TABULA RASA

Trusted Biometrics under Spoofing Attacks

http://www.tabularasa-euproject.org/

Funded under the 7th FP (Seventh Framework Programme)
Theme ICT-2009.1.4
[Trustworthy Information and Communication Technologies]

D3.1: Evaluation of baseline ICAO biometric systems

Due date: 30/09/2011
Submission date: 30/09/2011
Project start date: 01/11/2010
Duration: 42 months
WP Manager: Abdenour Hadid
Revision: 0

Author(s): Abdenour Hadid (UOULU), Neslihan Kose (EURECOM), Nesli Erdogmus (EURECOM), Sami Romdhani (MORPHO), Julian Fierrez (UAM), Dong Yi (CASIA) and Stan Z. Li (CASIA)

Project funded by the European Commission in the 7th Framework Programme (2008-2010)		
Dissemination Level		
PU	Public	Yes
RE	Restricted to a group specified by the consortium (includes Commission Services)	No
CO	Confidential, only for members of the consortium (includes Commission Services)	No

D3.1: Evaluation of baseline ICAO biometric systems

Abstract:

To gain insight into the performance of current biometric systems when not confronted to spoofing attacks, we report in this deliverable the results of some baseline systems. This will provide valuable baseline performance for later investigating the performance and the vulnerabilities of biometric systems when confronted to spoofing attacks. This document presents the evaluation and performance profiles for the ICAO biometrics addressed within the TABULA RASA project (i.e. 2D face, 3D face, multispectral face, fingerprint and iris). The baseline results presented here will form a cornerstone of all subsequent work related to spoofing and countermeasures, the performance of which will be compared to that presented here. All biometrics are assessed with a common, minimal protocol involving independent development and evaluation datasets and a common operating point, namely the equal error rate (EER). Detection error trade-off curves are also reported to illustrate the dynamic performance of all biometric systems for alternative operating points and different applications.

Contents

1 Introduction

The TABULA RASA project aims to assess the threat of direct, sensor-level spoofing to biometric systems and then to propose novel countermeasures. These activities require the comparison of system performance under spoofing conditions, first without and then with countermeasures, to baseline scores for standard, state-of-the-art biometric systems. The purpose of this deliverable (and also of the companion deliverable D3.2) is thus to establish the baseline scores to which all later experimental results with spoofing and countermeasures will potentially be compared. This particular document presents baseline scores for ICAO biometric modalities (i.e. 2D face, 3D face, multispectral face, fingerprint and iris) considered within the TABULA RASA project. Baseline scores for non-ICAO biometrics are presented within the companion document, D3.2. Both documents relate to the biometric database and system specifications previously reported in D2.2.

All results are presented in terms of detection error trade-off (DET) profiles which illustrate the dynamic behaviour of a biometric system as the decision threshold is changed, i.e. how the false acceptance rate varies according to the false rejection rate. While there are boundless variations on assessment approaches in the literature they all pertain to specific operating conditions, i.e. prior probabilities, false rejection and/or false acceptance costs. Most of these are furthermore dependent on specific applications and biometric modalities. The TABULA RASA project addresses numerous different use-cases and numerous mono-modal and multi-modal biometrics. Thus, in the absence of a single dominant application and multiple biometric combinations, all systems are optimised so as to minimise the equal error rate (EER) which is the sole, standard metric for all biometric modalities in the TABULA RASA project. Results reported here are therefore not necessarily directly comparable to other published results on the same dataset. This, however, does not detract from the impact of the work presented here since the focus in later deliverables, and of the wider project, is on the degradation in performance caused by spoofing relative to the baseline and thus the precise operating point is not of critical importance. Naturally, further work would be required to assess impacts on alternative operating points and specific applications but this is beyond the scope of the work in TABULA RASA. The illustration of DET plots nonetheless gives an idea of performance for other operating points even if they are not that used for optimisation. DET plots are presented for evaluation datasets only.

In all cases experiments relate to a standard, minimal specification which involves multiple sessions and independent development and evaluation datasets. Auxiliary datasets used for the learning of world or universal models and normalisation procedures etc. are independent from those used for development and evaluation. Independent datasets do not contain any overlap in terms of clients/subjects. Full details of database and system specifications are presented in D2.2 though a brief summary of both is provided here so that the document can be read independently.

Before starting with a biometric-by-biometric treatment of baseline results we present here a brief summary of future work in order to show how the material presented here fits into the wider picture and direction of the TABULA RASA project. Figure 1 illustrates three example DET profiles for any hypothetical mono-modal or multi-modal biometric and aims to distinguish the results in this deliverable from those which can be expected in future work. The lowest profile (solid black) is that of the baseline: a standard, state-of-the-art biometric system with no spoofing and no countermeasures. For this hypothetical biometric modality the EER is in the order of 10%. When spoofing attacks are applied performance is expected to degrade, perhaps considerably. Many such profiles will be derived for different spoofing attacks (whereas there is generally only one baseline). One such profile is illustrated and is the highest (dashed red) in Figure 1. It corresponds to an EER in the order of 40%. Finally, the third, middle profile (dashed blue) illustrates performance once spoofing countermeasures are applied. Once again, multiple profiles will be derived for different countermeasure strategies. That illustrated corresponds to an EER in the order of 20% thus showing an improvement over the higher spoofing profile but still a gap to the original baseline. The goal of the latter work is thus to reduce this gap as much as possible and thus the establishment of the lower baseline is a cornerstone of the project.

The only profile in Figure 1 which relates to the work presented in this document is the lowest, namely that of the baseline. Research related to other profiles is the subject of future work and deliverables. For example, D3.3 is the subject of performance under spoofing attacks (D3.4 for non-ICAO biometrics) due in M21 whereas D4.1 and D4.3 are the subject of first initial, and then advanced countermeasures (D4.2 and D4.4 for non-ICAO biometrics) due in M25 and M33 respectively.

In the following sections we present the baseline scores and DET profiles for each of the ICAO biometric modalities (i.e. 2D face, 3D face, multispectral face, fingerprint and iris).

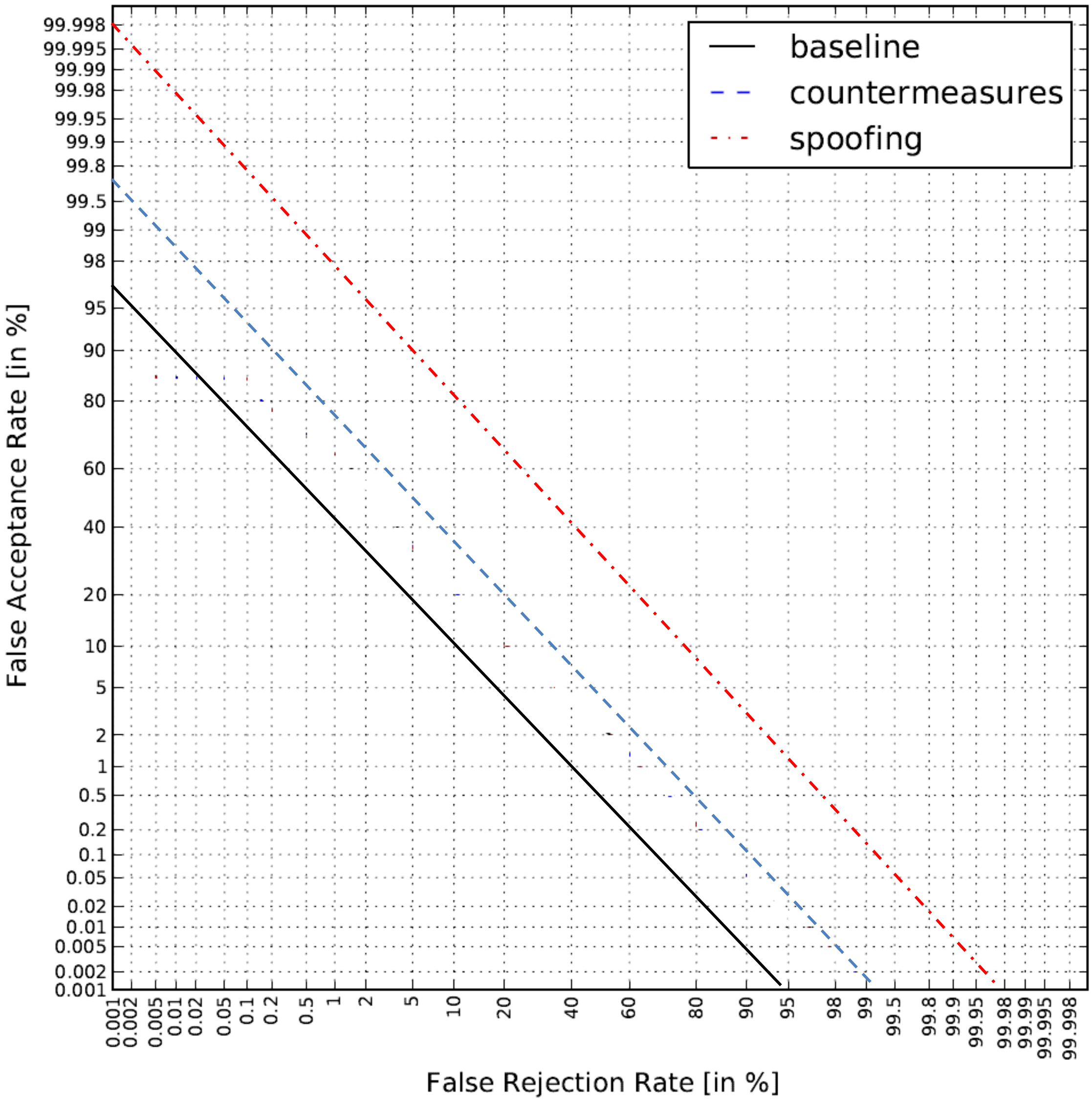

Figure 1: An overview of the different stages in the TABULA RASA project. The objective of the work reported in this deliverable is to establish the lower baseline performance of state-of-the-art biometrics systems for the ICAO modalities (i.e. 2D face, 3D face, multispectral face, fingerprint and iris)

2 2D Face Biometrics

Face recognition aims to uniquely recognize individuals based on their facial physiological attributes. It is a very active field of research because face recognition is natural and non-intrusive. Despite the significant progress in the field, the technology does not yet meet all security and robustness requirements needed by an authentication system for deployment in practical situations. In addition to difficulties related to robustness against a wide range of viewpoints, occlusions, ageing of subjects and complex outdoor lighting, face biometric techniques have also been shown to be particularly vulnerable to spoofing attacks.

There have been several efforts to monitor and evaluate the progress in face recognition. This includes various publically available databases (e.g. FRGC [10], FERET [9], XM2VTS [16], BANCA [5], CMU PIE [22] etc) and associated protocols, and many competitions. Most of these efforts are, however, devoted to evaluate the performance of face recognition methods under illumination changes, ageing of the subjects, occlusion, wide range of viewpoints etc. The problem of evaluating the systems under spoofing attacks is mostly unexplored.

To gain insight into the performance of current face biometric systems when not confronted to spoofing attacks, we report in this section the results of a baseline system (Parts-Based Gaussian Mixture Model) on a challenging publically available database (MOBIO database). This will provide valuable baseline performance for later investigating the performance and the vulnerabilities of face biometric systems when confronted to spoofing attacks.

2.1 Parts-Based Gaussian Mixture Model (PB-GMM)

The face verification system, developed at IDIAP, is chosen as the baseline system for face authentication. The system combines a part-based face representation and Gaussian mixture models (GMMs). The system divides the face into blocks, and treats each block as a separate observation of the same underlying signal (the face). A feature vector is thus obtained from each block by applying the Discrete Cosine Transform (DCT). The distribution of the feature vectors is then modeled using GMMs.

For feature extraction, the face is normalized, registered and cropped. This cropped and normalized face is divided into blocks (parts) and from each block (part) a feature vector is obtained. Each feature vector is treated as a separate observation of the same underlying signal (in this case the face) and the distribution of the feature vectors is modeled using GMMs. The feature vectors from each block are obtained by applying the DCT [20].

Once the feature vectors are calculated, feature distribution modelling is achieved by performing background model adaptation of GMMs [6, 13]. Background model adaptation first involves the training a world (background) model Ω_{world} from a set of faces and then the derivation of client models Ω^{i}_{client} for client i by adapting the world model to match the observations of the client. The adaptation is performed using a technique called mean only adaptation [21].

To verify an observation, x, it is scored against both the client (Ω^{i}_{client}) and world

(Ω_{model}) model. The two models, Ω^i_{client} and Ω_{world}, produce a log-likelihood score which is then combined using the log-likelihood ratio (LLR) to produce a single score. This score is used to assign the observation to the world class of faces (not the client) or the client class of faces (it is the client) based on a predefined threshold τ.

2.2 MOBIO Database

The MOBIO database[1] is chosen for the evaluation of the baseline face biometric system. The database was captured on a challenging acquisition platform (mobile phone), and has a large number of clients captured over a relatively long time period with many sessions. The diverse and challenging nature of the MOBIO database makes it a suitable choice for analyzing the performance of face biometric systems. The database is a publicly available bi-modal (audio and video) database captured at six different sites across five different countries. It was captured in two phases from August 2008 until July 2010 and consists of 150 participants with a female to male ratio of approximately 1:2 (99 males and 51 females). The database was recorded using two mobile devices: a mobile phone and a laptop computer (the laptop was only used to capture part of the first session). In total 12 sessions were captured for each client: 6 sessions for Phase I and 6 sessions for Phase II. The Phase I data consists of 21 recordings in each session whereas Phase II data consists of 11 recordings. The database was collected in natural indoor conditions. The recordings were usually made in offices at the various institutes. However, the same office was not always used. This meant that the recordings do not have a controlled background and nor is the illumination or acoustic conditions controlled. In addition to this, the client was free to hold the mobile phone in a comfortable way which meant that the acoustic quality and pose can vary significantly.

2.3 Performance Evaluation

The main protocol for using MOBIO database divides the database into three distinct sets: one for training, one for development and one for testing. The splitting was done so that each set is composed of the totality of the recording from two sites. This means that there is no information regarding the individuals or the conditions for a site between sets. The three sets, training, development, and testing, are used in different ways. For the training set the data can be used in any way deemed appropriate and all of the data is available. The development set can be used to derive fusion parameters. However, it must at least be used to derive a threshold that is then applied to the test data. To facilitate this, the development set and the test set both have the same style of protocol defined for them. The protocol for the development set and the test set are the same. The first five recordings from the first session are used to enrol the user. Testing is then conducted on each individual file for sessions two to twelve (eleven sessions are used for development and for testing) using 15 videos per session from Phase I and 5 videos per session from Phase

[1] `https://www.idiap.ch/dataset/mobio`

II. This leads to five enrolment videos for each user and 105 test client (positive sample) videos for each user. When producing impostor scores all the other clients are used.

2.3.1 Setup

Face Detection, Cropping and Normalisation :
We manually annotated one image per video by labeling the eye centers. This is used as a groundtruth for cropping the face. Hence, only one face image per video is processed. Using the eye positions, the face images are rescaled so that the eyes are aligned and resized to have 33 pixels between the two eyes. The face images are then cropped to be a 64×80 images and then illumination normalization is performed using the Tan & Triggs normalization method [24].

Feature Extraction :
In the experiments, the face is divided into 8×8 blocks using a sliding window with a 1-pixel shift. Each block is mean and standard deviation normalised and DCT features are then obtained by keeping the 28 lowest frequency coefficients of the DCT. Since the blocks are mean and standard deviation normalised the first coefficient is removed. Thus, $4,161$ blocks or observations are obtained from each facial image with each observation being described by a feature vector of 27 dimensions.

Enrolment :
Before enroling a user we derive a world or background model Ω_{world} to describe what a face looks like in general. This world model is formed using the data from the training set. This background model was trained to have 512 mixture components and is subsequently used to initialise the enrolment of a new user and for scoring. A new user is enroled by performing background model adaptation of GMMs. The new user is enroled by using mean only adaptation as described in [21] (with a relevance factor of 4) from the world model Ω_{world}. Thus for client i we obtain a new GMM Ω^i_{client} by adapting the world model Ω_{world} to match the observations of the client.

Verification :
Verification of an observation, $\boldsymbol{x}$, is performed by scoring against the claimed client model (Ω^i_{client}) and the world (Ω_{model}) model. The two models, Ω^i_{client} and Ω_{world}, both produce a log-likelihood score which are then combined using the log-likelihood ratio (LLR),

$$h(\boldsymbol{x}) = \ln(p(\boldsymbol{x} \mid \Omega^i_{client})) - \ln(p(\boldsymbol{x} \mid \Omega_{world})), \tag{1}$$

to produce a single score. Using a threshold τ this score is then assigned to be a true access when $h(\boldsymbol{x}) \geq \tau$ and false otherwise. We use an efficient approximation of this LLR referred to as linear scoring [8].

2.3.2 Results

The results of the experiments are shown in the DET curves in Figure 2 for male trails and in Figure 3 for female trails.

From the results, we can notice that the baseline system achieved acceptable performance (EERs of around 12.5% for male trails and 17.5% for female trails) but still far from being excellent. This is probably due to the challenging nature of the MOBIO database. The database indeed provides realistic and challenging conditions as it was captured on a mobile phone and in uncontrolled environment. Thus, this database provides a challenging test-bed for face authentication. The reported results can be seen as baseline performance for the MOBIO database when the system is not confronted with spoofing attacks. It is very interesting to gain insight into the performance of the system under spoofing attacks. This issue will be considered in deliverable D3.3.

Note that the baseline system uses only one frame per video for authentication. It is obvious that using more frames would enhance the performance of the system but at the cost of longer processing time. Exploiting the temporal information in the video sequences, e.g. through spatiotemporal representation would also enhance the performance of the system. However, the aim of the evaluation is to only provide baseline performance which of course can be improved.

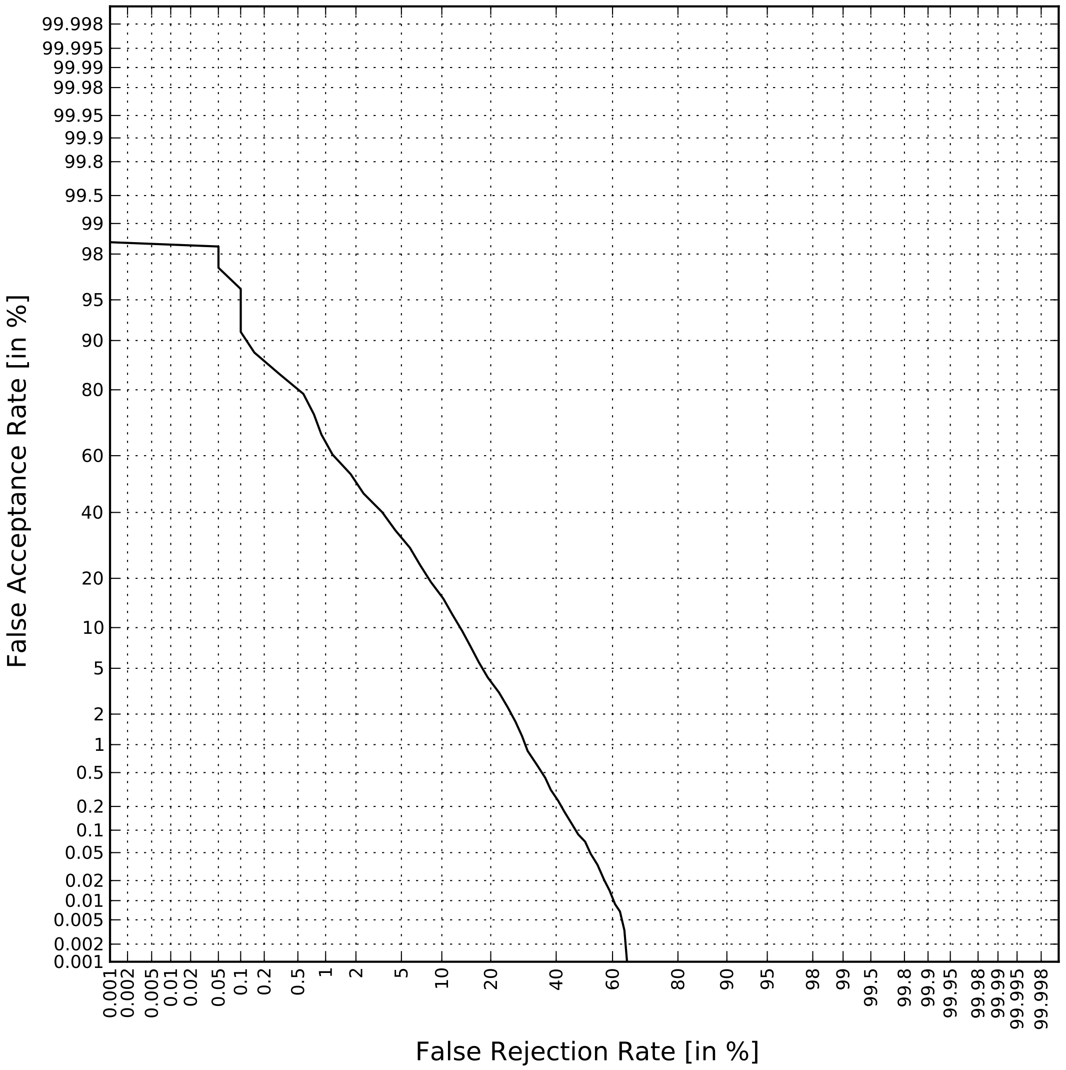

Figure 2: DET plot of the baseline face verification system on the MOBIO database for male trails

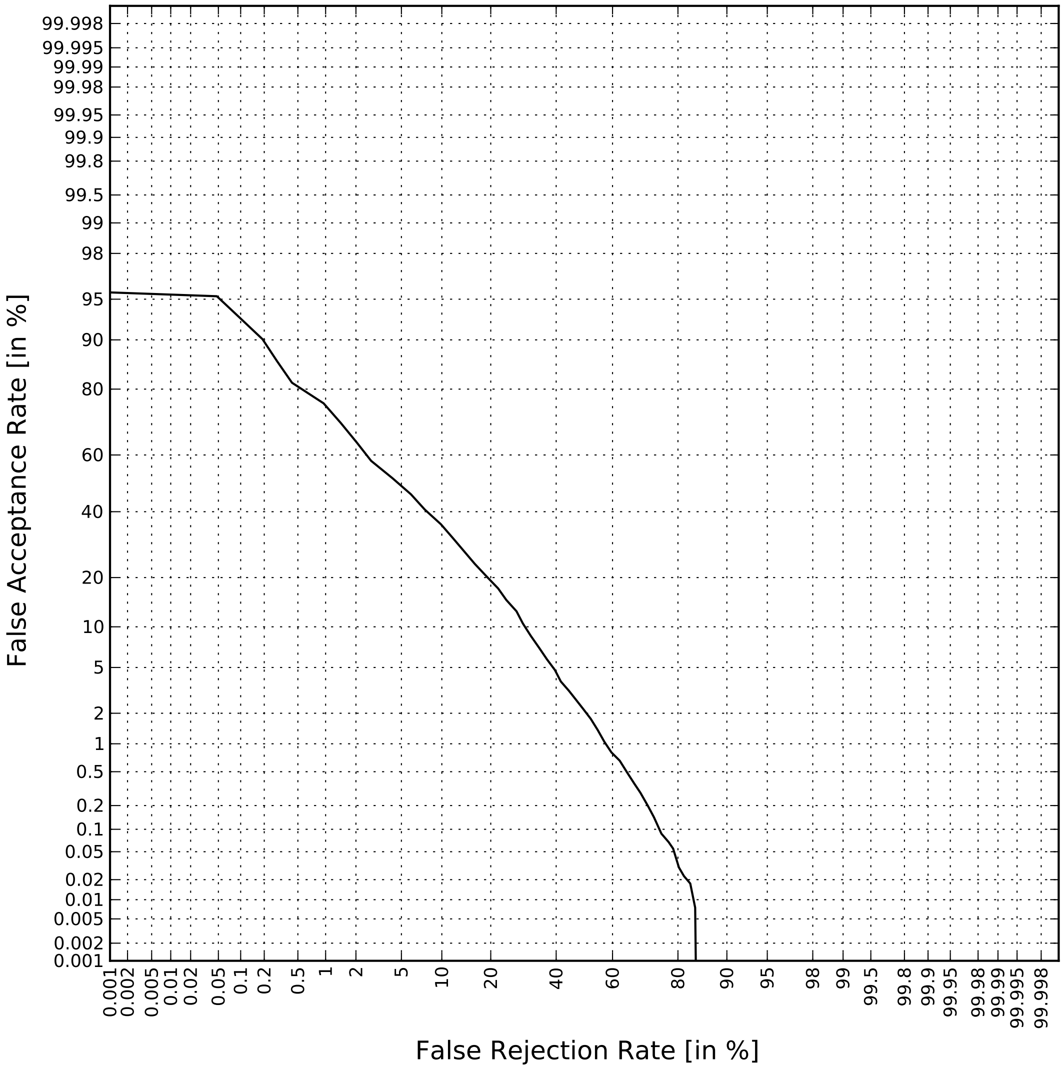

Figure 3: DET plot of the baseline face verification system on the MOBIO database for female trails

3 3D Face Biometrics

Until recently, research in the field of face recognition has been dominated by the use of 2D images. The proposed systems work well under controlled environments but tend to suffer under range of viewpoints, occlusions, aging of the subjects and complex outdoor lighting. To overcome some limitations (especially pose and illumination problems), researchers started to pay an increasing attention to the use of 3D face data [4]. Hence, 3D face recognition is currently a major research area. Traditional problems in 3D face recognition include facial expression changes, occlusion and missing parts, computational time and scalability to large databases.

There have been several studies to evaluate the progress in 3D face recognition. Hence various publically available databases (e.g. FRGC [19], 3D RMA [1], GavadDB [17] etc) and associated protocols have been created, and many evaluations have been conducted to assess the performance of 3D face recognition systems. However, these databases are not meant to analyze the performance of the systems under spoofing attacks, but mainly to test the performance under illumination changes, different pose angles, facial expression variations etc.

In this section, experimental results of a baseline system (namely EURECOM 3D Face Recognition System) on a challenging, publically available database (namely FRGC database) are reported. The goal of the experiments is to provide a baseline reference in terms of performances in the absence of spoofing attacks. This can be later used as a baseline for analyzing the vulnerabilities of 3D systems when confronted to spoofing attacks.

3.1 3D Face System

The 3D face recognition system for the baseline evaluation was developed in the Multimedia Image Group, EURECOM. This system uses a sparser representation of dense 3D facial scans and hence makes the comparison between faces easier for recognition. First a generic face is warped using the Thin Plate Spline (TPS) method for each 3D scan to remove the 'common' face shape information. For each face, 15 fiducial points are considered. Next, the generic face is aligned and scaled on to each face based on the set of points. Then it is coarsely warped to make the two surfaces as close as possible. Assuming that the two surfaces are in sufficient alignment and the correspondences are found as the closest vertices, 136 more point-pairs are obtained. Finally, the generic face is warped based on a total of 151 point-pairs. Thus, each 3D face model can be represented with the 3D vector of size 151×3 which is obtained from the warping parameters in x, y and z directions for each control point.

In order to measure the similarity between facial surfaces, the angle between the two warping vectors and the difference between their magnitudes and angles are calculated. This results in two distance vectors of size 151×1 for any pair of faces. Central tendencies of these vectors are fused and a decision is made based on the nearest neighbour approach.

3.2 The FRGC Database

The specification of the FRGC database can be found in [19]. A subject session consists of four controlled still images, two uncontrolled still images, and one three-dimensional image. The controlled images were acquired in a studio setting and are full frontal facial images taken under two lighting conditions. Facial expressions are neutral or smiling. On the other hand, the uncontrolled images were taken in varying illumination conditions. Each set of uncontrolled images contains smiling or neutral expressions. A Vivid 900/910 sensor is used to capture 3D images. It is a structured light sensor that captures a 640×480 range sampled and registered colour images. All 3D images were taken under controlled illumination conditions appropriate to the sensor. Subjects stood or were seated approximately 1.5 meters from the sensor.

The database consists of training and validation sets. The training set also consists of two parts which are a large still training set and a 3D training set. There are 222 subjects in the large still training set and 466 subjects in the validation set. The database includes $12,776$ images/videos in the large still training set, with 6,388 controlled still images and 6,388 uncontrolled still images. It includes 943×8 images/videos in the 3D training set that contains 3D scans, and controlled and uncontrolled still images. The 3D training set is for training 3D and 3D-to-2D algorithms. The validation set contains images from 466 subjects collected in 4007 subject sessions. In the validation set, there are 4007×8 images/videos. Finally, the database contains static and colourful subjects and consists of single faces. Images are in JPEG format and the resolution is 1704×2272 or 1200×1600.

The FRGC database can be obtained by contacting the FRGC Liaison at frgc@nist.gov. More information on how to access this database can be obtained from the FRGC website [3].

3.3 Performance Evaluation

The performance of the 3D face verification system developed at Eurecom is evaluated using the 4007 3D scans of the validation set of FRGC. This set contains scans of 466 subjects. Although most of the expressions are neutral, some few scans do contain some expression variations as the subjects are smiling or speaking, or have the mouth open. Both the target and query sets consisted of the all 4007 scans. Hence the number of tests amounts to about 8,002,565 impostors test and 23,456 true positive tests. For verification test, all faces in the database are compared with each other. When the two faces belong to same person, then the score is taken as true positive (genuine test). Otherwise it is an impostor test.

3.3.1 Setup

Fifteen landmark points are manually annotated on each 3D scan and on a template scan. These landmarks are selected according to the public landmark set which is explained in [23]. First the scans are pre-processed and the noise is removed. The purpose of the preprocessing is to crop the face region, to eliminate the spikes and the holes and to smooth the facial surface.

In order to crop the face region, robust nose tip detection is applied based on the method suggested in [12] and the facial surface is cropped by a sphere with radius 80mm, centered 10mm away from the nose tip in +z direction. Afterwards, spikes are removed by thresholding and a hole filling procedure is applied. Finally, a bilateral smoothing filter is employed to remove white noise while preserving edges. The template 3D scan is warped to each 3D scan by the Thin Plate Spline algorithm. Then, the warping parameters (feature vectors) that describe the non-global and non-linear transformations and represent the deviations from the common geometric structure are passed to the classifier for face recognition. These feature vectors are then compared using as metric the median of the cosine of the 151 angles between the corresponding deviation vectors as the distance metric.

3.3.2 Results

Figure 4 is a plot of the DET curve obtained by the Eurecom face verification system on the 3D facial scans of the FRGC dataset. The equal error rate (EER) is 11.2%. The FRGC report [19] details the system performances evaluated within the FRGC experiment at a FAR of 0.1%. As it is given in [2], the best algorithm evaluated obtains a VR of 87%, the worst algorithm obtains a VR of 40% and the baseline algorithm (using PCA) obtains a VR of 43%. At a FAR of 0.1%, the Eurecom's algorithm achieves a VR of 50%. Thus the algorithm is better than the FRGC baseline but there is room for improvement compared to the very best algorithms.

It would be interesting to test the performance of our system for different acquisition conditions, for instance, in the presence of illumination or pose variations, or to investigate the performance of our system for different expressions in detail. Unfortunately, there is no publicly available database of 3D scans with landmark annotations that display such variations. Hence, we used the FRGC v2 dataset that is, to the best of our knowledge, the most comprehensive 3D dataset publicly available that encompasses the most variations. It is worth mentionning that the goal of the experiments is only to provide baseline performance.

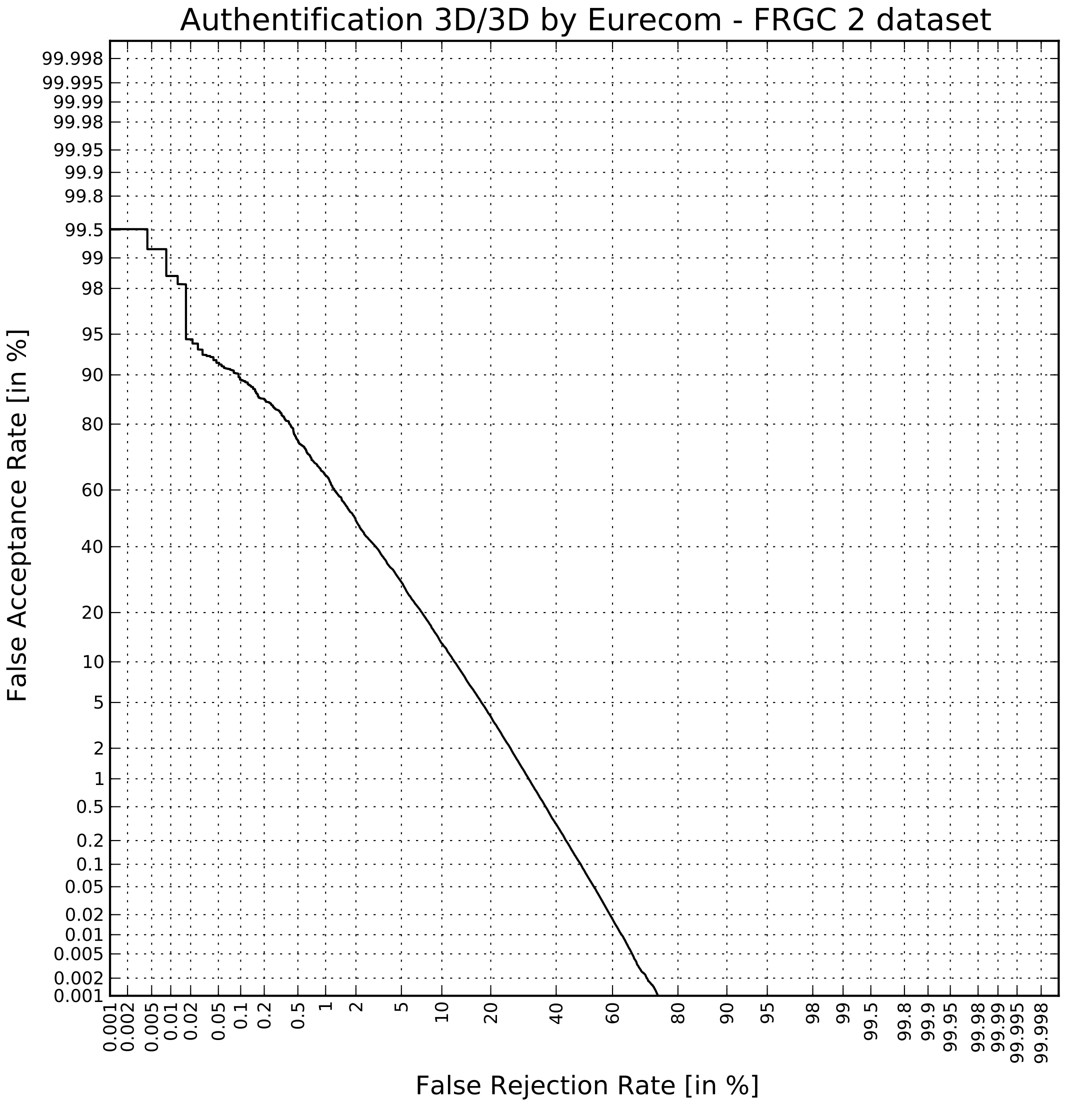

Figure 4: DET plot obtained by the Eurecom Face Verirication system on the 3D facial scans of the FRGC dataset

4 Multi-Spectral Face Biometrics

The use of multiple spectral bands for face biometrics permits to improve the performance and robustness of face recognition in realistic scenarios including uncontrolled illumination conditions. The infrared spectral band has become the most used due to several advantages: radiation which is harmless to health, good-quality images, improved sensitivity, low-cost cameras, etc. Furthermore, the infrared band is a wide spectral region that, using adequate filters, provides multiple images with different characteristics. The millimeter-waves spectral range has also been proposed but not so much research work has been developed with images acquired at such frequencies.

There have been very few papers studying multispectral face spoofing, among which two are most representative. In [18] Pavlidis and Symosek use light at two wavelengths, and a simple threshold method to detect genuine and fake faces. No experiments but only illustrations were reported in their paper. In [26], the authors select light at two different wavelength and then adopt LDA for final decision. However, this approach requires the distance between the user and the system to be exactly 30 cm, and utilizes the users forehead region to measure reflectance. Not only the forehead may be occluded, but also the exact distance is quite demanding and is impossible to execute in practice. The experiments that will be conducted in Tabula Rasa in multispectral face will overcome these limitations.

In this section, experimental results of a baseline multi-spectral system (using visual light and near infrared) developed by CASIA are reported on the challenging HFB database. Again, the goal of the experiments is to provide a baseline reference for later analyzing the vulnerabilities of multi-spectral face approaches when confronted to spoofing attacks.

4.1 The Multi-Spectral System

In multi-spectral face analysis, several spectrum bands can be available and hence exploited for recognition. This may result in different scenarios. In heterogeneous face recognition, for instance, face images in visible light (VIS) can be used for enrollment while near infrared (NIR) face images for testing. Other scenarios include matching between VIS and VIS or NIR and NIR. However, simultaneous use of various spectral bands (e.g. VIS and NIR) for enrollment and/or testing is of key interest as it can yield in best performance. Hence, instead of considering only heterogeneous face recognition scenario as initially planned in D2.2, we decided to investigate in this section a truly multi-spectral approach by fusing different spectral bands. A new multi-spectral baseline system is hence adopted for the baseline evaluation. It is developed by CASIA and uses Local Discriminant Analysis (LDA) for recognition. The system includes the following steps: multi-spectral image acquisition, face detection, eye localization, face alignment, subspace learning, feature extraction, matching and fusion.

The system uses a self-developed camera-device and captures 640 × 480 images in two bands: VIS (visual light) and NIR (near infrared). Then, faces and eyes are detected using an approach based on boosted multi-bloc local binary patterns [27] [25]. Subspace learning is then applied using LDA to obtain a projection matrix for computing low-dimensional discriminative features. In the matching phase, the cosine distance is adopted to evaluate the similarity of different samples and used for the classification. Furthermore, the "sum rule" is used for fusing the similarity scores generated in the two bands: VIS (visual light) and NIR (near infrared).

4.2 The HFB Database

The HFB dataset can be obtained from the Center for Biometrics and Security Research[2]. The database contains images corresponding to the VIS spectrum acquired using a Canon A640 Camera with an image resolution of 640 × 480 pixels. Four frontal images with neutral and smile expression (or with or without glasses) at two different distances were captured. A home-made device was also used to acquire images in the NIR band. For this acquisition NIR LEDs of 850nm were used as an active lighting source and a long pass optical filter was needed to cut off visible light, while allowing most of the 850nm light to pass. The images were also captured with different facial expressions and at different distances, having an image resolution of 640 × 480 pixels.

The HFB data corpus contains images from 100 subjects, with four images at VIS, and four at NIR for each subject. The database is divided into a training set and a test set. There is no intersection for both face images and persons between training and test sets, with half of the users in each set.

The first release of version 1 of the HFB database includes: (i) raw images in JPEG format for VIS and NIR, (ii) the eye coordinates of the two types of images manually labelled, and (iii) cropped versions of the raw images, in two sizes: 32 × 32 and 128 × 128 (based on the eye coordinates).

4.3 Performance Evaluation

For performance evaluation, the images of the first 40 subjects are selected to form the training set, and the images of the last 60 subjects are used as the testing set. To illustrate the performance of individual modality (VIS and NIR) and improvement of the multi-spectral fusion, three sub-experiments were conducted:

1. VIS vs. VIS
2. NIR vs. NIR
3. VIS+NIR vs. VIS+NIR

For the three experiments, six subsets are constructed from the HFB database:

1. VIS training set: all VIS images of the first 40 subjects, 4x40=140 images;

[2] http://www.cbsr.ia.ac.cn/english/Databases.asp

2. VIS gallery set: 2 VIS images of the other 60 subject are selected as gallery;
3. VIS probe set: the other 2 VIS images of the same 60 subjects in subset 2 form the probe;
4. NIR training set: all NIR images of the first 40 subjects, 4x40=140 images;
5. NIR gallery set: 2 NIR images of the other 60 subject are selected as gallery;
6. NIR probe set: the other 2 NIR images of the same 60 subjects in subset 5 form the probe;

For experiment 1, set 1 is used for training, and set 2 and 3 are used for testing. For experiment 2, set 4 is used for training, and set 5 and 6 are used for testing. For experiment 3, set 1 and 4 are used for training, and set 2, 3, 5, and 6 are used for testing.

Before training, all face images are processed by face detection and eye localization modules, aligned to 128x128 according to the eye positions. Then, LDA is applied on the aligned VIS and NIR training sets yielding in projection matrices PVIS and PNIR, respectively.

In the testing phrase of experiments 1 and 2, two aligned face images are firstly projected into their corresponding LDA subspace by the projection matrix PVIS and PNIR. Their similarity is evaluated by the cosine distance in the reduced subspace. For experiment 3, the input images include a VIS image pair and a NIR image pair, the similarity in VIS modality SVIS and the similarity in NIR modality SNIR are fused by the sum rule. The performance or DET curves could be illustrated according these similarity scores and their ground truth identities.

Note that there is no overlapping in subjects between training and testing sets. The system was not tuned or adjusted in any way for the HFB database other than the LDA procedure on the training set.

4.3.1 Setup

The baseline system includes the following steps: multi-spectral image acquisition, face detection, eye localization, face alignment, subspace learning, feature extraction, matching and fusion.

Multi-spectral image acquisition :
The system captures images in two bands using a self-developed device: VIS (visual light) and NIR (near infrared). The device consists of 18 NIR LEDs, an NIR camera, a color camera, and the box. The NIR LEDs and camera are for NIR face image acquisition. The color camera capture color face images are used for fusion with the NIR images. VIS and NIR images are synchronized by the system. The imaging device works at a rate of 30 frames per second with the USB 2.0 protocol for 640x480 images. One can refer to [11] for more details.

Face detection, eye localization and alignment :
Given an image, the MB-LBP+Boosting method [27] [25] is used to detect the rectangular of face and locate the precise positions of two eyes. Then, a face of 128x128 is cropped out according the eye positions by a similarity transformation.

Subspace learning :
Subspace learning methods are applied on a training set to obtain a projection matrix to reduce the noise, computation and enhance the discriminant of the raw face images. For the LDA, PCA is first performed to make the with-in class scatter matrix non-singular, giving a subspace of dimensionality of "sample num" minus "class num"; LDA is then applied on the reduced subspace to find the final c-1 LDA dimensions, where c is the class number.

Feature extraction and matching :
Before matching the face images, a projection matrix estimated on the training data is applied to the images in the test set. Therefore, the final dimensionality of the LDA feature is reduced much compared with the original face images. In the matching phase, the cosine distance is adopted to evaluate the similarity of different samples and used for the classification.

Multi-spectral score fusion :
The "sum rule" is used for fusing the similarity scores generated in multi-spectrums.

4.3.2 Results

The results of the experiments when matching the multispectral models with VIS and NIR face images are shown in the DET curves in Figures 5, 6 and 7.

From these results, we can notice that the baseline system achieved good performance based on VIS (around 3.5% EER), NIR (around 6% EER) and multi-spectral (VIS + NIR, around 1.8% EER) face images. The performance of VIS is better than NIR, which is probably due to the good quality of the VIS face images in the HFB database. When VIS and NIR face images are with the same quality, the VIS face images usually contain richer texture information than NIR face images. Therefore, under ideal environments, VIS is better than NIR for face recognition. However, as shown in [11], NIR performs better than VIS in real applications, because face imaging system is inevitably affected by the environments.

Although NIR is worse than VIS in the experiments, their fusion can still improve the performance significantly. From this, we can see that the information of the multi-spectral face images are complementary each other in the face recognition task. Fused similarity score are more discriminative than any single spectrum.

From all the experimental setups, multi-spectral face is the most adequate scenario to study the effect of spoofing on systems based on multi-spectral fusion (in this case VIS

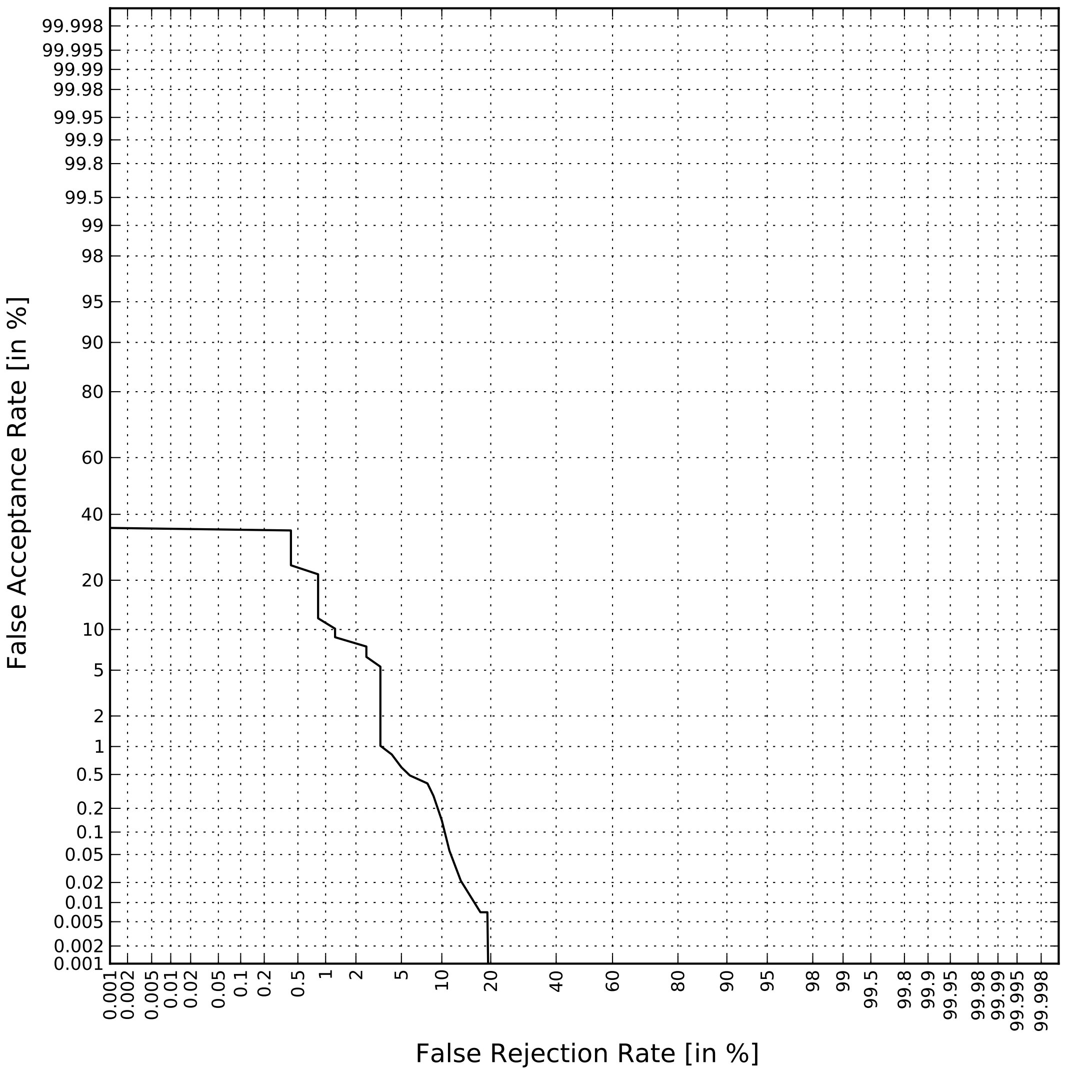

Figure 5: DET plot of the multi-spectral face verification system when matching VIS and VIS images

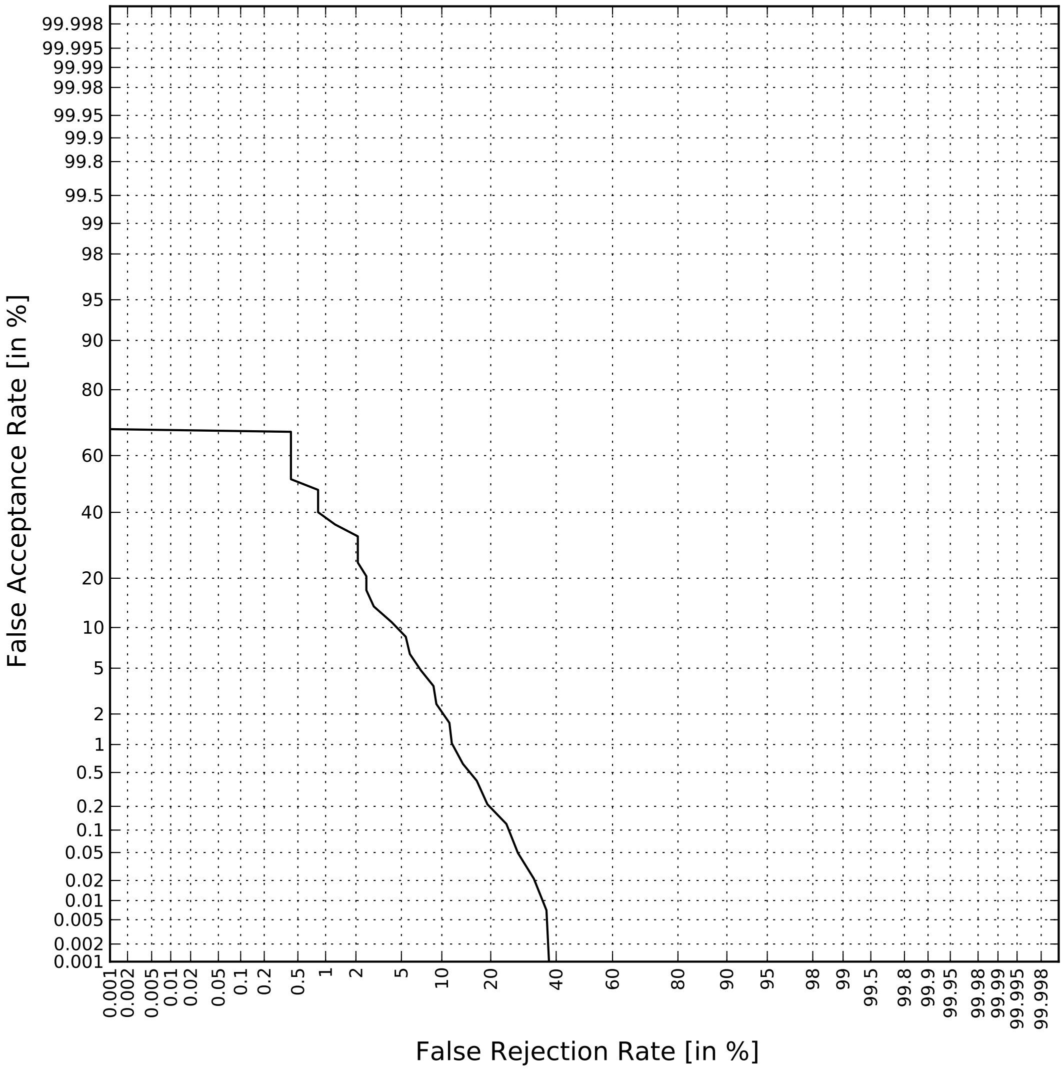

Figure 6: DET plot of the multi-spectral face verification system when matching NIR and NIR images

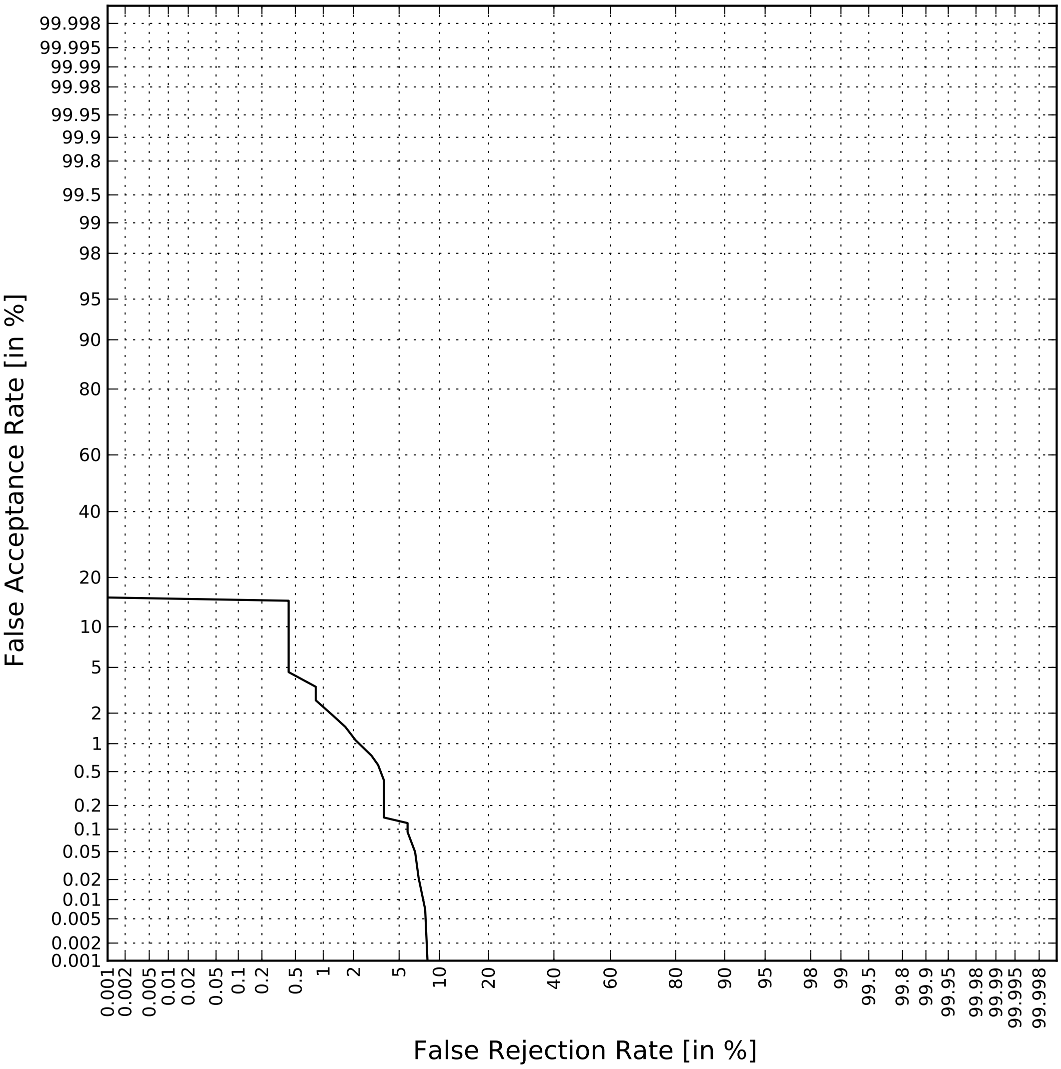

Figure 7: DET plot of the multi-spectral face verification system when matching VIS+NIR and VIS+NIR images

and NIR). Although we just use a simple method (LDA) as a baseline, the simple multi-spectral fusion already reflects its power to lower the error rate in the HFB database. The purpose is not to have a baseline with the best performance, but to have a baseline to which compare the effect of spoofing. This issue will be considered in deliverable D3.3.

5 Iris Biometrics

With fast development of iris image acquisition technology, iris recognition is becoming an important biometric technology, with multiple application areas in national ID cards, banking, e-commerce, biometric passports, etc. Since the 1990s iris image processing and analysis research has achieved great progress. However, performance of iris recognition systems in unconstrained environments is still far from perfect. Iris localisation, nonlinear normalisation, occlusion segmentation, liveness detection, large-scale identification and many other research issues all need further investigation.

The progress in this technology can be tracked through public benchmarks such as the Iris Challenge Evaluation (ICE) organized by NIST, which reports an updated state of the art when considering both controlled and uncontrolled noisy images. However, limitations of this technology related to spoofing have not been yet properly analyzed in the literature in open benchmarks, which is one of the objectives of TABULA RASA.

5.1 The Iris Recognition System

The iris recognition system used for the baseline evaluations was developed by L. Masek [15]. The system basically inputs an eye image, and outputs a binary biometric template. Matching is based on the Hamming distance between templates.

The system consists of the following sequence of steps: segmentation, normalisation, encoding and matching.

Segmentation and Normalisation
For the iris segmentation task, the system uses a circular Hough transform in order to detect the iris and pupil boundaries followed by eyelids and eyelashes removal. Iris boundaries are modelled as two concentric circles. For normalisation of iris regions, a technique based on Daugman's rubber sheet model is employed.

Feature Encoding and Matching
Feature encoding is implemented by convolving the normalised iris pattern with 1D Log-Gabor wavelets. The 2D normalised pattern is broken up into a number of 1D signals, and then these 1D signals are convolved with 1D Gabor wavelets. The rows of the 2D normalised pattern are taken as the 1D signal, each row corresponds to a circular ring on the iris region. The output of filtering is then phase quantised to four levels using the Daugman method, with each filtering producing two bits of data. The encoding process produces a bit-wise template containing a number of bits of information, and a corresponding noise mask which represents corrupt areas within the iris pattern.

For matching, the Hamming distance (HD) is chosen as a metric for recognition. The Hamming distance employed incorporates noise masking, so that only significant bits are used in calculating the Hamming distance between two iris templates.

5.2 The CASIA-IrisV3 Database

The Chinese Academy of Sciences (CASIA) Iris Image Database V3.0 (or CASIA-IrisV3 for short) covers a variety of iris capture situations, labelled as CASIA-Iris-Interval, CASIA-Iris-Lamp, or CASIA-Iris-Twins. It is publicly available and can be obtained from the Center for Biometrics and Security Research website[3]. CASIA-Iris-Interval will be used for baseline experiments. This database contains a total of 22,034 iris images from more than 700 subjects. Both eyes were captured for each subject. All iris images are collected under near infrared illumination using a high quality self-developed iris camera.

The two sets not used for experimentation here (CASIA-Iris-Lamp and CASIA-Iris-Twins) were collected using a hand-held acquisition device by OKI, with much lower image quality than the self-developed camera used for CASIA-Iris-Interval. Thus we select only for our experiments this latter set, which represent better a state of the art acquisition for physical access control.

The self-developed camera, as it is commonly done in state-of-the-art iris acquisition devices, uses near infrared illumination to better highlight the iris texture, which is specially helpful in dark irises.

5.3 Performance Evaluation

The iris recognition system described before was developed and tuned using as reference the first CASIA iris database (CASIA-IrisV1) [15].

For our experiments here, no development database has been used, all available information from CASIA-Iris-Interval was used for evaluation. Therefore, the system is used as initially developed for CASIA-IrisV1 and without change evaluated on CASIA-IrisV3. As explained in the results section, this lack of adaptation to the evaluated database is helpful towards evaluating the system performance and the effects of spoofing.

Each of the two eyes for all subjects was considered a different user, so more than $700 \times 2 = 1400$ users were considered in the experiments. For each user there are an average of 15 genuine images. Genuine scores were computed by matching for each user one genuine image to all the other genuine images of the given user. Impostor matchings were computed by matching the selected genuine image to all the images corresponding to the remaining users.

5.3.1 Setup

Segmentation and Normalisation

The range of search radius values for iris and pupil boundaries is set manually. A maximum value is also imposed to the distance between the circles centres. Eyelids are isolated first by fitting a line to the upper and lower eyelid using the linear Hough transform. Eyelash isolation is then performed by histogram thresholding. For normalization the centre of the pupil is considered as the reference point. Since the pupil can be non-concentric to the

[3] `http://www.cbsr.ia.ac.cn/IrisDatabase.htm`

iris, a remapping formula for rescale points depending on the angle around the circle is used. Normalisation produces a 2D array with horizontal dimensions of angular resolution and vertical dimensions of radial resolution, in addition to another 2D noise mask array for marking reflections, eyelashes, and eyelids detected in the segmentation stage.

Feature Encoding and Matching
In order to account for rotational inconsistencies, when the Hamming distance of two templates is calculated, one template is shifted left and right bit-wise and a number of Hamming distance values is calculated from successive shifts. This method corrects for misalignments in the normalised iris pattern caused by rotational differences during imaging. From the calculated distance values, the lowest one is taken.

5.3.2 Results

The results of the experiments are shown in the DET curve in Figure 8.

From these results, we can notice that the baseline system achieved acceptable performance but still it is far from being excellent (around 4% EER). This is probably due to the limitations of the system and the lack of adaptation to the specific kind of images tested (i.e., no development or tuning on CASIA-IrisV3 of the system). The main purpose of this baseline is to investigate the impact of spoofing, therefore using for testing a different kind of images (CASIA-IrisV3) to the ones for which the system was tuned for (CASIA-IrisV1 [15]) will help us in the comparison of the system working either with genuine or spoofed images (i.e., the system won't be tuned for any of these two kind of images). This comparison will be considered in deliverable D3.3.

Note that the baseline system uses only one image per user for enrollment and one image per user for authentication. It is obvious that using more images or a video would enhance the performance of the system but at the cost of longer processing time. However, the aim of the evaluation is only to provide a baseline performance with which to compare the effect of spoofing in a fair manner, not to obtain the lowest performance.

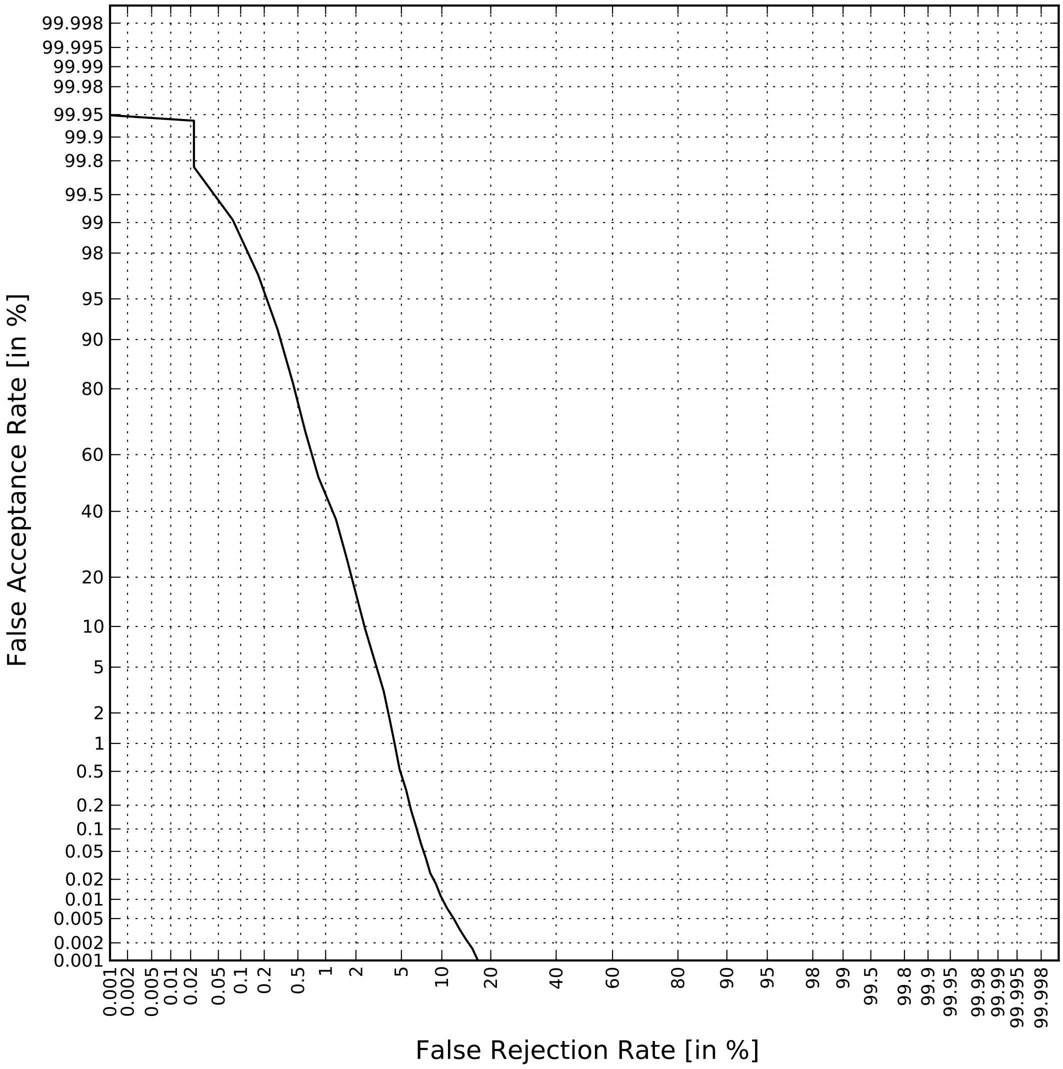

Figure 8: DET plot of the iris verification system

6 Fingerprint Biometrics

Fingerprint biometrics is one of the most developed biometric technologies, with multiple commercial products and adequate performance levels for applications such as physical access control, given that the population considered and the acquisition scenario are controlled and well behaved. This is corroborated by evaluation initiatives as the series of Fingerprint Verification Competitions (FVC) organized by Bologna University, or the various fingerprint evaluation campaigns organized by NIST (e.g., FpVTE 2003).

These solutions can be nevertheless inefficient or even impracticable when confronted with the varying quality of data encountered in some applications such as forensics in realistic scenarios, where latent fingerprints with low quality and partial data are usually encountered. This is also patent in public benchmarks such as the recent Evaluation of Latent Fingerprint Technologies (ELFT) organized by NIST.

The main focus of the TABULA RASA project regarding fingerprint biometrics is on evaluating state-of-the-art commercial systems on controlled data, evaluating its vulnerabilities, and finally developing adequate countermeasures against those vulnerabilities. Therefore we have selected as baseline a state-of-the-art commercial matcher provided by Morpho and a well known publicly available database (BioSecure). The application of automated fingerprint biometrics in forensics with latent data is out of the scope of the project.

As to previous works evaluating the impact of spoofing attacks against fingerprint systems, the main comparative benchmark conducted so far is the First International Fingerprint Liveness Detection Competition, organized in 2009 [14]. Although some interesting insights into the problem can be extracted from this initiative, the databases used and results are quite limited and insufficient for the objectives planned in Tabula Rasa.

6.1 Baseline Systems

6.1.1 The NFIS2 System

The minutiae-based NIST Fingerprint Image Software 2 (NFIS2) [7] is a minutiae-based fingerprint processing and recognition system formed from independent software, which constitutes a de facto standard reference system used in many fingerprint-related research contributions.

From the different software modules that are comprised within NFIS2, the most relevant for evaluation purposes are: MINDTCT for minutiae extraction, and BOZORTH3 for fingerprint matching.

MINDTCT
The MINDTCT system takes a fingerprint image and locates all minutiae in the image, assigning to each minutia point its location, orientation, type, and quality. Together with the minutiae set, a quality map of the input image is also generated using characteristics such as low contrast, incoherent ridge flow, and high curvature. Additionally MINDTCT

computes ridge counts between a minutia point and each of its nearest neighbors.

BOZORTH3
The BOZORTH3 matching algorithm computes a match score between the minutiae from two fingerprints to help determine if they are from the same finger. It uses only the location and orientation of the minutiae points to match the fingerprints, and it is rotation and translation invariant.

6.1.2 The MorphoKit System

MorphoKit is a fingerprint acquisition and processing SDK. The main features available in this SDK are the following:

- Coding of single-finger 500dpi grey-scale fingerprint images to create minutiae templates
- Authentication: 1:1 matching of single-finger fingerprint templates
- Template conversion of the proprietary CLV format to standard formats such as ANSI or ISO
- Live image acquisition with Morpho MSOXXX sensors

The fingerprint template format is CFV, which is a proprietary binary format (1300 bytes on average). Matching two templates provides a score that can be compared to a given threshold for the decision: MATCH/NO MATCH.

6.2 The BioSecure-DS2-Fingerprint Database

The acquisition of the BioSecure Multi-modal Database (BMDB) was jointly conducted by 11 European institutions participating in the BioSecure Network of Excellence. In the fingerprint related activities addressed in TABULA RASA the fingerprint sub-corpus comprised in the Dataset 2 of the BMDB will be used. The main characteristics are:

- Size: the full BMDB fingerprint dataset comprises data from 650 users. For the experiments in Tabula Rasa, we use the subset comprising 210 users publicly available in `http://biosecure.it-sudparis.eu/AB/`.
- Compatibility: it is fully compatible, in terms of sensors used and protocols followed, with other large, multi-modal databases such as BioSec or BioSecureID.
- Multi-modality: compared to other popular, mono-modal fingerprint benchmarks such as the datasets used in the series of FVC competitions, the use of the BMDB permits to perform real multi-modal fusion with other traits such as iris (recall fingerprint+iris is one of the multimodal scenarios studied in TABULA RASA).

Users	Fingers	Hands	Samples/finger	Total Samples/Session
210	Thumb/Index/Middle (3)	Rigth/Left (2)	2	$210 \times 3 \times 2 \times 2 = 2520$

Table 1: A summary of the fingerprint data in BioSecure DS2.

- Coverage: the BMDB was designed to be representative of the population that would make possible use of biometric systems. Thus, it presents both a balanced gender distribution, and also a balanced age distribution.

The BioSecure Multi-modal DB is publicly available through the BioSecure Foundation[4]. It comprises three different datasets acquired under different scenarios, namely: i) DS1, acquired over the Internet under unsupervised conditions, ii) DS2, acquired in an office-like environment using a standard PC and a number of commercial sensors under the guidance of a human supervisor, and iii) DS3, acquired using a mobile portable hardware under two acquisition conditions: indoor and outdoor.

For the fingerprint related activities of the TABULA RASA project, we will use two fingerprint sub-corpora comprised within DS2, the one captured with the optical sensor Biometrika FX2000 and the one captured with the Atmel sweeping thermal sensor.

Using these two databases permits us analyzing two very different scenarios of practical importance for Tabula Rasa: 1) physical access control using high a quality sensor (optical), and 2) authentication using a mobile device with a small size and low quality sensor (thermal sweeping).

6.3 Performance Evaluation

Similarly to the evaluation for 2D Face Biometrics, the BioSecure database has been divided into three distinct sets: training, development and testing, with 70 users in each of them. The training set is intended for system tuning and adjustment. The development set can be used to derive fusion parameters, and must be used to derive a threshold that is then applied to the test data. To facilitate the application of this threshold to the test data, the both sets have the same protocol defined for them.

The DS2 fingerprint sub-corpus used here was captured in two separate acquisition sessions. The data available for each session is summarised in Table 1.

Only the data from the right index of each subject was used, therefore the available information is 210×2 images for each of the two sessions. The three sets defined before (training, development, and testing) each contains 70 subjects randomly taken from the 210 available (this random selection is recorded so that these sets can be also generated in the same way for other modalities when fusing different biometrics). Each user is then enrolled by using the two images from the first session. A match score is generated by averaging the score computed against both enrolled images by the MorphoKit system. Genuine scores are generated for each of the two samples of the second session for all accounts in the considered set (training, development, or testing), and impostor scores are

[4] http://biosecure.it-sudparis.eu/AB/

generated by matching the enrolled account to both samples of the second session of the remaining accounts within the considered set.

6.3.1 Setup

No training or tuning of the MorphoKit system has been conducted on the BioSecure DB. The software was used as developed for Morpho sensors, but in this case applied to the data available in BioSecure DB, i.e., either from optical or thermal sensors by Biometrika and Atmel, respectively.

6.3.2 Results

The results of the experiments using the MorphoKit system are shown in the DET curves in Figures 9 and 10 for both sensors considered, optical and thermal, respectively.

From these results, we can notice that the MorphoKit system achieved excellent performance (around 0.02% EER for the optical sensor), which is especially relevant considering that the system was not tuned in any way for the kind of images used. On the other hand, the fingerprint data used was controlled during acquisition and is not as challenging as can be in applications such as forensics, out of the scope of this project. Anyway, the database used is a good indication of the kind of images found in some practical applications such as physical access control.

When comparing the results of the MorphoKit system for both sensors (optical and thermal, in Figures 9 and 10, respectively), we see an important performance drop from around 0.02% EER to around 0.7% EER, mainly due to the lower quality of the images acquired with thermal sweeping technology.

The results of the experiments using the NFIS2 system are shown in the DET curves in Figure 11 and 12 for both sensors considered, optical and thermal, respectively.

From this second set of results, we can notice that the baseline system from NIST achieves much lower performance than the MorphoKit system for the optical sensor (around 3% EER compared to 0.02% EER). This result highlights the extraordinary performance achieved by the MorphoKit system using the optical sensor.

When comparing the results for both baseline systems on the thermal sensor (Figures 10 and 12), we see that the performance of the MorphoKit system is in this case not so good as with the optical sensor (around 0.7% EER in this case compared to 0.02% before). This shows that the MorphoKit system, although having an extraordinary performance on the optical sensor, is significantly degraded on other kind of images such as those coming from the thermal one. On the other hand, the NIST system performs similarly on both kind of sensors (around 3% EER in both cases). This result demonstrates the robustness of the baseline system provided by NIST against changes in input images.

The reported results can be seen as baseline performance for the BioSecure database when the baseline systems are not confronted to spoofing attacks. It is very interesting to gain insight into the performance of the systems under spoofing attacks. This issue will be considered in deliverable D3.3.

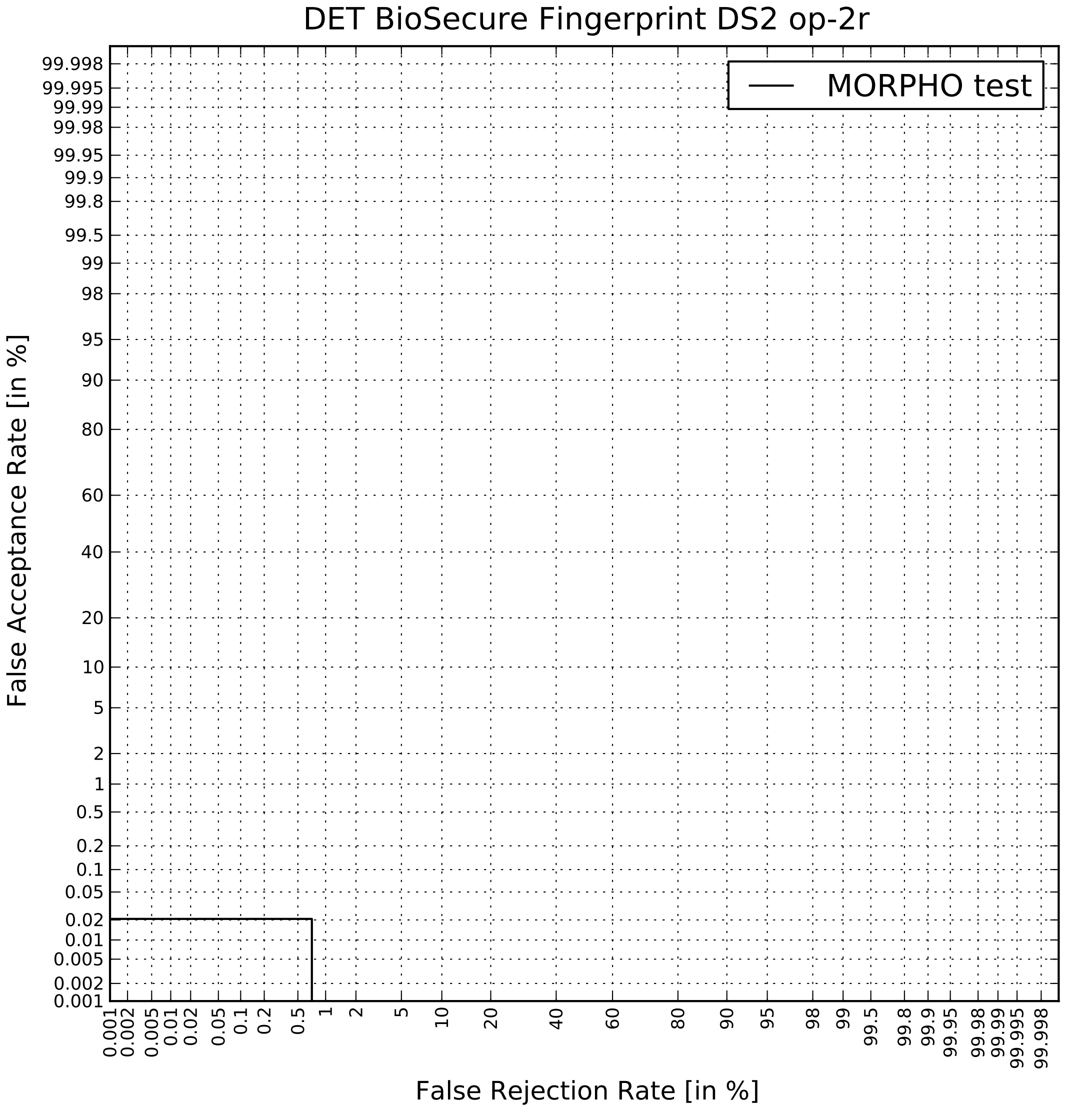

Figure 9: DET plot of the fingerprint verification system (MorphoKit) for the optical sensor

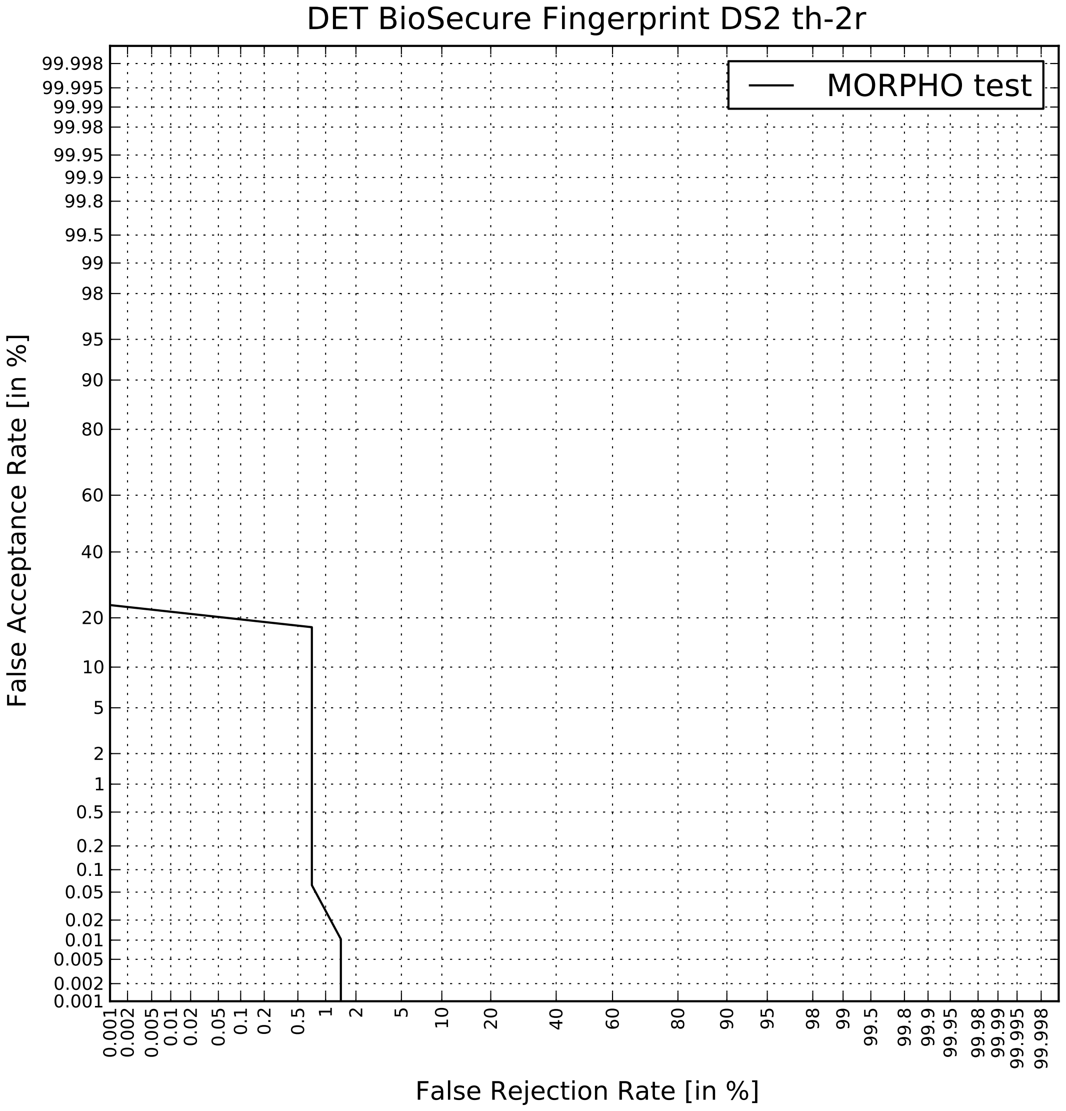

Figure 10: DET plot of the fingerprint verification system (MorphoKit) for the thermal sensor

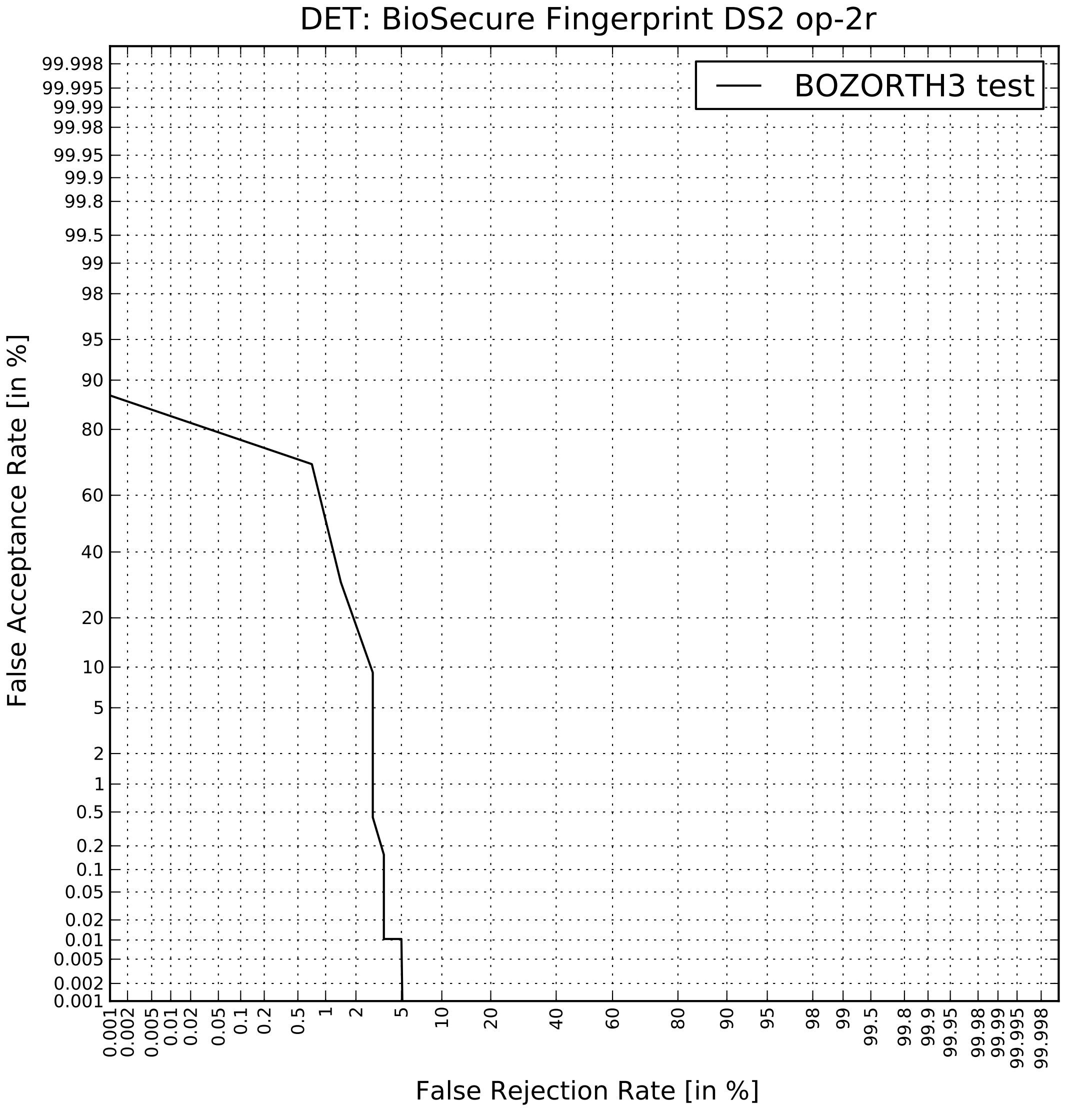

Figure 11: DET plot of the fingerprint verification system (NFIS2) for the optical sensor

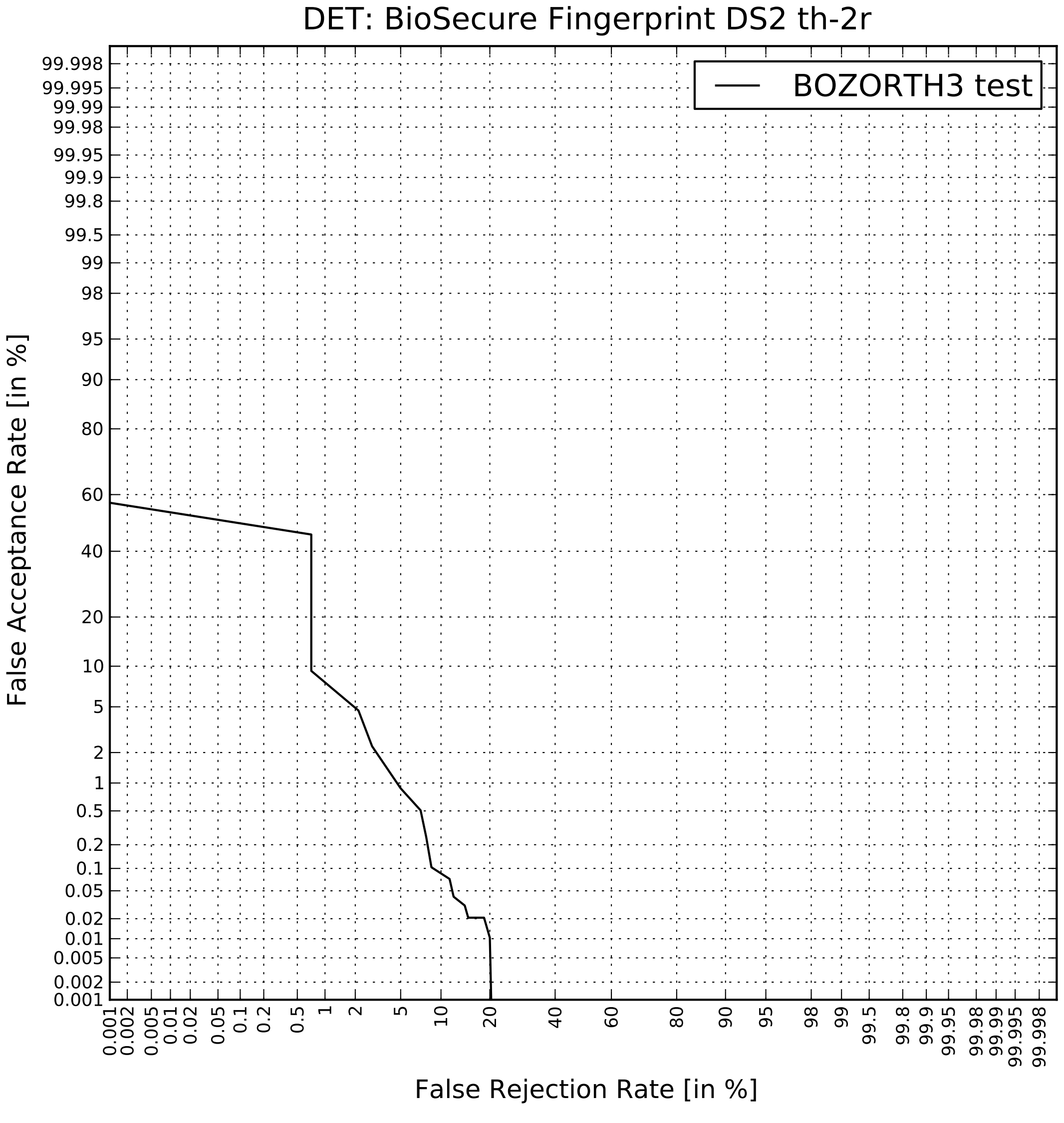

Figure 12: DET plot of the fingerprint verification system (NFIS2) for the thermal sensor

Note that the baseline systems use two images per user for enrolment and only one image for authentication. It is obvious that using more images for either enrolment or authentication would enhance the performance of the systems.

7 Summary

Presented in this document are equal error rates (EERs) and detection error trade-off (DET) profiles for ICAO biometrics assessed in the TABULA RASA project. Together they form the baselines for all future work in spoofing and countermeasures. We re-iterate that traditional biometrics research is not the goal in this project and thus we do not necessarily seek recognition performance which is superior to the baseline. We seek to show how (i) the baseline performance deteriorates in the face of spoofing and, more importantly, (ii) how countermeasures may be effectively harnessed to reduce the gap between performance under spoofing and the baseline performance presented here.

The different baseline systems show variations in performance from EER of 0.02% to 19.5%. While it is not the objective to compare the performance of each modality a summary of the EERs for each biometric is listed in Table 2.

Modality	Database	System	Subset	EER (%)
2D face	MOBIO	PB-GMM	male	12.5
			female	19.5
3D face	FRGC	3D Face System	-	11.2
Multi-spectral face	HFB	The Multi-Spectral System	VIS-VIS	3.5
			NIR-NIR	6.0
			VIS+NIR	1.8
Iris	CASIA-IrisV3	The Iris Recognition System	-	4.0
Fingerprint	BioSecure-DS2	MorphoKit	Optical	0.02
			Thermal	0.7
		NFIS2	Optical	3.0
			Thermal	3.0

Table 2: A summary of EERs for the ICAO biometrics addressed in TABULA RASA.

References

[1] 3d_RMA : 3D database. page http://www.sic.rma.ac.be/ beumier/DB/3d_rma.html.

[2] 3rd frgc workshop. page http://biometrics.nist.gov/cs_links/face/frgc/FRGC_WK3_Brief.pdf.

[3] NIST. face recognition grand challenge(FRGC). page http://face.nist.gov/frgc/.

[4] A. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2D and 3D face recognition: A survey. *Elsevier Pattern Recognition Letters 28*, pages 1885–1906, 2007.

[5] E. Bailly-baillire, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Marithoz, J. Matas, K. Messer, F. Pore, and B. Ruiz. The BANCA database and evaluation protocol. In *In Proc. Int. Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA03*, pages 625–638. Springer-Verlag, 2003.

[6] F. Cardinaux, C. Sanderson, and S. Marcel. Comparison of mlp and gmm classifiers for face verification on XM2VTS. In *Proc. International Conference on Audio- and Video-based Biometric Person Authentication*, pages 911–920, 2003.

[7] M. Garris, C. Watson, R. McCabe, and C. Wilson. User's guide to NIST fingerprint image software 2 (nfis2). Technical report, NIST, 2004.

[8] O. Glembek, L. Burget, N. Dehak, N. Brummer, and P. Kenny. Comparison of scoring methods used in speaker recognition with joint factor analysis. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4057–4060, 2009.

[9] P. Jonathon, M. Hyeonjoon, R. Syed, and R. Patrick. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[10] P. Jonathon, F. Patrick, S. Todd, B. Kevin, C. Jin, H. Kevin, M. Joe, M. Jaesik, and W. William. Overview of the face recognition grand challenge. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 01*, pages 947–954, 2005.

[11] S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):627–639, 2007.

[12] X. Lu and A. Jain. Multimodal facial feature extraction for automatic 3D face recognition. *Department of Computer Science, Michigan State University, East Lansing, Michigan, Tech. Rep. MSU-CSE-05-22*, August 2005.

[13] S. Lucey and T. Chen. A gmm parts based face representation for improved verification through relevance adaptation. In *Computer Vision and Pattern Recognition*, pages 855–861, 2004.

[14] G. L. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli, and S. A. C. Schuckers. First international fingerprint liveness detection competition - livdet 2009. In *Proc. ICIAP 09*, pages 12–23, 2009.

[15] L. Masek and P. Kovesi. Matlab source code for a biometric identification system based on iris patterns. Technical report, The School of Computer Science and Software Engineering, The University of Western Australia, 2003.

[16] K. Messer, J. Matas, J. Kittler, and K. Jonsson. XM2VTSDB: The extended M2VTS database. In *In Second International Conference on Audio and Video-based Biometric Person Authentication*, pages 72–77, 1999.

[17] A. Moreno and A. Snchez. GavabDB: a 3D face database. *In Workshop on Biometrics on the Internet*, pages 77–85, March 2004.

[18] I. Pavlidis and P. Symosek. The imaging issue in an automatic face/disguise detection system. In *Proc. IEEE workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, 2000.

[19] P. J. Phillips, P. J. Flynn, and T. Scruggs. Overview of the face recognition grand challenge. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*, 1:947–954, 2005.

[20] K. R. Rao and P. Yip. *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Academic Press, Boston, 1990.

[21] D. A. Reynolds. Comparison of background normalization methods for text-independent speaker verification. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 963–966, 1997.

[22] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1615–1618, 2003.

[23] P. Szeptycki, M. Ardabilian, and L. Chen. A coarse-to-fine curvature analysis-based rotation invariant 3d face landmarking. *International Conference on Biometrics: Theory, Applications and Systems (BTAS 09)*, 2009.

[24] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing*, 6(19):1635–1650, 2010.

[25] D. Yi, Z. Lei, and S. Z. Li. A robust eye localization method for low quality face images. In *Proc. International Joint Conference on Biometrics (IJCB 2011), Washington, D.C., USA, accepted*, 2011.

[26] S. Y. Youngshin Kim, Jaekeun Na and J. Yi. Masked fake face detection using radiance measurements. *Journal of the Optical Society of America A*, 26(4), 2009.

[27] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li. Face detection based on multi-block lbp representation. In *ICB*, pages 11–18, 2007.