|  |
| --- |
| N. Poh, R. Blanco-Gonzalo, R. Wong and R. Sanchez-Reillo, "Blind subjects faces database," in IET Biometrics, vol. 5, no. 1, pp. 20-27, 3 2016. doi: 10.1049/iet-bmt.2015.0016 |

The Blind Subject Face Database (BSFDB)

Abstract

Using your face to unlock a mobile device is not only an appealing security solution, but also a desirable or entertaining feature, such as taking selfies. It is convenient, fast, and does not require much effort, but only if you have no vision problems. For users with visual impairments, taking selfies could potentially be a challenging task. In order to study the usability and ensure the inclusion of mobile-based identity authentication technology, we have collected the Blind-Subjects Faces Data Base (BSFDB). Ensuring that technology is accessible to disabled people is important because they account for about 15% of the world population. The BSFDB database contains several individuals with visual disabilities who took selfies with a mock-up mobile device. The experimental settings vary in the image acquisition process or experimental protocol. Four experimental protocols are defined by a dichotomy of two controlled covariates, namely, whether or not a subject is guided by audio feedback and whether or not he/she has received explicit instructions to take the selfie. Our findings suggest that the importance of appropriate design of human computer interaction as well as alternative feedback design. The BSFDB database can be used to investigate topics such as usability, accessibility of the face recognition technology, or its algorithmic performance. All the gathered data is publicly available online including videos of the experiments with more than 70,000 face images of blind and partially blind subjects.

# Introduction

Biometric recognition refers to the automatic identification of individuals according to one or more unique physical or behavioural features such as iris, gait or face. Some of the desirable biometric properties for security-related applications are uniqueness, universality, permanence, performance, circumvention, acceptability and measurability [1]. Furthermore, the use of other traditional authentication schemes based on passwords is considered by users as cumbersome due to the necessity to remember a large variety of alphanumeric codes, which usually drives people to re-use the same password for several, if not all, authentication services. The use of biometrics allows user authentication through “something she/he is” or “something she/he does” – as in behavioural biometrics – in order to avoid the use of “something she/he knows”. Therefore, the biometrics technology offers a unique proposition for protecting sensitive data with applications from financial transactions to physical and remote access controls.

Biometric recognition is moving to mobile environments [2] and the range of possibilities for integrating biometrics is promising, with potential applications such as signing documents unequivocally, accessing to websites securely, executing administrative procedures and other electronic transactions. Indeed, according to a market report, end-users do prefer to have biometrics embedded in mobile devices, e.g. fingerprint or face recognition to unlock the smartphone. Indeed, for this reason, Apple has introduced fingerprint recognition in iPhone6.

Databases are crucial to test and improve the algorithms used in biometric systems. This is because biometrics is not an exact science; a biometric system is often affected by a number of environmental factors such as the lighting conditions. In addition, there are a number usability factors that may affect the system performance. Examples are how a mobile device is handled by the user; whether or not the user has physical or intellectual disabilities; and whether or not the user operates the system under normal or stressed conditions. As a result, it is important to develop databases to measure the technological progress, considering factors such as computers power, sensors design, and application scenarios.

Designing and collecting a biometrics database is a challenging and resource-consuming task. Often, it involves not only the database development team but also many volunteers.

We list below a number of aspects that one should consider when building a biometric database:

i) the recruited users should be representative of all the different main groups regarding age, gender and other human characteristics in order to obtain reliable and broad conclusions;

ii) people usually distrust a new piece of technology. When interacting with the technology, some users may consider the device to be intrusive;

iii) in order to recruit enough people for experimentation it is necessary to reward them so that their interest can be sustained throughout the data collection period;

iv) the participants need to be motivated: apathy or lack of interest can bias the experiments (especially in those regarding usability);

v) experts are needed for several tasks including: assessment design, experiments monitoring, data processing and storage, as well as validation; and

vi) biometric data need to be stored and processed according to data protection laws. Therefore legal implications should be considered.

Although many face databases have been developed in the past, none of these databases are aimed at evaluating the suitability of the face recognition technology for the blind or partially blind subject. Some of the most representative database examples are the FRVT (Face Recognition Vendor Test [3]), the FRGC (Face Recognition Grand Challenge [4]) and the FERET (Facial Recognition Technology [5]). Furthermore, the number of multimodal databases is increasing in recent years, e.g. BIOSECURE [6] or the MOBIO project [7]. All of these projects involve the work of several institutions and the collaboration of hundreds of participants and therefore they are considered references for the biometrics community. However, one of the main drawbacks of all of the above projects is the lack of individuals with visual-related disabilities.

This paper presents the Blind Subjects Faces DataBase (BSFDB), which is a novel database made up by face images of individuals covering all possible range of visual impairments During the process each subject was told to take self-videos of his/her face under different feedback modes. Each image is time-stamped so that subsequent studies can assess the efficiency of an interaction session. The camera used was a mock-up device which is very light and easy to handle and is connected to a PC. Four experiments per session were planned providing four different modes of feedback; thus allowing post-experimental analyses of usability, accessibility, and performance. The dataset collected includes videos of each user-session-experiment, personal data (gender, age, degree of blindness and opinions about the experiment) and more than 70 thousand face images. Each image was stored with the corresponding bounding boxes around a detected face (there may be several boxes per detected face), a time-stamp of the image, and the face detection confidence (FDC) which is the confidence of a face classifier that the region of interest is a face. Although there is a face detector used during the acquisition and interaction process in real time, a post-experimental evaluation may also include a more accurate face detection algorithm.

Our primary reason is to study if the current state-of-the-art face recognition can be used by blind or partially blind subjects or not; and if they are not adequate, how alternative forms of feedback (other than visual), such as audio and tactile can be used to assist them so that the technology is more usable. Ultimately, our goal is to render the face recognition technology available for everyone to use. For instance, it is

Other potential uses of the BSFDB include:

* Usability and accessibility studies under four controlled feedback modes.
* Usability in terms of efficiency. The inclusion of timestamps allows further studies such as the study of performance or users behaviour evolution along with the time.
* Ergonomics. Is the camera used the most usable and comfortable? Has it any positive or negative effects over the results or user's satisfaction?
* Face detection algorithms testing. This database contains several types of realistic face images captured under very challenging conditions. Many images are not well-aligned, contain partial faces, or faces that are blurred due to swift movements, or are out of focus. All these conditions make face detection and recognition very challenging.
* Face recognition algorithms testing. For instance, according to the images acquired where some of the main landmarks used in face recognition may be missed, the comparison in performance between alignment and alignment-free algorithms is interesting (i.e. they may be difficult to align).

The rest of the paper is organized as follows. Section 2 provides an overview of the existing database. Section 3 describes BSFDB in detail. In Section 3, potential uses of the database are suggested. Section 4 contains preliminary analysis of the database. Legal issues and database distribution details are given in Section 5 (ANEX?). Finally, in Section 6 are the conclusions of the paper.

# Related works

There are multiple examples of face databases in the literature and some of the most well-known are the following (a summary is in Table 1):

**FERET** database. The FERET image corpus was assembled to support government monitored testing and evaluation of face recognition algorithms using standardized tests and procedures. The FERET database was collected between 1993 and 1996 through 15 sessions. The database (collected in a semi-controlled environment) contains 1564 sets of images for a total of 14,126 images that includes 1199 individuals and 365 duplicate sets of images.

**BiosecurID** database. Contains 400 users and 8 biometric characteristics (voice, iris, face still, talking face, signature, handwriting, fingerprint, hand and keystroking) captured in an uncontrolled scenario. It was collected in 4 sessions distributed in a four-month time span.

**Biomet** database [8]. This multimodal database contains voice, face images (2D and 3D), hand images, fingerprints and signatures. The camera used for face recognition suppressed the ambient light influence. There were 3 sessions with 3 and 5 months spacing between them. The number of participants was 130 in the first session, 106 in the second and 91 in the last one.

**MOBIO** database. The MOBIO database was acquired with a mobile device and a laptop, and the modalities used were face and voice. More than 150 people participated in the process distributed in 12 sessions and in 5 countries (including native and non-native English speakers).

**BioSecure Multimodal DataBase (BMDB)**. More than 600 users are included in this database acquired with a mobile device and a laptop. The modalities used are voice, face, signature, hand, iris and fingerprint. There were 3 sessions and 3 different scenarios (over the internet, office environment and indoor/outdoor environments).

**Point and Shoot Face Recognition Challenge (PaSC)** [9]. This database includes 9,376 still images and 2,802 videos of 293 people. The images are balanced with respect to distance to the camera, alternative sensors, frontal versus not-frontal views, and different locations.

**Labelled faces in the wild (LFW)** [10]. The data set contains more than 13,000 images of faces collected from the web. Each face has been labelled with the name of the person pictured. 1680 of the people pictured have two or more distinct photos in the data set.

**Extended Multi Modal Verification for Teleservices and Security applications (XM2VTS)** [11]. This is a large multi-modal database captured onto high quality digital video. Contains 4 recordings of 295 subjects taken over a period of 4 months. Each recording contains a speaking head shot and a rotating head shot.

**BANCA** [12]. It is a multi-modal database intended for training and testing multi-modal verification systems. The BANCA database was captured in four European languages in two modalities (face and voice). For recording, both high and low quality microphones and cameras were used. The subjects were recorded in three different scenarios, controlled, degraded and adverse over 12 different sessions spanning three months.

Table 1. Popular face databases in the literature. \*(session1, session2, session3). '(one image per user, at least two images per user)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Year | #Users | #Sessions | Mobile | 3dfaces | #Traits |
| FERET | 1993-1997 | 1199 | 15 | No | No | 3 |
| BiosecurID | 2007 | 400 | 4 | No | No | 8 |
| Biomet | 2001-2002 | (130,106,91)\* | 3 | No | Yes | 6 |
| MOBIO | 2008-2010 | 152 | 12 | Yes | No | 2 |
| BMDB | 2006-2007 | (971,667,713)\* | 3 | Yes | No | 5 |
| PaSC | 2013 | 293 | 4 | No | No | 1 |
| LFW | 2007 | (5749, 1680)´ | - | No | No | 1 |
| XM2VTS | 2000 | 295 | 4 | No | No | 2 |
| BANCA | 2004 | 208 | 12 | No | No | 2 |

Although there are a large number of public face databases already, very few of them contain images taken with a mobile device. Even fewer are the databases which include disabled users [13]. To our best knowledge, none of the published works in the face recognition literature includes visually impaired users.

# The Blind Subjects Faces DataBase (BSFDB)

Although there are a lot of face databases, none of them are designed specifically to understand the need of face recognition for blind subjects. Technology inclusiveness should also be applied to all technologies including face recognition. According to the World Health Organization (WHO), approximately 15% of the world population has a significant physical or mental disability [14], which is a representative enough percentage of the whole population. In accordance, practically none of the state of the art face databases is representative of the real population in practice because they do not include subjects with disabilities.

The acquisition of the BSFDB was conducted by the B-lab, Department of Computing of the University of Surrey (United Kingdom). The database was collected at the St. Nicholas Centre for visually impaired people in Penang, Malaysia, during 2012. The work presented here shows the acquisition process and performance results of a database composed by self-captured face images by visually impaired individuals without external help.

*3.1 Users*

There were 40 participants in the evaluation (29 men and 11 women) covering representative wide range of age groups, gender, vision level and ability taking self-photos. None of the participants have ever used any biometric device previously. Their age distribution is as follow: 45% are under 25 years old, 42% are between 25 and 50, and 13% are over 50. Regarding the vision level, 16 of them have low vision, 14 can distinguish light and darkness, whereas 10 are completely blind, out of which 5 are born blind. Almost half of the participants (15) claimed the ability of taking self-pictures and 16 claimed the ability of taking other's pictures. The count of subjects’ age, gender, and level of visual impairment is shown in Figure 1.

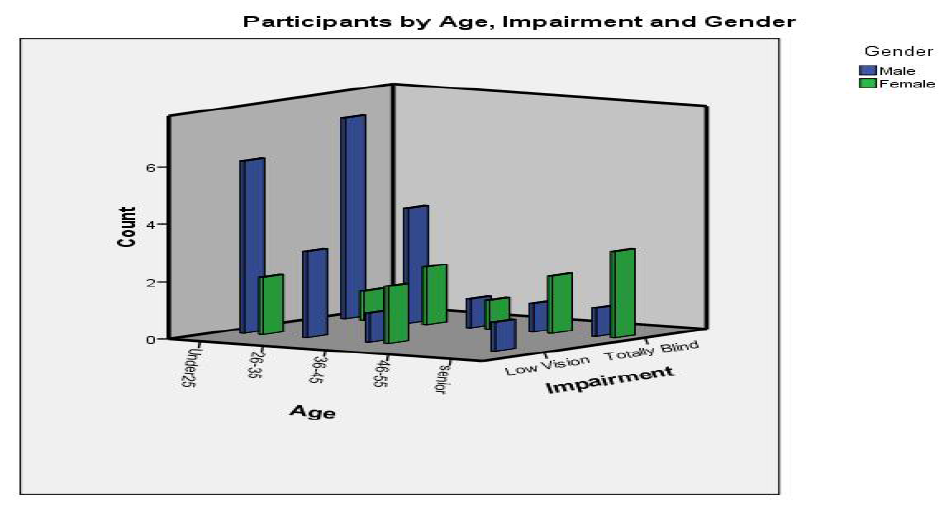


Figure 1. BSFDB participants by age, impairment and gender

*3.2 Usability Experiments*

In this section, we describe the design of four settings of usability experiments in order to understand how different feedback modes can improve the accessibility of the face recognition technology as well as its performance.

There are a number of controllable and uncontrollable factors that affect our usability experiments. Factors that affect the users such as moods and feelings are uncontrollable. However, we can partially control the experimental environment or the potential habituation with the device. In the experiments made during the database acquisition we have taken into account 3 of those factors: habituation, instructions received and audio feedback. The task is the same in each experiment setting: take a selfie emulating the scenario of unlocking a mobile device using face recognition. Examples of images taken in each experiment are shown in Figure 2. We measured the results in terms of accessibility and performance through four feedback modes, as enumerated below:

Experiment 1 (E1). The user receives no feedback or instructions when taking a selfie. This experiment is expected to be the worst in terms of the number of detected faces as well as the facial recognition performance because the user has not yet acquired the skill in taking selfies. Moreover, this experiment is the first one to be completed in the evaluation so that it does not bias the other experiments; since, by definition, the user has not accustomed to using the device.

Experiment 2 (E2). The user receives audio feedback just before taking his/her selfie. The audio feedback is set at 3 different frequency levels which depend on the face detection confidence (FDC) of the acquired image. The FDC is given by a Viola-Jones based face detector. The provided frequency is low (1.5 KHz) if the face detector does not detect any face, medium (4.5 KHz) if it detects a non-frontal face and high (7.5 KHz) if a frontal face is detected. The definition of non-frontal versus frontal face image is distinguished through a systematic experiment carried out offline. The audio feedback is intended to help the user to better point the camera to capture a face as frontal as possible. This experimental setting is to be completed right after E1 so that the user is expected to acquire the skill of holding the camera by appropriately adjusting its (position and distance from his/her face during this experiment. A detailed description of the above audio feedback mechanism can be found in [15] and [16].

Experiment 3 (E3). In this experimental setting, audio feedback is not provided but instead, the user receives information about how to take the selfie before starting the experiment. This information consists of a supervisor who helps the user to adjust the distance between the camera and the face so that a proper selfie face image is taken. Although the intention is to isolate E2 and E3 (completed right after E2), this is not completely possible because the user has already acquired the skill during E2. Therefore, he or she would have known approximately how to grab the camera to obtain the best self-image.

Experiment 4 (E4). This experiment is a combination of E2 and E3. In this experiment, the user receives previous instructions on how to grab the camera and also audio feedback during the capture process. Therefore, the face detection and recognition performance results are expected to be the best in this experiment because, apart from the audio feedback and the instructions, the user would have also acquired the skill and habituation needed in using the camera from E1, E2 and E3 settings.



Figure 2. Images of the same user in the E1, E2, E3 and E4 from left to right. Images are expected to improve in the last experiments (i.e. in distance and focus).

An ideal way to carry out the experiments is by using RCT (Randomized Controlled Trial), whereby a user is subject to one of the four modes of feedback. However, we were concerned with two issues with this approach:

The first is the fact that biometric performance is subject-dependent. Therefore, if the same experiment (with the same feedback mode) is conducted but on two different disjoint populations, two different results will be obtained. This is because the effect of the subject variability on the performance of face detection and recognition might be higher than the effect due to the mode of feedback. The impact of inter-subject variability on biometric performance is well known in the literature, and this phenomenon is referred to as Doddington’s zoo or biometric menagerie [17] [18] [19].

The second concern is that the number of blind subjects is limited, which is about 40. For this reason, deploying RCT would mean that one has to divide the population into a smaller set with 10 subjects for each mode of feedback. This is arguably not the best use of limited samples available to us.

We have therefore, opted for subjecting every volunteer to all the four modes of feedback, but doing so carefully so that the effect of one feedback does not influence that of another. One way to achieve this is by exploiting the natural ordering of the modes of feedback. For example, the E1 setting does not have any feedback and so should be carried out first. E2 and E3 are each independent of each other because the audio feedback does not convey any information about the instruction. However, the instruction (E3) mode should take place just after E2. A potential weakness of the above approach is that the volunteer may have become more familiar with the device after each experiment which is conducted sequentially. Fortunately, after a post experimental analysis, we found that this is not a concern. Finally, E4 should be conducted last because the user has to have the knowledge of interpreting the audio feedback and should have been given the instruction of taking a high quality selfie image.

*Hardware and Software in the BSFDB*

The camera used to capture the face images is an Advent Slim 300K web-cam, which is light and small. The device can be handled easily in a participant’s hand, as shown in Figure 3. It is connected to a laptop where a desktop application controls the process flow. The application captures a video stream from the camera and stores images with an image resolution of 640-by-480 pixels at 30 frames per second. The user has a timeout of 45s in order to obtain a frontal face image. Then, the application processes each image and returns the audio feedback as described in E2.



Figure 3. Example of the BSFDB acquisition, including the camera and the capture software

*BSFDB contents*

Each user underwent two sessions of experiment, each of which are separated by about two weeks apart. In each session, the user was then subject to the four experimental settings as described above. As a result, for each user, we have 8 unique experiment-session combinations; and hence 8 videos. Each video contains between 0 and 400 images. . The number of images varies from one video to another due to the time variability of the face detection. When an image is taken, it is saved with the corresponding bounding boxes around a detected face (noting that there may be several boxes per detected face), a time-stamp of the image, and the FDC associated with *each* bounding box. In addition to the raw images, BSFDB also contains pre-processed and extracted images. The pre-processing in to be explained in Section 4.1. Table 2 summarises the BSFDB content.

Table 2. Some statistics of BSFDB. Exp = Experiment, bb = bounding boxes, fdc = face detection confidence. Valid images are those which contain a face according to the face detector

|  |  |  |  |
| --- | --- | --- | --- |
| *Session* | *Exp* | *Snapshot, timestamp, bb, fdc* | *# of valid images and features (SIFT/ PCA)* |
| 1 | 1 | 17200 | 4303 |
| 2 | 17600 | 7386 |
| 3 | 17600 | 9069 |
| 4 | 17600 | 10816 |
| 2 | 1 | 16400 | 8478 |
| 2 | 16400 | 9150 |
| 3 | 16400 | 9911 |
| 4 | 16400 | 11065 |
| Total images | | 135600 | 70178 |

*Main concerns during the BSFDB acquisition and processing*

Having carefully analysed the database, we found some problems during the acquisition process as well as during the images processing stage. Both problems are mostly due to the difficulty of the blind or partially blind subjects in taking selfies. Several participants found it difficult to focus their face especially in the E1 setting where they had not audio feedback, for instance. The main problems are further elaborated below.

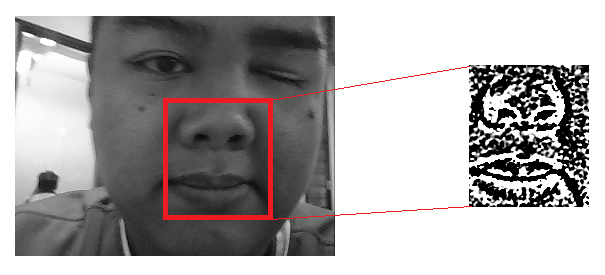


Figure 4. Example of false alarm due to the face detector. This is classified as an error in image processing. In this example, since the user is holding the camera too close to his face, the face detector has wrongly classified the nose and mouth as a face. The face detector is also operating at lower-than-usual face detection threshold in order to be able to tolerate for somewhat less frontal face images. As a result, false alarm is inevitable.

*During acquisition:* the user moved his/her head and the mobile device constantly when taking the selfies because of the lack of any reference to correctly focus his/her face. As a result partial face or faceless images were obtained. This increases the time spent taking the selfie (in some cases users exceeded the fixed timeout of 45s) and the presence of useless images in the database. Figure 5, illustrates six examples of error during acquisition:

a) the image is dark, the face is too close to the camera, decentred and not complete.

b) the user is too far from the camera and the face is not centred.

c) the user placed the hand between the camera and her face; there is too much light and the image is blurred.

d) the face image is not frontal.

e) the user is too close to the camera and there is too much light.

f) the user wears glasses and the image is blurred.

*During image processing*: due to the errors occurred during the images acquisition there were problems at the time to process the images: there are images without faces, with partial faces, rotated and blurred. Therefore the face detector does not work properly and has problems finding the faces or detecting wrong faces (e.g. confusing the nostrils with the eyes as in Figure 4).

# Preliminary analysis of the database

In this section we include a preliminary analysis of the BSFDB. First, we measure the performance obtained when applying 2 different types of face recognition algorithms: alignment and alignment-free algorithms. An alignment-free algorithm is used for face matching because it does not require precise matching of the face. Unlike alignment-based approach, an alignment-free algorithm can match two images containing only a partial face in each image. In the second analysis, we analyse the evolution of face detection confidence (FDC) over time. For both analyses, the BSFDB images are pre-processed; and this process is described next.



Figure 5. Examples of different problems found when acquiring face images

## *Pre-processing*

The pre-processing is made to reduce the adverse impact of noise; to correct for the varying lighting conditions and to ensure that each detected face image has the same size for the subsequent analyses. During the acquisition process, each image is subject to face detection and a bounding box is associated with each detected face. In the pre-processing stage, each image is cropped using those bounding boxes [15]. Afterwards, photometric normalization is applied to the cropped face image [20] in order to correct for the lighting variation, which is an important adverse factor when applying facial recognition. Finally all the images are resized to a common size of 120-by-142 pixels..



Figure 6. Image before and after the Pre-processing. The red square is the bounding box

*Experiment 1: Alignment versus Alignment-Free*

In this experiment we want to compare two kinds of algorithms, namely alignment and alignment-free face recognition. Since many of the BSFDB images are very noisy (due to occlusion, blurriness, partial faces, etc.), they often do not contain key facial landmarks that are commonly used in the traditional “alignment-based” algorithms. We have opted to use alignment-free algorithms because they can match face images without any prior alignment. We have selected the 300 best images (by FDC) for each user from all the experiments in session 1 and then matched all of between in session 2. The output of this process is a set of genuine scores from which False Reject Rate (FRR) can be estimated. In order to generate an impostor score set for a given target user, the images of the remaining users are used. Therefore, given 40 subjects, we have 40 sets of genuine scores, and 40x39 sets of impostor scores. The union of the 40 sets of genuine scores are used to calculate FRR whereas the union of the 40x39 sets of impostor scores are used to calculate False Acceptance Rate (FAR).

In addition, we also summarise the performance in terms of Equal Error Rate (EER). This performance reflects an operating decision threshold when FAR and FRR are equal. EER is a very useful performance metric to summarise a Receiver Operating Characteristic (ROC) curve because it enforces the prior class probabilities of false acceptance and false rejection to be equal.

The algorithms applied are the following:

Principal Component Analysis (PCA): The PCA is a method to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set [21]. For this experiment we used the PhD tool [22] [23] which implements the Eigenfaces approach to obtain the PCA vectors. We have used the ORL (AT&T) [24] as training set: 5 users, 10 images each one). Then, we built the PCA subspace through a linear subspace projection. Finally, the matching is performed by the dot product of two image samples represented in the PCA subspace

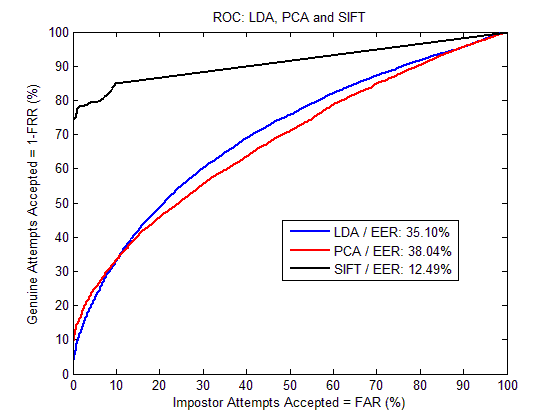


Figure 6. LDA, PCA and SIFT ROC curves with the obtained EER

Linear Discriminant Analysis (LDA): The LDA is used to find a linear combination of features which characterizes or separates two or more classes of objects or events [25]. For applying this method we have used also the PhD tool which implements the Fisherface approach. This means that a PCA step is performed prior to LDA to avoid singularity issues due to a small number of training samples (again the ORL (AT&T) was used as training set: 5 users, 10 images each one). Then, we built the LDA subspace through a linear subspace projection. Finally, the matching is also performed by computing a dot product between two samples in the LDA subspace

Scale-Invariant Feature Transform (SIFT) [26]: It is used to detect and describe local features in images and usually applied for object recognition purposes. SIFT does not use the face landmarks, so it is not based on the traditional alignment approach. In this work, we have used SIFT because it is resistant to occlusion, scale and orientation changes (convenient according to the BSFDB images). SIFT returns several descriptors (or keypoints) of an image. Then, the Euclidean distance between those descriptors and the descriptors of the image to compare with is obtained. If the Euclidean distance between 2 points is below a prefixed threshold it is considered a match. The result for each comparison was given as the division of the number of matches by the number of descriptors.

*Experiment 1 Results*

The performance results showed in Figure 6 in terms of ROC curves reveal that the BSFDB may be challenging for face recognition purposes due to the variations between the images in terms of lightning, blurriness and percentage of detected faces. Results are better when applying SIFT (EER: 12.49%) than in the traditional approaches. Furthermore, the results obtained with PCA (EER: 38.04%) and LDA (EER: 35.10%) show that the alignment methods are not reliable using this database due to the difficulty of find the main landmarks.

*Experiment 2: The evolution in time of the FDC*

In the second experiment we have measured the evolution of the FDC in time through the 4 experiments. It is expected that this value increases along with the time and the experiments (e.g. the FDC in E4 should be better than in E1), due to users are supposed to acquire skills taking selfies during the evaluation. The image acquisition time has been divided into 3 identical time slots (in order to better analyse the evolution in time) and the FDC was calculated in average in the 3 time slots.

*Experiment 2 results*

The results are shown in Figure 7 where there are 2 boxplot's groups: the first one (Figure 7 on the left) represents the first time slot (Time 1) and the second (Figure 7 on the right) represents the third time slot (Time 3). The differences in FDC are not highly significant between Time 1 and Time 3 in the Experiments 1 to 3 but there is an improvement in the Experiment 4 showing that the quality of the images improves when users receive instructions and audio feedback. This experiment reveals the differences in quality among images when users receive different feedback modes. These differences have also repercussion in the final performance as shown in [15]

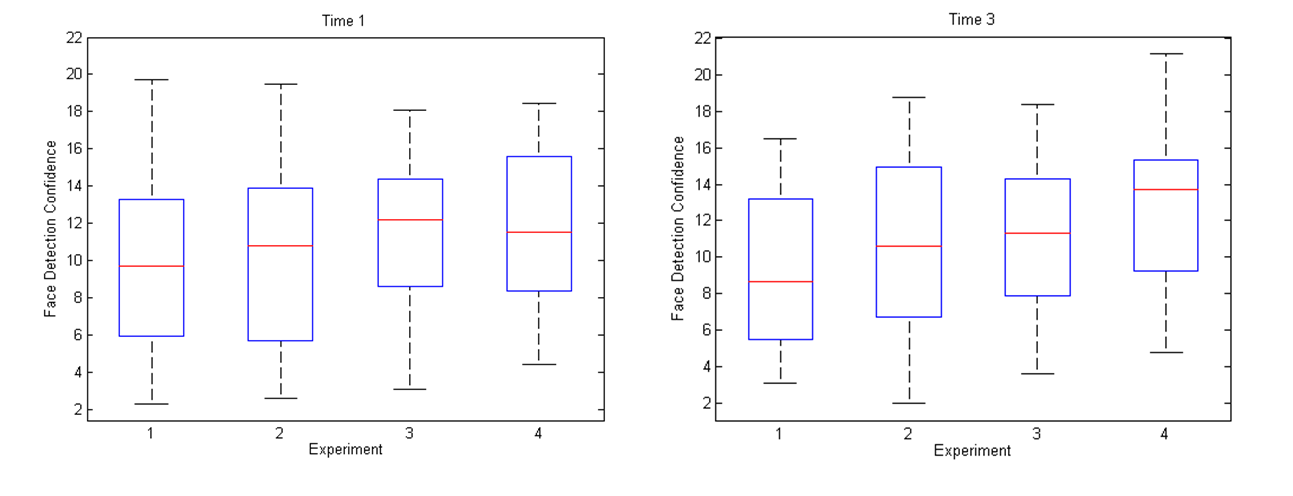


Figure 7. FDC boxplots by Experiments in Time 1 and Time 3

# Conclusions

The BSFDB is the first biometric database which aims to address the accessibility of the face recognition technology for the visually impaired. The database contains four experimental conditions following two dichotomies of factors: with and without audio feedback; and with and without instructions in taking selfie images. Moreover the database contains additional data sets such as the timestamps and face detection information (such as the bounding boxes and face detection confidence values). All these additional meta-data are useful to understand issues such as usability, accessibility in face recognition and/or face detection.

Our preliminary experiments on BSFDB can be summarised as follow. First, visually impaired users find that it is difficult to take selfies. The images so-obtained are often blurred, off centre, and partially visible/occluded. Because of the ill positioning of the face images, the alignment-free approach based on the SIFT features provide significantly better results than classical alignment-based approach. For instance, the SIFT approach attains an EER of 12.49% of EER whereas PCA attains 38.04% and LDA attains 35.10%.

Further experiments should be made using other databases in order to contrast these hypotheses. The analysis of based on the face detection confidence (FDC) over time shows an increase in the FDC when the user is more habituated to the system and receives feedback. Particularly, the detected face images are likely to be better when the user receives both instructions and audio feedback (the E4 setting). In either setting (E2 and E3), some performance is observed over the baseline default face recognition setting without any feedback or instruction (the E1 setting). Therefore, face recognition can be made more accessible with a careful and considerate design.

Appendix A: Legal issues and distribution

In this section, the legal issues and the distribution channels of the BSFDB are described. To participate in the evaluation, an end-user is required to sign the data protection form when requesting for the database. This form was prepared according to the European Directive 95/46/EC which regulates the processing of personal data within the European Union (biometrics information is considered personal and sensitive data). Moreover, we have included information regarding the experiment procedure and regarding how to proceed if a user wishes to remove their data from the BSFDB (this can be done anytime). Biometric and personal data are not linked with names and users are labelled numerically in the database. Once users have read and accepted the conditions they started with the data acquisition. All the biometric and personal data belong to the University of Surrey and it is publicly available online only for researching purposes in this web: [www.surreyfacesdb.com.ac.uk](http://www.surreyfacesdb.com.ac.uk).

*References*

[1] A.K. Jain, P. Flynn, and A.A. Ross, *Handbook of Biometrics*. NJ, USA: Springer-Verlag, 2007.

[2] S. R., Raul Blanco-Gonzalo, Ramon, “Biometrics on Mobile Devices,” in *Encyclopaedia of Biometrics*, Springer, 2014.

[3] P. J. Phillips, P. Grother, R. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, “Face recognition vendor test 2002,” in *IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003*, 2003, p. 44–.

[4] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, “Overview of the face recognition grand challenge,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, 2005, vol. 1, pp. 947–954 vol. 1.

[5] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, “The FERET database and evaluation procedure for face-recognition algorithms,” *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, Apr. 1998.

[6] J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M. R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J.-L. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, S. Garcia-Salicetti, L. Allano, B. Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, M. Ng, M. Fairhurst, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F. M. Sukno, S.-K. Pavani, A. Frangi, L. Akarun, and A. Savran, “The Multiscenario Multienvironment BioSecure Multimodal Database (BMDB),” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1097–1111, Jun. 2010.

[7] P. Tresadern, T. F. Cootes, N. Poh, P. Matejka, A. Hadid, C. Levy, C. McCool, and S. Marcel, “Mobile Biometrics: Combined Face and Voice Verification for a Mobile Platform,” *IEEE Pervasive Comput.*, vol. 12, no. 1, pp. 79–87, Jan. 2013.

[8] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. L. les Jardins, J. Lunter, Y. Ni, and D. Petrovska-Delacrétaz, “BIOMET: A Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities,” in *Audio- and Video-Based Biometric Person Authentication*, J. Kittler and M. S. Nixon, Eds. Springer Berlin Heidelberg, 2003, pp. 845–853.

[9] J. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer, P. J. Flynn, and S. Cheng, “The challenge of face recognition from digital point-and-shoot cameras,” in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1–8.

[10] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y.-W. Teh, E. Learned-Miller, and D. A. Forsyth, “Names and faces in the news,” in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*, 2004, vol. 2, pp. II–848–II–854 Vol.2.

[11] K. Messer, J. Matas, J. Kittler, J. Lüttin, and G. Maitre, “XM2VTSDB: The Extended M2VTS Database,” in *In Second International Conference on Audio and Video-based Biometric Person Authentication*, 1999, pp. 72–77.

[12] E. Bailly-Bailliére, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, and J.-P. Thiran, “The BANCA Database and Evaluation Protocol,” in *Audio- and Video-Based Biometric Person Authentication*, J. Kittler and M. S. Nixon, Eds. Springer Berlin Heidelberg, 2003, pp. 625–638.

[13] R. de Beni and C. Cornoldi, “Imagery limitations in totally congenitally blind subjects,” *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 14, no. 4, pp. 650–655, 1988.

[14] “WHO | World report on disability,” *WHO*. [Online]. Available: http://www.who.int/disabilities/world\_report/2011/report/en/. [Accessed: 29-Apr-2014].

[15] R. Wong, N. Poh, J. Kittler, and D. Frohlich, “Interactive quality-driven feedback for biometric systems,” in *2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, 2010, pp. 1–7.

[16] R. Wong, N. Poh, J. Kittler, and D. Frohlich, “Towards inclusive design in mobile biometry,” in *2010 3rd Conference on Human System Interactions (HSI)*, 2010, pp. 267–274.

[17] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, “Sheep, Goats, Lambs and Wolves A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation,” in *INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING*, 1998.

[18] N. Yager and T. Dunstone, “The Biometric Menagerie,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 220–230, Feb. 2010.

[19] N. Poh and J. Kittler, “A Biometric Menagerie Index for Characterising Template/Model-Specific Variation,” in *Advances in Biometrics*, M. Tistarelli and M. S. Nixon, Eds. Springer Berlin Heidelberg, 2009, pp. 816–827.

[20] X. Tan and B. Triggs, “Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions,” *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.

[21] “Principal Component Analysis and Factor Analysis,” in *Principal Component Analysis*, Springer New York, 2002, pp. 150–166.

[22] V. Štruc and N. Pavešić, “The Complete Gabor-fisher Classifier for Robust Face Recognition,” *EURASIP J Adv Signal Process*, vol. 2010, pp. 31:1–31:13, Feb. 2010.

[23] V. Štruc and N. Pavešić, *Gabor-Based Kernel Partial-Least-Squares Discrimination Features for Face Recognition ∗*. 2007.

[24] “http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.” .

[25] G. Dai and Y. Qian, “A Gabor direct fractional-step LDA algorithm for face recognition,” in *2004 IEEE International Conference on Multimedia and Expo, 2004. ICME ’04*, 2004, vol. 1, pp. 61–64 Vol.1.

[26] P. Kamencay, M. Breznan, D. Jelsovka, and M. Zachariasova, “Improved face recognition method based on segmentation algorithm using SIFT-PCA,” in *2012 35th International Conference on Telecommunications and Signal Processing (TSP)*, 2012, pp. 758–762.