THE VEDANA DATABASE

DAVID S. FREIRE OBREGÓN

A Thesis Submitted for the Degree of MSc. Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería (SIANI)

Under the supervision of
Modesto F. Castrillón Santana

Las Palmas de Gran Canaria - July 5, 2010

Acknowledgement

The research presented in this master thesis has been carried out at the SIANI Research Institute in Las Palmas de Gran Canaria, between the years 2009 and 2010, under the supervision of Dr. Modesto Castrillón Santana. I would like to thank him for his guidance, availability and enthusiasm during this master. It is great to work with him.

I would also like to thank the Fundación Universitaria for funding my research and Luis Antón for his encouragement and valuable advice. Special mention and thanks are due to Dr. Quoc Vuong for his expert advice on the usage of the RISE images.

Thanks to my officemates and colleagues at SIANI over this work, and especially to Javier, Antonio, Daniel, Enrique, Jorge and many others for the interesting discussions, for helping me when I was desperate recording people and generally for the nice working atmosphere.

I would like to thank the jury members of my master committee for their availability and their many interesting comments and criticism that will help me to improve this work.

Lastly and most importantly, I am deeply grateful to my lovely parents and my sister for their unconditional support, and to Cecilia for her encouragement, her smile and her constant love.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, the research topic of automatic facial analysis research topic has become a central topic in machine vision research. Nonetheless, there is a lack of a comprehensive set of face images that could be used as benchmark in the field. This lack of an easily accessible, suitable and common testing framework/resource represents the main obstacle to compare and extend the issues concerned with automatic facial expression analysis. Inside facial database's wide field there is a common goal which is to collect all possible and necessary data in a systematic way for a future use. Nowadays there are several databases publicly available, the choice of an appropriate database to be used should be made based on the task given (aging, expressions, gender, etc). Another way is to choose the data set specific to the property to be tested (e.g. how the algorithm behaves when images with lighting change or when different facial expressions are given). This chapter provides a comprehensive review of some publicly available databases for face recognition, face detection, and facial expression analysis. It is known that each face is affected by a large number of factors (identity, face pose, illumination, facial expression, age, occlusion, and facial hair). Another important issue that introduces complexity in automatic facial processing is the face's three-dimensional structure and its non rigidity. The development of robust algorithms that handle these variations requires large sets of sufficient size that include carefully controlled variations of these factors. Furthermore, common collections are necessary to comparatively evaluate algorithms. Even if the availability of public face databases is important for the advancement of the field, collecting a high quality database is a resource-intensive task. Thus, we present in this chapter an offline analysis of some existing databases as previous work. The theoretical concept of database and its relation with facial images collections will be evaluated in Sec. 1.1. As the focus of this work is to present a new database based on facial expressions, several aspects related to humans emotions and facial expression will be discussed in Sec. 1.2, while an analytical discussion about current facial collections, and their reliability and suitability to be used for evaluation will be evaluated in Sec. 1.3. The objectives of this work conclude this chapter.

1.1 Databases

As quoted from the encyclopedia of database systems *"a database is a collection of data for one or more multiple uses"* [27]. One way of classifying databases involves the type of content, for example: bibliographic, full-text, numeric, image. Other classification methods start examining database models or database architectures. Other models such as the hierarchical model and the network model use a more explicit representation of relationships.

Some advantages of databases are resumed as follows:

- Manageability. Images stored in the database can be directly accessed. It is possible to delete or modify (resize, copy, convert and rotate) database's elements.
- Security. Usually to grant access to any public database you must sign an end user license agreement. By signing this document the user, he or she who will make use of the database or the database interface, agrees to the following of some specific terms. These terms are related to commercial use, distribution or publications references. Confidentiality data about identities of the subjects that belong to the database is always safe.
- Backup/Recovery. If a downloaded database corrupts while it is being used, it is possible to recover it by just downloading again the database.
- Extensibility. This is an extremely important advantage. As it was already pointed, researchers can benchmark their own algorithms with the same datasets so they can compare, as fast and reliable as possible, their results.

But, there are also some concerns about these databases in use:

- Performance. The computer cost of accessing a huge amount of data can be really expensive depending on the data managed by the researchers (especially working with images).
- Database Size. This fact affects the database distribution. If the database is extremely heavy then network distribution could not be reachable and it would be necessary a CD distribution with additional costs and delay.

1.1.1 Facial Images Collections

It must be pointed out that facial images collections have sometimes been called databases. We must bear in mind, analyzing the definition of a database previously proposed, that this may in some cases be misleading because the term database hints at the existence of search and retrieval mechanisms for the material.

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 [42], and with significant pose and illumination variation [5], we felt that there was still a need for a collection consisting of a fairly large number of subjects, recorded 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 ability to benchmark they algorithms as fast 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 widest variety of datasets. They need databases that allow them to test their work under the most realistic environment 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 taken under controlled or uncontrolled condition by a mobile phone [43]. A similar situation happens with webcams. What if we want to identify ourselves or communicate with computers with just our facial expression, nowadays high resolution cameras are not very common, thus none of us have next to our computer this kind of camera or controlled conditions for identification or communication purposes. 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. In the past years, only a relatively small number of relevant face databases have been presented in the literature (See Section 1.3).

1.2 Face Interaction

The human face refers to the central sense organ complex, normally on the ventral surface of the head, and includes the hair, forehead, eyebrow, eyes, nose, ears, cheeks, mouth, lips, philtrum, teeth, skin, and chin. The face has uses of expression, appearance, and identity amongst others. At the highest level of the taxonomy, facial movements are either rigid or non-rigid. Rigid motions of the head include nodding, shaking, tilting, and rotating about the vertical axis. It is worth noting that all of these movements change the view of the face available to a stationary observer. It is also important to note that each of these movements can convey a social signal or connotation. Nodding the head can convey agreement, shaking back and forth can convey disagreement, and turns of the head, either toward or away from another person, can be used to

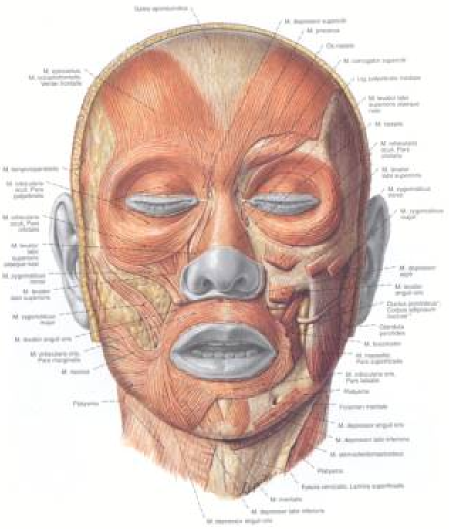


Figure 1.1: The human face. Image taken from [63].

initiate or break off a communication. Non-rigid movements of the face occur during facial speech, facial expressions, and eye-gaze changes. Again, these movements produce highly variable images of the person that can distort many of the identifying features of faces, like the relative distances of the features. The difference between the image of a person smiling and one of a person with a surprised expression can be quite strong. The relative position of the eyebrows, for example, with respect to the mouth and the other features, changes radically between these images. Facial speech, expression, and eye gaze movements can also convey a social message. When a person speaks, they rarely do so with a static or neutral expression. Eye gaze changes can signal boredom or interest and facial expression can yield an nearly limitless amount of information about a person's internal state. All these facts produce the ability to create thousands of different expressions. Facial behaviors are used for various functions, including

- Speech illustration. For instance, people often raise their brows when being inquisitive, and lower their brows when they lower their voices.
- Conversation regulation. We provide cues to others that we are either done talking and it's their turn, or not, through our faces (and voice).
- Cognition. People often furrow their brows when concentrating or are perplexed. They also purse their lips when conducting mental searches.
- Talking and eating. We use the muscles around the mouth area for talking and eating, and especially for speech articulation.
- Emotion signaling. We use the facial muscles to signal our emotional states.

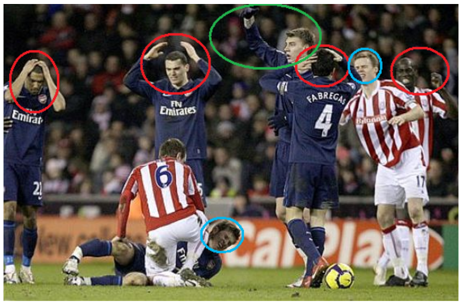


Figure 1.2: Different facial behaviours and gestures are remarkable in this scene. Inside the red circles symbolic and emblematic shock gestures can be seen, the green circle shows a deictic gesture calling the attention of the medical staff and blue circles show an expressive facial response of pain with high intensity. Image taken from [53].

- Expressive regulation. We also use the facial muscles to regulate our emotion signals.

- Emblematic gestures. These are movements that symbolically give verbal meaning that can be conveyed by words, such as the doubtful look produced by raising the upper lip and pushing the lower lip up. In fact, gestures have different functions and a taxonomy can be elaborated according to [44, 33, 88]:

  - Symbolic gestures: Gestures that, within a culture, have a single meaning. The OK gesture is one of such examples. Any Sign Language gestures also fall into this category (See Fig. 1.2).

  - Deictic gestures: These are the types of gestures most generally seen in HCI and are the gestures of pointing. They can be used for directing the listener's attention to specific events or objects in the environment, or for commanding.

  - Iconic gestures: These gestures are used to convey information about the size, shape or orientation of the object of discourse. These are the gestures made when someone says ”The eagle flew like this”, while moving his hands through the air like the flight motion of the bird.

  - Pantomimic gestures: These are gestures typically used to show the use of movement of some invisible tool or object in the speaker's hand. When a speaker says I moved the stone to the left, while mimicking the action of moving a weight stone with both hands, he is making a pantomimic gesture.

  - Beat gestures: The hand moves up and down with the rhythm of speech and looks like it is beating time.

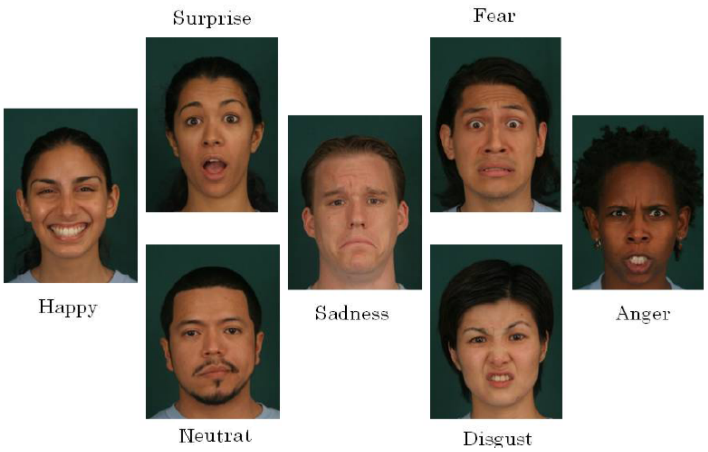


Figure 1.3: The six universal facial expression of emotions plus neutral. Facial expressions are particularly significant in interpersonal communication, interaction and relation; as they reveal information about the affective state, cognitive activity, personality, intention and psychological state of a person and this information is difficult to mask. Image taken from [58].

- Cohesive gestures: These are variations of iconic, pantomimic or deictic gestures that are used to tie together temporally separated but thematically related portions of discourse.

As it has been already pointed, emotions evolved as a rapid and coordinated response system that allows humans to respond quickly and efficiently to events that affect their welfare [15, 10, 26]. Facial expressions are part of that response system. The facial musculature has over 40 independent actions that can occur, which results in an extremely large number of possible expressions. But of this large potential repertoire, strong evidence now exists that a small number of specific facial configurations are universally and discretely produced when emotions are elicited [13].

Since Darwin, several researchers have studied basic emotions. According to [37], different basic emotion sets ranging from two to eighteen basic emotions have been proposed by different investigators. However, most of them agreeing at least on emotions anger, fear, happiness and sadness. One of the most influential researchers on emotional facial expressions and the theory of basic emotions has been Ekman, he concluded that the expressions associated with some emotions were basic or biologically universal to all humans [17]. The following is Ekman's (1972) list of basic emotions: anger, disgust, fear, happiness, sadness and surprise (See Fig. 1.3). However in the 1990s Ekman expanded his list of basic emotions [14], including a range of positive and negative emotions not all of which are encoded in facial muscles. The newly included emotions are: amusement, contempt, contentment, embarrassment, excitement, guilt, pride in

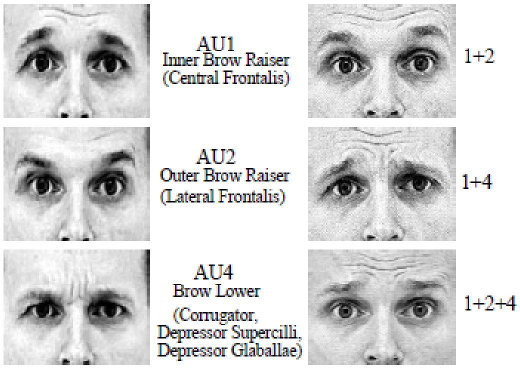


Figure 1.4: The Facial Action Coding System decomposes facial expressions into component actions. The three individual brow region actions and selected combinations are illustrated. When subjects pose fear they often perform 1+2 (top right), whereas spontaneous fear reliably elicits 1+2+4 (bottom right). Image taken from [58].

achievement, relief, satisfaction, sensory pleasure and shame.

The well-known Facial Action Coding System (FACS) by Ekman and Friesen [16] is the leading method for measuring facial movement in behavioral science. It is a human judgment system that is presently performed without aid from computer vision. In FACS, human coders decompose facial expressions into action units (AUs) that roughly correspond to independent muscle movements in the face. Ekman and Friesen described 46 independent facial movements, or facial actions (see Fig. 1.4). These facial actions are analogous to phonemes for facial expression. Over 7000 distinct combinations of such movements have been observed in spontaneous behavior. Advantages of FACS include:

- Objectivity. It does not apply interpretive labels to expressions but rather a description of physical changes in the face. This enables studies of new relationships between facial movement and internal state, such as the facial signals of stress or fatigue.

- Comprehensiveness. FACS codes for all independent motions of the face observed by behavioral psychologists over 20 years of study.

- Robust link with ground truth. There are two decades of behavioral data on the relationships between FACS movement parameters and underlying emotional or cognitive states. Automated facial action coding would be effective for human-computer interaction tools and low bandwidth facial animation coding, and would have a tremendous impact on behavioral science by making objective measurement more accessible.

Questions exist, however, concerning their source. There are at least two theoretical positions that can account for universality [32]. One suggests that universal expressions are produced by culture-constant learning. In this view, individuals around the world learn, through observational learning, modeling, and reinforcement, to associate the same facial configurations with the same emotional states or antecedent events. Facial expressions of emotion, thus, are universal because the same expressions are observed and modeled around the world in response to the same types of emotionally evocative situations. A second position suggests that universal expressions originate from an evolved emotion-response system. This position suggests that the facial configurations are genetically coded for all humans and are part of a larger response system involving cognitive, physiological, and phenomenological changes. For instance, the elicitation of anger would recruit a host of physiological responses that would prime an individual to fight (e.g., increased heart rate and respiration); gate mental activities to be alert for possible opponents; and produce threatening expressions that allowed for increased focus on objects, the baring of teeth in preparation for biting, and the perception of threat by others. According to this view, this coordinated response system is produced from a biologically resident source that requires little or no learning; facial expressions, therefore, are universal because they are a product of our evolutionary history.

Expressions and Recognition

Human expressions are straightly related to face recognition; motion of faces appears to facilitate subsequent recognition. Thus, Sinha established in his critical review [39] that dynamic cues aid face recognition only in some cases. According to Sinha, rigid motion (such as that obtained from a camera rotating around a motionless head) can facilitate recognition of previously viewed faces [3, 89], but there seems to be very little, if any, benefit of seeing these views during the learning phase. By contrast, non rigid motion (where the individuals exhibit emotive facial expressions or speech movements) plays a greater role. Experiments in [7], using subtle morphs of form and facial motion in novel (i.e., unfamiliar) faces, showed that non rigid facial motion from one face applied to the form of another face can bias an observer to misidentify the latter as the former (see Fig. 1.5). Experiments with famous (i.e., highly familiar) faces [25] again showed a facilitation in recognition with dynamic cues from expressive or talking movements, but not from rigid motion. Facilitation was most pronounced for faces whose movement was judged as "distinctive". Note also that facilitation comes from a natural sequence of moving images, not merely from having more views available: The facilitation is greatly lessened when the same frames are presented in random order or in a static array. Sinha concluded that *"these results suggest that face motion is more than just a sequence of viewpoints to the face recognition system. The dynamic cues from expressive and talking movements provide information about aspects of facial structure that transcends the gains of simply having multiple viewpoints."*

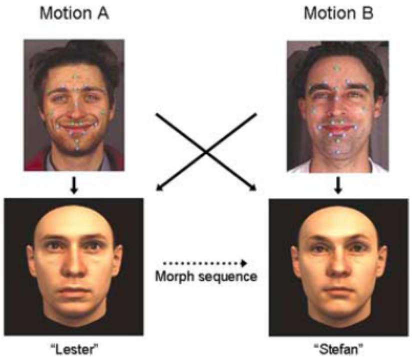


Figure 1.5: Facial motion from expressions and talking were morphed onto forms of "Lester" and "Stefan". Subjects could be biased to identify an anti-caricatured (morphed towards the average) form of Lester as Stefan when Stefan's movements were imposed onto Lester's form [7].

1.3 State of the Art

Available facial collections can be easily accessed via existing Internet resources [55, 60] or Google Internet search engine. In this section we review some publicly available databases for face recognition, face detection, and facial expression analysis. The scope of this section is limited to databases containing full face imagery.

Not all databases are discussed at the same level of detail because it is considered not suitable to compare them within this work. For example, Belfast Naturalistic Database [12] and PICS [71] collections contained only one emotional expression (smiling); and Yale Face Databases [6] (two separate collections) collection contained only three different emotional expressions (sad, sleepy and surprised; and smile, anger and screaming). Furthermore, some of the expressions in the latter two were not considered emotional (such as sleepy and screaming faces). Most of the collections are available free of charge for the research community by request; however, the availability of CMU (Carnegie-Mellon University Facial Expression Database) collection is usually restricted only to computer vision studies.

1.3.1 Collections for Face Recognition

Face recognition has benefited greatly from the many databases that have been produced to study it. Most of these databases have been created under controlled conditions to facilitate the study of specific parameters on the face recognition problem. These parameters include such variables as position, pose, lighting, expression, background, camera quality, occlusion, age, and gender. While there are many applications for face recognition technology in which one can

Collection	Year	Subjects	Pose	Facial Exp.	Availability	Website
AR	1998	126	1	4	Online	[72]
BANCA	2004	208	1	1	Commercial	[73]
CAS-PEAL	2004	1040	21	6	By request	[74]
CMU Hyper	2002	54	1	1	By request	[11]
CMU PIE	2000	68	13	3	By request	[21]
Equinox IR	2004	91	1	3	Commercial	[57]
FERET	2003	1199	9-20	2	By request	[75]
KFDB	2003	1000	7	5	By request	[9]
Harvard RL	2000	200	3	1	Online	[61]
MPI	1999	200	3	1	By request	[64]
ND HID	2002	300	1	2	By request	[68]
NIST MID	2000	1573	2	1	Commercial	[67]
ORL	1994	40	1	++	Online	[69]
Sheffield	1998	20	++	++	Online	[77]
U. Oulu	1999	125	1	1	By request	[78]
XM2VTS	1999	295	++	++	Commercial	[82]
Yale	1997	15	1	6	Online	[83]
Yale B	2000	10	9	1	Online	[84]

Table 1.1: 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 ++.

control the parameters of image acquisition, there are also many applications in which the practitioner has little or no control over such parameters. Table 1.1 gives an overview of all databases discussed in this section. Owing to space constraints not all databases are discussed at the same level of detail. The scope of this section is limited to databases containing full face imagery.

The FERET database [42] consists of monochrome images taken in different frontal views and in left and right profiles (See Fig. 1.6). Only the upper torso of an individual (mostly head and necks) appears in an image on a uniform and uncluttered background. The FERET database has been used to assess the strengths and weaknesses of different face recognition approaches [42]. Since each image consists of an individual on a uniform and uncluttered background, it is not suitable for face detection benchmarking. This is similar to many databases that were created for the development and testing of face recognition algorithms. The face database from AT&T Cambridge Laboratories (formerly known as the Olivetti database (ORL)) consists of 10 different images for forty distinct subjects [48]. The images were taken at different times, varying the lighting, facial expressions, and facial details (glasses).

The Harvard database consists of cropped, masked frontal face images taken

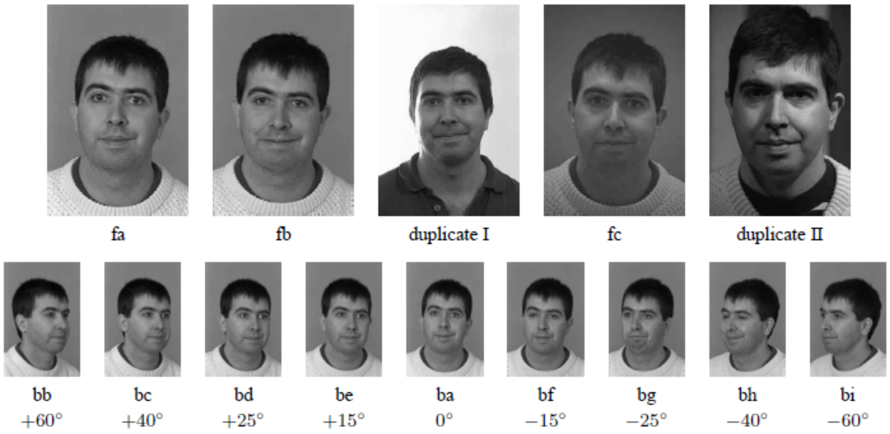


Figure 1.6: Frontal image and pose variation categories used in the FERET [42] evaluations. The duplicate I image was taken within one year of the **fa** image and the duplicate II and **fa** images were taken at least one year apart. Also, it can be appreciated that poses vary from +60. (**bb**) to full frontal (**ba**) and on to -60. (**bi**).

from a wide variety of light sources [20]. It was used by Hallinan for a study on face recognition under the effect of varying illumination conditions. With 15 individuals, the Yale face database contains 10 frontal images per person, each with different facial expressions, with and without glasses, and under different lighting conditions [38]. The M2VTS multimodal database from the European ACTS projects was developed for access control experiments using multimodal inputs [41]. It contains sequences of face images of 37 people. This database was replaced by the Extended M2VTS (XM2VTS) database [22]. This database was collected for research and development of identity verification systems using multimodal (face and voice) input data. The database contains 295 subjects, each recorded at four sessions over a period of 4 months. At each session two head rotation shots and six speech shots (subjects reading three sentences twice) were recorded. 3D models of 293 subjects are available as well. The AR database contains over 3,276 color images of 126 people (70 males and 56 females) in frontal view [31]. This database was designed for face recognition experiments under several mixing factors, such as facial expressions, illumination conditions, and occlusions. All the faces appeared with different facial expression (neutral, smile, anger, and scream), illumination (left light source, right light source, and sources from both sides), and occlusion (wearing sunglasses or scarf). The images were taken during two sessions separated by two weeks. All the images were taken by the same camera setup under tightly controlled conditions of illumination and pose.

1.3.2 Collections for Face Detection

In the last five years, face and facial expression recognition have attracted much attention though they have been studied for more than 20 years by psychophysicists, neuroscientists, and engineers. Many research demonstrations and commercial applications have been developed from these efforts. A first step of any face processing system is detecting the locations in images where faces are present. However, face detection from a single image is a challenging task because of variability in scale, location, orientation (up-right, rotated), and pose (frontal, profile). Facial expression, occlusion, and lighting conditions also change the overall appearance of faces.

As quoted from [34], a definition of face detection could be: Given an arbitrary image, the goal of face detection is to determine whether or not there are any faces in the image and, if present, return the image location and extent of each face. The challenges associated with face detection can be attributed to the following factors:

- Pose. The images of a face vary due to the relative camera-face pose (frontal, 45 degree, profile, upside down), and some facial features such as an eye or the nose may become partially or wholly occluded.

- Presence or absence of structural components. Facial features such as beards, mustaches, and glasses may or may not be present and there is a great deal of variability among these components including shape, color, and size.

- Facial expression. The appearance of faces is directly affected by a person's facial expression.

- Occlusion. Faces may be partially occluded by other objects. In an image with a group of people, some faces may partially occlude other faces.

- Image orientation. Face images directly vary for different rotations about the camera's optical axis.

- Imaging conditions. When the image is formed, factors such as lighting (spectra, source distribution and intensity) and camera characteristics (sensor response, lenses) affect the appearance of a face.

Common testing data sets are necessary to comparatively evaluate the performance of face detection algorithms. These data sets should be representative of real-world data containing faces in various orientations against a complex background. According to [34], popular choices are the previously mentioned (See Table 1.1) FERET, MIT, ORL, Harvard, and AR databases. Along with these public databases, independently collected, nonpublic databases are also often employed. The above mentioned databases are designed mainly to measure performance of face recognition methods and, thus, each image contains only one individual. Therefore, such databases can be best utilized as training

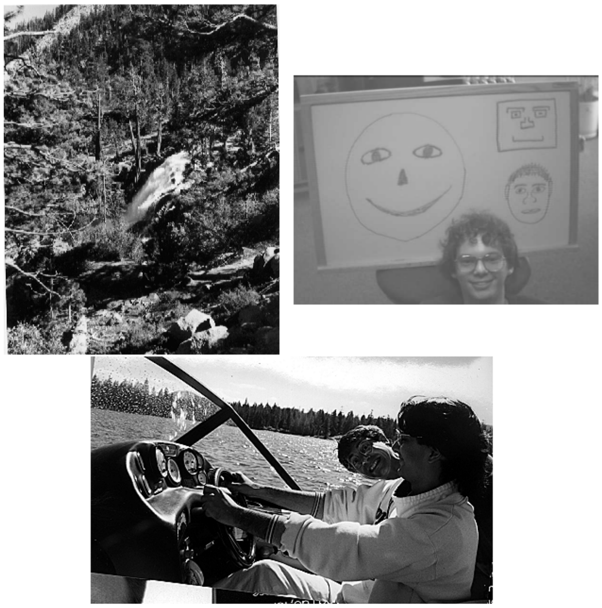


Figure 1.7: Sample images in MIT/CMU test set for frontal face detection [86]. Some images contain hand-drawn cartoon faces. Most images contain more than one face and the face size varies significantly.

sets rather than test sets. The tacit reason for comparing classifiers on test sets is that these data sets represent problems that systems might face in the real world and that superior performance on these benchmarks may translate to superior performance on other real-world tasks. Toward this end, researchers have compiled a wide collection of data sets from a wide variety of images.

Between collections for face detection, two public data sets emerged as quasi-standard evaluation test sets: the combined MIT/CMU test set for frontal face detection [86] and the CMU test set II for frontal and non frontal face detection [50]. In the following we describe both databases. They are available for download [59].

The combined MIT/CMU (See Fig. 1.7) data set includes 180 images organized in two sets:

- The first set consists of 301 frontal and near frontal mugshots of 71 different people. These images are high quality digitized images with a fair amount of lighting variation.
- The second set consists of 23 images with a total of 149 face patterns. Most of these images have complex background with faces taking up only a small amount of the total image area.

The CMU Test Set II dataset was collected to test face detection algorithms that are able to handle out-of-plane rotation. It contains 208 images with 441 faces out of which 347 are in profile view. The images were all collected from the Internet.

1.3.3 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 challenging 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 usage 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 [1] or not skilled [29, 19]. 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 elicit disgust by some subjects, while it could trigger fear by others. A more physiological point of view about emotional states was presented by Bradley in [8]. As Bradley pointed out,

"First of all the question of whether specific emotional states are related to specific physiological patterns neglects important facts that physiology will vary with action, and that actions associated with the same emotional state will also often vary. 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 bluntly, 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).".

Lacey [24] noted the great variety of contexts in which emotion can be induced in the laboratory. He confirmed the idea that the nature and direction of physiological change is dependent, to a great extent, on the experimental context. In the emotion literature, on the other hand, inferences regarding the physiology of fear, for example, are often made by comparing data from contexts as diverse as hearing loud noises, anticipating shock, imagining an intruder in the house, looking at a picture of an amputated leg, viewing a scary film, giving a speech, putting one's hand in cold water, or hearing an anguished scream. Conversely, responses to stimuli such as receiving money, listening to joyful music, looking at a picture of puppies, viewing an erotic film, imagining a day on the beach, receiving a good grade, thinking about winning the lottery, or anticipating a vacation are compared on the basis that they

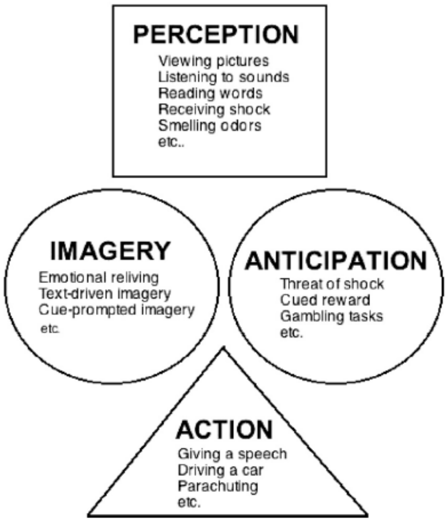


Figure 1.8: The most common induction paradigms in the psychophysiological study of emotion can be roughly classified as involving perception, imagination, anticipation, or action.

prompt a happy emotional state. The diverse sensory, cognitive, and motor processes elicited by these induction procedures may prompt quite different physiological profiles, irrespective of modulation by emotion. Taken together, understanding the psychophysiology of emotion will depend on clearly specifying the context of the emotional induction in the laboratory.

In the laboratory, contexts routinely used to induce affective reactions can be roughly organized into those that primarily target perception, anticipation, imagination, or action (See Fig. 1.8). While clearly not mutually exclusive (e.g., most anticipatory or imagery tasks include some perceptual input), these induction domains represent broadly similar contexts with respect to psychophysiological responses that may be task, but not emotion, related. For instance, physiological reactions in perception differ both quantitatively and qualitatively from those in imagination, due to the requisite physiology of these different tasks. Moreover, both within and between broad induction contexts, specific parameters can have strong effects on the degree and nature of emotional engagement i.e., the ease with which neural systems mediating emotional responses are activated and the resulting pattern of physiological change that can be associated with affective engagement.

Importantly, before 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 [40]) and University of Texas Database are the most used dynamic basic expression collections by researchers (See Table 1.2). Most of the collections have been created by asking

Collection	Year	Subjects	Facial Exp.	Availability	Website
JAFFE	1999	126	6 Basic Emotion & Neutral	Online	[76]
Cohn-Kanade	2000	100	6 Basic Emotion & Neutral	By request	[51]
DaFEx	2004	8	6 Basic Emotion & Neutral 3 Intensity Levels	By request	[52]
MMI	2005	61	6 Basic Emotion	By request	[66]
U. Texas	2005	284	++	By request	[81]

Table 1.2: 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 ++.

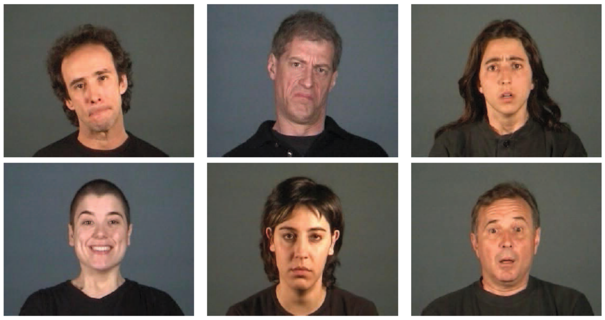


Figure 1.9: Six basic emotions presented by six different individuals; extracted from the DaFEx database [1]. The displayed emotions are, from left to right and top to bottom: anger, disgust, fear, happiness, sadness, and surprise.

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 [28]. Arrangements were such that the actors were able to monitor their faces and take the photographs themselves. In DaFEX collection [1], 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 [2], 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 [87] 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. Therefore, a new induced facial expression collection is presented in this work, employing a set of not skilled actors to generate our collection; there is a detailed description of the used data to generate this new collection in chapter 2 and the results obtained in chapter 3.

1.4 Objectives

The establishment of an easily accessible, comprehensive, benchmark database of facial expression exemplars has become an acute problem that needs to be resolved in order to open a new research in automatic facial expression analysis. A number of issues that make this problem complex were suggested by Pantic in [30].

The first issue takes care about the 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 motion images of faces showing prototypic expressions of emotion and various expressions. 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 [49]. Also, while few people are able to perform certain facial actions voluntarily, many are able to perform these actions spontaneously [87]. Because of the existence of many collections with acted category, this work just include spontaneously actions.

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 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. The labeling process determines not only whether a given computing system attempts to analyze or interpret the emotion associated with specific signals recorded in the database, but may also influence the achievable recognition accuracy [90]. As seen above, an individual's emotional state can be judged either indirectly, by the observer's judgments, or directly considering either the self-report or the measurement of facial activity. However, when it comes to the real-time application, both self-report and facial-measurements labeling are not sufficient. This is because the former suffers under the timing issue and the latter from describing the shown facial changes while neglecting what underlies them, being relatively time consuming, and demanding professional-trained observers for labeling FACS-based data. According to [16], 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. Hence, obtaining large collections of AU-coded facial-expression data is extremely tedious and, in turn, difficult to achieve.

The most convenient method for labeling such reliable data is the judgment approaches. Indeed, that is not surprising because

- inferring basic emotions is an intuitive process and matches the experience of the ordinary human in daily life,
- the simplicity of coping with a limited number of variables in contrast to other emotion models,
- their nature of being displayed and recognized universally, as proven in diversity of theoretical studies
- avoiding the possible loss of information caused by labeling data into 2D or 3D space, and raters can be just ordinary individuals rather than the professional trained observers demanded for judging FACS.

Judgment-based approaches are centered on the message conveyed by the considered cues (facial expressions, speech information). In order to categorizing affective state associated to one of these cues into a predefined number of emotion or mental activity classes, an agreement of a group of decoders is taken as ground truth. Each one of the two well-known emotion theories namely, basic emotion, dimensional emotion, has its own judgment approaches.

Regarding how the samples should 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. Because this work is not about actions performed, there is no need for subjects to be experts in production of expressions or individuals being instructed by such experts on how to perform the required facial expressions. This fact increases our number of potential subjects and reduces costs. 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.

A critical issue is how one facilitates 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, then a more comprehensive security strategy should be used.

Chapter 2

Experimental Setup

The facial expression analysis collection is constructed by the Artificial Intelligence division (GIAS) inside the research group SIANI [62], supported by the Universidad de Las Palmas de Gran Canaria [80]. The samples are all collected in Las Palmas de Gran Canaria, Spain. The goals to create this facial expression analysis collection include (1) providing the worldwide researchers of facial expression analysis community a large-scale face database for training and evaluating their algorithms; (2) facilitating the development of facial expression analysis by providing large-scale face videos with different sources of variations, especially pose, expression and accessories. Therefore an induction technique (in this case by watching selective videos) is used for recording facial expressions. 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. This chapter describes the complex procedure followed to collect an emotional database as a part of our research. The environment for the data acquisition process is described in Section 2.1, while the software developed and the equipment used for this aim is fully detailed in Section 2.2. This chapter concludes with a description of the video datasets used for the induction process in Section 2.3.

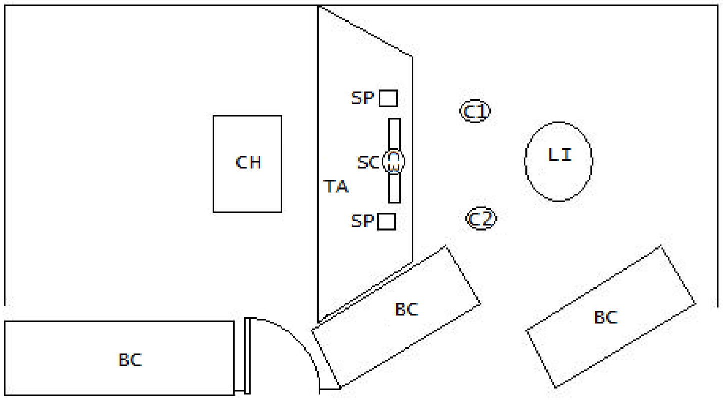


Figure 2.1: Basic scheme of the recording set and position configuration of the equipment.

Key	Item	Height
LI	Light	1,72 meters
TA	Table	0,75 meters
CH	Chair	0,50 meters
SC	Screen	20 inches
BC	Bookcase	1,97 meters
C1	Camera 1	1,40 meters
C2	Camera 2	1,40 meters
C3	Camera 3	1,25 meters
SP	Speakers	–

Table 2.1: For each item in Fig. 2.1 there is a key that identifies it. In this table the height of each item can be also appreciated.

2.1 Environment

In order to capture face images conveniently and efficiently, a special recording studio was set up in the basement of the SIANI facilities inside the Universidad de Las Palmas de Gran Canaria [80]. A general scheme of the recording set can be appreciated in Fig. 2.1 as well as the location of each item used in the experiment.

The dimensions of the room is about 4.33m x 1.9m x 1.97m (See Fig. 2.2). To record facial expressions with different poses, expressions or accessories, special equipment was configured in the room including three digital cameras and a special spotlight. For obtaining images of participants from three different poses, three different cameras recording simultaneously were used. The angles between the frontal camera C3 and cameras C1 and C2 are 50 and 41 degrees

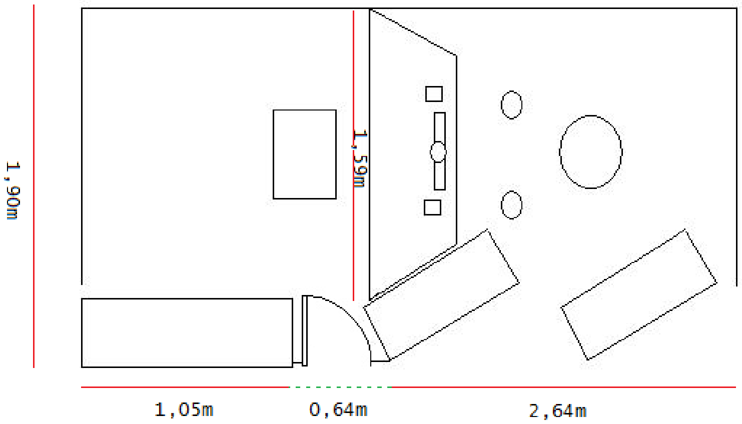


Figure 2.2: Studio layout dimensions.

respectively. There is no mechanism to activate all the cameras at the same time because cameras are not connected to and controlled by the same computer, but a standard procedure followed by the participants described in Section 2.2 solves this problem. Only one (C3) of the three digital cameras is connected to and controlled by the computer which is under the table (TA). Indeed, as it can be also appreciated in Fig. 2.1, there is a main entrance for the participants and a second access towards to allow the manipulation by hand of two (C2 and C3) of the cameras. About the illumination configuration, the set is occluded by the bookcases and a white curtain (See Fig. 2.3) with the aim to avoid distractions of participants during the acquisition process.

In order to take full advantage of the spotlight and to make background as invariant as possible, walls and internal face of bookcases are covered by reflective white cards (See Fig. 2.4). Thus, white cards are the default background of the record process. Also, in practical applications, many cameras are working under the auto-white balance mode, which may change the face appearance much in different scenes. Therefore, it is necessary to mimic this situation in the database.

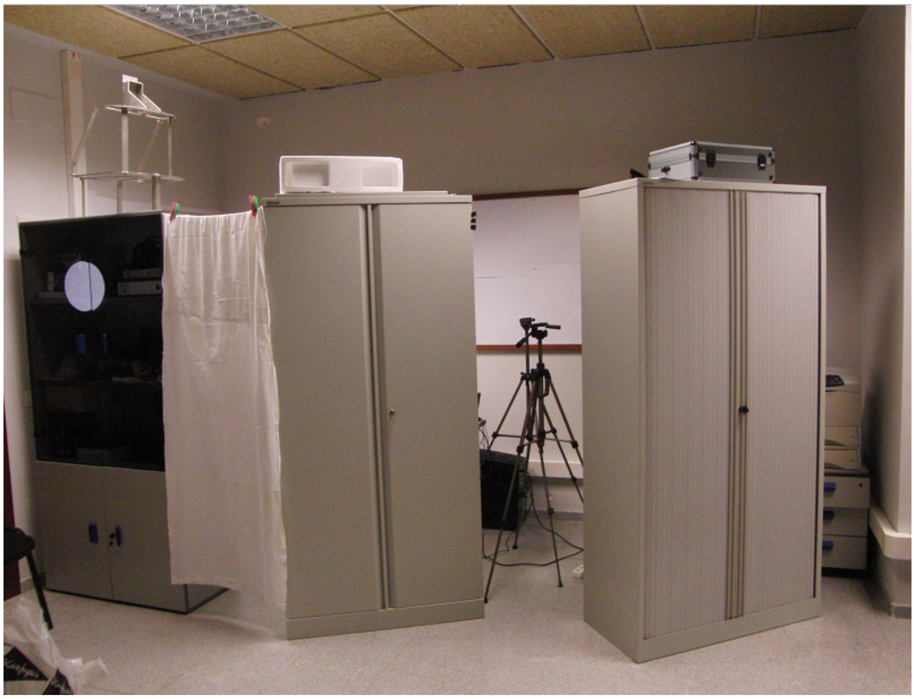


Figure 2.3: External view of the recording set. As it can seen, it is a main entrance for participants covered by a white curtain and, it is a second access in order to start/stop cameras by hand, as well as to explain to participants the standard procedure to follow in the experiment.

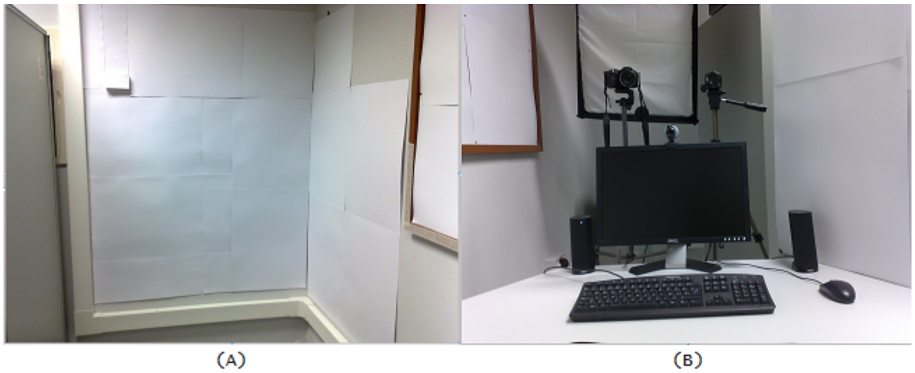
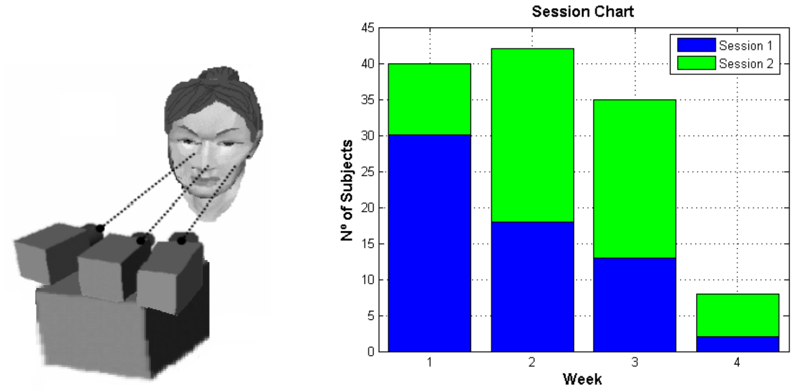


Figure 2.4: Internal view of the recording set. Image (B) shows the frontal view of the recording set, what participants see, and what is behind participants can be appreciable on image (A).

2.2 Experimental Setup

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 couple of days of separation between 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. The lamphouse is high temperature finished and a properly designed reflector, made of polished aluminum, ensured a high degree of reflectivity. For more detailed information about the equipment see Section 2.2.1.



(a) Basic scheme of the data acquisition process

(b) Our collection schedule represented in a bar graph, separated into sessions by color

Figure 2.5: Key factors related to the experimental setup.

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.

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 subjects face as a facial expression. For example, to have a subject assume a smiling expression, a funny content clip is shown to him/her (See Fig. 2.5(a)).

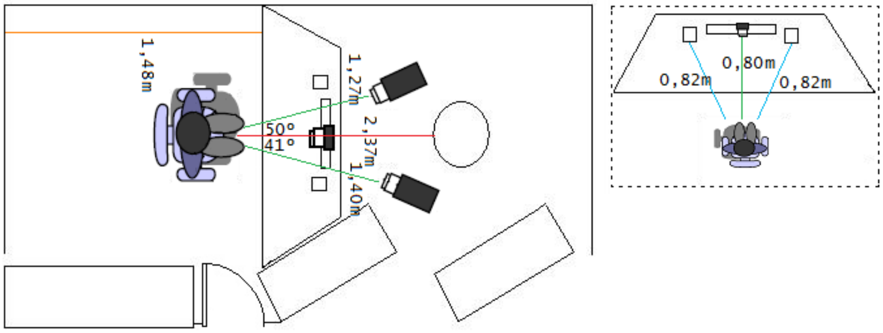


Figure 2.6: Configuration for the facial expression experiment. Each subject was presented with 17 successively displayed video sequences. Each sequence length was approximately 20 seconds and the viewing distance to the monitor was 80 cm. Furthermore, it is remarkable the distance between the user and the wall behind, avoiding the appearance of any shadow during the data acquisition process.

- Data acquisition process by the three cameras. Subjects were asked to watch the multimedia content while they were recording by the three cameras simultaneously.

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.

We captured data in two sessions, each session occurring during a different day. Our goal was to have, at least, a day separation between sessions for each participant. The weekly schedule for all three sessions is shown in Fig. 2.5(b). We began collecting data for session one in February, 2010. The total collection spanned 4 weeks, 12 sessions.

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 raised and lower their arms for 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).

Firstly, if it was the first time for the participant to be part of the experiment, they were asked to sign a contract for giving their image rights to the SIANI research group. For the facial expression experiment, the subjects were asked to sit in the chair and look at the screen situated in front of his position:

1. Before starting the recording session, the chair was adjusted to make the subject face horizontally to the camera C3 and to make the distance between subject and the camera C3 about 0,80 meters (See Fig. 2.6). For statistical purposes, subjects filled in by computer a form about their appearance in the session. The form asked if the subject was wearing glasses, a hat, etc.

2. Since we were aiming to synchronize all three cameras together for our database, and bearing in mind that cameras C1 and C3 do not have a sound channel, participants were asked to shake their heads up and down when they heard a high pitched sound.

3. Right from the start of the acquisition process, an assistant technician activates the cameras C1 and C2 by hand. After three seconds of delay, the high pitched signal sounded for 1 second and then the visual-stimuli process started.

4. The only constraint participants had during the experiment was to avoid covering their faces with hands. Motion and pose were left participant-dependent.

In all studies for the first experiment, stimuli were 17 short video sequences of different categories. Each category tried to generate facial expression responses that depict basic emotions (see Section 2.3). Stimuli were presented with developed software (See Section 2.2.2) in a randomized order. Between each video sequence, the subject was presented, for 3 seconds, a phase scrambling image (scrambled the phase of the fourier transform of the image for each color plane) of the last frame that removes any low-level adaptation and memory effects (see Section 2.3 for more details). The subjects were sitting approximately 80 cm from the monitor. Stimuli were shown on a 20 inches (18 inches viewable area) monitor with a resolution of 1024x768 pixels. The mean length of each stimulus is 20 seconds and each test took approximately 8 minutes. The subject's task was to watch these sequences trying not to fake any gestures, just like watching TV. The idea was to record naturalistic facial expressions while they were watching the sequences.

The body experiment took place after the facial expression. Only a pose was recordered for this second experiment, through camera C2 (See Fig. 2.7). Before starting this second 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.

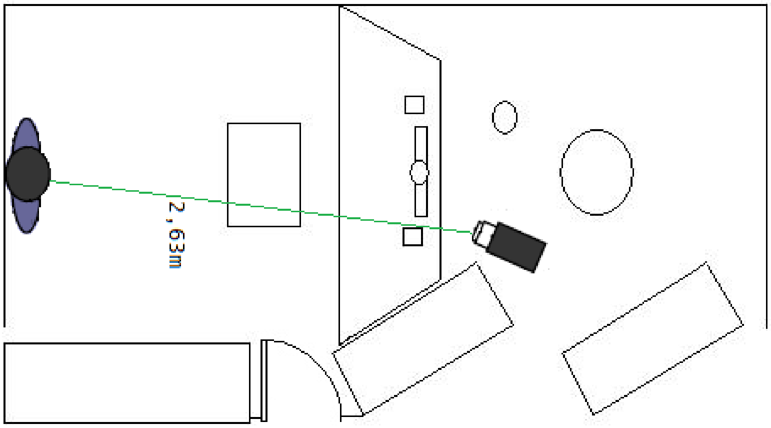


Figure 2.7: Configuration for the upper-body experiment. The experiment was recorded by just one camera, C2. This experiment took approximately 8 seconds.

2. After that, the subject must lower their arms back to the initial position next to their waist.

2.2.1 Equipment

Videos were collected using three digital cameras (See Fig. 2.6) without any synchronizing mechanisms but the standard procedure previously explained. The digital cameras C1, C2 and C3 are identified in Figure 2.8. The cameras C1 and C2 were put on both adjustable portable aluminum alloy tripods, which have the characteristic of stability and portability. While camera C3 was attached to the screen.

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 (tagged as camera C2 in Fig. 2.1) 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

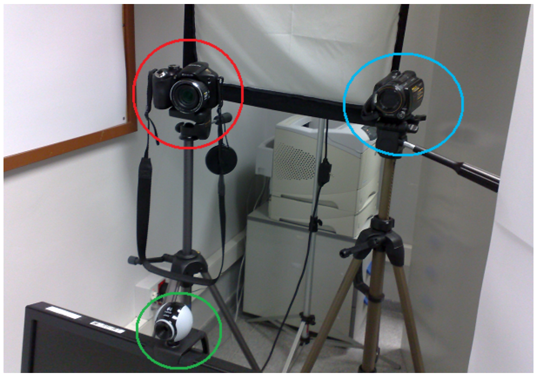


Figure 2.8: Digital cameras are highlighted by each circle. Inside the red one, camera C1 can be appreciated, a Casio Exilim EX-FH20. Inside the blue circle, camera C2 is a Sony Handycam model HDR-XR520VE and, inside the green circle it is camera C3, a Logitech Webcam 250C.

video that is 07:36 minutes length on cameras C2 and C3, it takes 53:53 length on camera C1. More detailed information can be visualized in Table 2.2.

The illumination approximated ambient lighting. Specifically, the recording setup consisted of one 1000 watt quartz halogen focusing open face floodlight (See Table 2.4), 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 camera C1's frame rate setting. The selected illumination avoids the flickering effect on camera C1 during the acquisition process. The light was 2,37 meters from the participant. In addition, a white card background was mounted on the walls behind and next to the participant.

The computer used for this experiment was a laptop Dell XPS M1330 connected to a 20" Screen, a keyboard, a mouse and the speakers. The XPS range is all about high performance, so the computer uses one of the fastest available Intel Core 2 Duo mobile CPU, the core on the T9300 run at 2.5GHz and is backed by 4MB of level 2 cache memory. See Table 2.3 for a detailed description.

Tech Specs	Camera C1	Camera C2	Camera C3
Picture			
Video Settings			
Resolution	480x360	1920x1080	640x480
Bitrate	7129 bps	3419 bps	1905 bps
Frequency	210 fps	30 fps	30 fps
Aspect Relation	4:3	16:9	4:3
Compression Video	IBM Motion JPEG	H.264/AVC	MPEG4 (DIVX/XVID)
Audio Settings			
Bitrate	-	448 bps	-
Channels	-	5	-
Frequency	-	48000 Hz	-
Format	-	AC3	-

Table 2.2: Cameras technical specifications. Cases marked with - means that conditions are non existent.

Tech Specs	Data
Processor Clock speed	2.5 GHz
Processor type	Core 2 Duo T9300
Processor manufacturer	Intel
RAM installed	4096 MB
RAM technology	DDR2 667MHz
Motherboard Chipset type	Intel 965PM/GM
Data bus speed	800 MHz
Hard drive type	7,800RPM SATA Hard Drive
Hard drive size	320 GB
Graphics processor	Nvidia GeForce 8400M GS
Video outputs	HDMI, D-Sub

Table 2.3: Computer technical specifications.

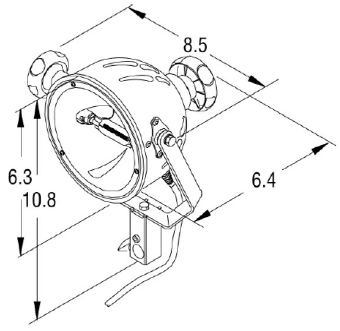
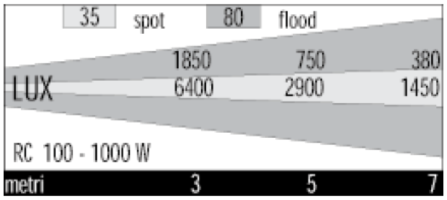
Tech Specs	Data
Dimensions (inches)	
Physical	Strong and lightweight aluminum construction. Focusable open face fixture. Lamphouse is high temperature finished. Flat aluminum yoke with standard 5/8” receiver.
Electrical	120-230V, 50/60Hz. 16/3 SO 90 type cable. Twelve foot (12) standard UL lead with 10 amp. Supplied with molded u-ground connector.
Lamp	1000w Maximum. R7s socket for linear double. Ended Tungsten halogen lamps. 
Optical	Designed reflector is made of polished aluminum. Extremely smooth focusing mechanism. Focusable from the rear.

Table 2.4: Illumination technical specifications. The reflector is made of polished aluminum to ensure a high degree of reflectivity. An extremely smooth focusing mechanism allows for a particularly even illumination and excellent power throughout all of the working positions.

2.2.2 Software

We have designed software to manage all the information about subjects and to control the camera C3. Stimuli were also presented by this software in a randomized order. The program code is divided into separate files that make the software modular and robust. An overview about the code's and file's hierarchy is supplied in this section.

The software is written in C++ combining OpenCV [70], Windows MFC and the Windows SDK [65]. Among the activities performed by the software, the activities that stand out are:

- User's register. The software saves all the information of each user and assigns a unique random ID to each user for privacy purposes.
- User's identification. It is able to identify anyone by a PIN code settled by the user during the registration process.
- Videos/Images presented in a randomized order and the software saves the timestamp between videos/images.
- Webcam controlled by software. Short videos are displayed by the software during each session. A separate thread keeps recording while contents are displayed.
- Manage subject's sessions. No video or image is displayed twice to the same subject.
- Configurable software. Timing issues are perfectly configurable.

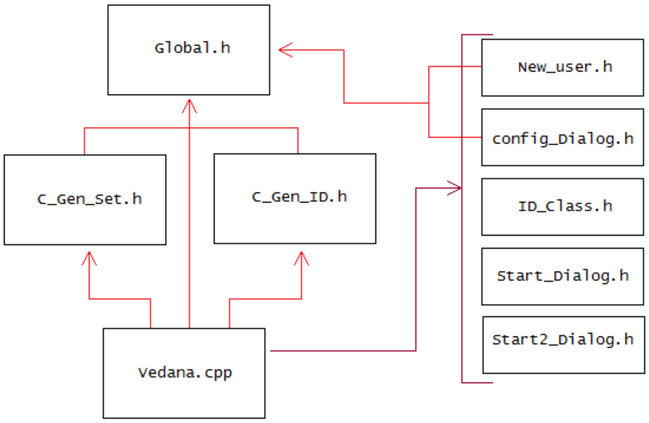
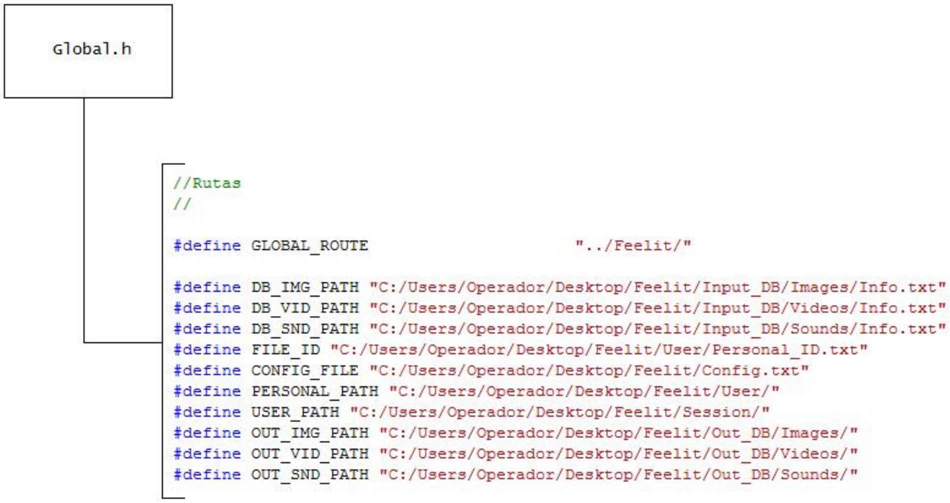
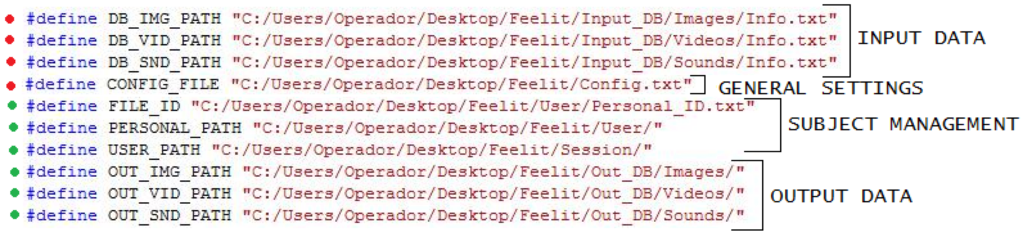


Figure 2.9: The program is divided into separate files making the software modular and robust. As it can be appreciated, Vedana.cpp is the main core of the program. Therefore, it is in charge to present stimuli and to record the data.



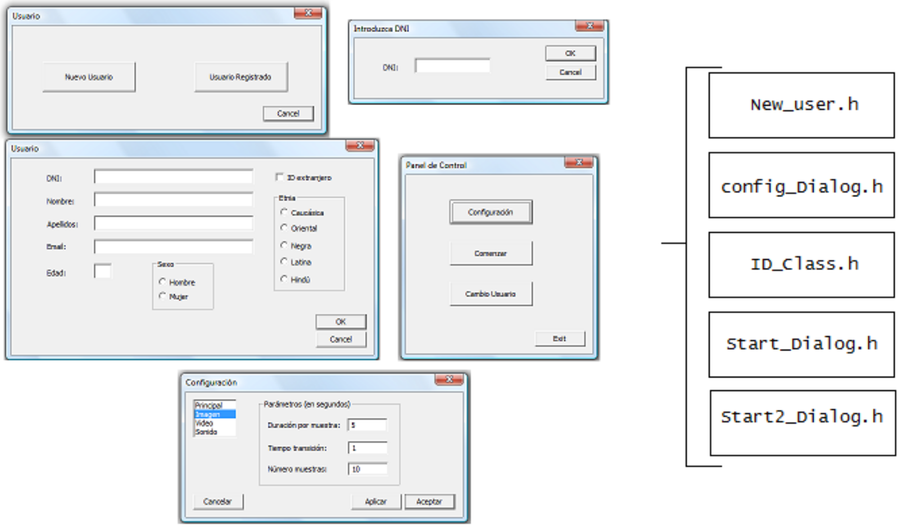
```
//Rutas
//

#define GLOBAL_ROUTE                      "../Feelit/"

#define DB_IMG_PATH "C:/Users/Operador/Desktop/Feelit/Input_DB/Images/Info.txt"
#define DB_VID_PATH "C:/Users/Operador/Desktop/Feelit/Input_DB/Videos/Info.txt"
#define DB_SND_PATH "C:/Users/Operador/Desktop/Feelit/Input_DB/Sounds/Info.txt"
#define FILE_ID "C:/Users/Operador/Desktop/Feelit/User/Personal_ID.txt"
#define CONFIG_FILE "C:/Users/Operador/Desktop/Feelit/Config.txt"
#define PERSONAL_PATH "C:/Users/Operador/Desktop/Feelit/User/"
#define USER_PATH "C:/Users/Operador/Desktop/Feelit/Session/"
#define OUT_IMG_PATH "C:/Users/Operador/Desktop/Feelit/Out_DB/Images/"
#define OUT_VID_PATH "C:/Users/Operador/Desktop/Feelit/Out_DB/Videos/"
#define OUT_SND_PATH "C:/Users/Operador/Desktop/Feelit/Out_DB/Sounds/"
```

Figure 2.10: The Global.h file specifies all the paths used during the experiment. There are paths for the input data (DB_XXX_PATH) and for the output data (OUT_XXX_PATH). There are also detailed the PERSONAL_PATH (where is placed all private information about participants) and the USER_PATH (place where session information is saved). FILE_ID is an important file that keeps the relation between each participant PIN and its ID (See Fig. 2.13).



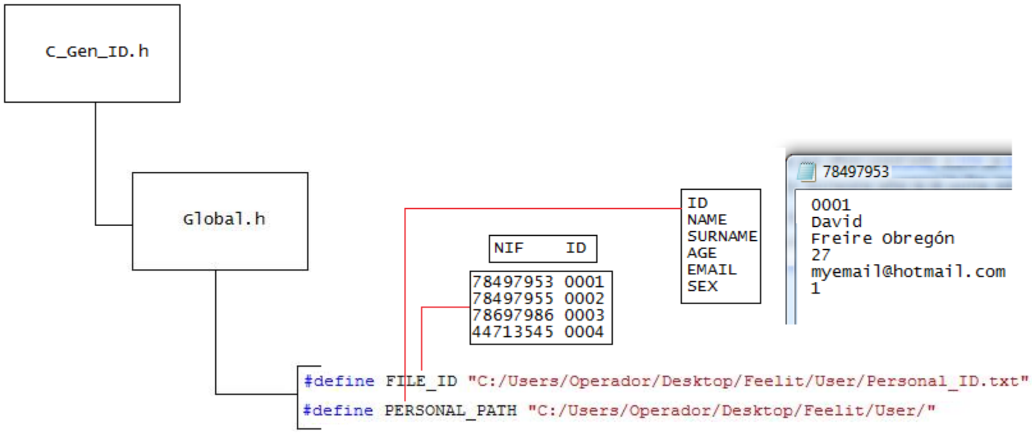
```
#define DB_IMG_PATH "C:/Users/Operador/Desktop/Feelit/Input_DB/Images/Info.txt"
#define DB_VID_PATH "C:/Users/Operador/Desktop/Feelit/Input_DB/Videos/Info.txt"
#define DB_SND_PATH "C:/Users/Operador/Desktop/Feelit/Input_DB/Sounds/Info.txt"
#define CONFIG_FILE "C:/Users/Operador/Desktop/Feelit/Config.txt"
#define FILE_ID "C:/Users/Operador/Desktop/Feelit/User/Personal_ID.txt"
#define PERSONAL_PATH "C:/Users/Operador/Desktop/Feelit/User/"
#define USER_PATH "C:/Users/Operador/Desktop/Feelit/Session/"
#define OUT_IMG_PATH "C:/Users/Operador/Desktop/Feelit/Out_DB/Images/"
#define OUT_VID_PATH "C:/Users/Operador/Desktop/Feelit/Out_DB/Videos/"
#define OUT_SND_PATH "C:/Users/Operador/Desktop/Feelit/Out_DB/Sounds/"
```

Figure 2.11: File hierarchy is divided into files generated by the user and files generated by the program. The user (not the subject, an assistant technician) can edit and modify the red dot files. Red dot files make reference to the general settings of the experiment and stimuli presented to subjects: videos/image/sound data information for each collection (collection ID, path and number of video/image/sound files in the collection). The program manages the green dot files. Subect's management information and output data are placed in the green dot files.



Duración por muestra: 5
Tiempo transición: 1
Número muestras: 10

Figure 2.12: Application programs in any language need some interactive screens to give inputs and obtain outputs. All windows based GUIs will definitely use dialogs for such user interactions. Some dialogs used by the software can be appreciated in this image.

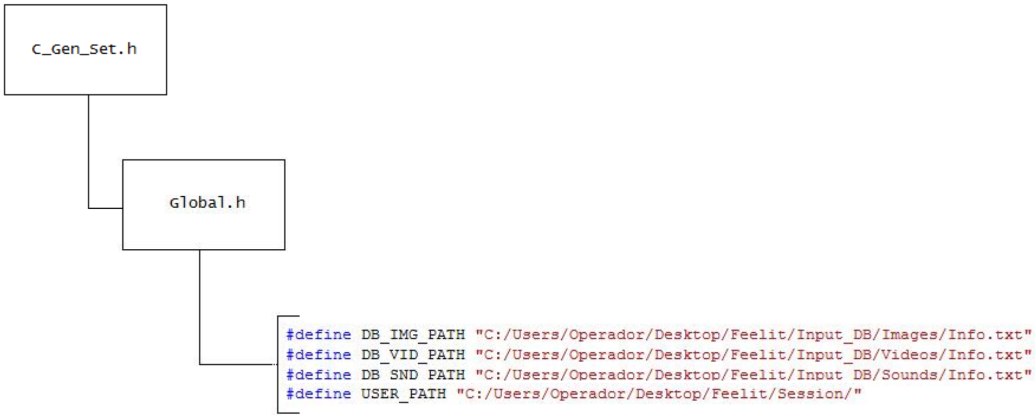


NIF	ID
78497953	0001
78497955	0002
78697986	0003
44713545	0004

```
0001
David
Freire Obregón
27
myemail@hotmail.com
1
```

```
#define FILE_ID "C:/Users/Operador/Desktop/Feelit/User/Personal_ID.txt"
#define PERSONAL_PATH "C:/Users/Operador/Desktop/Feelit/User/"
```

Figure 2.13: In order to guarantee the safety of the participant's privacy, no personal information is used during the data acquisition process. A random and unique ID is assigned to each participant. This ID is the key to manage the entire subject's information during the data acquisition process. As it can be appreciated, the FILE_ID file, keeps the relation between a given ID and the user's PIN (NIF). In this image it is also detailed how information is saved into the system.



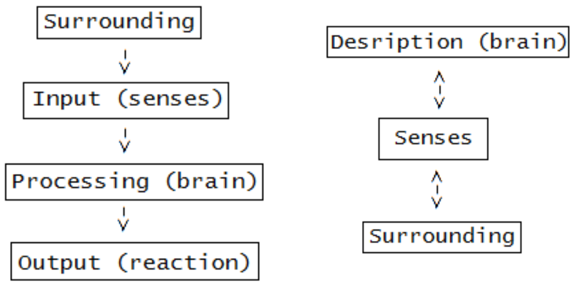
```
#define DB_IMG_PATH "C:/Users/Operador/Desktop/Feelit/Input_DB/Images/Info.txt"
#define DB_VID_PATH "C:/Users/Operador/Desktop/Feelit/Input_DB/Videos/Info.txt"
#define DB_SND_PATH "C:/Users/Operador/Desktop/Feelit/Input_DB/Sounds/Info.txt"
#define USER_PATH "C:/Users/Operador/Desktop/Feelit/Session/"
```

Figure 2.14: Videos/Images IDs and the timestamp between videos/images are saved. Information about timestamp is important for editing videos and synchronizing the record process when multiples cameras are used. All session data for each subject are stored for the purpose of avoiding repeating videos/images in future sessions.

2.3 Datasets

According to the elicitation method of emotions, databases can be categorized into three major classes: induced, acted, and naturalistic. This database is categorized into the induced class. Because of this reason, a thorough set of stimuli should be selected.

The first point to make clear is to justify the reason why video clips are the stimuli selected. Perception is the active process of acquiring, interpreting, selecting, and organizing sensory information that responds to a specific kind of physical phenomenon (stimuli) such as sound, temperature and taste [18]. In other words, it is a dynamic search for the best interpretation of available data. Some types of perception are depth perception, speech perception and haptic Perception. Cognitive science is an interdisciplinary study of the brain drawing from relevant fields including psychology, philosophy, neuroscience, linguistics, anthropology, computer science, biology, and physics.



(a) Direct Perception Scheme (b) Indirect Perception Scheme

Figure 2.15: Perception processes [18].

There are two possible perceptions: indirect or direct. In the indirect perception (top down perception) a lot of information reaches the eye but 90 % is lost before it reaches the brain. In this case, prior knowledge and past experience are crucial. Perception is a dynamic relationship between description (in the brain), senses and surrounding (See Fig. 2.15-b). To make a good perception the sensory information is combined with the past experience. The formation of incorrect interpretation will lead to errors of perception (Visual illusions) and perception can be ambiguous. On the other hand, in the direct perception (bottom up), there is no need for interpretation as the information that we receive is sufficient to interact directly with the environment. Sensory information is analyzed in one direction, beginning by analyzing the simpler input sensory data to the more complex tasks as processing them (See Fig 2.15-a). In the direct perception is enough helping information that aids our perception:

- Brightness: objects with brighter images are closer.
- Size: objects with smaller sizes are more distant.

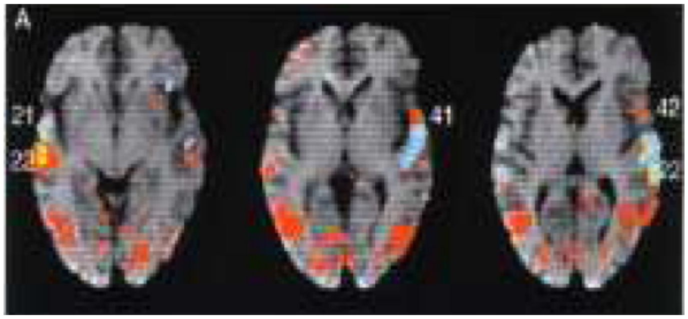


Figure 2.16: By using fMRI images, scientists were able to see how the the different regions of the brain were stimulated and also when they were stimulated while applying different stimuli such as audio or visual stimuli or both of them combined [23]. In this figure three colors, the red one stands for watching lips movement without sound, the blue one stands for listening a speech and the yellow is activated by both can be appreciated.

Still images only activate an unimodal perception (vision). For a good perception of a shape of an object the brain uses some several information. For example, for the shape of an object we need texture, shading and motion. If a single modality is not enough to make a good perception, information from several modalities can be combined. To integrate different Information two strategies are used:

1. The first is to maximize the incoming information.
2. The second strategy is to reduce the variance in the sensory estimate to increase its reliability.

In the case of different events happening simultaneously it appears that multi sensory integration favors the enhancement of signals derived from the same events and the suppression of signals derived from different events (See Fig. 2.16). Thus, video sequences provide a better stimulus because they maximize the incoming information to the subject and generate a multi-modal perception.

2.3.1 Video sequences

The stimuli dataset is composed by 230 of short video sequences downloaded from the Internet [85, 54, 56] and edited before the experiment. The edition process ensures the absence of redundant information, as well as keeping under control the execution time for each acquisition process. Although the main idea of the edition process was to make each sequence length about 20 seconds, it was not always possible because of avoiding the loss of the semantic meaning of the videos. Each of these 230 sequence was tagged into 17 categories/collections depending on their content. In fact, 17 video sequences, a sequence per category, were presented to each participant for each session. Because of the absence

Video Category	Number of Samples	Keywords
Animals - Positive	12	puppies, horses
Animals - Negative	30	larva, killing, cockroach
Extreme	17	fights, earthquakes
People - Babies	07	smiling, funny
People - Men	08	talking
People - Women	14	talking, dancing
Politics - Positive	04	funny
Politics - Negative	15	lies
Sex - Bizarre	14	pissing, tranny, muscles
Sex Female - Heterosexual	24	sexy
Sex Female - Homosexual	22	lesbian
Sex Male - Heterosexual	11	handsome
Sex Male - Homosexual	12	gay
Sports - Positive	03	F1, soccer
Sports - Negative	08	injuries
Nature	22	plants
People - Yawn	07	yawn, sleeping

Table 2.5: Video sequences categories. The order followed to present the stimuli in the experiments is in the same order of this table. Indeed this order responds to a deliberate strategy to induce responses on the subject. For example, people yawning's category is the last one with the intention to induce people either to yawn or to contempt.

of rights above those videos, it is impossible to provide which video sequences where presented to participants. However, other kind of information like how these videos were tagged or how they were presented to subjects, can be provided. First of all, video sequences were tagged depending on their contents (See table 2.5). Secondly a sequence per category is presented to the subject following the table 2.5 order. Inside each category, the selection of the presented video is totally random.

2.3.2 RISE images

Between videos a RISE image of the last frame is showed during three seconds. This technique called Random Image Structure Evolution (RISE) [47] is used in experimental investigations of high-level visual perception. Notably, these image sequences are created in a manner that strictly controls a number of important low-level image properties, such as luminance and frequency spectra, thus reducing confounds in the analysis of high-level visual processes. The main aim for using RISE images is the effect of this kind of images about removing any low-level adaptation and memory effects. Given these considerations, a preferred alternative approach to the very simple image scrambling implemen-

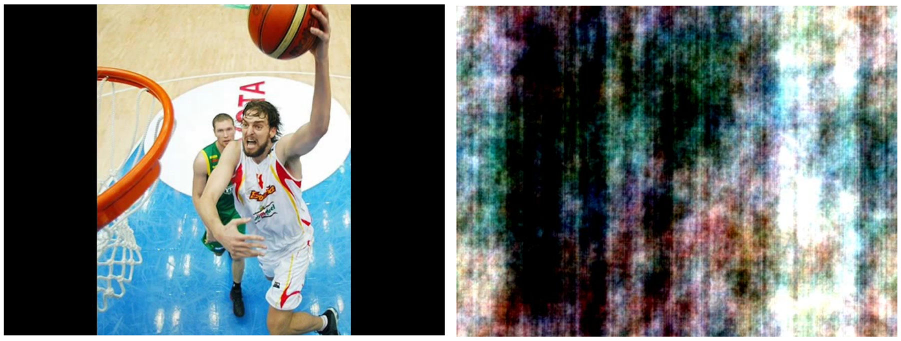


Figure 2.17: A sample RISE image generated by degradation of the phase matrix of the source image. Although the technique disrupts the spatial structure of the image, important low-level image properties of the original image, such as the spatial frequency spectrum and overall luminance, are perfectly preserved.

tation of RISE is to first perform an analysis of the frequency spectrum (i.e., a Fourier analysis) of the source image, then to manipulate the spatial structure of the image without altering its original power spectrum (as well as the overall luminance and contrast of the image). In the Fourier domain, this can be done by altering what is called the phase spectrum while retaining the amplitude (or power) spectrum. In fact, it has been shown that much of the information specifying natural image structure lies in the global phase spectrum [36], so that randomizing the phase spectrum has the effect of degrading the spatial structure of an image. (Further, replacing the phase spectrum of an image with that of another image results in an image that resembles the donor of the phase rather than of the amplitude spectrum; however, the lower-level attributes of the resulting image may better resemble the donor of the amplitude spectrum.) As such, an alternative to the scrambling technique described above is to manipulate the source image in the Fourier domain, progressively transforming the phase while holding constant the amplitude spectrum [45, 46]. In the case of the onset portion of a RISE sequence, then, the perfect image evolves from a random-seeming starting image that has been constructed using a random phase spectrum combined with the amplitude spectrum of the original image. The onset subsequence is achieved through progressive transformation of the random phase spectrum into that of the perfect image, and the offset subsequence is simply this process in reverse. (See Fig. 2.17 for an example.)

Chapter 3

Vedana Database

Selecting a name for a facial expression database is not an easy task. Contrary to most of researchers could think, the result of this task has an essential influence into the success of the dissemination process. Usually, the name for these databases is selected based on short abbreviations of the database's meaning, no more than two syllables. In many cases the choice is a "catchy name", easy to pronounce and remember (FERET [42], DAFEX [1], MMI [40]). Another common choice is to use the name of the place this database was made up (Harvard RL [20], U. Oulu [78], Sheffield [77]), but what happens if two databases are made in the same place?. Then, the answer could be to list these names (Yale, Yale B [6]), which can lead to researcher's confusion (Which is the last one? Yale or Yale B?). In contrast, other naming cases are difficult for researchers to remember (KFDB [9], M2VTS [22]). In this work the selection process of a name is far away from the conventional rules already mentioned. The main objective is to make the product to have its own identity and the choice should be a word with meaning, easy to pronounce, with no need of translation and elegant. **Vedana** is a word in sanskrit and pali traditionally translated as either 'feeling' or 'sensation'. In general, vedana refers to the pleasant, unpleasant and neutral sensations that occur when our internal sense organs come into contact with external sense objects and the associated consciousness. Because of this word's nature and our goal with this work, it is the chosen name for this collection. This chapter details a full description of the database. The results achieved as well as the statistical information about the subjects is described in Section 3.1, while a comparative analysis with other databases is fully detailed in Section 3.2. This chapter concludes with the complete process followed to disseminate Vedana Database in Section 3.3.

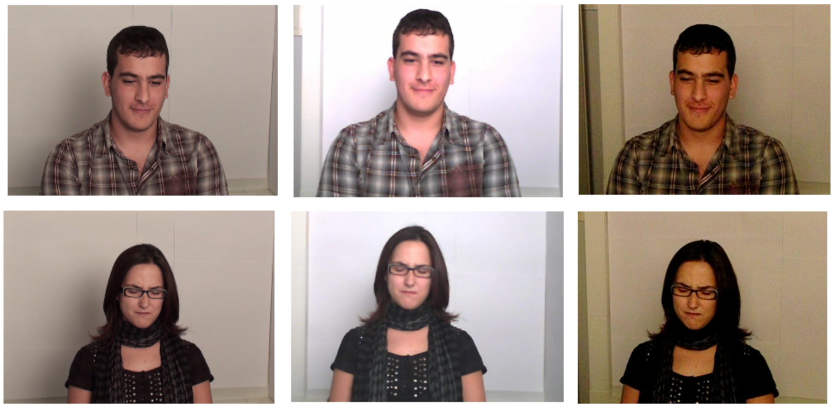


Figure 3.1: Different camera views with the subject looking right into the screen. It can be appreciated that recording settings are not the same for each camera according to the results this work is interested to achieve. The left pictures stand for the Sony Camera. This camera recorded 126 video sequences (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.

3.1 The 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 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). In Table 3.1 can be appreciated the distribution of these video sequences among the three cameras. The images on Figure 3.1 belong to the main set that is divided into three subsets depending on the camera used (see Fig. 2.1):

1. On the left view subset, all images are captured from camera C2, the Sony Camera.

2. On the frontal view subset, all images are captured from camera C3, the webcam, which is attached to the screen.

3. On the right view subset, all images are captured from camera C1, the Casio Camera, which captured 210 frames per second.

Experiment	Camera	Clips	Size
Facial Expression	C1	93	229,53 GB + 2,53 GB
Facial Expression	C2	126	24,87 GB + 0,07 GB
Facial Expression	C3	126	11,33 GB + 0,03 GB
Body	C2	126	852,5 MB + 4 MB
		Total Size	269,21 GB

Table 3.1: Database Size. It is important to mention that the size after the "+" is related to the background record session for each experiment and camera. Thus, the total number of video sequence of this database is 475.

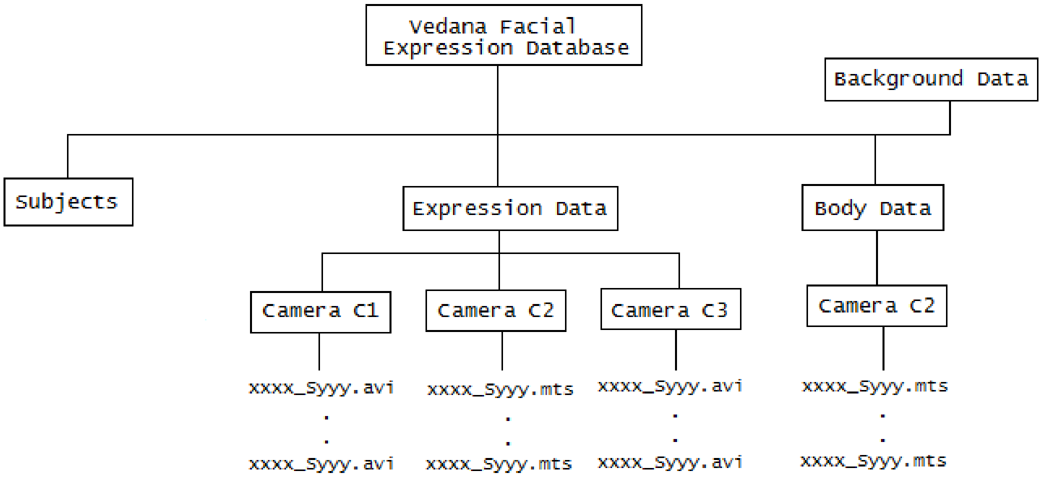


Figure 3.2: Data organization of the Vedana facial expression database.

According to the Table 3.1, the original Vedana database requires about 269,21 GB storage space. To facilitate the distribution, a directory structure is set as it is shown in Fig. 3.2. Thus, researchers can download files depending on their requirements. The subject directory contains public information related to the subjects as gender, accessories and the timestamp for each subject's session. The facial expression experiment and body experiment have their own directories too. Inside each of those directories the data is divided into several directories that belong to the cameras used in the experiment. Inside each camera directory, the database is structured by subjects. Each subject has two sequences. In addition, background video sequences have their own directory. For the video naming convention adopted in this database, the filename of each video sequence encodes the majority of the ground truth information of the sequence. Its format is described as follows: xxxx_Syyy.fff

It consists of two fields and it is 13 characters long. The fields are separated by underline marks as shown above. In fields, ”x”s, ”y”s and ”f”s represent two digital number sequences and a character type sequence respectively, which vary with the properties of each video sequence. The meaning, character type sequence and number sequences of each field are described in turn as follows:

1. The four digital number sequence tagged as ”x”s indicates a unique ID for each subject.

2. The three digital number sequence tagged as ”y”s indicates the session.

3. The three digital number sequence tagged as ”f”s indicates the format (.mts or .avi).

An example for the filename of the second session of subject 21 in camera C1 is 0021_S002.avi.

Another important feature of the video clips generated in this database is that clips begin with a head movement in order to synchronize the cameras. However, for camera C3 video sequences this procedure is not necessary. A green frame is inserted into the C3's video sequences at the beginning of each session. While RISE images are showed, a red frame is inserted into the sequences. When a subject is watching multimedia contents, no frame is inserted. For cameras C1 and C2 sequences the head movement means the start of the sequence (end of green frame in camera C3) and no frames are inserted during the acquisition process. For this reason, a file inside the timestamp directory contains the timestamp for the sessions, pointing when RISE images or contents are showed to the participant.

3.1.1 Statistics

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 (See Fig. 3.4). 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 [42] 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.

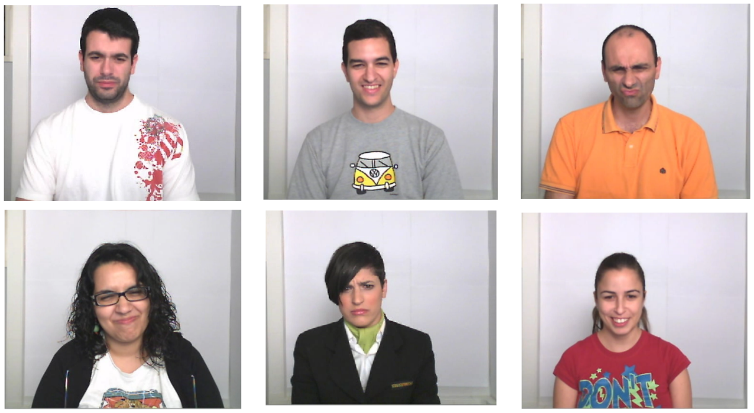


Figure 3.3: An example of a sequence of images of different subjects of both genders.

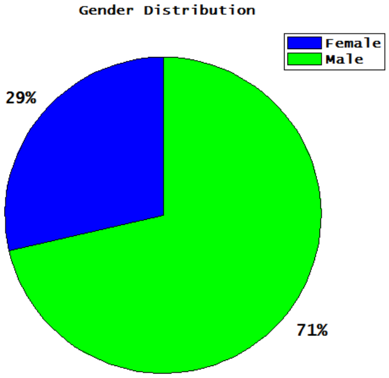


Figure 3.4: Database gender distribution.

Age range	Female	Male	Total
16 to 20 years old	01	11	12
21 to 30 years old	13	24	37
31 to 40 years old	01	07	08
41 to 50 years old	02	03	05
51 to 60 years old	01	00	01
Mean age	29,06	26,87	27,49

Table 3.2: Age distribution by ranges.

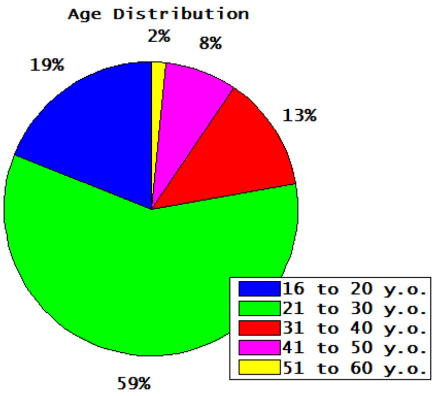
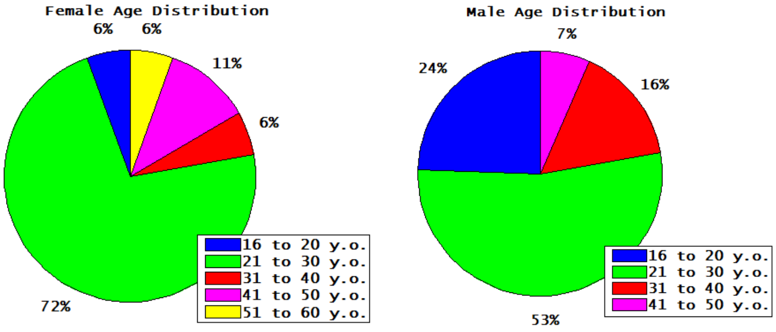


Figure 3.5: Database age distribution.



(a) Female age distribution

(b) Male age distribution

Figure 3.6: Age distribution by gender.

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. Table 3.3 resumes the accessories and whether or not each participant has facial hair along the sessions.

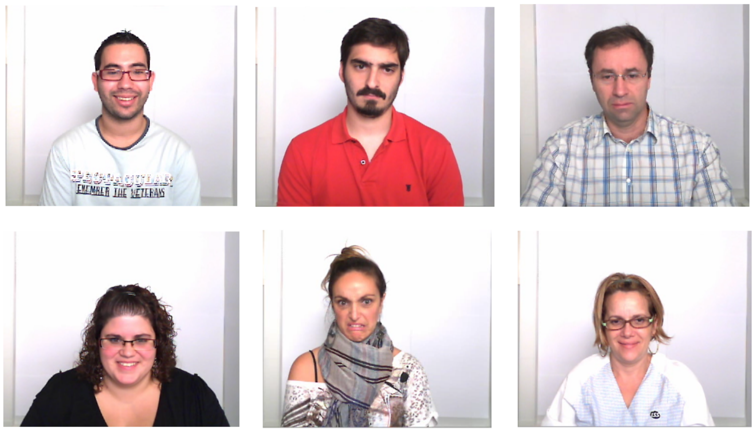


Figure 3.7: Example sequence that shows the age diversity of Vedana.

Item	Category	Female	Male
Glasses	None	22	56
Glasses	Eyeglasses	10	30
Glasses	Sunglasses	04	04
Hair	Buzz cut	00	02
Hair	Short	02	70
Hair	Medium	08	11
Hair	Long	13	01
Hair	Tied	12	03
Hair	Hood	01	03
Facial hair	None	36	27
Facial hair	Beard	00	51
Facial hair	Goatee	00	11
Facial hair	Moustache	00	01
Facial hair level	None	36	27
Facial hair level	Incipient	00	26
Facial hair level	Daily	00	28
Facial hair level	Weekly	00	09

Table 3.3: Accessories session scheme. This table shows the accessories or hair variations in the database.

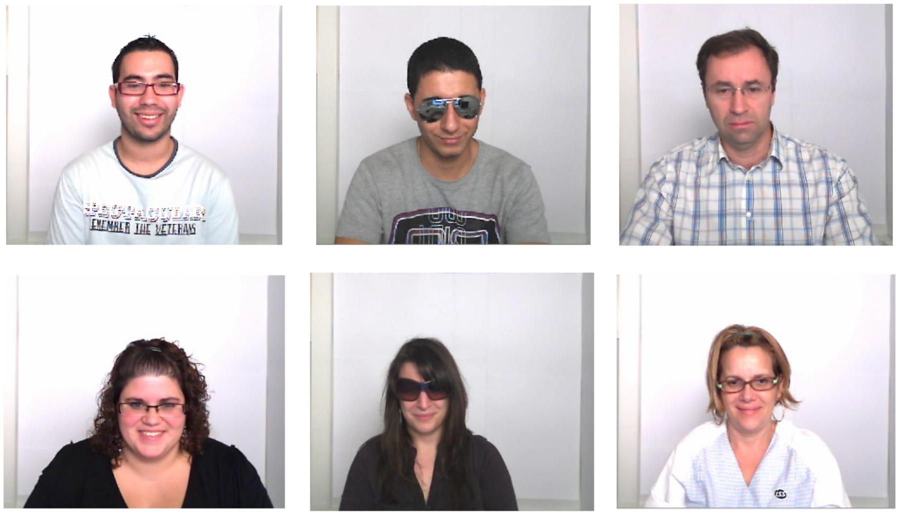


Figure 3.8: In this figure, different subjects wearing glasses can be appreciated. As they were left participant dependent, there is a great diversity of glasses; dark frame glasses, thin and white frame glasses, glasses without frame, etc.

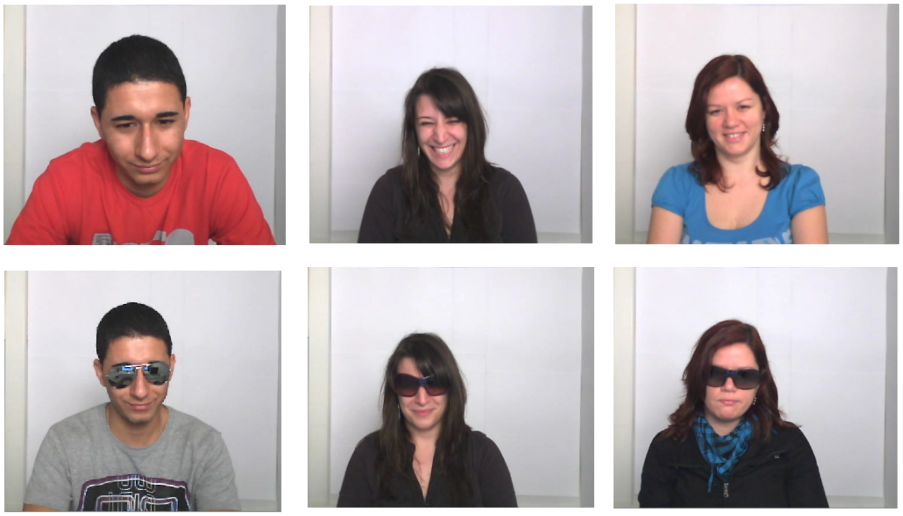


Figure 3.9: Those subjects that brought their own sunglasses were asked to make both experiments (facial recognition and body experiment) wearing the sunglasses for the first session and without sunglasses for the second one.

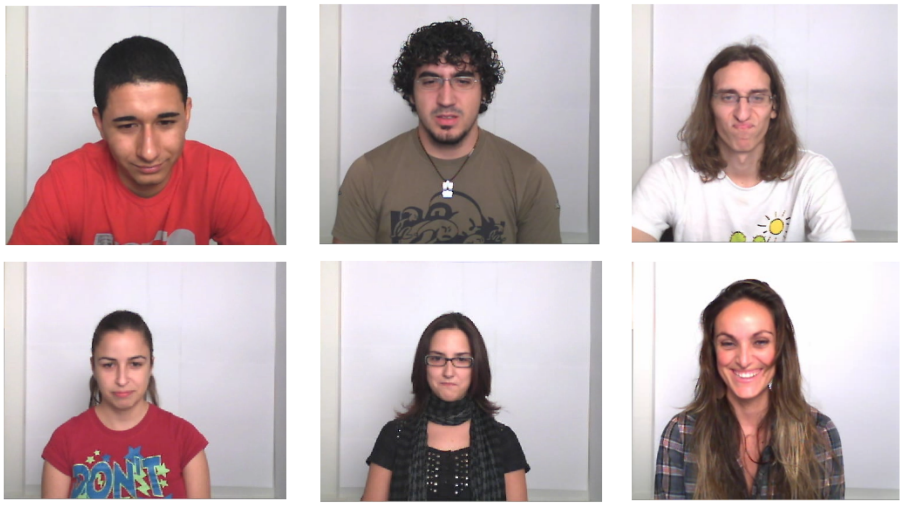


Figure 3.10: Hairstyle is also a strong point in Vedana's diversity.

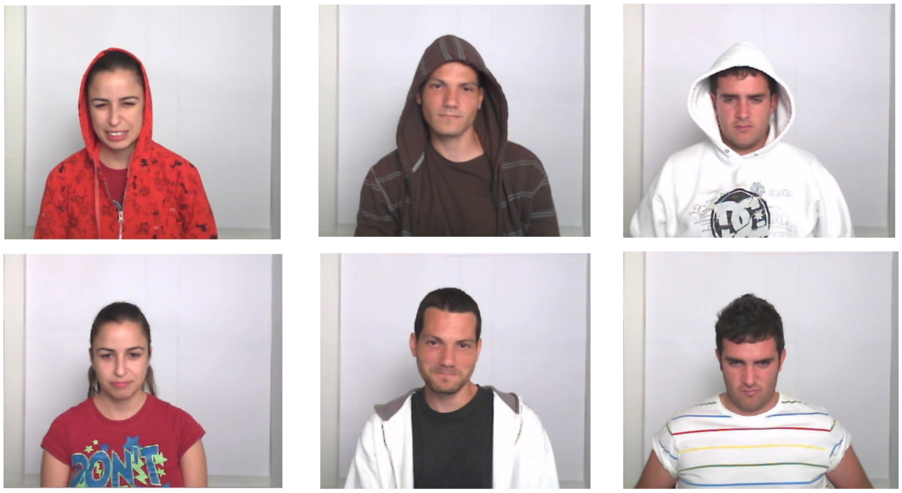


Figure 3.11: Those subjects that brought their own hoods were asked to make both experiments (facial recognition and body experiment) wearing the hood for the first session and without hood for the second one.

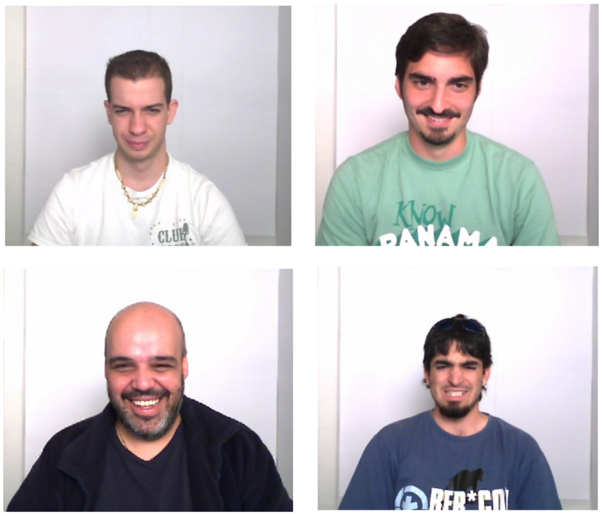


Figure 3.12: Facial hair introduces occlusion to the subjects. As it can be seen in this figure, there is a great diversity of facial hairstyle in the database.

3.1.2 Multimedia

The dynamic facial expression clips capture emotions such as happiness, sadness or disgust (See Fig. 3.13). 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 raised and lowered 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. 3.14).

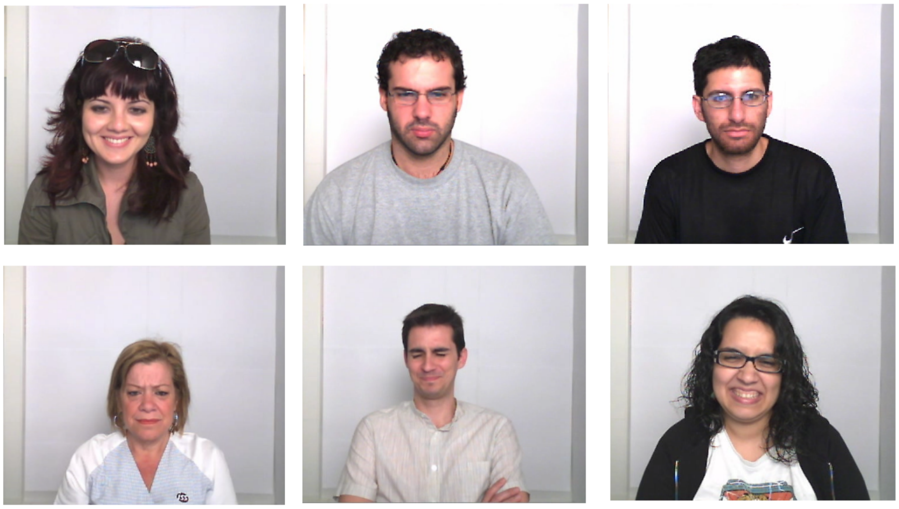


Figure 3.13: Facial expressions responses from different subjects.

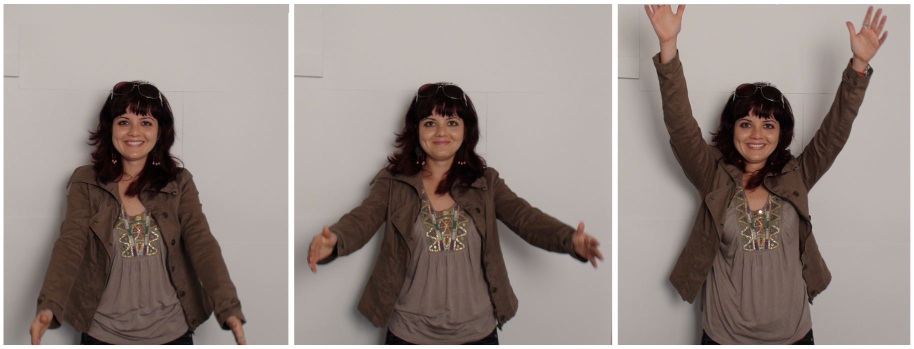


Figure 3.14: Example sequence of the body experiment.

	Cohn-Kanade	JAFFE	AR	PIE	MMI	**Vedana**
Reference	[87]	[28]	[31]	[21]	[40]	-
Static Images	No	219	4000	40000	740	No
Videos	2105	No	No	No	848	475
Emotion exp.	Yes	Yes	04	04	Yes	Yes
Multiple AUs exp.	Yes	Yes	04	04	Yes	Yes
AU-coded	Yes	No	No	No	Yes	No
No. subjects	210	10	126	68	19	63
Subject's age	18-50	?	?	?	19-62	18-60
Gender	Both	Female	Both	Both	Both	Both
Lighting	uni	?	var	var	uni	uni
Facial hair, glasses	No	No	Yes	Yes	Yes	Yes
Downloadable	No	Yes	No	No	Yes	Yes
Searchable	No	No	No	No	Yes	Yes

Table 3.4: Overview of some existing Face Databases.

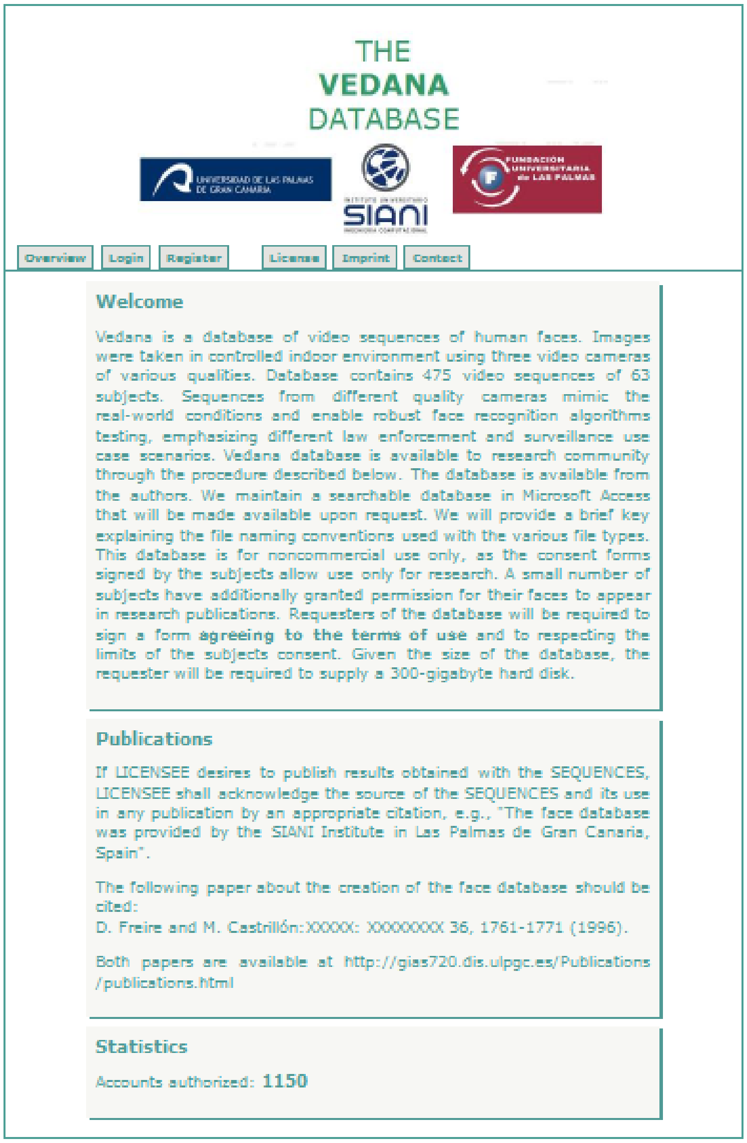
3.2 Comparative Analysis

The expression clips differ from previously available facial expression databases in several ways. First, as noted, the facial expressions have not been verified in a formal sense as being instances of one the primary expressions defined by Ekman and Friesen [16]. Second, most of the expressions are more subtle than those available previously, though see Pantic [30] for a review of the range of face images used in automatic analysis of facial expression. Third, because these are dynamic stimuli, head and eye movements often accompany the expressions. Finally, some clips contain more than one expression (e.g., a puzzled expression, which turns to surprise or disbelief, and ultimately laughter). Combined, the close-range videos provide test stimuli for face recognition and tracking algorithms that operate when the head is undergoing rigid and/or nonrigid transformations. The dynamic expression videos are likewise useful for computer graphics modeling of heads and facial animation.

An overview of the databases of face images that have been made publicly available is provided in Table 3.4. From these, the Cohn- Kanade facial expression database [87] 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 MMI [40] and Vedana databases, 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 this work's novel face image database.

3.3 Dissemination

The database is available from the authors (See Fig. 3.15). 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 (See Fig. 3.16) agreeing to the terms of use and to respecting the limits of the subject's consent. The database is available by download or by postage. Given the size of the database, the requester will be required to supply a 300-gigabyte hard disk and will be responsible for handling and postage.


UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
INSTITUTO UNIVERSITARIO SIANI
FUNDACIÓN UNIVERSITARIA de LAS PALMAS
Overview
Login
Register
License
Imprint
Contact

THE **VEDANA** DATABASE

Welcome

Vedana is a database of video sequences of human faces. Images were taken in controlled indoor environment using three video cameras of various qualities. Database contains 475 video sequences of 63 subjects. Sequences from different quality cameras mimic the real-world conditions and enable robust face recognition algorithms testing, emphasizing different law enforcement and surveillance use case scenarios. Vedana database is available to research community through the procedure described below. The database is available from the authors. We maintain a searchable database in Microsoft Access 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 subjects consent. Given the size of the database, the requester will be required to supply a 300-gigabyte hard disk.

Publications

If LICENSEE desires to publish results obtained with the SEQUENCES, LICENSEE shall acknowledge the source of the SEQUENCES and its use in any publication by an appropriate citation, e.g., "The face database was provided by the SIANI Institute in Las Palmas de Gran Canaria, Spain".

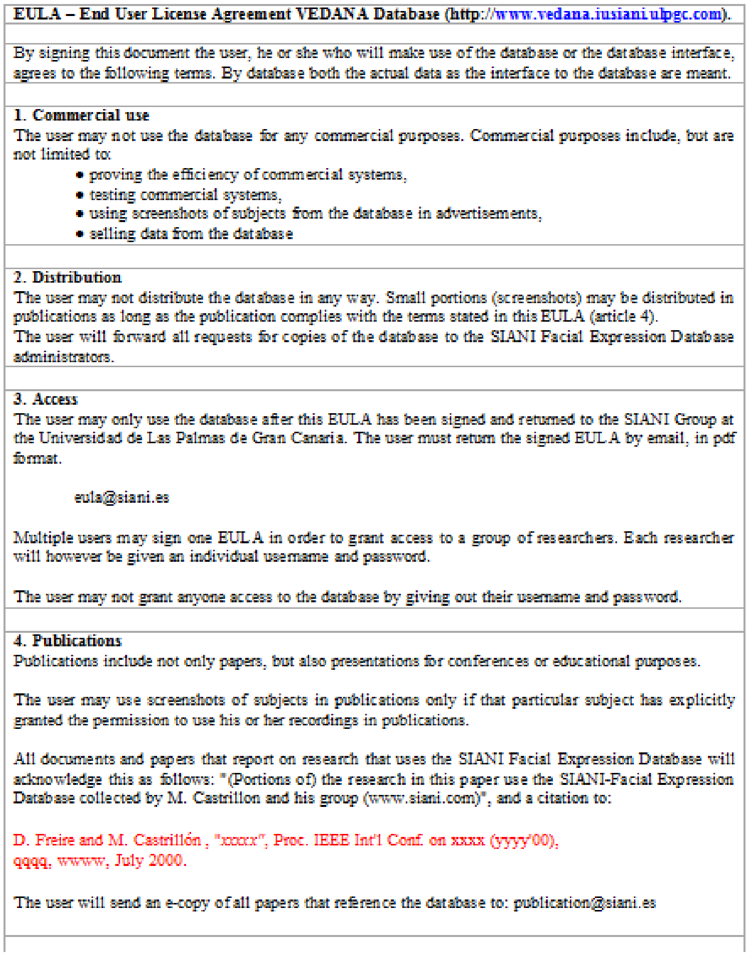
The following paper about the creation of the face database should be cited:
D. Freire and M. Castrillón:XXXXX: XXXXXXXX 36, 1761-1771 (1996).

Both papers are available at http://gias720.dis.ulpgc.es/Publications /publications.html

Statistics

Accounts authorized: **1150**

Figure 3.15: Vedana Database Webpage [79].



EULA – End User License Agreement VEDANA Database (http://www.vedana.iusiani.ulpgc.com).

By signing this document the user, he or she who will make use of the database or the database interface, agrees to the following terms. By database both the actual data as the interface to the database are meant.

1. Commercial use
The user may not use the database for any commercial purposes. Commercial purposes include, but are not limited to:

- proving the efficiency of commercial systems,
- testing commercial systems,
- using screenshots of subjects from the database in advertisements,
- selling data from the database

2. Distribution
The user may not distribute the database in any way. Small portions (screenshots) may be distributed in publications as long as the publication complies with the terms stated in this EULA (article 4).
The user will forward all requests for copies of the database to the SIANI Facial Expression Database administrators.

3. Access
The user may only use the database after this EULA has been signed and returned to the SIANI Group at the Universidad de Las Palmas de Gran Canaria. The user must return the signed EULA by email, in pdf format.

eula@siani.es

Multiple users may sign one EULA in order to grant access to a group of researchers. Each researcher will however be given an individual username and password.

The user may not grant anyone access to the database by giving out their username and password.

4. Publications
Publications include not only papers, but also presentations for conferences or educational purposes.

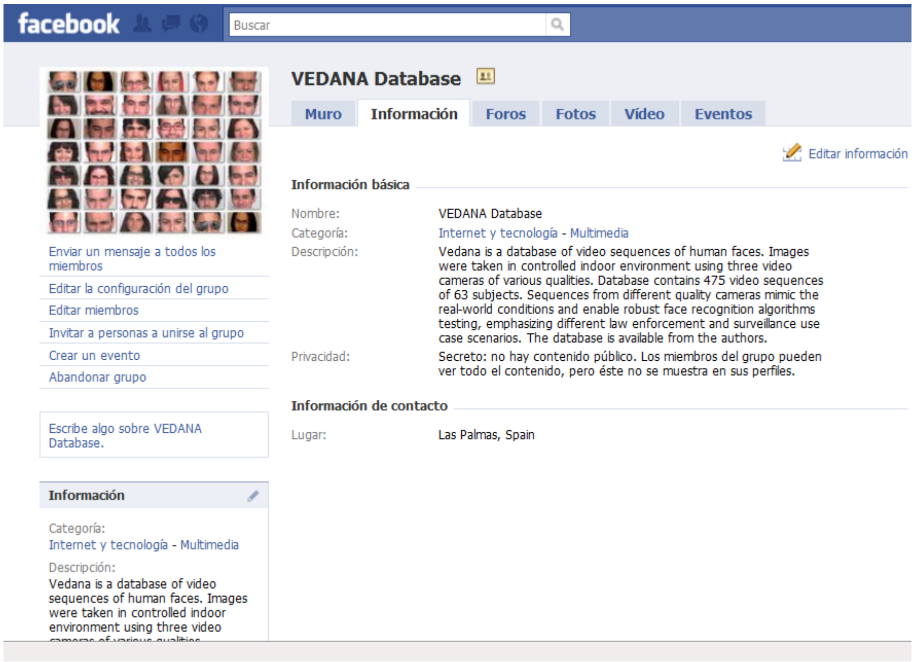
The user may use screenshots of subjects in publications only if that particular subject has explicitly granted the permission to use his or her recordings in publications.

All documents and papers that report on research that uses the SIANI Facial Expression Database will acknowledge this as follows: "(Portions of) the research in this paper use the SIANI-Facial Expression Database collected by M. Castrillon and his group (www.siani.com)", and a citation to:

D. Freire and M. Castrillón , "xxxx", Proc. IEEE Int'l Conf. on xxxx (yyyy'00),
qqqq, wwww, July 2000.

The user will send an e-copy of all papers that reference the database to: publication@siani.es

Figure 3.16: An extract of the End User License Agreement (EULA) for using the VEDANA Database.


facebook
Buscar
Muro
Información
Foros
Fotos
Vídeo
Eventos
Editar información
Enviar un mensaje a todos los miembros
Editar la configuración del grupo
Editar miembros
Invitar a personas a unirse al grupo
Crear un evento
Abandonar grupo
Escribe algo sobre VEDANA Database.

VEDANA Database

Información básica

Nombre:	VEDANA Database
Categoría:	Internet y tecnología - Multimedia
Descripción:	Vedana is a database of video sequences of human faces. Images were taken in controlled indoor environment using three video cameras of various qualities. Database contains 475 video sequences of 63 subjects. Sequences from different quality cameras mimic the real-world conditions and enable robust face recognition algorithms testing, emphasizing different law enforcement and surveillance use case scenarios. The database is available from the authors.
Privacidad:	Secreto: no hay contenido público. Los miembros del grupo pueden ver todo el contenido, pero éste no se muestra en sus perfiles.

Información de contacto

Lugar:	Las Palmas, Spain

Información

Categoría:
Internet y tecnología - Multimedia

Descripción:
Vedana is a database of video sequences of human faces. Images were taken in controlled indoor environment using three video cameras of various qualities

Figure 3.17: In order to provide and disseminate accurate, Vedana database must be inside new technologies.

Chapter 4

Conclusions

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.

Humans communicate effectively and are responsive to each other's emotional states. Computers must also gain this ability. This is precisely what the Human-Computer Interaction research community is focusing on: namely, Affective Computing. Expression recognition plays a significant role in recognizing one's affect and in turn helps in building meaningful and responsive HCI interfaces. The interested reader can refer to Zeng et al.'s comprehensive survey [91] to get a complete picture on the recent advances in Affect-Recognition and its applications to HCI. Apart from the two main applications, namely robotics and affect sensitive HCI, expression recognition systems find uses in a host of other domains like Telecommunications, Behavioral Science, Video Games, Animations, Psychiatry, Automobile Safety, Affect sensitive music juke boxes and televisions, Educational Software, etc. Practical real-time applications have also been demonstrated. Bartlett et al. have successfully used their face expression recognition system to develop an animated character that mirrors the expressions of the user (called the *CU Animate*) [35]. They have also been successful in deploying the recognition system on Sony's *Aibo* Robot and ATR's *RoboVie* [35]. Another interesting application has been demonstrated by Anderson and McOwen, called the *'EmotiChat'* [4]. It consists of a chat-room application where users can log in and start chatting. The face expression recognition system is connected to this chat application and it automatically inserts emoticons based on the user's facial expressions.

The availability of such an extensive database of color, high resolution face and body images can contribute to the development of new algorithms related to the face and body images processing. The limitations and issues discussed in Section 4.1 and above respectively, give rise to a number of new tasks for our future work so that our facial expression research could be moved towards a more realistic scenario.

4.1 Future work

There are still some limitations on the database in terms of data variety, data quality, data quantity, data usability, and its applicability for other applications. We will address the following aspects for the future development.

1. Variety: Limited by the storage capacity and the processing capability, the current system can only record three videos for each session. Each video recorded by a different camera with different settings, which makes it hard to develop a synchronization procedure for the captured data. We would like to expand the hardware setup using multiple cameras and larger storage space in an attempt to capture a greater variety of non-deliberate facial expressions for a longer period of time. The idea is to have three rows with, at least three identical cameras per row, each row with the same camera settings. As an option, for example, each subject can choose to perform mixed expressions freely so that various types of expressions can be included in a longer video sequence.

2. Quality: Limited by the range of capture of the camera C1 (Casio), which captures 210 frames per second, the quality of the video generated by such camera is very sensible to lighting conditions. Thus, a little bit

of noise can be visible on the camera C1's video sequences. To allow improvement for the quality of camera C1's video clips, we plan to record data systematically under multiple spotlights. Also, to improve the data quality, we will seek to expand the current system with more distributed cameras for a wider range of views.

3. Quantity and Applicability: Although the current database was designed for facial expression recognition task, it is applicable to testing face recognition algorithms. However, the size of the database (63 subjects) is still small with respect to the requirement from the face recognition task. We plan to expand the database to a larger scale in order to meet the requirement of real world applications.

4. Usability: One of the big challenges facing the Vedana database is the storage of a huge amount of data. The current database has the size of about 270 gigabytes. Bearing in mind that we will keep using high speed cameras, with the future expansion, the data size will be on the order of terabytes. The larger amount of the data increases the difficulty of data organization, annotation, distribution, processing, and evaluation. We will address this issue as a future work in order to make it easier to manage and search data. An automatic registration and annotation approach will be developed so that the temporal segments of facial expression sequences could be parsed and archived. Another issue to be investigated is the data representation or compression.

Bibliography

[1] F. Pianesi A. Battocchi and D. Goren-Bar. A first evaluation study of a database of kinetic facial expressions (dafex). *Proceedings of Int. Conf. on Multimodal interfaces.*, pages 214–221, 2004.

[2] S. Snow D. R. Hurst M. Pappas A. O'Toole, J. Harms and H. Abdi. A video database of moving faces and people. *Vision Res.*, 39(1):3145–3155, 2003.

[3] T. Vetter A.J. O'Toole and V. Blanz. Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: An application of three-dimensional morphing. *Vision Res.*, 39(1):3145–3155, 1999.

[4] K. Anderson and P.W. McOwan. A real-time automated system for recognition of human facial expressions. *IEEE Trans. Systems, Man, and Cybernetics Part B*, 36(1):96–105, 2006.

[5] P.N. Belhumeur A.S. Georghiades and D.J. Kriegman. From few to many: Generative models for recognition under variable pose and illumination. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.

[6] P.N. Belhumeur A.S. Georghiades and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 643–660, 2001.

[7] I. M. Thornton B. Knappmeyer and H. H. Bulthoff. The use of facial motion and facial form during the processing of identity. *Vision Res.*, 43(1):1921–1936, 2003.

[8] M. M. Bradley and P. J. Lang. *Motivation and emotion.* Cambridge University Press, 2006.

[9] M.C. Roh B.W. Hwang, H. Byun and S.W. Lee. Performance evaluation of face recognition algorithms on the asian face database, kfdb. *AVBPA 2003, LNCS 2688*, (2):557–565, 2003.

[10] C. Darwin. *The expression of emotion in man and animals.* 1998.

[11] Louis J. Denes, Peter Metes, and Yanxi Liu. Hyperspectral face database. Technical Report CMU-RI-TR-02-25, Robotics Institute, Pittsburgh, PA, October 2002.

[12] R. Cowie E. Douglas-Cowie and M. Schrder. A new emotion database: considerations, sources and scope. *Proceedings of ISCA Workshop on Speech and Emotion*, 2000.

[13] P. Ekman. Facial expression and emotion. *American Psychologist*, 48(1):384–392, 1993.

[14] P. Ekman. *Handbook of Cognition and Emotion*. T. Dalgleish and M. Power, 1999.

[15] P. Ekman. *Emotions revealed (2nd ed.)*. New York: Times Books, 2003.

[16] P. Ekman and W. Friesen. *Manual for the Facial Action Coding System*. Consulting Psychologists Press, 1977.

[17] P. Ekman and W.V. Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1):49–98, 1969.

[18] M. Ernst and H. Buelthoff. Merging the senses into a robust percept. *Cognitive Science*, (8):127–159, 2004.

[19] A. Savran et al. Bosphorus database for 3d face analysis. *Proceedings of First European Workshop on Biometrics and Identity Management Workshop*, pages 200–205, 2008.

[20] P. Hallinan. *A Deformable Model for Face Recognition Under Arbitrary Lighting Conditions*. PhD thesis, Harvard University, 1995.

[21] Illumination Internet source: Pose and Expression (PIE) Database. http://www.ri.cmu.edu/research_project_detail.html?project_id=418, 2000. [Online; accessed 07-April-2010].

[22] J. Kittler J. Luettin K. Messer, J. Matas and G. Maitre. Xm2vtsdb: the extended m2vts database. In *Proceedings of the 2nd International Conference on Audio and Video-based Biometric Person Authentication*, 1999.

[23] Y. Kamitani L. Shams and S. Shimojo. What you see is what you hear. *Nature*, 408(1):788, 2000.

[24] J.I. Lacey. *Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome*. Washington, DC: National Publishing., 1958.

[25] K. Lander and L. Chuang. Why are moving faces easier to recognize? *Vision Cognition*, 12(1):429–442, 2005.

[26] R. Lazarus. *Emotion and adaptation*. Oxford University Press, 1991.

[27] L. Liu and T. M. Ozsu. *Encyclopedia of Database Systems*. Springer, 2009.

[28] J. Budynek M. J. Lyons and S. Akamatsu. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12):1357–1362, 1999.

[29] M. Kamachi M. Lyons, S. Akamastu and J. Gyoba. Coding facial expressions with gabor wavelets. *Proceedings of Third IEEE Conf. on Face and Gesture Recognition*, pages 200–205, 1998.

[30] R. Rademaker M. Pantic, M. Valstar and L. Maat. Web-based database for facial expression analysis. In *Proceedings of IEEE International Conference Multmedia and Expo (ICME05)*, 2005.

[31] A.M. Martinez and R. Benavente. The ar face database. *CVC Technical Report*, 1998.

[32] D. Matsumoto and B. Willimgham. Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of Personality and Social Psychology*, 96(1):1–10, 2009.

[33] D. McNeill. *Hand and Mind: What Gestures Reveal about Thought.* University of Chicago Press, 1992.

[34] D. J. Kriegman Ming-Hsuan Yang and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.

[35] I. Fasel M.S. Bartlett, G. Littlewort and R. Movellan. Real time face detection and facial expression recognition: Development and application to human computer interaction. In *CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction*, 2003.

[36] A. V. Oppenheim and J. S. Lim. The importance of phase in signals. In *Proceedings of the IEEE*, 1981.

[37] A. Ortony and T.J. Turner. What's basic about basic emotions? *Psychological Review*, 97(3):315–331, 1990.

[38] J. Hespanha P. Belhumeur and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[39] Y. Ostrovsky P. Sinha, B. Balas and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(1):1948–1962, 2006.

[40] 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*, 36(1):433449, 2006.

[41] S. Pigeon and L. Vandendrope. The m2vts multimodal face database. In *Proceedings of the 1st International Conference Audio and Video Based Biometric Person Authentication*, 1997.

[42] P. Rauss P.J. Philips, H. Moon and S.A. Rizvi. The feret evaluation methodology for face-recognition algorithms. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.

[43] MOBIO Project. http://www.mobioproject.org/databases, 2008. [Online; accessed 31-March-2010].

[44] B. Rim and L. Schiaratura. *Fundamentals of Nonverbal Behaviour, chap. Gesture and Speech.* In R.S. Feldman and B. Rime (eds). Cambridge University Press, 1991.

[45] J. Sadr and P. Sinha. Exploring object perception with random image structure evolution. massachusetts institute of technology, artificial intelligence laboratory memo, 2001.

[46] J. Sadr and P. Sinha. Random image structure evolution (rise). *Vision Sciences Society Abstracts*, 83(1):128–142, 2001.

[47] J. Sadr and P. Sinha. Object recognition and random image structure evolution. *Cognitive Science*, 28(1):259–287, 2004.

[48] F.S. Samaria. *Face Recognition using Hidden Markov Models.* PhD thesis. Cambridge University Press, 1994.

[49] K. Schmidt and J. Cohn. Dynamics of facial expression: Normative characteristics and individual differences. In *Proceedings of IEEE International Conference Multmedia and Expo (ICME)*, 2001.

[50] H. Schneiderman and T. Kanade. A statistical method for 3d object detection applied to faces and cars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 746–751, November 2000.

[51] Internet source: Cohn-Kanade AU-Coded Facial Expression Database. http://vasc.ri.cmu.edu/idb/html/face/facial_expression/index.html, 2000. [Online; accessed 11-April-2010].

[52] Internet source: Database of Human Facial Expressions. http://tcc.itc.it/research/i3p/dafex/index.html, 2004. [Online; accessed 12-April-2010].

[53] Internet source: Diario AS. http://www.as.com/, 2010. [Online; accessed 14-May-2010].

[54] Internet source: El rellano. http://www.elrellano.com/, 2010. [Online; accessed 24-April-2010].

[55] Internet source: Emotion Research. http://emotionresearch.net/wiki/databases, 2010. [Online; accessed 09-April-2010].

[56] Internet source: Empflix. http://www.empflix.com/, 2010. [Online; accessed 24-April-2010].

[57] Internet source: Equinox Infrared Face Database. http://www.equinoxsensors.com/products/hid.html, 2004. [Online; accessed 07-April-2010].

[58] Internet source: Face and Emotion. http://www.face-and-emotion.com, 2002. [Online; accessed 15-May-2010].

[59] Internet source: Face Detection Databases. http://www.ri.cmu.edu/projects/project 419.html, 2002. [Online; accessed 10-April-2010].

[60] Internet source: Face Recognition Homepage. http://www.facerec.org/databases/, 2010. [Online; accessed 09-April-2010].

[61] Internet source: Harvard Robotics Lab (HRL) Database. ftp://cvc.yale.edu/cvc/pub/images/hrlfaces, 2000. [Online; accessed 07-April-2010].

[62] Internet source: Instituto universitario de Sistemas Inteligentes y Aplicaciones Numericas aplicadas a la Ingenieria (SIANI). http://www.siani.es, 2010. [Online; accessed 15-April-2010].

[63] Internet source: Interactive Human Face Modelling. http://www.cg.tuwien.ac.at/hostings/cescg/cescg-2001/gszijarto/index.html, 2004. [Online; accessed 15-May-2010].

[64] Internet source: Max Planck Institute for Biological Cybernetics Face Database. http://faces.kyb.tuebingen.mpg.de/, 1999. [Online; accessed 10-April-2010].

[65] Internet source: Microsoft MSDN. http://msdn.microsoft.com, 2010. [Online; accessed 24-April-2010].

[66] Internet source: MMI Facial expression database. http://www.mmifacedb.com/, 2004. [Online; accessed 10-April-2010].

[67] Internet source: NIST Mugshot Identification Database. http://www.nist.gov/srd/nistsd18.htm, 2000. [Online; accessed 08-April-2010].

[68] Internet source: Notre Dame HumanID Database. http://www.nd.edu/˜cvrl/cvrl/data_sets.html, 2002. [Online; accessed 07-April-2010].

[69] Internet source: Olivetti Research Lab (ORL) Database. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html, 1994. [Online; accessed 08-April-2010].

[70] Internet source: Open Computer Vision Library. http://sourceforge.net/projects/opencvlibrary/, 2010. [Online; accessed 24-April-2010].

[71] Internet source: Psychological Image Collection at Stirling (PICS). University of Stirling Psychology Department. http://pics.psych.stir.ac.uk/, 2010. [Online; accessed 07-April-2010].

[72] Internet source: The AR Face Database. http://cobweb.ecn.purdue.edu/˜aleix/aleix_face_db.html, 1998. [Online; accessed 09-April-2010].

[73] Internet source: The BANCA Database. http://www.ee.surrey.ac.uk/cvssp/banca/, 2004. [Online; accessed 07-April-2010].

[74] Internet source: The CAS-PEAL face database. http://www.jdl.ac.cn/peal/home.htm, 2004. [Online; accessed 07-April-2010].

[75] Internet source: The FERET Database. http://www.itl.nist.gov/iad/humanid/feret/, 2003. [Online; accessed 08-April-2010].

[76] Internet source: The Japanese Female Facial Expression (JAFFE) Database. http://kasrl.org/jaffe.html, 1999. [Online; accessed 11-April-2010].

[77] Internet source: The Sheffield Face Database. http://www.shef.ac.uk/eee/research/vie/research/face.html, 1998. [Online; accessed 07-April-2010].

[78] Internet source: The University of Oulu Physics-Based Face Database. http://www.ee.oulu.fi/research/imag/color/pbfd.html, 1999. [Online; accessed 07-April-2010].

[79] Internet source: The Vedana Database. http://www.siani.es/vedana, 2010. [Offline; under construction].

[80] Internet source: Universidad de Las Palmas de Gran Canaria (ULPGC). http://www.ulpgc.es, 2010. [Online; accessed 15-April-2010].

[81] Internet source: University of Texas Video Database. http://bbs.utdallas.edu/facelab/perceptionlab.htm, 2005. [Online; accessed 06-April-2010].

[82] Internet source: XM2VTSDB multi-modal face Database. http://www.ee.surrey.ac.uk/cvssp/xm2vtsdb/, 2003. [Online; accessed 09-April-2010].

[83] Internet source: Yale Face Database. ftp://plucky.cs.yale.edu/cvc/pub/images/yalefaces/, 1997. [Online; accessed 06-April-2010].

[84] Internet source: Yale Face Database B. ftp://plucky.cs.yale.edu/cvc/pub/images/yalefacesb/tarsets/, 2000. [Online; accessed 06-April-2010].

[85] Internet source: Youtube. http://www.youtube.com, 2010. [Online; accessed 24-April-2010].

[86] K.K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):39–51, 1999.

[87] J. Cohn T. Kanade and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the 4th IEEE Int Conf on Automatic Face and Gesture Recognition*, 2000.

[88] M. Turk. *Gesture Recognition.* Lawrence Erlbaum Associates, 2001.

[89] L. Banka W. Schiff and G. de Bordes. Recognizing people seen in events via dynamic mug shots. *Amer. J. Pychol*, 99(1):219–231, 1986.

[90] T. Kanade Y. Tian and J. Cohn. Facial expression analysis. *Handbook of Face Recognition*, (2):247–275, 2005.

[91] G.I. Roisman Z. Zeng, M. Pantic and T.S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009.