

© [2006] IEEE. Reprinted, with permission, from [Hatice Gunes and Massimo Piccardi, A Bimodal Face and Body Gesture Database for Automatic Analysis of Human Nonverbal Affective Behavior, EPattern Recognition, 2006. ICPR 2006. 18th International Conference on (Volume:1 ), 20-24 Aug. 2006]. This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Technology, Sydney's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org). By choosing to view this document, you agree to all provisions of the copyright laws protecting it

# **A Bimodal Face and Body Gesture Database for Automatic Analysis of Human Nonverbal Affective Behavior**

Hatice Gunes and Massimo Piccardi

*Computer Vision Research Group, Faculty of IT, UTS, Australia  
{haticeg,massimo}@it.uts.edu.au*

## **Abstract**

*To be able to develop and test robust affective multimodal systems, researchers need access to novel databases containing representative samples of human multi-modal expressive behavior. The creation of such databases requires a major effort in the definition of representative behaviors, the choice of expressive modalities, and the collection and labeling of large amount of data. At present, public databases only exist for single expressive modalities such as facial expression analysis. There also exist a number of gesture databases of static and dynamic hand postures and dynamic hand gestures. However, there is not a readily available database combining affective face and body information in a genuine bimodal manner. Accordingly, in this paper, we present a bimodal database recorded by two high-resolution cameras simultaneously for use in automatic analysis of human nonverbal affective behavior.*

## **1. Introduction**

A firm perceptual coupling between the communicating partners is mandatory if a successful and truly bidirectional communication between the users and the computers is to be achieved. In today's technology, the users are able to combine various modalities to communicate their intentions to the system, e.g., textual input (i.e. with keyboard and mouse), spoken, and more recently, multimodal input. Current research in affective computing and intelligent interaction also enables systems to utilize additional sensory channel information to sense and communicate with the user using perceptual techniques such as visual, auditory and physiological emotion recognition.

Automatic visual emotion recognition has attracted the interest of the pattern recognition and computer vision research communities for the past decade. As evidence of that, we note the invited session held at the recent ICPR 2004 on computers that recognise and respond to human emotion. To date, significant results have been reported in recognition of emotions using facial expressions (e.g. [3]). Emotion recognition via body movements and gestures has only recently started attracting the attention of computer science and human-computer interaction (HCI) communities [10]. However, the interest is growing with works similar to these presented in [2] and [12]. So far, most of the work in affective computing has focused on only a single channel of information (e.g. facial expression). However, reliable assessment typically requires the concurrent use of multiple modalities (i.e. speech, facial expression, gesture, and gaze) occurring together [10].

Despite the common usage of multiple modalities in human-human interaction (HHI), relatively few works have focused on implementing emotion recognition systems using affective multimodal data [10]. The most common approach has been to combine facial expression with audio information [13]. Our focus, instead, is more on emotion recognition from different video modalities capturing the affective state of the user. Our motivation is three-fold. Firstly, studies such as that by Ambady and Rosenthal [1] reported that humans are able to judge behaviors most accurately when they are able to observe both the face and the body. In their experiments, ratings based jointly on face and body proved 35% more accurate than those based on the face alone [1]. Therefore, a reliable automatic affect recognition system should attempt to combine facial expression and body gestures. Secondly, the existing automatic affect analyzers have generally shown a performance improvement when using multimodal information [2, 12]. Finally, to our best knowledge, there is not a publicly available - readily accessible database combining expressive face and body display in a truly bimodal manner. Therefore, in this paper, we present the first bimodal face and body gesture database suitable for use in automatic analysis of human nonverbal expressive behavior.

## 2. Background for expressive face and expressive body gesture

The leading study of Ekman and Friesen [7] formed the basis of visual facial expression recognition. Their studies suggested that anger, disgust, fear, happiness, sadness and surprise are the six basic prototypical facial expressions recognized universally. Brave and Nass also provide details of the facial cues for displayed emotions in [4]. Most of the researchers argue that six universal emotion categories are not sufficient to describe all facial expressions in detail. However, most of the existing automated facial expression analyzers still use the Ekman and Friesen theory [2, 3, 4, 11, 13].

Human recognition of emotions from body movements and postures is still an unresolved area of research in psychology and non-verbal communication. Ambady and Rosenthal found out that humans rely on the combined visual channels of face and body more than any other channel when they make judgments about human communicative behavior [1]. In his paper [6], Coulson presented experimental results on attribution of six emotions to static body postures by using computer-generated figures. From his experiments he concluded that human recognition of emotion from posture is comparable to recognition from the voice, and some postures are recognized as well as facial expressions. Burgoon et al. clearly discuss the issue of emotion recognition from bodily cues and provide useful references in a recent publication in the context of national security [5]. They focus on the identification of emotional states such as positivity, anger and tension in videos from body and kinesics cues. In section 4, we provide a table based on the cues described by Brave and Nass [4], Burgoon et al. [5] and Coulson [6] with the list of the expressive face and body gestures and the correlation between the gestures and the emotion categories used in the recordings of our database.

## 3. Existing databases and their limitations

### 3.1. Facial expression databases

There have been some attempts to create comprehensive test-beds for comparative studies of facial expression analysis. The most famous and commonly used of these are [11], [16] and [17], chosen here as representative examples, rather than an exhaustive survey. We provide the details of these databases in Table 1. To date, the Cohn-Kanade AU-Coded Facial Expression Database is the most commonly used database in research on automated facial expression analysis [3].

**Table 1. Examples of existing facial expression databases.**

Name	JAFFE [16]	Cohn-Kanade [11]	MMI [17]
of data	219 images	2105 videos	740 images and 848 videos
Subjects	10 Japanese female	100 subjects; 65% female, 15% African-American, and 3% Asian/Latino, 18 to 30 year old	19 subjects, 3 ethnicities, 19-62 year old
Expressions	6 basic facial expressions + 1 neutral	facial expressions and combination of facial action units	6 basic facial expressions single action units combination of action units
features	Each image rated on 6 emotion adjectives by 60 Japanese subjects.	in-plane and limited out-of-plane motion beginning with a neutral face FACS coding for each subject and expression codes refer only to the final frame	Data from frontal and profile view are provided
Limitations	only gray scale images ethnically not diverse	limited image data from the frontal camera are available for distribution data available for public use is gray scale mostly single action units are rarely included	-

### 3.2. Body gesture databases

Gesture databases exist for static hand postures, dynamic hand postures and dynamic hand gestures, mostly for command entry purposes [19].



Figure 1. Examples of gestures from [18].

Existing databases mainly consist of non-affective one hand gestures only, and do not take into consideration the relationship between body parts (i.e. between hands; hands and the face; hands, face and shoulders etc.) [18, 19]. For instance, the Massey Hand Gesture Database is one such database [18]. Examples of its images are shown in Fig. 1. In general, these databases lack expressiveness of the body and its parts and therefore cannot be used for analysis of human nonverbal communicative behaviors.

### 4. Created database: FABO

To cope with the existing limitations, we created a bimodal database consisting of combined face and body expressions recorded simultaneously. According to five factors that were defined by Picard in [13] as influencing the affective data collection, our system consists of affective data that are (a) posed; (b) obtained in lab settings; (c) with an emphasis on expression rather than feelings; (d) openly recorded and (e) obtained with an emotion purpose. This is consistent with the characteristics of existing facial databases [14].

*Settings:* We created a set-up with a simple background to ease background removal and processing of upper body features. Prior to the video recordings of each subject, we recorded a video of the empty background for possible future use in the processing stage.



Figure 2. The settings we used for data acquisition.

The laboratory we used for the recordings was exposed to varying day light conditions. Therefore, to achieve consistent light conditions we decided to use artificial lights. We used two halogen spot lights to create a natural lightning environment, namely, an LTM Pepper 200 W pointed at the white ceiling of the room and an LTM Pepper 300 W pointed at the white wall next to the subject as shown in Figure 2. We avoided using the lights straight on the face of the subject to prevent erosion of the fine details of facial features required for facial information analysis.

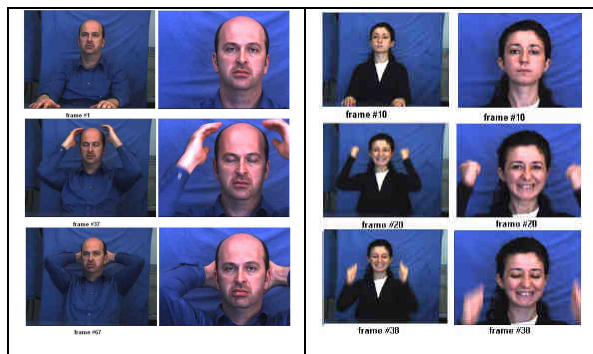
This setting was created with the help of an expert from the Film and Video Media Centre at the University of Technology, Sydney, to obtain the best illumination and natural colors possible. Additionally, the limited natural daylight was used for diffusing the effect of the artificial lights on the subject's body and face. The details of the subjects and cameras used for the recordings are given in Table 2.

**Table 2. FABO: subjects and cameras.**

<b>Number of subjects</b>	Total 23: Europe (10), Middle East (2), Latin America (3), Asia (7), Australia (1) .
<b>Subjects age range</b>	18-24 year-old (9), 25-29 (7), 30-40 (4), 40-50 (1). 12 of the subjects were female and 11 male.
<b>Cameras</b>	2 SONY XCD-X710CR, (1024x768 pixels), 15 fps, progressive scan CCD with square pixels Lenses: (a) Fuji lens 1:1 6/35 mm HF35HA-1 (for the face) (b) Avenir CCTV lens 16 mm F1.4 (for the body). The adequate level of zooming for the face and the body was obtained with subjects at a distance of about 4 meters.

*Recordings:* Prior to the recordings, the subjects were instructed as follows:

1. take a neutral position in the beginning, facing the camera and looking straight to it with hands visible and placed on the table;
2. do not to move the body when asked to create facial expressions (e.g. back and forth);
3. between each recording go back to the neutral position.



**Figure 3. Example sequences from FABO obtained from body (left columns) and face (right columns) cameras.**

The recordings for each subject took approximately over one hour. All recordings were initially stored to disk as raw data. However, their total size of more than 150 GB would have hindered online distribution, thus limiting the usefulness of our database. Therefore, we decided that “gentle” data compression was a needed requirement. For compression, we used the DivX 5.x MPEG-4 video codec with a bitrate of 4 Mbit/s. No visible artifacts were introduced in the decompressed videos and the database size was reduced to approximately 9 GB.

**Table 3. List of the face and body gestures performed for the recordings of FABO.**

expression	Face gesture	Body gesture
<b>neutral</b>	no expression	hands on the table, relaxed
<b>uncertainty</b>	lip suck lid droop eyes closed eyes turn right/left/up/down	head tilt left/right/up/down palms up shoulder shrug palms up+ shoulder shrug right/left hand scratching the head/hair right/left hand touching the right ear right/left hand touching the right nose right/left hand touching the chin right/left hand touching the neck right/left hand touching the forehead both hands touching the forehead right/left hand below the chin, elbow on the table two hands behind the head
<b>anger</b>	brows lowered and drawn together lines appear between brows lower lid tense/ may be raised upper lid tense/ may be lowered due to brows' action lips are pressed together with corners straight or down or open nostrils may be dilated	open/expanded body hands on hips/waist closed hands / clenched fists palm-down gesture lift the right/ left hand up finger point with right/left hand, shake the finger/hand crossing the arms
<b>surprise</b>	brows raised skin below brow stretched, not wrinkled horizontal wrinkles across forehead eyelids opened jaw drops open or stretching of the mouth	right/left hand going to the head both hands going to the head moving the right/left hand up two hands touching the head two hands touching the face, mouth right/left hand touching the face, mouth both hands over the head right/left hand touching the face selftouch two hands covering the cheeks selftouch two hands covering the mouth head shaking body shift- backing
<b>fear</b>	brows raised and drawn together forehead wrinkles drawn to the center upper eyelid is raised and lower eyelid is drawn up mouth is open lips are slightly tense or stretched and drawn back	body contracted closed body/closed hands / clenched fist body contracted, arms around the body selftouch (disbelief)/ covering the body parts/ arms around the body/shoulders body shift- backing, hand covering the head body shift- backing, hand covering the neck body shift- backing, hands covering the face. both hands over the head selftouch (disbelief)/ covering the face with hands
<b>anxiety</b>	lip suck lip bite lid droop eyes closed eyes turn right/left/up/down	hands pressed together in a moving sequence tapping the tips of the fingers on the table biting the nails head tilt left/right/up/down
<b>happiness</b>	corners of lips are drawn back and up mouth may or may not be parted with teeth exposed or not cheeks are raised lower eyelid shows wrinkles below it, and may be raised but not tense wrinkles around the outer corners of the eyes	body extended hands clapping arms lifted up or away from the body with hands made into fists
<b>disgust</b>	upper lip is raised lower lip is raised and pushed up to upper lip or it is lowered nose is wrinkled cheeks are raised brows are lowered	hands close to the body body shift- backing orientation changed/moving to the right or left backing, hands covering the head backing, hands covering the neck backing, right/left hand on the mouth backing, move right/left hand up
<b>boredom</b>	lid droop eyes closed lip suck eyes turn right/left/up/down	body shift, change orientation, move to the right/left hands behind the head, body shifted hands below the chin, elbow on the table
<b>sadness</b>	inner corners of eyebrows are drawn up upper lid inner corner is raised corners of the lips are drawn downwards	contracted/closed body dropped shoulders bowed head body shift- forward leaning trunk covering the face with two hands selftouch (disbelief)/ covering the body parts/ arms around the body/shoulders body extended + hands over the head hands kept lower than their normal position, hands closed move slowly two hands touching the head move slowly one hand touching the neck, move slowly hands together closed and head to the right

*Expressions Performed:* In our database, most face and body expression data have been collected by asking subjects to perform a series of expressions. When performing expressions:

1. the subjects were given the freedom to perform according to their liking by asking them how they use their face and body for a particular expression;
2. in some cases the subjects came up with a variety of combinations of face and body gestures;
3. when the subjects could not think of any specific gesture combination they were directed by the experimenter.

As a result of the feedback and suggestions obtained from the subjects, the number and combination of face and body gestures performed by each subject varies slightly.

In the first stage, subjects were asked to perform the following facial expressions: neutral, happiness, surprise, fear, anger, sadness, disgust, anxiety, uncertainty, positive\_surprise and negative\_surprise. In the second stage they performed the same expressions using their face and body simultaneously. Examples of the data recorded by camera 1 (for body) and camera 2 (for face) can be seen in Fig. 3. The complete list of the gestures the participants performed and/or were asked to perform is given in Table 3.

## 5. Discussion

After the recordings, the subjects were asked to fill in a survey about the most difficult and the easiest expressions to perform and their daily usage of face and body gestures. Questions were a combination of closed-ended (given choice between explicit alternatives – i.e. which expression was the most difficult to perform) and direct (soliciting a

specific piece of information - i.e. if they use their facial expressions in daily life). The details of the answers provided are given in Table 4.

**Table 4. Results from the FABO survey.**

Survey question	Summary of replies
The most difficult expressions to perform	anger (7 votes) and sadness (6 votes), followed by fear (4 votes) and disgust (2 votes); reason given: the subject "does not do it often in daily life" (12 votes) ; the subject does not know how to perform that particular facial expression deliberately (6 votes);
Use of the facial expressions together with the body/head/shoulders/arms/hands	facial expressions with head motions (up/down/right/left/rotate) (8 votes); with arms/hands (move arms, hands) (11 votes); facial expressions with hands and head (touching the head, face) (11 votes); facial expressions with body (changing posture, moving closer if interested) (7 votes) ;
what do subjects look at in HHI to decode what the other person feels.	the facial expression (16 votes); facial expression together with body gestures (9 votes); the body movement (changing posture) (6 votes); arm/hand movement and gestures (2 votes) head motions (1 vote);
how do subjects use their body/head/shoulders/arms/hands	together with their speech (11 votes); rarely use them (4 votes) in HHI; not aware of how often they use those (5 votes);
The easiest expression to perform	happiness (9 votes), neutral expression (6 votes); Reason given the subject does it often (15 votes) ; the subjects like doing the "happiness" expression;
Daily Usage of Face and Body Expressions in Human-Human Interaction	Subjects make extensive use of the facial expressions/actions together with their speech (14 votes); Subjects are not aware of how often they use their facial expressions and body gestures although they acknowledge their use (7 votes)

According to these answers the following interpretations can be drawn:

1. When recording visual affective data in a lab setting, the most difficult expressions to perform are the so-called negative expressions according to the "emotion wheel" defined by Plutchik [15].
2. It is difficult to produce particular expressions "artificially" without feeling the emotion that causes the expression
3. It is difficult to pass from one expression to another with instructions in the given short time.
4. The easiest expressions to perform are the so-called positive expressions according to the "emotion wheel" defined by Plutchik [15].

This answer might be explained by the fact that, in HHI as social beings, we are trained to smile in front of cameras from childhood while we tend to conceal our negative feelings most of the time.

In general, creating a natural setting for multimodal data recording and processing is challenging [13]. As confirmed by many researchers in the field, directed affective facial and bodily action tasks differ in appearance and timing from spontaneously occurring behavior [11, 13]. Moreover, as concluded from our survey results, without training as actors few people can perform these actions deliberately. Differences between spontaneous and deliberate facial actions may then be significant [11]. However, creating a spontaneous face and body gesture database is a involves ethical and privacy concerns together with technical difficulties (illumination, consistency, repeatability etc.). Given these restrictions, databases of directed expressions have been the only alternative possible to date [13].

Currently, the annotation in our database consists of the subjects' labeling and evaluation of their own performance; and experimenters' labeling as experts. The database is in the process of being annotated further by

independent human observers, uploaded on the website and obtaining internal clearance for public distribution. It will be made available through our research group's website<sup>1</sup>.

The videos contained in the FABO database enable a number of different tasks in human nonverbal expression research including, amongst others:

- recognition of emotions from individual frames or entire expressive sequences;
- automatic detection of the time boundaries of expressive sequences (time segmentation);
- automatic labeling of the various stages of an expressive sequence (i.e. recognition of the neutral, onset, offset, and apex states);
- stochastic temporal modelling of expressive sequences;
- moreover, all of the above tasks can be performed for either separate or combined face and body modalities.

The FABO database has been successfully used for the validation of the approaches proposed in [8] and [9] which could not have been possible with any existing databases due to their lack of combined affective face and body displays.

## 6. Conclusion

In this paper, we presented FABO, a bimodal face and body gesture database suitable for use in automatic vision-based analysis of human nonverbal communicative behavior. The FABO database contains approximately 1900 videos of facial expressions recorded by the facial camera, and face and body expressions recorded by the face and body cameras, simultaneously. This database is the first to date to combine facial and body displays in an organised manner, hence enabling significant future progresses in affective computing research.

## 7. References

- [1] Ambady, N. and Rosenthal, R., "Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis", *Psychological Bulletin*, 111, 2 (Feb. 1992), pp. 256-274.
- [2] Balomenos, T., et al. "Emotion Analysis in Man-Machine Interaction Systems", Proc. of MLMI 2004, *LNCS 3361*, pp. 318 - 328.
- [3] Bartlett, M.S., et al., "Machine learning methods for fully automatic recognition of face expressions and face actions", Proc. of IEEE SMC 2004, pp. 592-597.
- [4] Brave, S. and Nass, C., "Emotion in HCI", In J. Jacko & A. Sears (Eds.), *The HCI Handbook*, Hillsdale, NJ: Lawrence Erlbaum Associates, 2002.
- [5] Burgoon, J. K., et al., "Augmenting Human Identification of Emotional States in Video", Proc. of Int. Conference on Intelligent Data Analysis, 2005.
- [6] Coulson, M., "Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence", *J. of Nonverbal Behavior*, 28 (2) (2004).
- [7] Ekman, P. and Friesen, W. V., *Unmasking the face: a guide to recognizing emotions from facial clues*, Imprint Englewood Cliffs, N.J.: Prentice-Hall, 1975.
- [8] Gunes, H., and Piccardi, M., "Affect Recognition from Face and Body: Early Fusion vs. Late Fusion", Proc. IEEE SMC 2005, Hawaii, USA, Oct. 2005.
- [9] Gunes, H., and Piccardi, M., "Fusing Face and Body Display for Bi-Modal Emotion Recognition: Single Frame Analysis and Multi-Frame Post Integration", Proc. of the 1st Int'l Conf. on Affective Computing and Intelligent Interaction (ACII 2005), Beijing, China, Oct. 2005.
- [10] Hudlicka, E., "To feel or not to feel: The role of affect in human-computer interaction", *Int. J. Hum.-Comput. Stud.*, 59 (1-2), (2003), pp. 1-32.
- [11] Kanade, T., Cohn, J., and Tian, Y., "Comprehensive database for facial expression analysis", Proc. of the Int'l Conference on Face and Gesture Recognition, 2000.

---

<sup>1</sup> <http://research.it.uts.edu.au/cvrg/>



- [12] Pantic, M., Sebe, N., Cohn, J.F., Huang, T., "Affective Multimodal Human-Computer Interaction", Proc. ACM Int'l Conf. on Multimedia 2005, pp. 669-676.
- [13] Kapoor, A., Picard, R. W. and Ivanov, Y., "Probabilistic Combination of Multiple Modalities to Detect Interest", Proc. IEEE ICPR 2004.
- [14] Picard, R. W., *Affective Computing*, MIT Press, Cambridge, 1997.
- [15] Plutchik, R., *Emotion: A psychoevolutionary synthesis*. New York: Harper and Row, 1980.
- [16] Japanese Female Facial Expression (JAFFE) Database, [www.mic.atr.co.jp/~mlyons/jaffe.html](http://www.mic.atr.co.jp/~mlyons/jaffe.html).
- [17] MMI Face Database, <http://www.mmifacedb.com>.
- [18] Massey Hand Gesture Database, [http://www.massey.ac.nz/~fdadgost/xview.php?page=hand\\_image\\_database/default](http://www.massey.ac.nz/~fdadgost/xview.php?page=hand_image_database/default)[http://www.massey.ac.nz/~fdadgost/xview.php?page=hand\\_image\\_database/default](http://www.massey.ac.nz/~fdadgost/xview.php?page=hand_image_database/default)
- [19] FGNet Facial Emotions and Expressions Database, <http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html>.