual method to convey emotions and one of the most powerful means to relate to each other. A typical automatic system for the recognition of facial expressions is based on a representation of the expression, learned from a training set of pre-selected meaningful features. The learning process relies on the labels associated by an expert or a group of experts to the training samples. The experts are asked to associate each images in the training set to one of the expressions we are dealing with. In other words we must have label makers (the experts) reliable enough and who have *strong knowledge* of the problem in order to ensure the correctness of what we are trying to learn. What is really important is to *how get and use this knowledge*. The facial expressions evaluation survey has been created in order to find a way to extract this knowledge directly from the experts. In the issue of expressions evaluation every single human can be considered as an expert and gives his/her contribution in building this "common sense knowledge". The survey aims at collecting a dataset created by a population of human observers, from all around the world, doing different jobs, having different cultural backgrounds, ages and gender, belonging to different ethnic groups, doing the survey from different places (work, home, on travel ...). This heterogeneity in the respondent population will give researchers the opportunity to investigate what are (part of) the human factors which play different roles in the perception of human expressions. At the same time, it will provide hints to understand what facial features are important and what are their impact on the expression recognition task performed by different people. This is important for most of the human-human interactions, given that

> *"... the face is the most extraordinary communicator, capable of accurately signaling emotion in a bare blink of a second, capable of concealing emotion equally well... "*
>
> *Deborah Blum*

Finally, the analysis of the survey data will be able to provide insights for Human-Computer Interaction applications. Indeed, any prior model built on real data can be employed in order to improve the design of an automatic human expression recognition system. We include in the dataset several facial measures computed on the images presented to the participants and detailed in Section 4. This is designed to provide researchers with a common set of features for a fair comparison of algorithms. The rest of the paper is structured as follows: in Section 2 we describe the facial expressions images used in the survey. In Section 3 a detailed description of the web-based survey is given. The set of provided facial features is tackled in Section 4 and a summary of the collected data is reported in Section 5. Conclusions are finally reported in Section 6.

2 Images Database

Construction of a good database of facial expressions requires time and training of subjects. Only a few of such databases are available, such as the Cohn-Kanade Database [8], JAFFE [9] and most recently the MMI database [11]. The images used in the survey comes from the Cohn-Kanade Database [8]. The database consists of expression sequences of subjects, starting from a neutral expression and ending most of the time in the peak of the facial expression. Subjects are university students enrolled in introductory psychology classes. They ranged in age from 18 to 30 years. Subjects were instructed by an experimenter to perform a series of 23 facial displays. Six of the displays were based on descriptions of prototypic emotions (i.e, happiness, anger, fear, disgust, sadness and surprise). There are 104 subjects in the database and only 11 of them gave the consent for publications. The subset of the Cohn-Kanade Database used in this survey consists of the 1274 images of these 10 subjects (9 women and 2 men).

3 On-line survey

The survey is available in three languages (English, Italian and French). At the beginning of the survey and only once, the participant has to create a new account and insert a few personal information, as shown in Figure 1(a). The socio-economics fields are important in order to segment the labeller population based on different background

(a) (b)

Figure 1: On line survey interface a)Socio-economic form;b)Image annotation interface

knowledge, age, occupation and education. The complete list and description of the socio-economics characteristics is reported in Table 1. The user can guarantee her own privacy choosing freely his own username and password. The data are treated confidentially and only for scientific purposes. Anyway, most of the fields include a “None” option for those responders who do not want to answer. The annotation process consists in associate an expression label (among a set of available human expressions) to each image presented to the survey’s participant. In the list of the available expressions we included, in addition to the seven prototypic emotions (happiness, surprise, fear, anger, disgust, sadness) postulated by Ekman [3], the “I don’t know” and “Other” options. The last two options have been introduced in order to deal with images extremely ambiguous to the participant. A simple and intuitive interface, see Figure 1(b), has been designed in order to facilitate the annotation process. The survey can be stopped whenever the participant wants by logging off and restarted from the first unlabelled image at her next login. Each participant can take part to the survey as many times as she wants.

4 Features : description and extraction

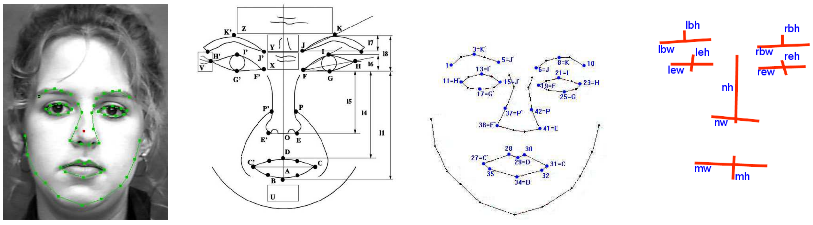
. The survey, described in the previous paragraph, provides the raw data capturing the participants perception of facial expressions. This raw data consists on a set of facial expressions images (the Cohn-Kanade images) and the set of participants choices among the nine options. In order to provide a valuable set of features together with the participants choices we identify and extract some facial visual cues helping in describing an expression. The Facial Action Coding System (FACS) [4] is nowadays the *de facto* standard to describe changes in facial expressions in terms of facial muscle actions (i.e., facial action units, AUs). Inspired by the FACS and by the EMFACS [6], the Ekman dedicated system for emotion-specified expressions, we compute the first set of features included in the database. Zhang and Ji [7] group AUs of facial expressions as primary AUs and auxiliary AUs, see Table 2. The primary AUs refer to those AUs or AU combinations that univocally describe one of the 6 expressions. The auxiliary AUs provide an additional support to the expression description. Additionally, changes in facial transient features, such as wrinkles and furrows, also provide support cues to infer certain expressions. In order to transform the AUs in a set of quantitatively measures Zhang and Ji translate these appearance changes descriptors in a set of

Variable	Description
UserID	Unique identifier for each participant.
UserGender	1 if male, 0 otherwise
UserBirthDate	Age in years
UserOccupation	Occupation (00 = None, 01 = Medical, 02 = Educational, 03 = Management, 04 = Scientific, 05 = Engineering, 06 = Technical, 07 = Rural, 08 = Other)
UserFormation	Education (04 = High School, 05 = University, 06 = PhD, 07 = Other)
UserEthnic	Ethnic (00 = None, 01 = White, 02 = Black, 03 = Asian, 04 = Mixed White-Black, 05 = Mixed White-Asian, 06 = Mixed Asian-Black , 07 = Other)
UserRegion	Continent partecipant belongs to (00 = None, 01 = Africa, 02 = Antartica, 03 = Asia, 04 = Australia, 05 = Europe, 06 = North America, 07 = South America)
UserScienceKW	Participant scientific knowledge (00 = None, 02 = Behavioral Science, , 03 = Social Science, 04 = Computer Science, 05 = Cognitive Science, 06 = Otehr)
UserLanguage	Web Interface chosen language (01 = French, 02 = English, 03 = Italian)
UserLocation	Participant location (01 = Home, 02 = Work, 03 = Other)

Table 1: *Description of Participant Socio-Economic Variables.*

Emotional Category	Primary Visual Cues					Auxiliary Visual Cues					
	AU	AU	AU	AU	AU	AU	AU	AU	AU	AU	Transient Feature(s)
Happiness	6	12				25	26	16			Wrinkles on outer eye canthi, presence of nasolabial furrow
Sadness	1	15	17			4	7	25	26		
Disgust	9	10				17	25	26			Presence of nasolabial furrow
Surprise	5	26	27	1+2							Furrows on the forehead
Anger	2	4	7	23	24	17	25	26	16		Vertical furrows between brows
Fear	20	1+5	5+7			4	5	7	25	26	

Table 2: The association of six emotional expressions to AUs, AU combinations, and Transient Features (from [7])



(a) (b) (c) (d)

Figure 2: a) Facial landmarks (55 points); b) the geometrical relationship of facial feature points, where the rectangles represent the regions of furrows and wrinkles; c)the corresponding points on the face mask obtained with the AAM (from [7]); d) Featural descriptors used in the definition of the EDUs

geometrical relationships of some facial feature points. We use an Active Appearance Model (AAM) [2] with 55 landmarks, as the ones in Figure 2(a) and 2(c), to represent and detect the face in images. AAM is a statistical-based method for matching a combined model of shape and texture to unseen faces. Figure 2(c) shows the relations between the features points suggested by Zhang (Figure 2(b)) and the landmarks automatically extracted by AAM. Table 4 lists the set of distances and angles linguistically reported in Table 2 in terms of landmarks points. The last 4 entries of Table 4 refer to the four transient features and their related measures. Transient wrinkles and furrows are the result of facial muscles movements. These movements produce small ridges perpendicular to the muscular motion direction in certain face regions. The regions of facial wrinkles and furrows are indicated by rectangles in Fig. 2(b). The change of wrinkles in the region $\square X$ is directly related to AU9 (Nose Wrinkler). The furrows in the regions $\square Z$, $\square Y$, $\square V$, $\square U$ provide diagnostic information for the identification of AU2 (Outer Brow Raiser), AU4 (Brow Lowerer), AU6 (Cheek Raiser), and AU17 (Chin Raiser), respectively. The presence of furrows and wrinkles on an observed face image can be determined by edge feature analysis in the areas where transient features appear. In order to detect these features, an edge detection with embedded confidence, proposed by Meer and Georgescu [10], is used. The detection is successively refined by analysing the direction of the extracted edges. Referring to Figure 2(b), wrinkles in regions $\square Z$ and $\square X$ should be mostly horizontal while those in region $\square Y$ mostly vertical . Figure 3 shows examples of transient feature detection. The ratio between edge pixels (wrinkles) and background pixels (skin) is used to measure wrinkles in regions $\square X$ and $\square Y$. For the forehead wrinkles in $\square Z$ and for for nasolabial furrow binary presence variables are adopted: these variables are equal to 1 if the corresponding fur-

AUs	Facial Visual Cues
AU1	$\angle FHJ$, $\overline{JF}$ increased OR $\overline{JF}$ increased, $l8$ nonincreased
AU2	$l8$ increased and $\overline{JF}$ nonincreased furrow in $\square Z$ increased
AU4	$l8, \overline{FJ}, \overline{JJ'}, \overline{FP}, \overline{F'P'}$ decreased, $\angle HFI$ increased and wrinkle in $\square Y$
AU5	$l6, \overline{JF}$ and $\overline{JJ'}$
AU6	nasolabial furrow presence and wrinkle in $\square V$
AU7	$\angle HFI$ nonincreased and $\angle HGF$ increased
AU9	wrinkle increased in $\square X$ nasolabial furrow presence OR $\overline{PF}, \overline{FJ}$ decreased
AU10	$l4$ decreased and $\lvert\overline{FC} - \overline{F'C'}\rvert$ increased, nasolabial presence OR $\overline{OD}$ decreased, $\overline{DB}, \overline{C'C}$ increased
AU12	$\overline{FC}, \overline{F'C'}$ decreased, $\overline{CC'}$ increased, $\overline{GI}$ nonincreased
AU15	$\overline{FC}, \overline{F'C'}, \overline{CC'}$ increased
AU16	$\overline{OD}$ nonchange, $\overline{DB}$ decreased
AU17	$\overline{OB}$ decreased and wrinkle in $\square U$ presence
AU20	$\overline{CC'}$ increased and $\overline{FC}, \overline{F'C'}$ nonchange
AU23	$\overline{DB}, \overline{CC'}$ decreased
AU24	$\overline{DB}$ decreased, $\overline{CC'}$ nonchange
AU25	$\overline{DB}$ increased, $\overline{DB} < T_1, \overline{CC'}$ nonincreased
AU26	$T_1 < \overline{DB} < T_2, \overline{CC'}$ nonincreased
AU27	$\overline{DB} > T_2, \overline{CC'}$ nonincreased

Table 3: *Linguistic description of the AUs of Figure 2 (from [7])*

rows are present in the face, and zero otherwise. We decide to discard the measures on $\square V$ and $\square U$ for two main reasons: 1)the related wrinkles are not always detectable in subjects; 2)they are redundant, since strictly linked to wrinkle and furrows in the retained regions.

In the visual perception community there is a general agreement on the fact that face

FACS Measures	**Measures on mask** 2(c)
$\overline{JJ'}$	6-5
$\overline{JF}$	6-19
$\overline{J'F'}$	5-15
$\overline{KG} \equiv l8$	8-25
$\overline{K'G'}$	3-17
$\overline{GI} \equiv l6$	25-21
$\overline{G'I'}$	17-13
$\overline{PF}$	42-19
$\overline{P'F'}$	37-15
$\overline{FC}$	19-31
$\overline{F'C'}$	15-27
$\overline{FD} \equiv l4$	25-29
$\overline{F'D}$	17-29
$\overline{OD}$	$\left(\frac{39+40}{2}\right)$-29
$\overline{OB}$	$\left(\frac{39+40}{2}\right)$-33
$\overline{DB}$	29-33
$\overline{C'C}$	31-27
$\angle FHJ$	angle between 19, 23 and 6
$\angle F'H'J'$	angle between 15, 11 and 5
$\angle HFI$	angle between 23, 19 and 21
$\angle H'F'I'$	angle between 11, 15 and 13
$\angle HGF$	angle between 23, 25 and 19
$\angle H'G'F'$	angle between 15, 17 and 11
Nose Wrinkles 3(a)	Presence Detection
Eyes Wrinkles 3(b)	Presence Detection
Forehead Wrinkles 3(c)	Presence Detection
Nasolabial Fold 3(d)	Presence Detection

Table 4: *Correspondences between measures on masks 2(b) and 2(c)*

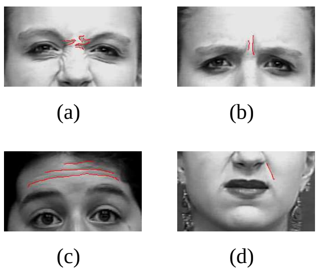


Figure 3: *Transient feature detection: (a) vertical furrows between brows, (b) horizontal wrinkles between eyes, (c) horizontal wrinkles on the forehead, and (d) nasolabial fold.*

recognition is the result of two main sources of information: the featural one coming from individual facial features (mouth, nose, etc.) and the configural one related to the facial layout and configuration of the previous features [5]. The measures extrapolated by the FACS give information about isolated components in a face, providing a featural contribution to face representation. According to the hypothesis of configural encod-

EDU1	$\frac{lew+rew}{leh+reh}$	**EDU8**	$\frac{leh+reh}{lbh+rbh}$
EDU2	$\frac{lbw}{lbh}$	**EDU9**	$\frac{lew}{nw}$
EDU3	$\frac{rbw}{rbh}$	**EDU10**	$\frac{nw}{mw}$
EDU4	$\frac{mw}{mh}$	**EDU11**	EDU2 / EDU4
EDU5	$\frac{nh}{nw}$	**EDU12**	EDU3 / EDU4
EDU6	$\frac{lew}{mw}$	**EDU13**	EDU2 / EDU10
EDU7	$\frac{leh}{mh}$	**EDU14**	EDU3 / EDU10

Table 5: Expressions Descriptive Units

ing, the spatial relationships between facial components provide additional sources of information in the analysis of facial expressions. In order to exploit the combination of these two useful sources we have decided to add a group of measures encoding the interactions among the featural descriptors showed in Figure 2(d). For that purpose we extract the set of measures, called Expression Descriptive Unit (EDU), reported in Table 5 and introduced by Antonini et al. in [1]. The first 5 EDUs represent, respectively, the eccentricity of eyes, left and right eyebrows, mouth and nose. The EDUs from 7 to 9 represent the eyes interactions with mouth and nose, while the 10th EDU is the nose-mouth relational unit. The last 4 EDUs relate the eyebrows to mouth and nose. The EDUs can be intuitively interpreted. For example, in a face displaying a surprise expression, the eyes and the mouth are usually opened and this can be captured by EDU7 ($eye_{height}/mouth_{height}$). FACS and EDU provide measures of local

Figure 4: Examples of synthesized faces obtained varying the first 5 c parameters from the mean face ($\pm 3std$). The mean values and standard deviations are with respect to the training set of the AAM algorithm.

facial features or areas that are prone to change with facial expressions, but they do not provide a description of a face as a global entity. This information can be obtained considering the appearance vector *c* matching the face in the processed image. Figure 4 shows the effect of varying the first 5 appearance model parameters, included in the features set, showing changes in identity and expression.

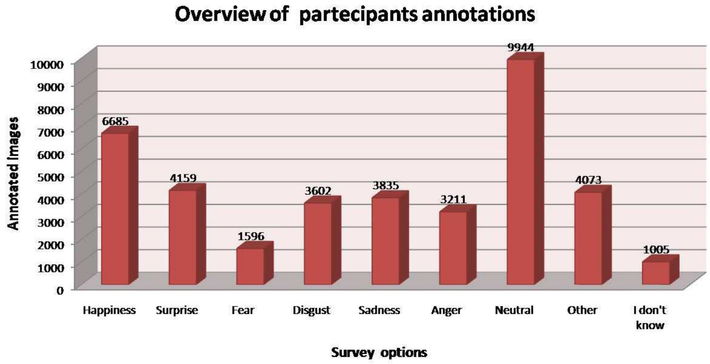


Figure 5: Overview of participants' choices over the whole set of images

5 Collected data

Until now *1784* participants took part to the survey for a total of around *40684* annotated images. In Figure 6 we reported some statistics on the participants. The 6 pie charts show how they are distributed based on their personal information. We can observe that the majority of participants lives in Europe and the "White" group is the most numerous one. However, we have representatives from all the populated continents and from all the ethnic groups.
Concerning participants' cultural background, almost half of the sample has a University Education and all the "Occupation" categories are quite well represented. Computer science and other not listed science branches are the two biggest groups for "Scientific Knowledge". Anyway, a good number of participants with social, behavioural and cognitive science background took part in the survey as well. Figure 5 shows the choice distribution of the participants annotations over the whole set of observations.

6 Conclusions

The goal of the web-based survey presented in this paper is to provide a valuable and complementary dataset to the existing facial expressions databases. Differently from them, the survey does not focus on enlarging the number of facial images, it rather investigates the human perception of expressions. A set of measures on different facial descriptors has also been described. These measures should provide the researchers a test-bed set of features for comparing different expressions analysis algorithms. The survey is still online and we will continue to increase the number of annotations to the supplied data.

References

[1] G. Antonini, M. Sorci, M. Bierlaire, and J. Thiran. Discrete choice models for static facial expression recognition. In J. Blanc-Talon, W. Philips, D. Popescu, and P. Scheunders, editors, *8th International Conference on Advanced Concepts for Intelligent Vision Systems*, volume 4179 of *Lecture Notes in Computer Science*, pages 710–721, Berlin, 2006. Springer Berlin / Heidelberg. ISBN: 978-3-540-44630-9.

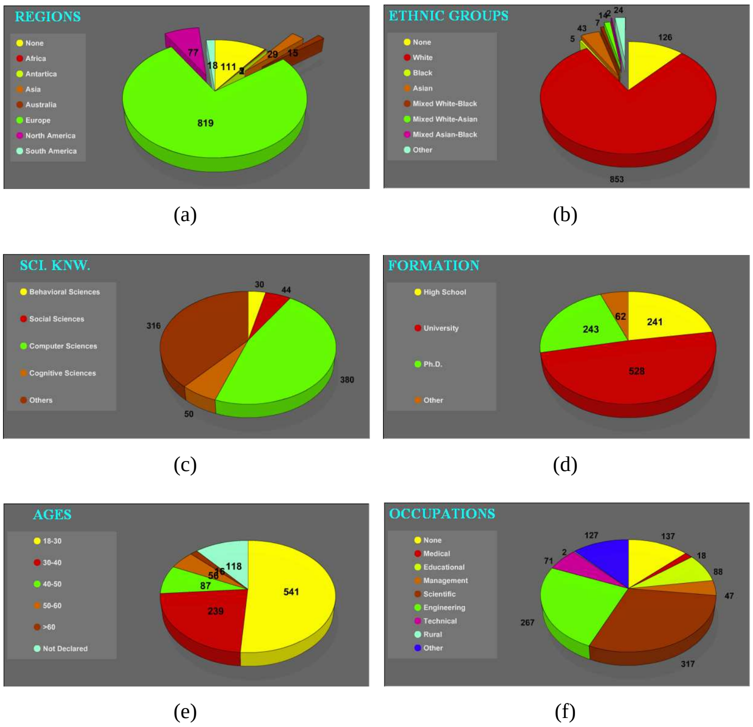


Figure 6: Survey statistics

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23:681–685, June 2001.

[3] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, (17):124–129, 1971.

[4] P. Ekman and W. V. Friesen. *Facial Action Coding System Investigator's Guide*. Consulting Psycologist Press, Palo Alto, CA, 1978.

[5] M. J. Farah, K. D. Wilson, M. Drain, and J. N. Tanaka. What is ”special” about face perception? *Psychol Rev*, 105(3):482–498, Jul 1998.

[6] W. V. Friesen and P. Ekman. Emotional facial action coding system. 1984.

[7] Y. Z. Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. *Transactions on Pattern Analysis and Machine Intelligence*, 27(5):699–714, May 2005.

[8] T. Kanade, J. Cohn, and Y. L. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, pages 46 – 53, March 2000.

[9] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets, 1998.

[10] B. Meer, P.; Georgescu. Edge detection with embedded confidence. *Transactions on Pattern Analysis and Machine Intelligence*, 23(12):1351–1365, Dec 2001.

[11] M. Pantic, M. F. Valstar, R. Rademaker, and L. Maat. Web-based database for facial expression analysis. In *ICME*, pages 317–321. IEEE, 2005.