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]

D2.2: Specifications of Biometric Databases and Systems

Due date: 29/04/2011
Submission date: 29/04/2011
Project start date: 01/11/2010
Duration: 42 months
WP Manager: Javier Acedo
Revision: 1

Author(s): N. Kose (EURECOM), R. Vipperla (EURECOM), N. Evans (EURECOM), J.-L. Dugelay (EURECOM), A. Riera (STARLAB), A. Soria-Frisch (STARLAB), J. Fierrez (UAM), A. Hadid (UOULU), J. Bustard (USOU), S. Li (CASIA), S. Brangoulo (MORPHO), G.-L. Marcialis (UNICA)

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

D2.2: Specifications of Biometric Databases and Systems

Abstract:

This document defines the different biometric databases and systems that will be used within the TABULA RASA project. The range of biometrics considered includes: 2D face, 3D face, multi-spectral face, iris, fingerprint, voice, gait, vein and electro-physiology, in addition to multi-modal biometrics. Since, for any specific biometric, databases and systems are not necessarily provided by the same partner it is essential that both database and systems providers share a common understanding of each component. The document also provides a basis for common evaluation strategies and protocols which will serve to ensure quality and that all evaluation results may be meaningfully and reliably interpreted. The same databases and systems described here will furthermore be used for baseline, spoofing and countermeasure assessements. It is thus critical that formal specifications are defined.

Contents

1 Introduction

This document details the biometric databases and systems that will be used within the European Union (EU) 7th Framework Programme (FP7) Small or Medium-Scale Focused Research Project (STREP) entitled 'Trusted Biometrics under Spoofing Attacks' (TABULA RASA).

Biometrics included within the focus of the project include: 2D face, 3D face, multi-spectral face, iris, fingerprint, voice, gait, vein and electro-physiology. The project will also consider various multi-modal biometric combinations. Thus, in addition to specific databases for mono-modal biometrics, specific multi-modal collections are also required. This document describes the different databases that were identified through the TABULA RASA kick-off meeting and first Technical Meeting. Also described are the numerous biometric systems, including approaches to multi-modal fusion/scoring. In most cases there is one database and system per biometric, however, in some cases there are either multiple datasets and/or systems. This is mostly due to the need for compatibility with multi-modal biometrics for which performance will be compared to their respective mono-modal counterparts. The selection of standard, large databases, where possible, will help to ensure quality, statistical significance and that all evaluation results may be meaningfully and reliably interpreted. It is therefore important that the context of each evaluation is defined.

This document does not describe any evaluation work, baseline or otherwise. This will be reported in following deliverables. Evaluation work, both of baseline systems, spoofing threats and countermeasures will, however, be based upon the same databases and systems described in this document.

The remainder of this document is organised as follows. Each mono-modal biometric is first discussed in Sections 2 to 10. ICAO-biometrics[1] are described in Sections 2 to 6 whereas non-ICAO biometrics are described in Sections 7 to 10. In each case the database and corresponding system are described in turn with a common structure. Finally, multi-modal databases and systems are described in Section 11.

[1]face, fingerprint and iris biometrics, as per the International Civil Aviation Authority

2 2D face biometrics

Face recognition is a preferred biometric in identity recognition since it is natural, robust and non-intrusive. Face recognition aims to uniquely recognise individuals based on their facial physiological attributes. Unfortunately, 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 where a person tries to masquerade as another one by falsifying data.

In this section, we describe the baseline face database and authentication system that will be considered in the TABULA RASA project for investigating the performance and the vulnerabilities of face biometric systems.

2.1 Database

There are several benchmark face databases (e.g. FERET [89], FRGC [2], CMU PIE [91] etc) and also multi-modal databases (e.g. XM2VTS [87], BANCA [88], BIOSECURE [92] etc.) that can be used for evaluating face recognition systems. Some of these databases are publicly available for research purposes while others are not. Below is a description of some major, publicly available databases.

2.1.1 Existing databases

FERET:
The FERET database [89] consists of a total of 14051 grey-scale face images representing 1199 individuals. The images contain variations in lighting, facial expressions, pose angle etc. The frontal face images are divided into five sets as follows: *fa* set, used as a gallery set, contains frontal images of 1196 people; *fb* set contains 1195 images of clients who were asked for an alternative facial expression than in the fa photograph; *fc* set contains 194 images taken under different lighting conditions. *dup I* set contains 722 images taken later in time; and *dup* II set of 234 images consisting of a subset of the dup I set containing those images that were taken at least a year after the corresponding gallery image. The FERET database is widely used in evaluating face recognition methods.

XM2VTS:
The XM2VTS (extended M2VTS) database [87] consists of audio recordings and video sequences of 295 clients uttering three fixed phrases, two ten-digit sequences and one seven-word sentence, with two utterances of each phrase, in four sessions taken at one month interval. Fig. 1 shows examples of face images from XM2VTS database. The main drawbacks of this database are its limitations to uniform background and controlled illuminations. XM2VTS database has been frequently used in the literature for comparison

of different biometric systems.

BANCA:
The BANCA database [88] consists of audio recordings and video sequences of 208 clients (half men and half women) recorded in three different scenarios, controlled, degraded and adverse, over a period of three months. The clients were asked to say a random 12-digit number, their name, their address and date of birth, during each recording. The BANCA database was captured in four European languages (English, French, Italian and Spanish) but only the English part was made publicly available. Both high- and low-quality microphones and cameras were used for recording. The BANCA database provides realistic and challenging conditions and allows for comparison of different systems with respect to their robustness.

Although publicly available, the databases described above do not contain realistic and common environmental variations associated with the usage and performance evaluation of face biometric systems in challenging settings. A suitable database should, for instance, fulfil the following criteria: it should be publicly available for research purposes, recorded in natural environments during a long time period, contain a large number of clients with several recordings (shots) per client, etc. Fortunately, the recently recorded MOBIO database meets most of the required criteria as it can be used to address several important issues in face biometrics and enables a fair comparison of future mono-modal and multi-modal biometric authentication systems. The diverse and challenging nature of the MOBIO database has motivated its usage in TABULA RASA.

2.1.2 The MOBIO database

Among the most recent face databases is the MOBIO database which 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 analysing the performance of face biometric systems. The MOBIO database[2] is publicly available for research purposes and can be downloaded after completing and signing an End User License Agreement (EULA).

The MOBIO database is a publicly available bi-modal (audio and video) database captured at six different sites across five different countries. The database 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

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

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.

There are many possible protocols that can be used with the MOBIO database. The main protocol 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 II (see Table 1 for a description of the usage of data for the Testing and Development splits). 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. For instance, if in total there were 50 clients then the other 49 clients would perform an impostor attack.

Development and Testing Splits		
Session number	Usage	Questions to use (number of questions)
Session 1	Enrollment	**Set questions only** (5)
Session 2	Test Scores	**Free speech only** (15)
Session 3	Test Scores	**Free speech only** (15)
Session 4	Test Scores	**Free speech only** (15)
Session 5	Test Scores	**Free speech only** (15)
Session 6	Test Scores	**Free speech only** (15)
Session 7	Test Scores	**Free speech only** (5)
Session 8	Test Scores	**Free speech only** (5)
Session 9	Test Scores	**Free speech only** (5)
Session 10	Test Scores	**Free speech only** (5)
Session 11	Test Scores	**Free speech only** (5)
Session 12	Test Scores	**Free speech only** (5)

Table 1: Table describing the usage of data for the Testing and Development splits of the MOBIO database.

2.2 System

Many 2D face recognition methods have been proposed in literature. In order to better gain insight into the performance and the vulnerabilities of the existing systems, it is suitable to consider a state-of-the-art method as the baseline system. Among state-of-the-art systems are those based on local binary patterns (LBPs), developed at the University of Oulu, and IDIAP's face verification system combining part-based face representation and Gaussian Mixture Models (GMMs). IDIAP's system is chosen to be used as the baseline system for face authentication in the TABULA RASA project.

2.2.1 Existing systems

Face recognition as a research field has existed for more than 30 years and has been particularly active since the early 1990s [82]. Researchers from many different fields (from psychology, pattern recognition, neuroscience, computer graphic and computer vision) have attempted to create and understand face recognition systems [82]. This has led to many different methods and techniques in this field. These techniques have often been divided into two groups (1) holistic matching methods and (2) feature-based matching methods [82]. Holistic approaches use the whole face as one input while feature-based methods extract multiple features (eye position, nose position, angles between physical features or local frequency responses) and analyse them separately before fusing the different results [82].

The state-of-the-art for face recognition is currently dominated by two themes: using parts or partitions of the face (which is often not strictly a feature-based nor a holistic technique) and the use of Local Binary Patterns (LBPs) [83]. These two themes are exemplified by the local Gabor binary histogram sequences (LGBPHS) technique [84]. This technique obtains local histograms of LBPs from non-overlapping blocks and then concatenates these histograms to form a single sequence or feature vector; this can be considered to be both feature-based and holistic. These two themes are not unique to the LGBPHS technique, with several other methods making use of them as the basis for their systems, this includes: region-based LBP histograms [85] with adaptation [86] and feature distribution modelling of the local discrete Cosine transform (DCT) [93].

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

IDIAP's face verification system, which combines a part-based face representation and GMMs, will be used as a baseline system for face authentication.

Parts-based approaches divide the face into blocks, or parts, and treat each block as a separate observation of the same underlying signal (the face). According to this technique, a feature vector is obtained from each block by applying the Discrete Cosine Transform (DCT) and the distribution of these feature vectors is then modelled using GMMs. The PB-GMM face authentication system consists of three steps: feature extraction, feature distribution modelling and face verification.

Feature Extraction
The feature extraction algorithm is described by the following steps. The face is normalised, registered and cropped. This cropped and normalised 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 modelled using GMMs. The feature vectors from each block are obtained by applying the DCT [93].

Feature Distribution Modelling
Feature distribution modelling is achieved by performing background model adaptation of GMMs [79, 80]. 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 [81] as this requires only few observations to derive a useful approximation for adapting the means of each mixture component.

Verification
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),

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

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) and consequently a threshold τ has to be applied to the score $h(x)$ to declare (verify) that x matches to the i^{th} client model Ω^i_{client}, i.e if $h(x) \geq \tau$.

3 3D face biometrics

It is observed that variations in pose, illumination and expression limit the performance of 2D face recognition techniques and, in recent years, 3D face recognition has shown promise to overcome these challenges [1]. 3D-face recognition will therefore also be considered in the TABULA RASA project.

3.1 Database

3D capturing processes are becoming cheaper and faster, and for this reason recent works attempt to solve the problem directly on a 3D face model [1]. Currently there are several available 3D databases with different specifications. In TABULA RASA a 3D database which provides sufficient and high quality 3D data is needed.

3.1.1 Existing databases

In fact, there are very few 3D face databases and mostly they contain very little data. In this section we describe the most important 3D face databases with brief details of their specifications that were obtained from [1]. All of the databases described here are publicly available.

The 3D Royal Military Academy (RMA) of Belgium database is a cloud of points database. Data size is 4000 points and the database contain data from 120 clients (106 male, 14 female). For each client, there are three 3D models. There are no texture images. For a long time the 3D RMA database has been the only publicly available database, although its quality is rather low.

SAMPL is a range image type database. Data size is 200*200 from 10 clients. For each client there are 33 3D models (for 2 sub) and one 3D model (for 8 sub). Texture images are also provided.

The University of York 1 3D database is another range image type database which contains data from 97 clients. For each client, there are ten 3D models but there are no texture images. The University of York 2 3D database is also a range image type database but, with data from 350 clients, it is significantly larger. For each client there are 15 3D models but again there are no texture images.

Finally we consider the GavadDB 3D database. It is a tri-mesh database and contains data from 61 clients (45 male, 16 female). For each client there are 9 3D models but again no texture images.

As for all biometric databases one of the most important aspects relates to the amount of data. The Face Recognition Grand Challenge (FRGC) database has a relatively large number of samples and consists of 50,000 recordings divided into training and validation partitions [2] making it the largest currently available dataset of 3D faces [3]. Another important aspect is the quality of the data. The FRGC data corpus contains high resolution still images taken under controlled lighting conditions and with unstructured illumination, 3D scans and contemporaneously collected still images [2]. Also in the FRGC database 3D images consist of both range and texture channels, which is another significant advantage of this database. When compared to the other databases, due to its advantages, the FRGC database is more popular, hence in the TABULA RASA project we will also use the FRGC database for 3D face biometrics.

3.1.2 The FRGC database

To obtain the FRGC data set one should contact the FRGC Liaison at frgc@nist.gov. On FRGC website [4] it is mentioned that the request must come from a full-time employee or faculty member of the requesting organisation/university. Furthermore, it is mentioned that data and software licenses will need to be signed by legal authorities who are approved to sign such licenses on behalf of the given organisation [4].

Information concerning the specification of the FRGC database is obtained from [2]. For the FRGC database 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 (two or three studio lights). Facial expressions are neutral or smiling. The uncontrolled images were taken in varying illumination conditions. Each set of uncontrolled images contains two expressions, smiling or neutral. The 3D images were taken under controlled illumination conditions appropriate for the Vivid 900/910 sensor which is a structured light sensor that takes a 640×480 range sampled and registered colour image [2]. Subjects stood or were seated approximately 1.5 meters from the sensor.

The database consists of two sets: training and validation. The training set is also divided into two parts which are a large still training set and a 3D training set. There is controlled and uncontrolled illumination in the database. There is also 3D data in the database, which is the reason of choosing the database for our 3D face system.

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 4,007 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.

3.2 System

Many criteria can be adopted to compare existing 3D face recognition algorithms by taking into account the type of problems they address or their intrinsic properties; such as robustness to expressions or sensitiveness to size variation. For example, approaches exploiting a curvature-based presentation cannot distinguish between two faces of a similar shape, but can differentiate between faces of different size. In order to overcome this problem, some methods are based on point-to-point comparison or on volume approximation.

3.2.1 Existing systems

We analyse the different approaches under 3 main categories: 2D image based, 3D image based and multi-modal.

2D
The 2D-based methods mainly work on 2D images while being supported by some 3D data. A first example can be given as the 3D Morphable Models by Blanz and Vetter [5] where facial variations are synthesised by a morphable model which is a parametric model based on a vector space representation of faces. Using the proposed method a recognition rate of 95% on the CMUPIE dataset and 95.9% on the FERET dataset is reported. Another interesting approach is given by Lu et al. [6] where a 3D model is used to generate various 2D facial images. They performed experiments on a dataset of 10 subjects building 22 synthesized images per subject with different poses, facial expressions and illuminations. The method achieves a recognition rate of 85%, outperforming methods based on principal component analysis (PCA) on the same dataset.

3D
Approaches based on 3D data commonly encounter the problem of misalignment. One possible solution is to use a morphable model for 3D acquisition from a profile and a frontal view image [7]. In another approach an Iterative Closest Point (ICP) algorithm is often used to align facial surfaces. In [8] ICP is utilised to establish correspondences between 3D faces which are then compared by using GMM. A recognition rate of 97.33% is reported on the 3D RMA database. ICP-based methods treat the 3D shape of the face as a rigid object. Segmentation processes are proposed to treat the face recognition problem as a non-rigid object recognition problem to improve robustness to variations in facial expression. Chua et al. [9] exploit 'rigid' facial regions such as the nose, eye sockets and forehead by using a Point Signature two-by-two comparison which achieves a recognition rate of 100% on a dataset of 6 subjects and 4 facial expressions. In another approach, 3D facial data is treated as a cloud of points and PCA is applied to determine new axes that best summarise the variance across the vertices [10]. A recognition rate of 100% is claimed to be reached on a dataset of 222 range images of 37 subjects with different facial expressions.

Multi-modal
Multi-modal methods are based on both 2D and 3D data. In [11], PCA is separately performed on intensity and range images and experiments conducted on a dataset of 275 subjects. A recognition rate of 89.5% is reported for the intensity images, 92.8% for range images and 98.8% for the combined solution. Papatheodorou and Rueckert [12] proposed a 4D registration method based on ICP by adding textural information.

3.2.2 3D Face system

The 3D face recognition system to be used for the baseline evaluations was developed in the Multimedia Image Group, EURECOM. In summary it introduces a sparser representation for the dense 3D facial scans and hence makes the comparison step much easier for recognition.

In order to remove the 'common' face shape information a generic face is first warped using the Thin Plate Spline (TPS) method for each 3D scan. 15 fiducial points are assumed to be available for each face. Before starting the warping step the generic face is aligned and scaled to each face firstly based on these 15 points only. Afterwards it is coarsely warped to make the two surfaces as close as possible. 141 more point-pairs are obtained assuming that the two surfaces are in a sufficiently good alignment and the correspondences are found as the closest vertices. Finally, the generic face is warped based on total 156 point-pairs and, as a result, each 3D face model can be represented with the 3D vector of size 156×3 which is obtained from the warping parameters in x, y and z directions for each control point.

In order to measure the similarity (distance is more appropriate for our case) between facial surfaces, the angle between the two warping vectors and the difference between their magnitudes and angles for each point are calculated. This results in two distance vectors of size 156×1 for each compared face pair. A weighted sum of their central tendency is utilised for matching which is based on the nearest neighbour approach.

4 Multi-spectral face biometrics

In order to circumvent nuisance factors such as pose, illumination, occlusion, and facial expression, all of which are common in realistic scenarios, the research community has proposed not only the use of 3D face images, but also the use of face images acquired from the non-visible spectrum. 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 images with different characteristics. The millimetre-waves spectral range has also been proposed but not so much research work has been developed with images acquired at such frequencies.

4.1 Database

Currently, several face databases have been captured at various spectral ranges within the visible (VIS) and infrared (IR) bands. Some of them are presented bellow together with the HFB database.

4.1.1 Existing databases

EQUINOX HID (Human Identification at a Distance)
The EQUINOX HID database [95] was collected by Equinox Corporation under DARPA's HumanID program. It contains images in the following modalities: co-registered broadband-visible/LWIR (Long Wave Infrared: 8-12 microns), MWIR (Medium Wave Infrared: 3-5 microns), SWIR (Short Wave Infrared: 0.9-1.7 microns). The database consists of over 25000 frames from 91 distinct subjects. Unfortunately this database is no longer available.

The University of Notre Dame Database
This database [96, 97] has a large collection of visible and thermal facial images acquired with a time-gap. It consists of 2294 images acquired from 63 subjects during 9 different sessions under specific lighting (a central light turned off, all lights on) and facial expression conditions (neutral expression, smiling expression). The number of users in this case is quite small.

UH (University of Houston) Database
This database [97] contains thermal facial images of varying expressions and poses from 300 subjects. The images were captured using a high quality Mid-Wave Infra-Red (MWIR) camera. In spite of the high number of subjects, this database presents an important disadvantage: it contains images acquired at only one spectral range.

WVUM (West Virgina University Multi-spectral) Face Database
This database [98] consists of VIS and SWIR face images of 50 subjects. Ten types of image were acquired from each subject: one at VIS and nine at IR spectrum (one with no

filter and eight using band pass filters at different wavelengths). For the SWIR part of the database the face images were acquired with different poses (frontal, left and right). This database would be a good candidate if it contained more subjects.

PolyU NIR Face Database

The Hong Kong Polytechnic University Near Infrared Database [99] contains NIR normal face images and images with expression, pose and scaling variation from 335 subjects. For multi-spectral face recognition research images acquired from at least two spectral bands are needed, however this database only contains images acquired at NIR.

PolyU-HSFD Face Database

The Hong Kong Polytechnic University Hyper-spectral Face Database [100] includes 300 hyper-spectral image cubes from 25 volunteers with an age range from 21 to 33. For each individual, several sessions were collected. Each session consists of three hyper-spectral cubes: frontal, right and left views with neutral-expression. The spectral range is from 400nm to 720nm with a step length of 10nm, producing 33 bands in all. Again the number of users is very small.

All the previous face databases include only one or two types of images acquired from IR and/or VIS spectrums. Furthermore, the number of subjects in most of the presented databases is small. The CASIA HFB (Heterogeneous Face Biometrics) database has a considerable amount of users (100) and, in addition to IR and VIS face images, it contains 3D images, which are a topographic map of the face. These different image types are said to be heterogeneous because their formation principles are different. This heterogeneity enables the design of a robust face biometric recognition system, offering new challenges at the same time.

4.1.2 The HFB database

To obtain the HFB dataset one should follow the instructions from the Center for Biometrics and Security Research website[3]. Specifications of version 1 of the HFB database are obtained from [101].

Images corresponding to the VIS spectrum were acquired using a Canon A640 Camera with an image resolution of 640×480. Four frontal images with neutral and smile expression (or with or without glasses) at two different distances were captured.

A home-made device was used to acquire the 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. Again, the images were captured with different facial expressions and at different distances, having an image resolution of 640 × 480.

The 3D images were obtained using a Minolta Vivid 910 laser scanner, which provides

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

the depth (z) of every point in the plane (x, y) of the face. Optimal environmental conditions were set in order to obtain good quality 3D images (black background, no hair on the face, etc.).

The HFB data corpus contains images from 100 subjects: 57 males and 43 females. Four images at VIS, four at NIR and 1 or 2 3D images were acquired per subject. For the 3D faces there are two face images per subject for 92 subjects and only one per subject for the remaining 8 subjects. This results in a total of 992 images from 100 subjects.

The first release of version 1 of the HFB database includes: (i) raw images in JPEG format for VIS and NIR and wrl format for 3D images, (ii) the processed 3D faces, (iii) the eye coordinates of the three types of images manually labelled, and (iv) cropped versions of the raw images, in two sizes: 32×32 and 128×128 (this was done base on the eye coordinates).

4.2 System

Different systems have been developed in multi-spectral face recognition depending on the particular application considered. One of the most popular cases is the matching between different spectral bands or modalities, which is referred to as heterogeneous face biometrics (HFB), such as between visual face images and NIR, or between 3D and NIR [101].

4.2.1 Existing systems

The main characteristic of systems based on HFB is the fact that direct appearance based matching is no longer appropriate to solve the problem, and a method to normalise the images from the different spectral bands or modalities is needed. Different approaches have been proposed. A system based on canonical correlation analysis (CCA) was proposed in [102] for NIR-VIS face image matching. Then, dimensionality was reduced using principal component analysis (PCA) and linear discriminant analysis (LDA). Recognition results obtained with LDA-CCA were much better than those without applying CCA.

In [103], using the data from the Multiple Biometric Grand Challenge (MBGC) portal challenge where NIR face videos are used as the probe set and VIS images in the target set, Laplacian of Gaussian (LoG) features were extracted and a non-learning based method was proposed. Results were compared to those obtained by other methods such as PCA or LDA in combination with CCA obtaining very significant improvements in performance.

In [104], Difference-of-Gaussian (DoG) pre-processing filtering was adopted to obtain a normalised appearance for all heterogeneous faces. Then, multi-scale block local binary pattern (LBP) was applied to encode the local image structures in the transformed domain, and further learn the most discriminant local features for recognition. Experiments show that the proposed method significantly outperforms existing ones in matching between VIS and NIR face images. The verification rate of this proposed method at 0.1% false alarm rate (FAR) is 67.5%, and 87.5% at 1% FAR.

4.2.2 The multi-spectral face system

From the existing systems for multi-spectral face recognition, the one to be used for the baseline evaluations is similar to the system described in [104]. DoG filtering and LBP is applied to normalise the face images from different spectral bands.

In the pre-processing stage, Difference-of-Gaussian (DoG) is applied to the raw images (VIS, NIR and 3D) to normalise the appearance. Also, DoG filtering helps to reduce illumination variation, image noise and aliasing, while preserving enough details for recognition. Then, the Local Binary Pattern (LBP) approach is used to learn discriminant local structures for further recognition. 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 in order to construct an open-set test protocol. Linear Discriminant Analysis (LDA) is applied on the training set to construct a universal subspace. This transformation is applied to the images in the test set before matching. Thanks to the richness of the database, different matching configurations such as VIS-NIR, VIS-3D or NIR-3D can be studied.

5 Iris biometrics

With fast development of iris image acquisition technology, iris recognition is expected to play a strong role in the future of biometric technology, with wide application areas in national ID cards, banking, e-commerce, welfare distribution, biometric passports and forensics, 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.

5.1 Database

Currently, several iris databases have been captured. Some of them are presented below together with the CASIAv3 database.

5.1.1 Existing databases

MBGC (Multiple Biometric Grand Challenge) is a project with the goal to investigate, test and improve performance of face and iris recognition technology [94]. Among the technology development areas within the MBGC, the MBGC portal challenge database provides Near Infrared (NIR) iris still images and videos.

The iris image datasets used in the Iris Challenge Evaluations (ICE) in 2005 and 2006 [105] were acquired at the University of Notre Dame and contains iris images of a wide range of quality, including some off-axis images. Both databases are currently available. One unusual aspect of these images is that the intensity values are automatically contrast-stretched by the LG 2200 to use 171 grey levels between 0 and 255.

The Multimedia University has released two iris databases. The MMU1 iris database [106] is comprised of a total number of 450 iris images which were collected using a LG IrisAccess2200 semi-automated camera and operating at the range of 7-25 cm from the user to the camera. On the other hand, the MMU2 iris database is comprised of 995 iris images collected using a Panasonic BM-ET100US Authenticam with an operating range of 47-53 cm away from the user. These iris images are contributed by 100 volunteers of different age and nationality. Each of them contributes 5 iris images for each eye.

The UBIRIS.v1 database (2004) [107] is comprised of 1877 images collected from 241 persons in two distinct sessions. This database incorporates images with several noise factors, simulating less constrained image acquisition environments. This enables the evaluation of the robustness of iris recognition methods. A new version of this database, UBIRIS.v2 (2006) [108], was collected under non-constrained conditions (at-a-distance, on-the-move and on the visible wavelength), with corresponding more realistic noise factors. The major purpose of the UBIRIS.v2 database is to constitute a new tool to evaluate the feasibility of visible wavelength iris recognition under far-from-ideal imaging conditions. In this scope, the various types of non-ideal images, imaging distances, subject perspectives and lighting conditions existing on this database could be of strong utility in the specification of the visible wavelength iris recognition feasibility and constraints.

The BATH iris database [109] was designed to obtain very high quality iris images. The initial objective was to capture 20 images from each eye of 800 subjects. The commercially available database is now twice this size. A majority of the database is comprised of students from 100 different countries and staff from the University of Bath. The images are of a very high quality taken with a professional Machine Vision Camera with infrared illumination and a consistent image capture setup.

The BioSecure database [110] is a multi-modal database which includes data from face, voice, iris, fingerprint, hand and signature modalities, within the framework of three datasets corresponding to real multi-modal, multi-session and multi-environment situations. Moreover, in order to increase the representativeness of the database, BioSecure participants agreed to collect the above mentioned data in a variety of sites (11 at the end) involving a number of countries spread over Europe. The iris database contains data from 210 persons in two sessions in which two images where taken per eye. Further details of the BioSecure database are given in Sections 6 and 11.

5.1.2 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 publically available and can be obtained from the Center for Biometrics and Security Research website[4]. It will be used for baseline experiments.

CASIA-IrisV3 contains a total of 22,034 iris images from more than 700 subjects. All iris images are 8 bit grey-level JPEG files, collected under near infrared illumination. Almost all subjects are Chinese except a few in CASIA-Iris-Interval. The three data sets were collected at different times in which CASIA-Iris-Interval and CASIA-Iris-Lamp have a small overlap in subjects.

Iris images of CASIA-Iris-Interval were captured with our self-developed close-up iris camera. The most compelling feature of our iris camera is that we have designed a circular NIR LED array, with suitable luminous flux for iris imaging. Because of this novel design, our iris camera can capture very clear iris images. CASIA-Iris-Interval is well-suited to study the detailed textual features of iris images.

CASIA-Iris-Lamp was collected using a hand-held iris sensor produced by OKI. A lamp was turned on/off close to the subject to introduce more intra-class variations. Elastic deformation of iris texture due to pupil expansion and contraction under different illumination conditions is one of the most common and challenging issues in iris recognition. So CASIA-Iris-Lamp is good for studying problems of non-linear iris normalisation and robust iris feature representation.

CASIA-Iris-Twins contains iris images of 100 pairs of twins which were collected during

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

the Annual Twins Festival in Beijing using OKI's IRISPASS-h camera. The iris is usually regarded as a kind of phenotypic biometric characteristics and, as such, even twins should have their unique iris patterns. It is thus interesting to study the similarity and dissimilarity between iris images of twins.

The unique filename of each image in CASIA-IrisV3 denotes some useful properties associated with the image such as subset category, left/right/double, subject ID, class ID, image ID etc.

5.2 System

Iris recognition has become a popular research topic in recent years. Due to its reliability and nearly perfect recognition rates, iris recognition is used in high security areas. A literature review of the most prominent developed algorithms is showed below.

5.2.1 Existing systems

Looking at different approaches to analysing the texture of the iris has perhaps been the most popular area of research in iris biometrics. One body of work effectively looks at using something other than a Gabor filter to produce a binary representation similar to Daugman's iris code. Many different filters have been suggested for use in feature extraction. Sun et al. [111] use a Gaussian filter. Here the gradient vector field of an iris image is convolved with a Gaussian filter, yielding a local orientation at each pixel in the unwrapped template. They quantise the angle into six bins. This method was tested using an internal CASIA dataset of 2,255 images obtaining an overall recognition rate of 100%. Another interesting approach with very good results is given by Monro et al. [112] where the discrete cosine transform is used for feature extraction. They apply the DCT to overlapping rectangular image patches rotated 45 degrees from the radial axis. The differences between the DCT coeffcients of adjacent patch vectors are then calculated and a binary code is generated from their zero crossings. In order to increase the speed of the matching, the three most discriminating binarized DCT coeffcients are kept, and the remaining coeffcients are discarded.

Another body of work looks at using different types of filters to represent the iris texture with a real-valued feature vector. Ma et al. [113] use a variant of the Gabor filter at two scales to analyse the iris texture. They use Fisher's linear discriminant to reduce the original 1,536 features from the Gabor filters to a feature vector of size 200. Their experimental results show that the proposed method performs nearly as well as their implementation of Daugman's algorithm, and is a statistically significant improvement over other algorithms they use for comparison. The experimental results shown a correct recognition rate of 94.33% across the 2245 images of CASIA database.

A smaller body of work looks at combinations of these two general categories of approach. Here, it is important note the work of Hollingsworth et al. [114] where they acquire multiple iris codes from the same eye and evaluate which bits are the most consistent bits in the iris code. They suggest masking the inconsistent bits in the iris code to improve performance reaching equal error rates (EERs) of 0.068% under different subsets selected from the ICE database using still images and video recordings.

5.2.2 The iris recognition system

The iris recognition system to be used for the baseline evaluations was developed by L.M̃asek [115, 116]. The system basically inputs an eye image, and outputs a binary biometric template and it is based on the Hamming distance between templates. The system consists of the following sequence of steps that are described next: 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. Iris boundaries are modelled as two concentric circles. The range of search radius values is set manually. A maximum value is also imposed to the distance between the circle's centre. An eyelids and eyelashes removal step is also performed in the system. 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 normalisation of iris regions, a technique based on Daugman's rubber sheet model is employed. The centre of the pupil is considered as the reference point, and radial vectors pass through the iris region. Since the pupil can be non-concentric to the 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
Feature encoding is implemented by convolving the normalised iris pattern with ID Log-Gabor wavelets. The 2D normalised pattern is broken up into a number of ID signals, and then these ID signals are convolved with ID Gabor wavelets. The rows of the 2D normalised pattern are taken as the ID signal, each row corresponds to a circular ring on the iris region. It uses the angular direction since maximum independence occurs in this direction [116]. The output of filtering is then phase quantised to four levels using the Daugman method [117], with each filtering producing two bits of data. The output of phase quantisation is a grey code, so that when going from one quadrant to another, only 1 bit changes. This will minimise the number of bits disagreeing, if say two intra-class patterns are slightly misaligned, and thus will provide more accurate recognition [116]. 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, since bit-wise comparisons are necessary. The Hamming distance employed incorporates noise masking, so that only significant bits are used in calculating the Hamming distance between two iris templates. 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 [117]. 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.

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. 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. 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. The application of automated fingerprint biometrics in forensics with latent data is out of the scope of the project.

6.1 Database

Due to the large variety in existing sensors of different technologies, the acquisition process for fingerprints is relatively cheap, easy and fast. This has resulted, in the last few years, in the collection of multiple fingerprint datasets, most of them comprised in larger multi-modal databases containing as well other biometric traits.

6.1.1 Existing databases

As mentioned above, most of the fingerprint corpora currently available are part of larger multi-modal databases which are the result of collaborative efforts in recent research projects. Examples of these joint efforts include previous European projects such as BioSec [20] or the BioSecure Network of Excellence [21] in addition to national projects such as the Spanish Ministerio de Ciencia y Tecnología (MCYT) [25] or BioSecureID [19] databases.

Apart from the fingerprint datasets included in multi-modal databases, other efforts have been directly conducted to the acquisition of just fingerprint data, from which we highlight the datasets used in the series of Fingerprint Verification Competitions (FVCs) in 2000, 2002, 2004, and 2006.

6.1.2 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. This sub-corpus presents several characteristics which do not possess the other fingerprint datasets available nowadays, and which make it specially suited for the performance and security objectives defined within the TABULA RASA project:

Users	Fingers	Hands	Samples/finger	Total Samples/Session
667	Thumb/Index/Middle (3)	Rigth/Left (2)	2	$667 \times 3 \times 2 \times 2 = 8004$

Table 2: A summary of the fingerprint data characteristics in BMDB.

- Size: the BMDB dataset comprises fingerprint data from over 650 users.
- Compatibility: it is fully compatible, in terms of sensors used and protocols followed, with other large, multi-modal databases such as BioSec [20] or BioSecureID [19].
- 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 or face (also relevant for TABULA RASA).
- 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 (around 45%-55% of women/men), and also a balanced age distribution: about 40% of the subjects present in the database are between 18 and 25 years old, 20-25% are between 25 and 35, 20% are between 35-50 years old, and the remaining 15-20% are above 50.

The BioSecure Multi-modal DB is publicly available through the BioSecure Foundation[5]. 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 the fingerprint sub-corpus comprised within DS2 and captured with the optical sensor Biometrika FX2000. This sub-corpus was captured in two separate acquisition sessions. The data available for each session is summarised in Table 2. The data is stored as bmp images of 296×560 pixels captured at a resolution of 569 dpi. Further details are presented in Section 11.

6.2 Systems

As the market leading biometric many different fingerprint recognition systems have been proposed in the literature. From a general point of view all of them may be included in one of these three categories: i) correlation-based, ii) minutiae-based, or iii) based on features of the ridge pattern.

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

6.2.1 Existing systems

Correlation-based methods
These systems compute and maximise the cross correlation between pixels of the stored and input samples. As an effect of the displacement and the rotation that exists between samples of the same fingerprint, the similarity cannot be computed by simply evaluating the correlation but it has to be maximised for different vertical and horizontal offsets of the fingerprint, and for different rotations. These operations entail a huge computational cost and, except for very good quality samples, these methods do not present in general comparable results to those obtained with the other two types of approaches. Different algorithms have been developed that are able to substantially decrease the computational cost of the matching process by computing the local correlation at specific areas of the fingerprint such as the core, or close to very good quality minutiae points. However, none of these techniques provide a clear performance improvement over the general method.

Minutiae-based methods
This is the most popular and widely used technique as it presents the best performance results, and is the basis of the fingerprint comparison made by fingerprint examiners. Minutiae are extracted from the two fingerprints and stored as sets of points in the two-dimensional plane. Minutiae-based matching essentially consists of finding the alignment between the template and the input minutiae sets that result in the maximum number of minutiae pairings.

Ridge feature-based methods
Minutiae extraction is difficult in very low-quality fingerprint images. However, whereas other features of the fingerprint ridge pattern (e.g. local orientation and frequency, ridge shape, texture information) may be extracted more reliably than minutiae, their distinctiveness is generally lower. The approaches belonging to this family compare fingerprints in term of features extracted from the ridge pattern. Two main reasons induced researchers to look for other fingerprint discriminative features beyond minutiae:

- Reliably extracting minutiae from poor quality fingerprints is very difficult. Although minutiae may carry most of the fingerprint discriminatory information, they do not always constitute the best trade-off between accuracy and robustness.

- Additional features may be used in conjunction with minutiae (and not as an alternative) to increase system accuracy and robustness.

The more commonly used alternative features are the size and shape of the fingerprint, number, type and position of singularities, spatial and geometrical attributes of the ridge lines, shape features, sweat pores, or global and local texture information.

6.2.2 System 1: The MorphoKit system

The system used for the baseline evaluation will be the MorphoKit. This software development kit (SDK) developed by Morpho is including Morpho proprietary algorithms for generating minutiae templates and 1:1 or 1:N matching. A detailed description of the underlying algorithm cannot be disclosed and the following represents a brief specification of the SDK.

MorphoKit is a fingerprint acquisition and processing SDK. It is primarily intended to be used in the development of biometric application by private companies outside SagemDS, and will not include any of the specific features required for AFIS development (classification, segmentation of slap images, flatbed scanner management ...). It is designed to be used in the development of small to medium scale biometric applications: it will included the best coding and 1:1 matching technology available today, but a limited version of our 1:many matching technology, limited to small databases (3000 records for the standard product, no more than 100,000 records in any case), without the matching speed improvements specific to AFIS products. It is not designed for AFIS enrolment/identification or forensic applications and will not support 1000 dpi images or multi-finger images. 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
- Identification: 1:many matching of reference template against a memory template database
- 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 self-describing proprietary binary format. From the user's point of view, it is just a binary buffer of variable length (1300 bytes on average). Authentication and identification functions will only accept templates in CFV format. Matching templates is providing a score that can be compared to a given threshold for the decision: MATCH/NO MATCH. Reference thresholds are provided to meet a specific performance target. In other words, a target false accept rate (FAR) can be reached and guaranteed with a given fixed threshold.

6.2.3 System 2: The NFIS2 system

The minutiae-based NIST Fingerprint Image Software 2 (NFIS2) [26] 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.

NFIS2 contains software technology, developed for the Federal Bureau of Investigation (FBI), designed to facilitate and support the automated manipulation and processing of fingerprint images. Source code for over 50 different utilities or packages and an extensive User's Guide are distributed on CD-ROM which is available free of charge[6].

From the 50 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. The architecture of MINDTCT can be divided in the following stages: 1) Generation of image quality map; 2) Binarization; 3) Minutiae detection; 4) Removal of false minutiae, including islands, lakes, holes, minutiae in regions of poor image quality, side minutiae, hooks, overlaps, minutiae that are too wide, and minutiae that are too narrow (pores); 5) Counting of ridges between a minutia point and its nearest neighbours; 6) Minutiae quality assessment.

Because of the variation of image quality within a fingerprint, MINDTCT analyses the image and determines areas that are degraded. Several characteristics are measured, including regions of low contrast, incoherent ridge flow, and high curvature. These three conditions represent unstable areas in the image where minutiae detection is unreliable, and together they are used to represent levels of quality in the image. The image quality map of stage 1 is generated integrating these three characteristics. Images are divided into non-overlapping blocks, where one out of five levels of quality is assigned to each block.

The minutiae detection step scans the binary image of the fingerprint, identifying local pixel patterns that indicate the ending or splitting of a ridge. A set of minutia patterns is used to detect candidate minutia points. Subsequently, false minutiae are removed and the remaining candidates are considered as the true minutiae of the image. Fingerprint minutiae matchers often use other information in addition to just the points themselves. Apart from minutia's position, direction, and type, MINDTCT computes ridge counts between a minutia point and each of its nearest neighbours. In the last stage, MINDTCT assigns a quality/reliability measure to each detected minutia point. Even after performing the removal stage, false minutiae potentially remain in the list. Two factors are combined to produce a quality measure for each detected minutia point. The first factor is taken directly from the location of the minutia point within the quality map generated in stage 1. The second factor is based on simple pixel intensity statistics (mean and standard deviation) within the immediate neighbourhood of the minutia point. A high quality region within a fingerprint image is expected to have significant contrast that will cover the full grey-scale spectrum.

[6] `http://fingerprint.nist.gov/NFIS/`

BOZORTH3

The BOZORTH3 matching algorithm computes a match score between the minutiae from any 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. For fingerprint matching, compatibility between minutiae pairs of the two images are assessed by comparing the following measures: i) distance between the two minutiae and ii) angle between each minutia's orientation and the intervening line between both minutiae.

7 Voice biometrics

With the growth in telecommunications and vast related research effort voice-based authentication has emerged over the last decade as a popular and viable biometric. Speaker recognition is generally the preferred or even only mode of remote verification over the telephone, for example. Speaker recognition also has obvious utility in multi-modal biometric systems where it is commonly used with face recognition. Speaker verification will be assessed within the scope of the TABULA RASA project both as a single biometric and combined with face verification. Both are vulnerable to various kinds of spoofing attacks from impersonation, replay attacks, voice morphing and synthesised speech etc.

7.1 Database

As with any biometric, speech databases suited to the speaker verification task should in general have a large, representative number of speakers. Since speech characteristics from the same person can vary significantly from one recording to another there is a requirement for multi-session data which should reflect differences in acoustic characteristics related not only to the speaker, but also to differing recording conditions and microphones. It has been suggested that collection over a period of three months [42] is a minimum in order to reliably capture variations in health and fatigue for example. Since the information in a speech signal is contained in its variation across time, i.e. it is a dynamic signal, speaker verification performance also varies significantly depending on the quantity of data used both for training and testing. Especially since the countermeasures to be developed later in the TABULA RASA project may require large amounts of data in order to model suprasegmental features, the provision for an evaluation condition containing large amounts of training data is essential.

7.1.1 Existing databases

Databases such as TIMIT [27], Aurora [28] and Switchboard [29] are all used widely for speech technology research. Though these databases have also been used for speaker recognition evaluations to some extent, they are designed primarily for speech recognition research. In addition these corpora are somewhat limited in the variety of microphone and recording conditions and also well defined development, evaluations, training and testing subsets targeted for speaker recognition experimentation. Existing corpora, such as the CHAINS [30], YOHO [31] and CSLU [32] datasets are specific to speaker recognition, but have a limited number of speakers. The EVALITA [33] evaluations provided for a dedicated speaker recognition task in 2009 but did not feature in 2007 and is not included in the evaluation plan for 2011.

The Speaker Recognition Evaluation (SRE) [37] datasets collected by the National Institute of Standards and Technology (NIST) are presently the de facto standard evaluation platform for speaker recognition research and are the only realistic means of gauging the state-of-the-art. The recent 2010 evaluations attracted over 50 research institutes from 5 continents. Since they provide for large, multi-session datasets with different evaluation conditions and since they facilitate comparisons to the existing state-of-the-art, the NIST SRE datasets will be used for speaker verification work within the TABULA RASA project. Since not all NIST datasets are publicly available we have decided to restrict those used in TABULA RASA to datasets that are or will soon be publicly available through the Linguistic Data Consortium[7] (LDC). We also note that the same datasets have been used previously in related work [63].

Finally we refer to two multi-modal datasets that include a speech component. The three datasets from the BioSecure Multi-modal Evaluation Campaign (BMEC) [34] contain seven different modes including speech. They are described in Section 11. The MOBIO dataset [35] contains both face and voice modalities and will be used in TABULA RASA for 2D-face and voice. For comparative purposes baseline mono-modal speaker recognition experiments will also be conducted on the MOBIO dataset which is described in Section 11.

7.1.2 The NIST SRE datasets

The 2003, 2004, 2006 and 2008 NIST SRE datasets will all be publicly available later this year through the LDC. They contain several hundreds of hours of speech data collected over the telephone including some calls made using mobile telephones. Further details of each dataset are available from the LDC website with additional information available from the NIST SRE website[8]. Each evaluation involves one compulsory, 'core' experiment and several other optional experiments. The differences between each experiment or condition entail mostly different quantities of training and/or test data and possibly varying channel conditions. Training and testing protocols are defined and allow for different systems and technologies to be readily and meaningfully compared according to standard experimental and evaluation protocols and metrics. We note, however, that this may not be entirely the case for TABULA RASA work since our system may not necessarily be optimised for the same operating conditions (costs).

A typical speaker recognition system requires an independent development set in addition to independent auxiliary data which is needed for background model training and the learning of normalisation strategies. This data typically comes from other NIST datasets, such as the 2003 and 2004 datasets, and will be the case for all TABULA RASA work. All NIST SRE datasets have a very similar specification and in the following we outline specifically the 2008 NIST SRE dataset.

The 2008 evaluation dataset is composed of data from the Mixer [36] corpora and contains a total of 13 different conditions which correspond to 6 different training and 4 different test scenarios. The main differences between them relates to the quantity or duration of speech data in addition to other differences in microphone characteristics, recording environments, language and the level of vocal effort, for example. The datasets are composed from subsets of the Mixer 3, 4 and 5 corpora from the LDC. They contain speech conversations recorded over the telephone, from multiple microphones within a

[7] `http://www.ldc.upenn.edu/`

[8] `http://www.itl.nist.gov/iad/mig/tests/sre`

		Test			
		10-sec	short3	long	summed
Train	10-sec	optional			
	short2	optional	**core**		optional
	3conv		optional		optional
	8conv	optional	optional		optional
	long		optional	optional	
	3summed		optional		optional

Table 3: Train/test condition matrix for the NIST 2008 speaker recognition evaluation, reproduced from [43]. The 'short2-short3' condition was the core condition in 2008.

room and from interview sessions with both conversational and read speech. The standard experimental protocol defines various rules for participation in addition to lists of trials (e.g. file lists for training and testing) for each evaluation condition.

Basic evaluation rules relate to the independence of trial decisions, permitted normalisation procedures, human interaction and the use of additional data, such as speech transcripts, etc. Full details can be found in the evaluation plan [43] and official evaluation workshop presentations [44]. The 13 different evaluation conditions are illustrated in Table 3 which highlights the core 'short2-short3' condition involving 2 or 3 different recordings for training and testing respectively and where each recording contains approximately 3 or 5 minutes of speech data. Initial plans are to use the core condition for TABULA RASA research, in addition to an appropriate extended data task, but will be subject to revision according to spoofing and countermeasure technology to be developed later in the project.

For each required trial, participants are required to determine whether or not the given target speaker is active in the given test segment. This involves determining an appropriate likelihood score and, according to an empirically optimised threshold, a positive or negative decision. Although it is not the only metric the standard, core evaluation metric is defined as follows:

$$C_{Norm} = \frac{C_{Miss} \times P_{Miss/Target} \times P_{Target} + C_{FA} \times P_{FA/NonTarget} \times P_{NonTarget}}{C_{Default}} \qquad (2)$$

where the cost of a miss and of a false alarm (FA) are 10 and 1 respectively, where the probability of a target and non-target are 0.01 and 0.99 respectively and where the normalisation factor $C_{default} = 0.1$ is defined in order that a system which always returns a negative decision obtains a score of $C_{Norm} = 1$. While the C_{Norm} metric is the default, dynamic performance, including a comparison of minimum and actual costs (i.e. with regard to optimised and actual thresholds), are compared according to standard detection error trade-off (DET) curves [45].

7.2 System

Given the complexity of standard NIST speaker recognition datasets it is desirable that the system adopted is well-adapted and suited to running such evaluations. Although initial experiments will involve only a small number of trials, later experiments will be automated and involve many thousands of trials. Computational efficiency is thus also a requirement. Speaker recognition experiments will also be performed in a multi-modal setting (see Section 11) and thus it is also sensible that the system may be used in conjunction, or fused with a face recognition system. In the following we review some existing tools that are appropriate in this case and then describe in more detail the system adopted for the TABULA RASA project.

7.2.1 Existing systems

Speaker recognition systems have advanced rapidly over the last few decades and there exist some useful software packages and libraries that can be used to build state-of-the-art speaker recognition systems with relative ease.

SPro[9], the open-source speech signal processing toolkit, provides for highly configurable feature extraction. The Hidden Markov Model Toolkit (HTK)[10] and the Hidden Markov Model Synthesis Toolkit (HTS)[11] provide a set of tools for building statistical speaker models and can also be used for feature extraction. Matlab[12] from Mathworks Inc. has various toolkits for statistical pattern recognition and is an excellent tool to prototype quickly a speaker recognition system and to develop advanced algorithms. Octave[13], its open source equivalent, also provides powerful features. The ALIZE/Mistral platform[14] [38, 39] is a library for biometric authentication and provides a comprehensive set of functions related to the task of statistical speaker recognition. LIA-RAL[15] is a set of tools for speaker recognition and is built using the ALIZE/Mistral library. libsvm[16] is a library which provides for support vector classification and has been integrated into LIA-RAL. The Torch toolkit[17] also has robust implementation of Support vector machine based classifiers. Finally FoCal[18], a set of Matlab functions for the fusion and calibration of multiple classifiers, has proven very popular in the speaker recognition community.

SPro, ALIZE, LIA-RAL and FoCal are arguably the most popular tools for speaker recognition. They are all open-source, are used in combination by many independent teams and have achieved state-of-the-art performance in the NIST speaker recognition

[9] http://www.gforge.inria.fr/projects/spro
[10] http://htk.eng.cam.ac.uk/
[11] http://hts.sp.nitech.ac.jp/
[12] http://www.mathworks.com/products/matlab/
[13] http://www.gnu.org/software/octave/
[14] http://www.lia.univ-avignon.fr/heberges/ALIZE/
[15] http://www.lia.univ-avignon.fr/heberges/ALIZE/LIA_RAL
[16] http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[17] http://www.idiap.ch/scientific-research/resources/torch
[18] http://sites.google.com/site/nikobrummer/focal

evaluations. Furthermore the ALIZE/MISTRAL toolkits have been used for voice transformation in order to demonstrate the threat from spoofing. This combination will be used for all TABULA RASA work.

7.2.2 The ALIZE speaker recognition system

The 'ALIZE' speaker recognition system is something of a misnomer since ALIZE is really a library, not a toolkit. Even so, the open-source LIA-RAL toolkit, which does provide a set of executable for speaker recognition, has inherited the name of the library on which it is based. TABULA RASA work in speaker recognition will be based upon the implementation described in [40].

While ALIZE has native support for most standard feature file formats, SPro is the most popular. SPro provides for both Mel [46] and linear scaled frequency cepstral coefficients in addition to linear prediction coefficients, static and dynamic features. Features generally encompass some channel characteristics which manifests as convolutional noise. Under conditions of mismatched training and testing these effects can lead to significant degradations in performance and some means of channel compensation generally prove beneficial. In the standard baseline setup this includes cepstral mean and variance normalisation. ALIZE also provides a comprehensive suite of different feature normalisation strategies including feature warping [47], feature mapping [48] and factor analysis eigenchannel compensation [49]. Given that the spoofing attacks to be considered in TABULA RASA involve recording and replaying at the sensor level it will be necessary to investigate the effect of channel compensation approaches since they may inadvertently assist spoofing attacks.

The standard approach to statistical speaker modelling is based on Gaussian mixture models (GMMs) [50] and is the approach adopted in ALIZE. First, a world model [52] or universal background model (UBM) is trained using expectation maximisation (EM) [53] and large amounts of data from a pool of background speakers. Due to the common lack of speaker-specific data, target speaker models are generally adapted from the UBM during enrolment through maximum a posteriori (MAP) adaptation [54]. Although all the parameters of the UBM can be adapted, the adaptation of the means only has been found to work well in practice [50] and is the approach in the largely standard baseline system.

Scores correspond to the log-likelihood ratio of the target model and the test segment, normalised with respect to the background model. Various score-level normalisation procedures are also generally employed with the most popular being test-normalisation (TNorm) [51]. TNorm is used to normalise the score with respect to a set of cohort, impostor speakers and generally leads to significant improvements in performance. Additional normalisation strategies include zero normalisation (ZNorm) and handset normalisation (HNorm) [56]. Final decision logic is based on a threshold which is empirically determined using a large, representative development set. False alarm and false rejection rates can be traded-off in the usual manner by varying the threshold.

The ALIZE framework also provides for more recent approaches which harness the power of SVMs and joint factor analysis (JFA). Support vector machines (SVMs) [57] have become a popular approach to pattern classification and speaker verification is no exception. The more recent SVM-based approaches such as the generalised linear discriminant sequence kernel (GLDS) [58] and the GMM super-vector linear kernel (GSL) [59] are capable of outperforming the standard GMM-based approach and are supported in ALIZE. The GSL approach is one example where the input to the SVM classifier comes from a conventional GMM and is formed from the concatenation of the GMM mean vectors into the so-called GMM super-vector [59]. Other approaches supported in ALIZE include nuisance attribute projection (NAP) [60] and joint factor analysis (JFA) [61].

Finally we note that the ALIZE system has been used in the past to investigate vulnerabilities to spoofing attacks coming from voice transformation [63]. This work, however, did not consider strictly sensor-level attacks. The ALIZE system and NIST speaker recognition datasets described above are not sufficient on their own to assess the spoofing threat at the sensor-level. Initial plans to consider specifically sensor-level attacks will involve the re-recording of speech data to simulate replay-style attacks and then an automated approach using artificial microphone/room/loudspeaker impulse responses. Whereas the former will involve only a small number of trials it will better reflect the practical scenario. The latter setup, while rather artificial, will involve many more trials and will provide for greater statistical significance. We will conduct a comparison of the two approaches to assess differences in recognition/spoofing performance of both approaches.

8 Gait biometrics

Biometric gait recognition refers to the recognition of people from the way they walk. This has recently become a topic of great interest in biometrics research. Compared to some other biometric modalities, gait is potentially beneficial since it can be acquired from a distance and it does not require contact or client cooperation. Furthermore, gait is potentially difficult to hide or spoof. To gain insight into the vulnerability of biometric gait recognition systems to spoofing attacks, the TABULA RASA project will consider USOU Gait Database and evaluate two baselines systems developed at the universities of Southampton and Oulu. The chosen database and systems are described below.

8.1 Database

There are many databases suitable for studying gait recognition. In the case of TABULA RASA a dataset is needed which has sufficient subjects and enough calibration information that novel spoofing and countermeasure techniques can be investigated using it.

8.1.1 Existing databases

The earliest databases comprised of only tens of subjects, sometimes wearing specified clothing. More recent databases include outdoor as well as indoor data with uncontrolled illumination and with variation in camera viewpoint. The two early gait databases, HiD (NIST, US) [70] and Soton (Southampton UK) [71], both resulting from the DARPA Human ID at a Distance programme have been accompanied by the newer CASIA database (CAS, China) [72]. These databases primarily afford sequences of 2D images.

More recently, Southampton has recorded a new database (The USOU gait database) using a new multi-view gait tunnel which has 8 synchronised cameras viewing a subject walking in a controlled environment [68]. The USOU gait database is one of the largest gait databases and crucially contains multiple views and detailed camera calibration information. This enables 3D reconstruction from the data, and as such provides valuable information that can be used for examining potential spoofing and countermeasure techniques. Because of this it will be used as the basis for this project.

8.1.2 The USOU gait database

The USOU Gait Database consists of 2705 separate recordings taken from 227 Subjects. Each recording consists of 8 synchronised video sequences of approximately 140 frames. Each subject is recorded walking through the tunnel at least 9 times. In addition, 36 of the subjects are recorded on two separate dates to enable the investigation of the effect of time and clothing variation on recognition performance. The database will be available for download soon[19]. Access requires the completing and signing of a Database Release Agreement.

[19] http://www.gait.ecs.soton.ac.uk/

The Database was produced using the University of Southampton Multi-Biometric Tunnel [68], a constrained environment that is designed with airports and other high throughput environments in mind. It is able to acquire a variety of non-contact biometrics in a non-intrusive manner. The system uses eight synchronised IEEE1394 cameras to capture gait and additional cameras to capture images from the face and one ear, as an individual walks through the tunnel.

Whilst a subject is inside the tunnel their gait is recorded by eight calibrated Point Grey Dragonfly cameras, allowing the reconstruction of 3D volumetric data. The gait cameras all have a resolution of 640×480 and capture at a rate of 30 FPS, they are connected together over an IEEE1394 network employing synchronisation units to ensure accurate timing between cameras.

Video is also captured of the subjects face and upper body using a high resolution 1600×1200 IEEE1394 camera, enabling face recognition. A 1600×1200 snapshot is taken of the subjects side of the head, which can be used for ear biometrics.

8.2 Systems

This section provides a summary of a number of the main systems that have been developed for gait recognition. This is followed by a description of the two systems that will be used to provide baseline performance measures.

8.2.1 Existing systems

Current state-of-the-art systems show that it is possible to recognise people using gait recognition by using silhouette or model based approaches [73]. Both approaches start by analysing video data to detect the walking subject. Silhouette-based approaches have enjoyed the most success. In particular, those which use the averaged silhouette have proved most popular [74]. Feature set selection has been deployed to identify which features of the silhouettes contribute most to recognition [69]. Early model-based approaches used pendular models to calculate the variation in thigh inclination, when walking [75]. This work has been extended to a unified approach which can model running or walking simultaneously [76]. More recently, these models have been employed in conjunction with vertex based location. This approach tracks and describes people by the motion of their joints. Such model based approaches have also been subject to feature set selection, which revealed that motion components can have greater discriminative ability than the structural components. A further area of research is the effect of the camera viewpoint relative to the subject's walking direction. By assuming that the human is a solid object, walking in a periodic fashion along a linear path (for two gait periods), recognition can be achieved which is invariant to the direction of path relative to the camera [77]. The covariate factors are of equal concern and there is interest in the degree that external factors impede recognition by gait. Analysis of covariate factors including clothing and footwear has shown that wearing a trench coat or the wearing of flip flops can significantly affect recognition performance [78]. This is not surprising in that the wearing of clothes that obscure the whole body will naturally obviate any gait biometric. In addition, the walking style consistent with flip flops is known to differ from that when wearing normal shoes, though no research has parametrised this effect. As such, the current state-of-art has been to investigate the nature of the within-class variation, and the between-class variation in gait as a biometric. It is interesting that the earliest approaches achieved recognition rates exceeding 90% and this is matched by the most recent approaches on databases extending to 300 subjects. Much of the earlier work was conducted on data acquired using controlled conditions whereas later recognition has used data acquired outdoors, which has resulted in slightly lower recognition performance.

8.2.2 System 1: USOU gait recognition system

To provide a baseline recognition performance USOU will provide an average silhouette based gait recognition system. The system uses shape from silhouette 3d reconstruction to synthesise a profile silhouette sequence from which an average silhouette is constructed.

3D volumetric data is used to synthesise silhouettes from a fixed viewpoint relative to the subject. The resulting silhouettes are then passed to a standard 2D gait analysis technique; in this case the average silhouette. The advantage of using three-dimensional data is that silhouettes from any arbitrary viewpoint can be synthesised, even if the viewpoint is not directly seen by a camera.

Silhouettes are taken from a side-on orthogonal viewpoint. This view is not seen by any camera and so must be synthesised. The use of a side-on viewpoint facilitates comparison with previous results. To generate the average silhouette images the centre of mass is found for each frame [69]. The average silhouette is then found by summing the centre of mass aligned silhouettes.

The derived average silhouette is scale normalised so that it is 64 pixels high, whilst preserving the aspect ratio. The average silhouette is treated as the feature vector and used for leave-one-out recognition, using nearest-neighbour classification and the euclidean distance as the distance metric between samples.

8.2.3 System 2: UOULU gait recognition system

UOULU has proposed to use dynamic texture descriptors, Local Binary Patterns from Three Orthogonal Planes (LBP-TOP), to describe human gait in a spatio-temporal way.

The dynamic texture based gait recognition system of the University of Oulu works as follows [66, 67]. Firstly, a video sequence of a person's walking can be thought as spatio-temporal volume. The volume is partitioned into sub-volumes. Using the sub-volume representation, motion and shape are encoded on three different levels: pixel-level (single bins in the histogram), region-level (sub-volume histogram) and global-level (concatenated sub-volume histograms). Secondly, LBP-TOP description is formed by calculating the LBP features from XY, XT and YT planes of volumes and concatenating the histograms to catch the transition information in spatio-temporal domain. The LBP-TOP features from each sub-volume are extracted and concatenated to encode motion and shape characteristics.

Thirdly, to use the multi-resolution information, original uniform patterns are improved with ordering sampling points according to the sampling angle, by which they will also produce codes that satisfy the bit transition condition and any number of sampling points can be used on different LBP kernels. Fourthly, the length of the LBP-TOP histogram representation can be quite large depending on the number of sampling points and number of sub-volumes that are used. A better and more compact representation can be obtained by using feature selection methods. Gentle AdaBoost was used to perform feature selection and to build a strong classifier. Instead of building a classifier that gives the identity of the person from one sample, a two-class classifier was trained, which classifies weather two samples come from the same person or not.

Experiments on CMU MoBo dataset and USF dataset, and comparison with other methods showed the suitability of the proposed representation for gait recognition.

9 Vein and fingerprint biometrics

The vein-based biometric is an emerging technology which uses penetration of near infrared light through the skin to create an image of the vein network. Two acquisition modes are possible with this technology: a reflection mode and a transmission mode. Identification of the vein network is one of the rare biometric techniques that uses almost invisible in-vivo features of the human body. It has appealing properties regarding universality criteria, discriminatory ability, permanence, acceptability and vulnerability and the technology is known to be very precise and difficult to fake.

The work in TABULA RASA for vein biometrics is based on the Morpho product FingerVP. This sensor is dedicated to the acquisition of the vein network and fingerprint of a given finger. Because of legal issues in the exploitation agreement of the Hitachi vein sensor technology, the vein biometric cannot be used as a standalone component and will necessarily be associated with the corresponding fingerprint biometric. In the context of this document it is considered as a pseudo-multi-modal biometric and is detached from the work presented in Section 11 where it is assumed that individual scores are available for each mode. This is not the case here; a single score is produced and corresponds to simultaneous verification based on both the vein and fingerprint biometrics. The vein-fingerprint biometric is thus here treated independently from the material in Section 11.

9.1 Database

A multi-modal biometric using vein and fingerprints is very attractive. Is it among the rare multi-modal biometrics where different modalities are acquired simultaneously from the same body part (other examples are iris/face and 2D/3D face). Therefore, it is possible to go beyond multi-modal fusion methods to defeat spoofing attacks [118] and it is possible to study the correlation between intrinsic characteristics of the two modalities. The databases used for performance evaluation should allow studying different approaches and evaluating their performances.

9.1.1 Existing databases

Even if NIST intend to collect a FingerVein database there is currently no public database of real finger vein + finger print samples. Therefore an internal database will be acquired by Morpho for the project.

9.1.2 The TabulaRasaVP database

This database will be acquired using the MORPHO FingerVP scanner. This sensor allows simultaneous acquisition of fingerprint and second phalanx finger vein patterns. The database should contain data from at least 30 volunteers. 6 samples per participants will be acquired, representing a total of 180 samples. Data will be acquired in 2 sessions, one week apart (hopefully with the same amount of volunteers) and with 5 acquisitions per

Number of Persons	Number of Fingers	Number of acquisitions/finger	Number of acquisition sessions
30	6	10	2

volunteer. The collection of samples during two sessions aims to take into account some variability in the acquisition process: finger pose, ambient temperature, previous activity of the persons from whom samples are acquired. The setup is summarised in Table 9.1.2

9.2 System

This section introduces vein and fingerprint biometric systems and provides a description of the system that will be used for the TABULA RASA project to establish baseline performance measures.

9.2.1 Existing systems

We can find a few FingerVein systems (i.e. addressing only the vein modality of fingers) available in the market. Hitachi is the sensor technology industry leader. Their sensor technology is used in L-1 Identity's 4G FingerVein station.

There are, however, not many vein + fingerprint sensors. Among these, we find NEC's Fingerprint + Finger vein system and Morpho's FingerVP. Concerning algorithms, there is a rich bibliography concerning fingerprint and finger vein [119–121] and multi-modality fusion [122].

9.2.2 The FingerVP system

The system that will be used for TABULA RASA work is Morpho's FingerVP. To develop this brand new product, Morpho has partnered with Hitachi, an engineering and information technology giant, to develop a multi-modal biometric recognition module. Developed and produced by Morpho, this module combines the best of Hitachi's vein imaging technology (VeinID) and Morpho's fingerprint identification technology. The complementary nature of these two identification methods, namely the recognition of the pattern of minute blood vessels under the skin, and the simultaneous processing of fingerprint data allows the module developed by Morpho to offer levels of security and accuracy that are unrivalled worldwide. Designed to be easily integrated into any type of identification system, this module meets requirements for a wide range of applications, including access control, identity checks and secure payments.

The Morpho FingerVP system was developed for the Morpho FingerVP scanner. It performs an adapted fusion of fingerprint and finger vein biometrics. This fusion allows to maximise accuracy [123] and guaranties the chosen security level (FAR). This means that some trade-off can be found to maximise spoofing resistance while minimising the false rejection rate. The system only outputs the fusion result and no access to individual scores of each modality can be provided. The product is designed to work on sets of images acquired from the vein sensor and from one image acquired from the fingerprint sensor. From these images, a template is created in a proprietary binary format. The template matching algorithm provides a consolidated score that can be compared to a given threshold in order to obtain a decision. Reference thresholds are provided to meet a specific performance target, i.e. a target False Accept Rate can be reached and guaranteed with a given fixed threshold.

The main features available in this system are the following:

- Coding of single-finger 500dpi grey-scale fingerprint images and vein images to create templates
- Authentication : 1-to-1 matching of fingerprint + vein templates
- Identification : 1-to-many matching of reference template against a memory template database

10 Electro-physiology biometrics

In this section we describe the existing databases and systems developed in Starlab regarding the electro-physiological biometric module. More precisely, Starlab has developed 2 electro-physiological biometric systems, one based in electro-encephalogram (EEG) and another based on electro-cardiogram (ECG). Both systems are based on the ENOBIO electro-physiological recording device, also developed by Starlab.

10.1 Database

Here we highlight some existing databases and then focus on the data that has been collected in Starlab.

10.1.1 Existing databases

There are some publicly available datasets on the web, but we have not found any that were recorded for biometric purposes. Most of them deal with pathologies (epilepsy or alcoholism) or with event-related potentials (ERPs). As an example, the Swartz Center for Computational Neuroscience, a centre of the Institute for Neural Computation from the University of California, San Diego, offers several dataset such as the ones described below:

- Psychophysics (4Mb): One subject (80 trials) from a visual attention task (32-channel; Matlab format).
- Psychophysics (30Mb): About 10 EEG files in different binary format. Task specification are not provided.
- Psychophysics (450Mb): 5 subjects and 2 conditions (64 channels, Matlab format).
- Psychophysics (700 Mb): 122 subjects recorded using 64 channel (Alcoholic and Controls performing a visual matching task).
- Epilepsy data: A very comprehensive database of epilepsy data files.
- Sleep data: Sleep EEG from 8 subjects (EDF format).
- Motor imagery data: Motor imagery data for BCI project (Matlab files).
- P300 data: P300 data used for BCI project (Matlab format).
- Animal and human EEG: few trials of EEG data from rats, visual evoked potential, epilepsy, and rest.
- Continuous EEG: few seconds of 64-channel EEG recording from an alcoholic patient.

There are also some EEG datasets available form the Brain Computer Interface (BCI) competitions. There have been 4 competitions so far, each one with different datasets. These datasets are quite diverse and they include magneto-encephalogram, electro-corticogram and several ERP tasks such as P300 paradigm and motor imagery. Usually only data from few subjects is available, making those datasets useless for biometric purposes.

There are no publicly available database of both EEG **and** ECG samples for biometrics purposes and thus we have chosen to use our own recorded databases for the TABULA RASA project. They include those recorded within the HUMABIO and ACTIBIO projects. Although the associated protocols for each database are not the same, sensor and electrode locations are identical and thus they are both useful for the purposes of evaluation. Both databases are proprietary. Starlab wishes to retain the right of deciding case-by-case on its availability for other partners in the consortium. Given the case, the sharing conditions will be agreed between the two involved partners.

10.1.2 The Starlab databases

Starlab has participated in two European projects related to biometrics. In both of them, the main role of Starlab was to implement a biometric system based on EEG and ECG, although other types of electro-physiological signals were also explored such as ERPs, electro-oculography (EOG) and electro-myography (EMG). A short description of these two project follows:

- HUMABIO is an EU co-funded 'Specific Targeted Research Project' (STREP) where new types of biometrics are combined with state-of-the-art sensorial technologies in order to enhance security in a wide spectrum of applications such as transportation safety and continuous authentication in safety critical environments such as laboratories, airports or other buildings.

- ACTIBIO aims to research and develop a completely new concept in biometric authentication, i.e. the extraction of biometric signatures based on the response of the user to specific stimuli while performing specific work-related activities. The novelty of the approach lies in the fact that the measurements used for authentication correspond to the response of the person to specific events being however, fully unobtrusive and also fully integrated in an ambient intelligence infrastructure.

Concerning the electro-physiological biometric systems, the main difference between the 2 projects is that in HUMABIO, the users had to follow a protocol for both enrolment and authentication (seated comfortably during some time with the eyes closed) while in ACTIBIO the users did not need to be seated and they did not need to keep their eyes closed. In the later, the artifact removal/correction stage was an important challenge to be solved for the success of the project.

As mentioned earlier the EEG/ECG recording device is ENOBIO (see Figure 1), a product developed at STARLAB BARCELONA SL. It is a wireless 4-channel device (plus the common mode) with active electrodes. It is therefore quite unobtrusive, fast and easy

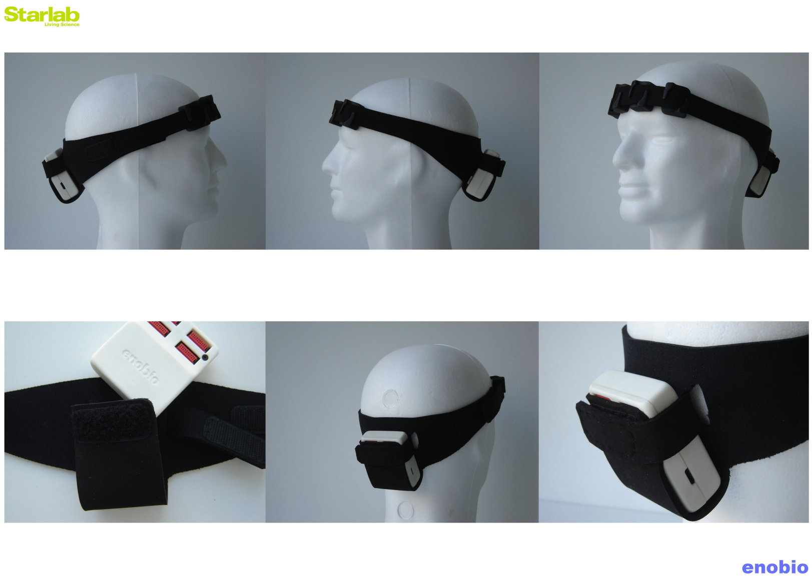

Figure 1: ENOBIO electro-physiological sensor

to place. Even though ENOBIO can function with dry contacts, in this study a conductive gel was used. The sampling rate of ENOBIO is 250 Hz, but the version used in HUMABIO had a sampling rate of 256 Hz. The electrode placement for both data sets is as follows:

- two on the forehead (FP1 and FP2) for EEG recording;
- one on the left wrist for ECG recording;
- one on the right earlobe as reference and
- one on the left earlobe as the hardware common mode.

This configuration allows the ENOBIO sensor to record EEG and ECG data at the same time. The databases collected through the HUMABIO and ACTIBIO projects are described in further detail below.

HUMABIO:
The recordings are carried out in a controlled environment. The subjects are asked to sit in a comfortable armchair, to relax, be quiet and close their eyes. Then three 3-minute takes are recorded to 32 subjects and four 3-minute takes are recorded to the 8 subjects, preferably on different days, or at least at different moments of the day. The 32-subject set is used as reference subject in the classification stage and the 8 subjects are the ones that are enrolled in the systems. Then several 1-minute takes are recorded afterwards to

these enrolled subjects, in order to use them as authentication tests. Both the enrolment takes and the authentication takes are recorded under the same conditions.

This was the initial work carried out in HUMABIO, but with time we were able to collect more data. In total we have 50 subjects enrolled (that is 50 subjects had 4 3-minute takes recorded) and 12 subjects underwent the authentication process several times. We have a total of 190 1-minute authentication takes. We also recorded 3 authentication takes from 7 subjects who just had a coffee and then 3 surther authentication takes from 7 other subjects after smoking a cigarette.

ACTIBIO:
The data was recorded in two different sessions in Barcelona in the SmartRoom at UPC. The first session was recorded in mid January while the second was recorded at the end of February. A total of 29 subjects were recorded although 2 of them have only one session recorded.

In the first session, 2 2-minute enrolment takes were recorded. We also recorded 1-minute takes while the subject was walking and finally 2 office repetitions. In the second session, we recorded again 2 2-minute enrolment takes, a 1-minute take while the subject was walking and finally, in this case, we only recorded 1 office repetition. The office repetition consisted in several actions that were annotated using video, such as seated in a chair, playing minesweeper on a PC, typing in a keyboard, writing on a sheet of paper, drinking water, etc.

As in the HUMABIO protocol, the ENOBIO is placed in a head band where we have 2 electrodes in the forehead (for EEG and EOG) and the ENOBIO transmitter in the back of the head. 2 other electrodes are placed in both earlobes (common ground in left ear lobe and active reference in right ear lobe). Finally another electrode is placed in the left wrist in order to record ECG. This configuration is the same for all the sessions, takes and modalities (enrolment, office and walking). We can see in Figure 2 an example of a user performing the 'answering the phone' task.

The enrolment recordings consist of four 2-minute takes. The subject is asked not to move while he/she is seated in a chair watching a movie. That means that the subject keeps his eyes open and blinks during the enrolment.

In summary, regarding the data recorded in the SmartRoom, we had 29 subjects enrolled (that is 29 subjects had 4 2-minute take recordings). We have 2 office takes, 2 walking takes and 2 extra takes (subject was not performing any particular action) for this set of 29 subjects. We also had 43 subjects enrolled during the pilots and 35 subjects authenticated, with a total of 57 authentication takes.

10.2 System

The are several papers describing EEG and ECG biometric based systems. In this section we briefly describe some of these systems and then focus on StarFast, a system developed by Starlab.

Figure 2: ENOBIO setup

10.2.1 Existing systems

The system presented herein attains an improvement in classification performance by combining feature fusion, classification fusion and multi-modal biometric fusion strategies. This kind of multi-stage fusion architecture has been presented in [145] as an advancement for biometry systems. Here we describe a ready-to-use authentication biometric system based on EEG and ECG. This constitutes the first difference with already presented works [127, 128, 130–132, 137–141, 148]. The system presented here undertakes subject authentication, whereas a biometric identification has been the target of those works.

A reduced number of electrodes have been already used in past works [127, 128, 130–132, 148] in order to reduce system obtrusiveness. This feature has been implemented in our system. There is however a differential trait. The two forehead electrodes are used in our system, while in other papers other electrodes configurations are used, e.g. [128] uses electrode P4. Our long-term goal is the integration of the biometric system with the ENOBIO dry and wireless sensory unit [146, 147, 155].

10.2.2 The StarFast system

The StarFast system was developed by Starlab Barcelona. It is a biometric system based on EEG and ECG signals. It is based on the ENOBIO sensor, described above and also developed by Starlab. It can record both signals at the same time and thus allows both biometrics to be collected simultaneously. Of course it can also work using only EEG or ECG. The details of the system are provided in the next section.

Authentication algorithm based on EEG

As with all the other biometric modalities, our system works in two steps: enrolment and authentication. This means that for our system to authenticate a subject, this subject needs first of all to enrol into the system. In other words, their biometric signature has to be extracted and stored in order to retrieve it during the authentication process. Then the sample extracted during the authentication process is compared with the one that was extracted during the enrolment. If they are similar enough, then they will be authenticated.

First of all, a pre-processing step is applied to the two EEG channels. They are both referenced to the right earlobe channel in order to cancel the common interference that can appear in all the channels. This is a common practice in EEG recordings. Since the earlobe is a position with no electrical activity, and since it is very easy and unobtrusive to place an electrode there with the help of a clip, this site appeared the better one to reference the other electrodes. After referencing, a second-order pass band filter with cut off frequencies of 0.5 and 40 Hz is applied. Once the filters are applied, the whole signal is segmented into 4-second epochs. Artifacts are kept, in order to ensure that only one minute of EEG data will be used for testing the system. We remind the reader that the subject is asked to close his/her eyes in order to minimise eye related artifacts.

We conducted an intensive preliminary analysis on the discrimination performance of a large initial set of features, e.g. Higuchi fractal dimension, entropy, skewness, kurtosis, mean and standard deviation. We chose the five features that showed a higher discriminative power. These five different features were extracted from each 4-second epoch and input into our classifier module. All the mentioned features are simultaneously computed in the biometry system presented herein. This is what we denote as the multi-feature set. The features are detailed in the following. We can distinguish between two major types of features with respect to the number of EEG channels employed in their computation. Therefore we can group features in single-channel features (auto-regression coefficients and Fourier Transform in our case) and two-channel features (the synchronicity features used in our system are: correlation, mutual information and coherence).

The work presented herein is based on the classical Fisher's Discriminant Analysis (DA). DA seeks a number of projection directions that are efficient for discrimination, i.e. separation into different classes. For an n-class problem, the DA involves n-1 discriminant functions (DFs). Thus a projection from a d-dimensional space, where d is the length of the feature vector to be classified, into an (n-1)-dimensional space, where $d \geq n$, is achieved. Note that in our particular case, the subject and class are equivalent. In our algorithm we work with 4 different DFs: linear, diagonal linear, quadratic and diagonal quadratic. The interested reader can find more information about DA in [136]. Taking into account the 4 DFs, the 2 channels, the 2 single-channel features and 3 synchronicity features, we have a total of 28 different classifiers. By 'classifier' we here mean each of the 28 possible combinations of feature, DF and channel. We use an approach that we refer to as a 'personal classifier', which is explained herein, for the identity authentication case: the 5 best classifiers, i.e. the ones with more discriminative power, are used for each subject. When a test subject claims to be, for example, subject 1, the 5 best classifiers for subject 1 are used to do the classification. By fusion the outputs of the best 5 personal classifiers we obtain a more reliable classification result. The final output of our system is:

- binary decision (authentication result)
- score (probability of the claimed subject)
- confidence level (an empiric function that maps the difference between threshold and score to a percentage)

Authentication algorithm based on ECG
We reference the ECG channel placed in the left wrist to the right earlobe reference channel. A first difference with the EEG pre-processing is that, in this case, we are not using 4-seconds epochs. Now, we segment each single heart beat waveform from the ECG signal.

We use the heart beat waveform as the input feature in our classifiers, since it is the one that showed the higher discriminative power between subjects. As previously said, from each minute of data we extract each single heart beat waveform. For defining the heart beat waveform feature, we decimate to a 144 length vectors. All these vectors in their totality are the heart beat waveform features. Thus, the total number of feature vectors, in this case, depends on the number of heart beats in a one-minute interval, i.e. on the heart beat rate.

The authentication methodology is very similar to the one used in EEG. The difference is that now we only have one feature, but we still have 4 DFs, so at the 'best classifier selection' stage, we select the best DF for each subject. In this modality there is no data fusion. Once the best DF is found, then the classification is made for the 'heart beat shape' feature and for the selected DF. The outputs for this modality are the same as in the EEG section.

11 Multi-modal biometrics

A multi-modal biometric system [193] results from the simultaneous inclusion of different sensors or the results of their associated analysis modules, which are called biometric modalities in this context, in a computer system used for subject identification or authentication. The application of this concept in a biometric system is expected to improve the performance of the overall biometry recognition system [179, 188].

11.1 State-of-the-art in data fusion

Once the modality data is integrated in the biometric system it can be processed in different manners according to one of the three following options [184]:

1. Data fusion is the operation whereby a multi-sensory data set is transformed into a unique representational form [167, 186, 196].
2. One further option is the parallel employment of data in the system.
3. Lastly the data of one modality can be used in order to guide the processing of another one.

The simplest way of fusing data is putting them in a common reference system, whereby the resulting data dimensionality is the sum of the individual ones, e.g. [177, 198]. In this way a general purpose processing or classification algorithm can be used in the larger dimensional feature space. However this configuration results in the disadvantage that pattern recognition systems present more counter-intuitive behaviours in large feature spaces than in smaller ones, what has been called the curse of dimensionality in pattern recognition [176]. Beyond this fact, some works emphasise the importance of developing special data fusion algorithms for applications where data fusion is involved [173] in order to take full advantage of this processing stage. [173] claims that the most important step in fusion algorithms is to acquire consistent data sets, co-register them, and develop appropriate data fusion techniques. This contradicts most of frameworks, where general purpose pattern recognition techniques, e.g. [199, 200], are used for fusion as well.

The most basic operators developed in mathematics are the sum and the product. These operators have been used together with some other lightly evolved ones like the ordinal operators maximum, median and minimum and the majority voting operator in data fusion from an early stage of research [181]. They are still used in schemes including data fusion methodologies together with light modifications and further simple ones like the average operator [168, 170].

However all of these operators are just the starting point from which more advanced fusion operators have evolved, particularly in the field of soft-computing and fuzzy operator research [196]. Different families of operators were already theoretically compared in [172], i.e. T- and S-norms [182], means in a generalised sense (f-mean, OWA [201], Choquet Fuzzy Integral), MYCIN operators, the Dempster orthogonal sum, possibility fusion operators, Bayesian based fusion operators, and symmetrical sums. A further study [183] compares fuzzy aggregation operators with non-fuzzy ones. It compares, on the one hand the weighted majority voting, the minimum, the maximum, the average, the product, and the Naïve-Bayes operators, and on the other hand, the fuzzy integral and so-called decision templates in six benchmark pattern recognition problems. The authors finally state that fuzzy fusion outperforms non-fuzzy operators in these six cases. To the best of our knowledge the work in [194] undertook the most recent review on fuzzy aggregation from a theoretical point of view. Although not being so complete as [172], it includes some of the most recent developments in the field, e.g. uni-norms and absorbing norms, together with interesting aspects on the topic. Some other operators used are the generalised means [174] or all variants of fuzzy integrals [196]. The most recent survey on data fusion can be found in [171]. It discusses the topic by grouping the operators on main families: min and max, means, medians, ordered weighted averaging, fuzzy integrals, T- and S-norm, a.k.a. conjunctive and disjunctive functions, mixed aggregation, MYCIN, uni-norms, null-norms.

11.2 Use of chimeric/virtual users

Chimeric or virtual-subject databases are currently being used for the evaluation of multi-modal biometric systems [187,189,190]. This experimental practice consists in the artificial creation of new 'virtual' users by integrating uni-modal data of different users, i.e. if we have two subjects 1 and 2 in a database with modalities A and B, we can create two more users presenting 1A-2B and 1B-2A as modalities. There are, however, some concerns on the usage of such databases. They can be summarised in the following points:

- Conceptual problems arise first. The goal of biometrics is to determine the identity of users presented to the system. As such the value in adapting a system to recognise non-existent subjects is questionable. If they do not exist, then there is no need for the system to identify them.

- Other reasons are theoretical. Modalities (or data streams) can be characterised in a multi-modal system as complementary, independent, or redundant [196]. Fusion parametrisation lays on this relationship. For creating a chimeric database independency among modalities has to be supposed a priori. Then the system cannot adapt to other relationships.

- Lastly there are two practical issues. Chimeric users can be generated for compiling an artificial evaluation database. It is argued that in this case the database can be used just to generate some figures that characterise performance of the fusion system. This characterisation is exclusive for this database and no extrapolation into real-world application of the system can be derived from here. The only usage of such a performance evaluation is the comparison among different fusion frameworks in an artificial scenario. Such a comparison has no value with respect to generalisation in the real-world application of the compared fusion frameworks. This comparison is of very limited usage beyond publication of results.

- Databases compiled in such a way contain users that do not exist in the real world. Therefore the system will never encounter such a user in its real-world application. Moreover one of the goals of the fusion system is to characterise the relationship among the different biometric modalities (see above). The system is wrongly characterising this relationship by training the system with artificial data. Furthermore this training may lead the system to adapt to relationships opposite to the ones found in a real-world situation, since the 'real' users are the minority in the chimeric users database. This becomes worse, the more users we consider. Therefore such a database cannot be used to characterise relationships among modalities in a real-world situation, and it is thus argued that fusion performance is false w.r.t. these conditions.

On the other hand, other experts think that multi-modal biometric performance evaluation allows to:

- Assume that multiple biometrics are independent.
- Assume 'chimeric data' are synthesised following principles of the nature that generate biometric data (e.g. appearances).
- No prior knowledge is available how multi-modal data are combined in biological human beings.
- Given the points above, there is no extra bias beyond the capability of modelling with finite data. Bias is inevitable with finite data.

In summary it is of potential interest to address these issues further during our project because answering to this question could have a great scientific and practical impact, since:

- This question received little attention so far.
- One of the few works that already addressed it can be found in [189, 190]. Their arguments and reported results questioned seriously the use of chimeric data. One of the conclusions was that the performances assessed using chimeric databases do not provide a good estimation of real performances. All in all, the authors of [189, 190] claim that the use of chimeric data is questionable and can be useless.
- However the use of chimeric data can be acceptable for comparing different algorithms, of course taking into account the implicit assumptions that are necessary when using chimeric data, i.e. under artificial conditions.
- There is some consensus that anti-spoofing/liveness detection evaluation cannot be investigated with a database which contains chimeric users.

It can be summarised that this is a fundamental issue to address further during the project. This could be one of the contributions (or sub-products) of the TABULA RASA project, even if it might be a little bit out of the main scope of the project. Therefore we should collect both chimeric and real data, in order to address some fundamental questions like:

- Whether and when the use of chimeric data is useful/motivated.
- Are performances achieved with chimeric data reasonable approximations of performances achieved with real multi-modal data? Under which conditions?
- Which are proper and theoretically grounded uses of chimeric data?

Ultimately, however, since its use remains somewhat contentious and since it became clear that it could be easily avoided, it was decided that chimeric data will not be allowed for the evaluation of baseline systems, neither uni-modal nor multi-modal. Later in the project, when spoofing and countermeasures are to be investigated, the issue will be readdressed where necessary.

11.3 Databases

We do not provide a comprehensive review of alternative, existing databases for multi-modal biometrics since the number of possible biometric combinations and appropriate databases is too large to be addressed here. In the following we review the various multi-modal databases that are to be considered in the TABULA RASA project. Not addressed here is the MOBIO database which is to be used for multi-modal face-voice biometrics and is described in Section 2.

11.3.1 Database 1: The BMDB database

To obtain the multi-scenario, multi-environment BioSecure Multimodal Database (BMDB) the reader is referred to the BioSecure Association[20]. Information about the specifications of the BMDB database is obtained from [64]. Some further information specific to fingerprints is provided in Section 6. The acquisition of the BMDB was jointly conducted by 11 European institutions participating in the BioSecure Network of Excellence [65]. BMDB is comprised of three different data sets, which are:

- Data Set 1 (DS1), acquired over the Internet under unsupervised conditions (i.e. connecting to an URL and following the instructions provided on the screen). The modalities included in DS1 are voice and face. DS1 includes data from 971 subjects in two different sessions.

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

MODALITY	DATA SET 1 (DS1)	DATA SET 2 (DS2)	DATA SET 3 (DS3)	# SAMPLES
Common AV - indoor	**11 samples**	**11 samples**	**11 samples**	**33**
- Audio-video	4 PIN 4 sentences 1 digits sequence	4 PIN 4 sentences 1 digits sequence	4 PIN 4 sentences 1 digits sequence	
- Face still (webcam)	2 frontal face images	2 frontal face images	2 frontal face images	
Common AV - outdoor	-	-	**11 samples**	**11**
- Audio-video			4 PIN 4 sentences 1 digits sequence	
- Face still (webcam)			2 frontal face images	
Signature	-	**25 samples** 15 genuine 10 imitations	**25 samples** 15 genuine 10 imitations	**50**
Fingerprint - thermal	-	**12 samples** (3 × 2 × 2)	**12 samples** (3 × 2 × 2)	**24**
Fingerprint - optical	-	**12 samples** (3 × 2 × 2)	-	**12**
Iris	-	**4 samples** (2 × 2)	-	**4**
Hand - digital camera	-	**8 samples** (2 × 4)	-	**8**
Face still - digital camera	-	**4 samples** (2 + 2)	-	**4**
# SAMPLES	**11**	**76**	**59**	**146**

Table 4: Biometric data acquired per subject at each session of collection of the database.

- Data Set 2 (DS2), acquired in an office environment (desktop) using a standard PC and a number of commercial sensors under the guidance of a human acquisition supervisor. The modalities included in DS2 are voice, face, signature, fingerprint, hand, and iris. DS2 includes data from 667 subjects in two different sessions.

- Data Set 3 (DS3), acquired using mobile portable hardware under two acquisition conditions: indoor and outdoor. Indoor acquisitions were done in a quiet room, whereas outdoor acquisitions were recorded in noisy environments (office corridors, the street, etc.), allowing the donor to move and to change his/her position. The modalities included in DS3 are face, voice, fingerprint and signature. DS3 includes data from 713 subjects in two different sessions.

The three data sets of BMDB include a common part of audio and video data, which is comprised of still face images and talking face videos. Also, signature and fingerprint data have been acquired both in DS2 and DS3. Additionally, hand and iris data were acquired in DS2. Table 1 summarises this information and describes the number of samples acquired per subject in each session.

Data Set 2 (DS2) is of special interest for TABULA RASA because of the biometric modes included, in particular face, optical fingerprints and iris. The biometric sensors employed to capture the signals in Data Set 2 (DS2), as well as some examples of valid and invalid acquisition samples are shown in Figure 1.

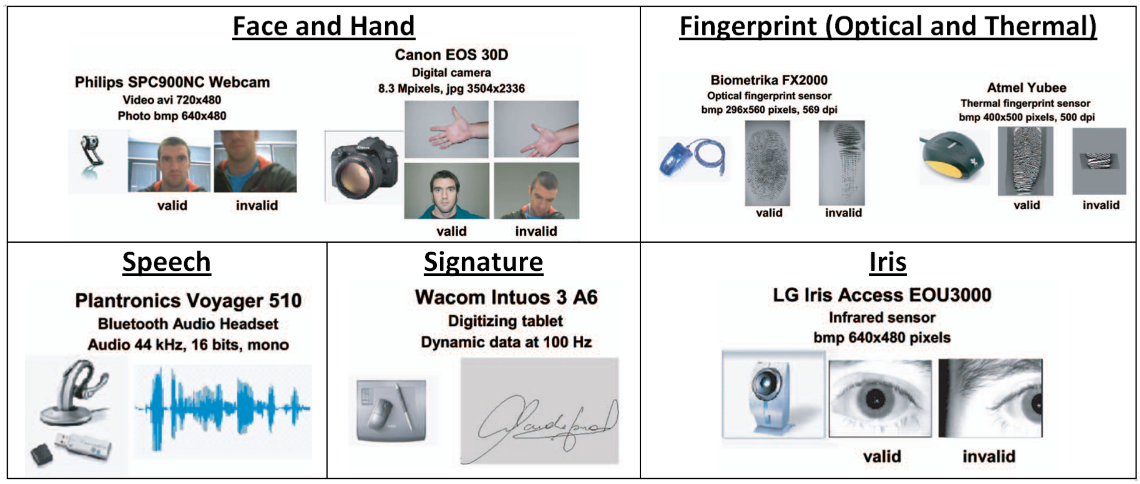

Figure 3: Hardware devices used in the acquisition of DS2 together with acquisition samples.

11.3.2 Database 2: The BEED database

The BEED data set constitutes a subset of the data set generated in the ACTIBIO project, which was acquired within an ambient intelligence facility. Up to 29 subjects go through a data acquisition protocol within two different scenarios denoted as workplace and office. In the first one the subject walks around the workplace, whereas in the second one, a seated subject realises different office related activities, e.g. answering the phone, watching a video, typing a document on the computer. As a consequence different modalities are applied to the different activities, i.e. a modality like gait can not be extracted when the subject is sitting.

The ambient intelligence facility where the ACTIBIO database was acquired took into account the signals of following sensors:

- Five calibrated cameras: 2 USB cameras (lateral, zenithal for gait recognition), 1 stereo camera (used for gait recognition as well), 1 high resolution Fire-Wire camera (used for face related modalities), and 1 Pan-Tilt-Zoom Camera (used for office activity recognition).
- A sensing seat based on strain sensors realised by means of Conductive Elastomers (CE) composites.
- An ENOBIO® wire-less electro-physiological sensor.

The biometric modalities included in the ACTIBIO database are the following:

- EEG and ECG [192].
- Face Recognition, face dynamic patterns [185].

- Activity based biometrics [169].
- Sensing seat dynamics [178].
- Gait recognition (different approaches) [180, 195].

The BEED database includes data of just 12 subjects, whose data was acquired on two sessions with 8 repetitions each. Moreover the data to be taken into account includes only the electro-physiological modalities the authentication results based on the Electroencephalogram (EEG) and Electrocardiogram (ECG) signals [192].

The number of data points in the BEED database is formed by:

- 12*A positive examples (also known as legal authentication examples), and
- 66*A negative examples (also known as impostor authentication examples),

where 66 comes from the number of possible pair comparisons among subjects, i.e. 12*11/2, and A is the number of analysed activities of each scenario, i.e. around 70 for office. It is worth mentioning that in this second scenario each modality has its own authentication rate, i.e. how frequent is the modality able to issue an authentication. Therefore the two electro-physiological modalities are co-registered before undergoing its fusion. Each of the examples in the database includes three types of values: the classification score (as a real-valued number), the decision result obtained after applying a modality dependent threshold (as a binary value), and the confidence of the classification (as a real-valued number).

A detailed specification of the BEED database can be found in [202].

11.3.3 Selected databases and multi-modal combinations

Given the different use cases to be implemented involving multi-modal databases, we have decided to select different databases depending on the corresponding use case. In the following we list the specific combinations that will be investigated in a multi-modal context and the database which has been selected in each case.

- 2D face + voice: this work will be performed using the MOBIO database which is described in Section 2. The relevant use case in this instance relates to the access control.
- 2D face + fingerprint: to be undertaken using the BMDB database which is described in Section 11.3.1. The use case relates to border control.
- 2D face + 3D face: again related to border control and to be investigated using the FRGC which is described in Section 3.1.2.
- EEG + ECG: to be undertaken with the BEED database described in Section 11.3.2. The use case relates to a telepresence brain computer interface.

Full details of use case scenarios is reported in D2.1.

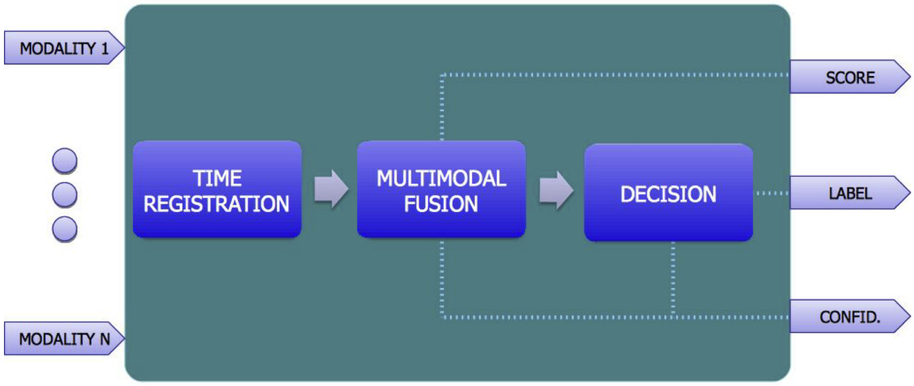

Figure 4: Block diagram of the fusion module for the design phase.

11.4 Systems

Once again, there exist many different systems that are suitable for score-level, late fusion. They are not addressed here. Instead we concentrate on the two different fusion systems that will be considered in the TABULA RASA Project which are both described below.

11.4.1 System 1: The MITSfusion system

Here we describe the Multi-modal system based on Iterative Tree of Soft data fusion operators (MITSfusion) system for authentication [197] which was developed and implemented during the ACTIBIO[21] project. The main goal of the fusion module is the generation of a fusion result from a multi-modal input set following a pattern recognition approach. Its basic functionality is mapping the outputs of the modalities set into an output characterising the authentication of the subject, whose identity is being evaluated.

The fusion module presents two different architectures, which are devoted to the phases denoted in the following paragraphs as design and recall. While the design phase architecture includes training, the other structure is recalled on-line when the authentication system based on MITSfusion is working. The assessment undertaken in the design phase is used for setting up the fusion module recall structure. The aspects that need to be assessed in the design phase include: which fusion operator is going to be implemented, which are the optimal parameters of this operator, whether the optimal parameter set need to be adapted either to the subject or not by taking into account the achieved performance levels, and which additional operations besides the fusion do have to be implemented. Once these questions have been answered, a fusion module can be implemented for recall.

As it can be observed in Figure 4 the fusion module employed in the design phase is formed by three different sub-modules. The first one undertakes the time registration of the modality inputs. Time registration is necessary in order for these values to be

[21] http://www.actibio.eu:8080/actibio

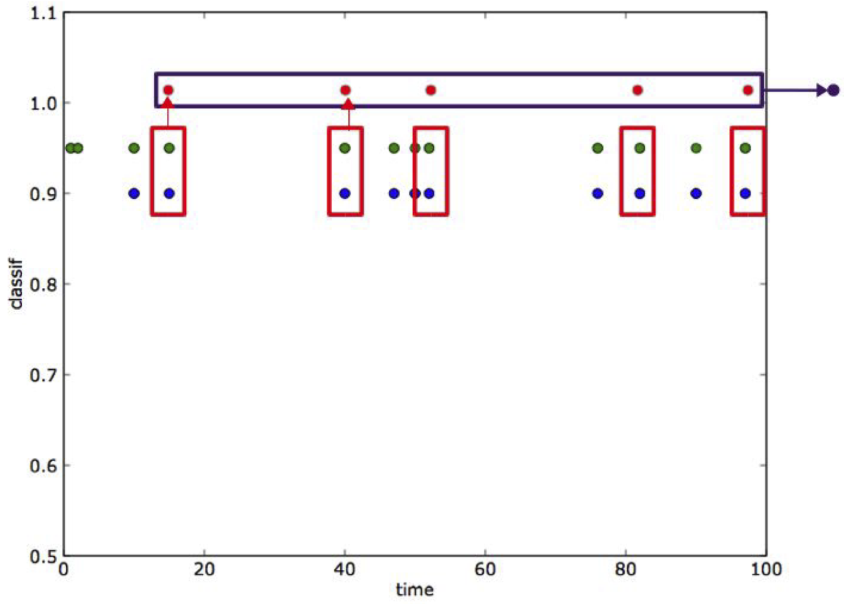

Figure 5: Illustration of the difference between the multi-modal fusion fusion (in red) and the time domain fusion (in magenta) on the exemplary case of EEG (green) and ECG (blue) registered data.

synchronised. Since in the final application the fusion request is issued by the central unit with already synchronised data, this stage is only needed in the design phase.

Once the time registration has been undertaken, we can realise the fusion itself. There are two types of fusion: multi-modal and time domain fusion. In the first case we fuse the results of the modality authentications for each time sample (see Figure 5). This operation is expected to take advantage of the redundancy or complementarity [196] among the different modalities in order to improve the robustness of the fused authentication. In the fusion across activities the fusion tries to improve the authentication robustness based on another rationale. On the course of time it is expected that the identity of the user will not change. This will only take place if the user has been substituted (normally in a forced manner), what should be detected by the system. Therefore the time domain fusion should ensure the smoothness of a change in the authentication result. The MITSfusion is focused in the multi-modal fusion. For this purpose a fusion methodology has been developed and implemented based on the application of fusion operators (see Section 11.1).

The multi-modal fusion has been implemented by a so-called iterative fusion operator tree. The application of this iterative tree attains solving the problem of receiving inputs of changing dimensionality, i.e. because of the aforementioned dependence of the authentication capability of the different modalities on the activity realised by the user. This problem can be solved through the associativity of the corresponding aggregation operators as first stated in [172]. The fulfilment of such a property ensures that an operator can be applied on vectors of changing dimensionality from a mathematical point of view. However from a practical point of view this is not always necessary (or possible). Being the applied operators defined in the unit hypercube, it is possible to iteratively apply the selected fusion operator on pairs of the input variables. Furthermore the iterative fusion operator tree presents the property that the iterative computation can be stopped when a missing value appears, i.e. when a modality is not capable of authenticating the user for a particular activity.

This methodology will be denoted herein as building an iterative fusion operator tree. An iterative fusion operator tree is a generalisation of the recursive extended aggregation functions [171, 175], i.e. it does not always result from the application of the recursive extended aggregation. This will apply only if the fusion operator taken into account is associative. If the operator is associative, the result will be equivalent to this resulting from the application of the operator to the input variables all at once. However if the overall result is performing with sufficient reliability, we can ignore the mathematical theoretical aspects. Nevertheless the iterative free operator tree presents a property that complicates the fulfilment of the associativity. The parameters of each node ($\mathbf{p}_i$) are selected for each pair of input variables. The selection results from an extensive search in the parameter space of the corresponding fusion operator. Here the search targets the maximisation of the Area Under the Curve (AUC) after the fusion of each pair. While the design phase includes this extensive search procedure, the authentication phase includes an iterative fusion operator tree and the optimal parameters. In the following we describe the interface and functionality of the MITSfusion system.

Input

All biometric modalities will provide the same output that will be used by the fusion module as input:

- Binary decision (0 not authenticated or 1 authenticated).
- Authentication score (real value between 0 and 1).
- Confidence (real value between 0 and 1, or -1 in the case the modality does not provide data. A modality might not provide data because it is not used, as the sensing seat if the subject is walking, or because the activity was too short for a given modality to perform the authentication).

If the fusion is subject based or based on a temporal parameter (e.g. which activity the subject is undergoing) we may need a connection with a database in order to retrieve the multi-modal signature, which include the fusion parametrisation. Alternatively these parameters could be stored in a file in the same fusion module and thus the connection with the database would not be necessary. An alternative for implementing a fusion based on a temporal parameter relies on the usage of the confidence measures delivered by the modalities, since these measures might change on the temporal domain.

Output
The fusion data centre will deliver the following values:

- Binary decision (binary value: 0 not authenticated or 1 authenticated).
- Authentication Score (real value between 0 and 1).
- Confidence (real value between 0 and 1 or -1 in the case the fusion can not be computed, e.g. no available input data).

Functionality
For each authentication procedure each modality provides 3 values to the fusion module. The modalities that cannot perform the authentication provide a confidence value of -1 (the decision and classification score can be set to -1 as well). In other words, all modalities will provide some values to the fusion module even if the modality itself is not working or cannot perform the authentication in the given amount of time. This condition implies that the dimensionality of the fusion operator will change over time, so that we should ensure that fusion operators are capable to cope with inputs of changing dimensionality.

The fusion will finally output a binary decision, a confidence level and the authentication score based on the application of an iterative fusion operator tree. The fusion operator to be used will be selected based on performance evaluation in the design phase. This evaluation will be conducted on off-line data provided by the modalities.

11.4.2 System 2: Multi-modal biometric verification systems

Fusion of multiple matchers is a reasonable and powerful solution aimed to increase the performance of personal verification systems based on biometrics [157]. The general claim is that different biometrics exhibit complementary and non correlated characteristics which can be exploited at different levels. The most widely adopted scheme is the so-called 'score-level', based on the combination of their scores. To this aim, a normalisation step is often necessary to better adapt the genuine and impostor distributions of different matcher.

These systems generally assume the 'independence' or 'uncorrelation' among matching scores of biometric involved. Since it is difficult to derive a 'physical' correlation among different biometrics of the same individual, neither theoretically, nor experimentally, this assumption has shown to work in many cases. This aspect has found several experimental confirmations in the last years [157, 161–164].

Another point is that multi-biometric systems are generally claimed 'intrinsically robust' against spoof attacks: an attacker would need to replicate all biometrics in order to crack the system.

In this Section, we present several parallel and serial score-level fusion rules which UNICA will use for testing ICAO and non-ICAO multi-modal biometric systems, with and without spoof attacks, in WP3 and WP4.

Likelihood ratio rule

The so-called Likelihood Ratio Rule (LLR) [161] is based on the estimation on the 'joint' probability of genuine users and impostors class, if multi-modal biometrics are used. Ratio between genuine and impostors distributions defines a novel value, usually indicated with λ, Final decision is made as follows: if $\lambda > \lambda*$ genuine users class is associated to the input biometrics, otherwise impostors class is associated. $\lambda*$ is the threshold value defining the operational point adopted.

This LLR can be modified in order to become a serial fusion rule [166], by progressively estimating LLR of first i^{th} biometric scores.

Fixed score-level fusion rules

A set of so-called 'fixed' score-level fusion rules will be implemented by UNICA. These rules are called 'fixed' because they are non-parametric rules [162]. Therefore, they do not require separate data sets for estimation of parameters. The aim is to compute a fused matching score in order to take the final decision by standard thresholding.

Adopted fixed rules by UNICA are [162]:

- *Simple average rule (SA).*
- *Simple product rule (SP).*
- *Bayes rule-like fusion (B).*
- *Max rule (MAX).*
- *Min rule (MIN).*

Trained Score-level fusion rules

In general such rules require the estimation of a set of parameters, each one associated to a certain matching score provided by the related biometric system. The estimation of above set depends on the availability of a separate data set. This motivates the fact that these rules are called 'trained'.

Adopted trained rules by UNICA are [162, 163]:

- *Weighted average rule (WS).*
- *Weighted product rule (WP).*
- *Logistic-based fusion (Perceptron) (P).*

Serial fusion of multiple matchers (SF)

In this rule, the subject submits to the system the first biometric which is processed and matched against the related template. If the resulting score is more than a predefined upper threshold, she/he is accepted as a genuine user. If the score is less than a predefined lower threshold, she/he is rejected as an impostor. Otherwise, the system requires a second biometric. The same approach is followed until the final matcher is reached. On the basis of it, the subject is finally accepted or rejected by a simple threshold [164, 165].

12 Summary

This document describes the numerous different biometric databases and systems that will be used within TABULA RASA for evaluation purposes. They will be used for the evaluation of baseline systems, spoofing attacks and countermeasures later in the project.

Databases have been selected according to current trends for each biometric and on account of their familiarity. In most cases, particularly for ICAO-biometrics, there are standard, large and dedicated, publicly available databases. They have been used where appropriate and will provide for reliable, meaningful results which are easily compared with others' work with minimal effort being spent on adapting existing systems to new data and protocols. This will allow us to concentrate effort on the spoofing threat and countermeasures which form the heart of the TABULA RASA project. In some cases, however, particularly for the more experimental biometrics, publicly databases either do not exist or are not of sufficient size or appropriate for this work. In this case the document outlines how proprietary datasets will be used or how new data will be collected.

Biometric systems are those used by each partner in previous work and are state-of-the-art. The document outlines each biometric system with a brief account of the setup in each case. No evaluation work is reported in this document; an assessment of their baseline performance will follow in Deliverables 3.1 and 3.2.

Also reported are the databases and fusion systems that will be used for multi-modal work, together with an account of the state-of-the-art in data fusion. Arguments related to the use of chimeric data from so-called virtual users are presented. All of the baseline evaluation work aims to use non-chimerical, i.e. genuine users, and the multi-modal datasets have been carefully selected in accordance with the use cases and various multi-modal combinations in order to avoid the use of chimeric data.

References

[1] A. F. Abate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3D face recognition: A survey. In *Pattern Recognition Letters*, vol. 28, pp. 1885-1906, 2007.

[2] P. J. Phillips, P. J. Flynn, T. Scruggs, et al. Overview of the Face Recognition Grand Challenge, In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*, vol. 1, pp. 947–954, 2005.

[3] J. A. Cook, M. D. Cox, V. Chandran, and S. Sridharan. Robust 3D Face Recognition from Expression Categorisation. In *International Conference on Biometrics*, September, Seoul, Korea, 2007.

[4] NIST. Face Recognition Grand Challenge(FRGC). http://face.nist.gov/frgc/

[5] V. Blanz, T. Vetter. Face recognition based on fitting a 3D morphable model, *IEEE Trans. Pattern Anal. Machine Intell*, vol. 25 no.9, pp. 1063–1074.

[6] X. Lu, R. Hsu, A. Jain, B. Kamgar-Parsi. Face recognition with 3D model-based synthesis, In *Proc. Internat. Conf. on Biometric Authentication (ICBA)*, Hong Kong, pp. 139-146.

[7] A. Ansari, M. Abdel-Mottaleb. 3-D face modeling using two views and a generic face model with application to 3-D face recognition, In *Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS 2003)* Miami, FL, USA, pp. 37-44.

[8] J. Cook, V. Chandran, S. Sridharan, C. Fookes. Face recognition from 3D data using iterative closest point algorithm and Gaussian mixture models, In *Proc. 2nd Internat. Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2004)*, Thessaloniki, Greece, 6-9 September.

[9] C. S. Chua, F. Han, Y.K. Ho. 3D human face recognition using point signature, In Proc. *4th IEEE Internat. Conf. on Automatic Face and Gesture Recognition (FG 2000)* Grenoble, France, pp. 233-238.

[10] C. Hesher, A. Srivastava, G. Erlebacher. Principal component analysis of range images for facial recognition, In *Proc. Internat. Conf. on Imaging Science, Systems and Technology (CISST 2002).*

[11] K. Chang, K. Bowyer, P. Flynn. An evaluation of multi-modal 2D + 3D face biometrics, *IEEE Trans. Pattern Anal. Machine Intell. (PAMI 2004).*

[12] T. Papatheodorou, D. Rueckert. Evaluation of automatic 4D face recognition using surface and texture registration, In *Proc. Sixth IEEE Internat. Conf. on Automatic Face and Gesture Recognition* Seoul, Korea, pp. 321-326, 2004.

[13] D. Maltoni, D. Maio, A.K. Jain and S. Prabhakar, Handbook of Fingerprint Recognition (Second Edition), Springer (London), 2009.

[14] B.G. Sherlock, D.M. Monro and K. Millard, Algorithm for enhancing fingerprint images, Electronics Letters, vol. 28, no. 18, pp. 1720, 1992.

[15] O. Trier and A.K. Jain, "Goal-directed evaluation of binarization methods," IEEE Transactions on Pattern Analysis Machine Intelligence, vol. 17, no. 12, pp. 1191-1201, 1995.

[16] A. M. Bazen and S.H. Gerez, Fingerprint matching by thin-plate spline modelling of elastic deformations, Pattern Recognition, vol. 36, no. 8, pp. 1859-1867, 2003.

[17] R. Cappelli, D. Maio and D. Maltoni, Modelling Plastic Distortion in Fingerprint Images , in proceedings Second International Conference on Advances in Pattern Recognition (ICAPR2001), Rio de Janeiro, pp.369-376, March 2001.

[18] A. K. Jain, S. Prabhakar, L. Hong, S. Pankanti. FingerCode: A Filterbank for Fingerprint Representation and Matching, In Proceedings of CVPR'1999. pp.2187 2187

[19] J. Fierrez, J. Galbally, J. Ortega-Garcia, M. R. Freire, F. Alonso-Fernandez, D. Ramos, D. T. Toledano, J. Gonzalez-Rodriguez, J. A. Siguenza, J. Garrido-Salas, E. Anguiano, G. Gonzalez-de-Rivera, R. Ribalda, M. Faundez-Zanuy, J. A. Ortega, V. Cardenoso-Payo, A. Viloria, C. E. Vivaracho, Q. I. Moro, J. J. Igarza, J. Sanchez, I. Hernaez, C. Orrite-Urunuela, F. Martinez-Contreras and J. J. Gracia-Roche, "BiosecurID: A Multimodal Biometric Database", Pattern Analysis and Applications, vol13, n. 2, pp. 235-246, 2010.

[20] J. Fierrez, J. Ortega-Garcia, D. Torre-Toledano, and J. Gonzalez-Rodriguez. BioSec baseline corpus: A multimodal biometric database. Pattern Recognition, 40:1389-1392, 2007b.

[21] 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, R. Ng, M. Fairhust, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F. Sukno, S.-K. Pavani, A. Frangi, L. Akarun, A. Savran, "The multi-scenario multi-environment Biosecure Multimodal database (BMDB)," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 32, n. 6, pp. 1097-1111, 2010.

[22] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. L. les Jardins, J. Lunter, Y. Ni, and D. Ptrovska- Delacretaz. BIOMET: A multimodal person authentication database including face, voice, fingerprint, hand and signature modalities. In Proc. IAPR Audio- and Video-Based Biometric Person Authentication (AVBPA), pages 845-853. Springer LNCS-2688, 2003

[23] B. Dumas, J. Hennebert, A. Humm, R. Ingold, D. Petrovska, C. Pugin, and D. V. Rotz. MyIdea - Sensors specifications and acquisition protocol. Computer Science Department Research Report DIUF-RR 2005.01, University de Fribourg in Switzerland, 2005.

[24] D. Dessimoz, J. Richiardi, C. Champod, and A. Drygajlo. Multimodal biometrics for identity documents (MBioID). Forensic Science International, 167:154-159, 2007.

[25] J. Ortega-Garcia, J. Fierrez, D. Simon, M. F. Gonzalez, V. Espinosa, A. Satue, I. Hernaez, J. J. Igarza, C. Vivaracho, D. Escudero, and Q. I. Moro. MCYT baseline corpus: a bimodal biometric database. IEE Proc. Vision, Image and Signal Processing, 150(6):391-401, 2003.

[26] M. Garris, C. Watson, R. McCabe, and C. Wilson. User's guide to NIST Fingerprint Image Software 2 (NFIS2). National Institute of Standards and Technology, 2004.

[27] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett N. Dahlgren and V. Zue. TIMIT Acoustic-Phonetic Continuous Speech Corpus, Linguistic Data Consortium, Philadelphia, 1993

[28] D. Pearce and H. Hirsch. The Aurora Experimental Framework for the Performance Evaluation of Speech Recognition Systems under Noisy Conditions, ISCA ITRW ASR2000, pp. 29–32, 2000

[29] J. Godfrey and E. Holliman. Switchboard-1 Release 2. Linguistic Data Consortium, Philadelphia, 1997

[30] F. Cummins, M. Grimaldi, T. Leonard and J. Simko. The CHAINS corpus: CHAracterizing INdividual Speakers. Proc of SPECOM, pp. 431–435, 2006.

[31] J. Campbell and A. Higgins. YOHO Speaker Verification, Linguistic Data Consortium, Philadelphia, 1994

[32] CSLU: Speaker Recognition Version 1.1, Linguistic Data Consortium, Philadelphia, 2006

[33] G. Aversano, N. Brmmer and M. Falcone, EVALITA 2009 Speaker Identity Verification Application Track - Organizer's Report, in Poster and Workshop Proceedings of the 11th Conference of the Italian Association for Artificial Intelligence, Reggio Emilia, Italy, ISBN 978-88-903581-1-1, 2009

[34] J. Ortega-Garcia et al. The Multiscenario Multienvironment BioSecure Multimodal Database (BMDB). IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE VOL. 32, NO. 6, 2010

[35] S. Marcel, C. McCool, P. M. Ahonen, et al. Mobile biometry (mobio) face and speaker verification evaluation. In Proceedings of the 20th International Conference on Pattern Recognition, 2010.

[36] C. C. Joseph, J. P. Campbell, H. Nakasone, D. Miller and K. Walker. The Mixer Corpus of Multilingual, Multichannel Speaker Recognition Data. Proc. 4 th International Conference on Language Resources and Evaluation. 2004

[37] A. Martin and C. Greenberg. The NIST 2010 speaker recognition evaluation, Interspeech, pp. 2726–2729, 2010

[38] J. Bonastre , N. Scheffer , D. Matrouf , C. Fredouille , A. Larcher , R. Preti , G. Pouchoulin , N. Evans , B. Fauve and J. Mason. ALIZE/SpkDet: a state-of-the-art open source software for speaker recognition. In Proc. Odyssey: the Speaker and Language Recognition Workshop, 2008

[39] S. Meigner, T. Merlin, C. Lévy, A. Larcher, E. Charton, J. Bonastre, L. Besacier, J. Farinas and B. Ravera. Mistral: Plate-forme open source d'authentification biométrique, JEP, Avignon, 2008

[40] B. G. B. Fauve, D. Matrouf, N. Scheffer, J. Bonastre and J. S. D. Mason. State-of-the-Art Performance in Text-Independent Speaker Verification Through Open-Source Software. IEEE Transactions on Audio Speech and Language processing, Vol.15, No. 7 , pp.1960–1968, 2007

[41] N. Brümmer and J. du Preez. FoCal: Fusion and Calibration of automatic speaker detection systems. http://www.dsp.sun.ac.za/~nbrummer/focal/index.htm

[42] S. Furui. An overview of speaker recognition technology. ESCA Workshop on Automatic Speaker Recognition, Identification and Verification. pp. 1-9, 1994.

[43] NIST. The NIST Year 2008 Speaker Recognition Evaluation Plan. http://www.itl.nist.gov/iad/mig/tests/sre/2008/sre08_evalplan_release4.pdf. 2010

[44] A. Martin, C. Greenberg. 2008 NIST Speaker Recognition Evaluation. Slides From Evaluation. Workshop Presentation. http://www.itl.nist.gov/iad/mig/tests/sre/2008/official_results/NIST_sre08_webpres.pdf. 2008.

[45] A. Martin, G. Doddington, T. Kamm, M. Ordowski and M. Przybocki. The DET curve in assessment of detection task performance. In Proc. EuroSpeech, pp. 1895-1898. 1997.

[46] P. Mermelstein. Distance measures for speech recognition, psychological and instrumental. Pattern recognition and artificial intelligence.Academic Press, New York, 1976.

[47] J. Pelecanos and S. Sridharan. Feature warping for robust speaker verification. Speaker Odyssey, The Speaker Recognition Workshop, pp. 213-218, 2001

[48] D. A. Reynolds. Channel robust speaker verification via feature mapping. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2:53-56, 2003.

[49] P. Kenny, G. Boulianne, P. Ouellet and P. Dumouchel. Joint Factor Analysis versus Eigenchannels in Speaker Recognition. IEEE Transactions on Audio, Speech and Language Processing, vol.15, no.4, pp.1435-1447, May 2007

[50] D. A. Reynolds and T. F. Quatieri and R. B. Dunn. Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing, 10(1-3):19-41,2000.

[51] R. Auckenthaler and J. S Mason. Gaussian Selection Applied to Text-Independent Speaker Verification. In Proc. Speaker Odyssey, pp. 83–88, 2001

[52] M. Carey, E. Parris, and J. Bridle, A speaker verification system using alphanets. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, pp.397–400, May 1991.

[53] A. P. Dempster and N. M. Laird and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, 39(1):138,1977.

[54] J.-L Gauvain and C. H. Lee. Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. IEEE Transactions on Speech and Audio Processing, 2(2):291-298,1994.

[55] C. Auckenthaler, Lloyd-Thomas, "Score Normalization for Text-independent Speaker Verification Systems," Digital Signal Processing, vol. 10 No 1-3, 2000.

[56] R. B. Dunn, T. F. Quatieri, D. A. Reynolds and J. P. Campbell. Speaker recognition from coded speech and the effects of score normalization. Signals, Systems and Computers, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on.2:1562-1567,2001.

[57] V. N. Vapnik, The Nature of Statistical Learning Theory. Springer, 1995.

[58] W. M. Campbell, J. P. Campbell, D. A. Reynolds, E. Singer and P. A. Torres-Carrasquillo. Support vector machines for speaker and language recognition. Computer Speech & Language. 20(2-3):210-229,2006.

[59] W. M. Campbell, D. E. Sturim and D. A. Reynolds. Support vector machines using GMM supervectors for speaker verification. Signal Processing Letters, IEEE. 13(5):308-311,2006.

[60] W. M. Campbell, D. E. Sturim, D. A. Reynolds and A. Solomonoff. SVM Based Speaker Verification using a GMM Supervector Kernel and NAP Variability Compensation. IEEE International Conference on Acoustics, Speech and Signal Processing. 1:1, May 2006.

[61] P. Kenny. Joint factor analysis of speaker and session variability: Theory and algorithms CRIM-06/08-13. Technical report, CRIM, 2006.

[62] Z. Karam and W. Campbell. A new kernel for SVM MLLR based speaker recognition. Interspeech. pages 290-293, 2007.

[63] D. Matrouf, J.-F. Bonastre and C. Fredouille, Effect of voice transformation on impostor acceptance, Proc. of ICASSP 2006, Toulouse, France, 2006

[64] 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, R. Ng, M. Fairhust, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F. Sukno, S.-K. Pavani, A. Frangi, L. Akarun, A. Savran, "The multi-scenario multi-environment Biosecure Multimodal

database (BMDB)," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 32, n. 6, pp. 1097-1111, 2010.

[65] BioSecure "Biometrics for Secure Authentication, FP6 NoE, IST- 2002-507634," http://www.biosecure.info, 2004.

[66] V. Kellokumpu, G. Zhao, S. Li, and M. Pietikäinen. Dynamic texture based gait recognition. In: Proc. IAPR/IEEE International Conference on Biometrics 2009

[67] V. Kellokumpu, G. Zhao and M. Pietikäinen. Dynamic textures for human movement recognition. Proc. ACM International Conference on Image and Video Retrieval (CIVR 2010), Xian, China, 2010.

[68] R. D. Seely, S. Samangooei, L. Middleton, J. N. Carter and M. S. Nixon. The University of Southampton Multi-Biometric Tunnel and introducing a novel 3D gait dataset. Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems, BTAS 08, September 2008.

[69] G. V. Veres, L. Gordon, J. N. Carter, and M. S. Nixon. What image information is important in silhouette-based gait recognition? Proc. CVPR, volume 2, pages 776-782, 2004.

[70] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The humanid gait challenge problem: Data sets, performance and analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(2):162-177, 2005.

[71] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter. On a large sequence-based human gait database, special session on biometrics,. In in 4th Int. Conf. Recent Advances in Soft Computing, pages 66-72, 2002.

[72] S. Yu, T. Tan, and K. Huang. A study on gait-based gender classification. IEEE Transactions on Image Processing, 18(8):1905-1910, 2009.

[73] M. S. Nixon and J. N. Carter. Automatic recognition by gait. Proceedings of the IEEE, 94(11):2013-2024, 2006.

[74] M. S. Nixon, T. Tan, and R. Chellappa. Human ID Based on Gait. Springer-Verlag New York, 2006.

[75] D. Cunado, M. S. Nixon, and J. N. Carter. Automatic extraction and description of human gait models for recognition purposes. Computer Vision and Image Understanding, 90(1):1-41, 2003.

[76] C.-Y. Yam, M. S. Nixon, and J. N. Carter. Automated person recognition by walking and running via model-based approaches. Pattern Recognition, 37(5):1057-1072, 2004.

[77] I. Bouchrika, M. Goffredo, M. S. Nixon, and J. N. Carter. Viewpoint invariant gait recognition. IEEE Transactions on Systems, Man and Cybernetics (B), 2010. Accepted for publication.

[78] I. Bouchrika, M. Goffredo, M. S. Nixon, and J. N. Carter. Covariate analysis for viewpoint independent gait recognition. In The 3rd IAPR/IEEE International Conference on Biometrics, 2009.

[79] 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 10581059, 2003.

[80] S. Lucey and T. Chen. A GMM parts based face representation for improved verification through relevance adaptation. In Proc. International Conference on Computer Vision and Pattern Recognition (CVPR), volume 2, pages 855861, 2004.

[81] D. A. Reynolds. Comparison of background normalization methods for text independent speaker verification. In Proc. European Conference on Speech Communication and Technology (EuroSpeech), pages 963966, Rhodes, Greece, September 1997.

[82] [215] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. ACM Computing Surveys, 35(4):399459, 2003.

[83] [146] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell., 24(7):971987, 2002.

[84] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition. In Proceedings of the 2005 International Conference on Computer Vision, pages 786791, 2005.

[85] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In Proceedings of the 2004 European Conference on Computer Vision, pages 469481, 2004.

[86] Y. Rodriguez and S. Marcel. Face authentication using adapted local binary pattern histograms. In Proceedings of the 2006 European Conference on Computer Vision, pages 321332, 2006.

[87] K. Messer, J. Matas, J. Kittler, J. Luttin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In Audio- and Video-based Biometric Person Authentication, AVBPA99, pages 7277, 1999. Washington, D.C., March 1999. 16 IDIAPRR 99-02.

[88] V. Popovici, J. Thiran, E. Bailly-Bailliere, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariethoz, J. Matas, K. Messer, B. Ruiz, and F. Poiree. The BANCA Database and Evaluation Protocol. In 4th International Conference on Audio- and Video-Based Biometric Person Authentication, Guildford, UK, volume 2688 of Lecture Notes in Computer Science, pages 625638. SPIE, 2003.

[89] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, The feret evaluation methodology for face recognition algorithms, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1090 1104, 2000.

[90] 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, IEEE International Conference on Computer Vision and Pattern Recognition, 2005.

[91] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. IEEE T-PAMI, 25(12):16151618, 2003.

[92] J. Ortega-Garcia et al. The Multi-scenario Multi-environment BioSecure Multimodal Database (BMDB), PAMI(32), No. 6, June 2010, pp. 1097-1111

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

[94] http://www.nist.gov/itl/iad/ig/mbgc.cfm

[95] http://www.equinoxsensors.com/products/HID.html#db

[96] X. Chen, P. J. Flynn, K. W. Bowyer, PCA-Based Face Recognition in Infrared Imagery: Baseline and Comparative Studies, pp.127, IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003

[97] P. Buddharaju, I. Pavlidis, Physiological face recognition is coming of age, Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on , pp.128-135, 20-25 June 2009

[98] T. Bourlai, N. Kalka, A. Ross, B. Cukic, L. Hornak, Cross-Spectral Face Verification in the Short Wave Infrared (SWIR) Band, Pattern Recognition (ICPR), 2010 20th International Conference on, pp.1343-1347, 23-26 Aug. 2010

[99] http://www4.comp.polyu.edu.hk/b̃iometrics/polyudb_face.htm

[100] http://www4.comp.polyu.edu.hk/b̃iometrics/hyper_face.htm

[101] S. Z. Li, L. Zhen, A. Meng, The HFB Face Database for Heterogeneous Face Biometrics research, Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on, pp.1-8, 20-25 June 2009

[102] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Z. Li. Face Matching Between Near Infrared and Visible Light Images. In Proceedings of IAPR/IEEE International Conference on Biometrics, pages 523-530, 2007.

[103] D. Yi, S. Liao, Z. Lei, J. Sang, and S. Z. Li. Partial Face Matching Between Near Infrared and Visual Images in MBGC Portal Challenge. In Proceedings of IAPR/IEEE International Conference on Biometrics, 2009.

[104] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li. Heterogeneous face recognition from local structures of normalized appearance. In Proceedings of IAPR/IEEE International Conference on Biometrics, 2009.

[105] National Insitute of Standards and Technology. Iris challenge evaluation data. http://iris.nist.gov/ice/.

[106] MMU - Multimedia University iris database. http://pesona.mmu.edu.my/c̃cteo/.

[107] H. Proenca and L. A. Alexandre. UBIRIS iris image database. 2004. http://iris.di.ubi.pt

[108] H. Proenca, S. Filipe, R. Santos, J. Oliveira, L. A. Alexandre; The UBIRIS.v2: A Database of Visible Wavelength Iris Images Captured On-The-Move and At-A-Distance, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1529-1535, vol. 32, num. 8, August, 2010.

[109] University of Bath Iris Image Database. Available from: http://www.bath.ac.uk/elec-eng/research/sipg/irisweb/

[110] 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.W.R.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 Multi-Scenario Multi-Environment BioSecure Multimodal Database (BMDB)", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 32, n. 6, pp. 1097-1111, June 2010. http://biosecure.it-sudparis.eu/

[111] Z. Sun, T. Tan, and Y. Wang. Robust encoding of local ordinal measures: A general framework of iris recognition. In Proc. BioAW Workshop, pp. 270-282, 2004.

[112] D. Monro, S. Rakshit, and D. Zhang. DCT-based iris recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No.4, pp.586-595, April 2007.

[113] L. Ma, T. Tan, Y. Wang, and D. Zhang. Personal identi?cation based on iris texture analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No.12, pp. 1519-1533, 2003.

[114] K. Hollingsworth, K. Bowyer, and P. Flynn. All iris code bits are not created equal. In Biometrics: Theory, Applications, and Systems, Sept 2007.

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

[116] L. Masek et al., Recognition of human iris patterns for biometric identification, Ph.D. thesis, Technical Report, The school of Computer Science and Software Engineering, The University of Western Australia, 2003.

[117] J. Daugman, "How iris recognition works," Proceedings of 2002 International Conference on Image Processing, vol. 1, pp. 22-25, 2002.

[118] Rodrigues et al, Robustness of multimodal biometric fusion methods against spoof attacks, Journal of Visual Languages and Computing.

[119] Miura et al, Automatic Feature Extraction from non-uniform Finger Vein Image and its Application to Personal Identification, MVA2002.

[120] Zhang et al, Multiscale Feature Extraction of Finger-Vein Patterns Based on Curvelets and Local Interconnection Structure Neural Network, ICPR2006

[121] Yang et al, Efficient Finger Vein Localization and Recognition, ICPR 2010

[122] A. Ross, Information fusion in biometrics, Pattern Recognition Letters 24 (2003)

[123] H. Jarosz, Biomtrie et multi-modalit : un saut technologique vient-il d'tre franchi, ExpoProtection2010

[124] S. Eischen , J. Luctritz and J. Polish (1995) Spectral analysis of EEG from Families. Biological Psycholology, Vol. 41, pp. 61-68.

[125] N. Hazarika, A. Tsoi and A. Sergejew (1997) Nonlinear considerations in EEG signal Classification. IEEE Transactions on signal Processing, Vol. 45, pp. 829-836.

[126] S. Marcel, J. Mill (2005) Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. IDIAP Research Report 05-81, 11 pp.

[127] G. Mohammadi et al. (2006) Person identification by using AR model for EEG signals. Proc. 9th International Conference on Bioengineering Technology (ICBT 2006), Czech Republic, 5 pp.

[128] R. Paranjape et al.(2001) The electroencephalogram as a biometric. Proc. Canadian Conf. on Electrical and Computer Engineering, pp. 1363-1366.

[129] M. Poulos et al. (1998) Person identification via the EEG using computational geometry algorithms. Proceedings of the Ninth European Signal Processing, EUSIPCO'98, Rhodes, Greece. September 1998, pp. 2125-2128.

[130] M. Poulos et al. (1999) Parametric person identification from EEG using computational geometry. Proc. 6th International Conference on Electronics, Circuits and Systems (ICECS '99), v. 2, pp. 1005-1008.

[131] M. Poulos et al. (2001) On the use of EEG features towards person identification via neural networks. Medical Informatics & the Internet in Medicine, v. 26, pp. 35-48.

[132] M. Poulos et al. (2002) Person identification from the EEG using nonlinear signal classification. Methods of Information in Medicine, v. 41, pp. 64-75.

[133] A. Remond, Ed. (1997) EEG Informatics. A didactic review of methods and applications of EEG data processing, Elsevier Scientific Publishing Inc., New York, 1997.

[134] N. Sviserskaya, T. Korolkova (1995) Genetic Features of the spatial organization of the human cerebral cortex. Neuroscience and Behavioural Physiology, Vol. 25, N. 5, pp. 370-376.

[135] M. Deriche, A. Al-Ani (2001) A new algorithm for EEG feature selection using mutual information. Acoustics, Speech, and Signal Processing, 2001. Proceedings. '01, pp. 1057 - 1060 vol.2

[136] R. Duda et al., Pattern Classification, Wiley, New York, 2001.

[137] L. Biel et al. (2001) ECG analysis: a new approach in human identification. IEEE Transactions on Instrumentation and Measurement, Vol. 50, N. 3, pp. 808-812.

[138] C.K. Chang (2005) Human identification using one lead ECG. Master Thesis. Department of Computer Science and Information Engineering. Chaoyang University of Technology (Taiwan).

[139] S. Israel et al. (2005) EGC to identify individuals. Pattern Recognition, 38, pp. 133-142.

[140] M. Kyoso (2001) Development of an ECG Identification System. Proc. 23rd Annual International IEEE Conference on Engineering in Medicine and Biology Society, Istanbul, Turkey.

[141] R. Palaniappan and S.M. Krishnan (2004) Identifying individuals using ECG beats. Proceedings International Conference on Signal Processing and Communications, 2004. SPCOM '04, pp. 569-572.

[142] G. Winterer et al. (2003) Association of EEG coherence and an exonic GABA(B)R1 gene polymorphism. Am J Med Genet B Neuropsychiatr Genet, 117:51-56.

[143] M. Kikuchi et al. (2000) Effect of normal aging upon interhemispheric EEG coherence: analysis during rest and photic stimulation. Clin Electroencephalogr, 31: 170-174.

[144] R. Moddemeijer (1989) On estimation of entropy and mutual information of continuous distributions. Signal Processing vol.16 nr.3 pp.233-246

[145] A. Ross, A. Jain (2003) Information fusion in biometrics. Pattern Recognition Letters 24 pp.2115-2125

[146] G. Ruffini et al. (2006) A dry electrophysiology electrode using CNT arrays. Sensors and Actuators A 132 34-41

[147] G. Ruffini et al. (2007) ENOBIO dry electrophysiology electrode; first human trial plus wireless electrode system. 29th IEEE EMBS Annual International Conference.

[148] A. Riera et al. (2007) Unobtrusive Biometric System Based on Electroencephalogram Analysis. Accepted at EURASIP Journal on Advances in Signal Processing.

[149] Hogans et al. (1961) Effects of ethyl alcohol on EEG and avoidance behavior of chronic electrode monkeys. Am J Physio, 201: 434-436

[150] J. Sorbel et al. (1996) Alcohol Effects on the Heritability of EEG Spectral Power alcoholism: clinical and experimental research

[151] S. Jin et al (2004) Effects of total sleep-deprivation on waking human EEG: functional cluster analysis. Clinical Neurophysiology, Volume 115, Issue 12, Pages 2825-2833

[152] K. Takahashi (2004) Remarks on Emotion Recognition from Bio-Potential Signals . 2nd International Conference on Autonomous Robots and Agents.

[153] A. Haag et al. (2004) Emotion Recognition Using Bio-sensors: First Steps towards an Automatic System. Springer-Verlag Berlin Heidelberg, ADS, LNAI 3068, pp. 36-48.

[154] J. Llobera (2007) Narratives within Immersive Technologies. arXiv:0704.2542.

[155] G. Ruffini et al. (2006) First human trials of a dry electrophysiology sensor using a carbon nanotube array interface. arXiv:physics/0701159

[156] http://prag.diee.unica.it/pra/eng/research/biometrics/multimodal/datasets

[157] A. Ross, K. Nandakumar, A.K. Jain (Eds.), Handbook of multibiometrics, Springer, 2006.

[158] L. Hong, A.K. Jain, Integrating faces and fingerprints for person identification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 (12) 1295-1307, 1998.

[159] M. Indovina, U. Uludag, R. Snelick, A. Mink, A.K. Jain, Multimodal Biometric Authentication Methods: A COTS Approach, Proc. MMUA 2003, Workshop on Multimodal User Authentication, pp. 99-106, Santa Barbara, CA, December 11-12, 2003.

[160] S. Prabhakar and A.K. Jain, Decision-level Fusion in Fingerprint Verification, Pattern Recognition, 35 (4) 861-874, 2002.

[161] K. Nandakumar, Y. Chen, S.C. Dass, and A.K. Jain, Likelihood ratio-based biometric score fusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30 (2) 342-347, 2008.

[162] G.L. Marcialis and F. Roli, Score-level fusion of fingerprint and face matchers under stress conditions, IEEE 14th International Conference on Image Analysis and Processing ICIAP 2007, September, 10-14, 2007, Modena (Italy), R. Cucchiara Ed., ISBN 978-0-7695-2877-9, pp. 259-264.

[163] G.L. Marcialis and F. Roli, Fusion of multiple fingerprint matchers by single-layer perceptron with class-separation loss function, Pattern Recognition Letters, 26 (12) 1830-1839, Elsevier, 2005.

[164] G.L. Marcialis, F. Roli, and L. Didaci, Personal identity verification by serial fusion of fingerprint and face matchers, Pattern Recognition, 42 (11) 2807-2817, 2009.

[165] G.L. Marcialis, P. Mastinu, and F. Roli, Serial fusion of multi-modal biometric systems, Proc. of IEEE International Workshop on Biometric Measurements and Systems for Security and Medical Applications (BioMS2010), September, 9, 2010, Taranto (Italy), ISBN: 978-1-4244-6302-2, DOI: 10.1109/BIOMS.2010.5610438, pp. 1-7

[166] K. Takabashi, M. Mimura, Y. Isobe, and Y. Seto, A secure and user-friendly multimodal biometric system. Biometric Technology for Human Identification, A.K. Jain, N.K. Ratha (Eds.), Proc. of SPIE, Vol. 5404, pp. 12-19, 2004.

[167] M. A. Abidi and R. C. Gonzalez. *Data fusion in robotics and machine intelligence.* Academic Press Professional, Inc., San Diego, CA, USA, 1992.

[168] F. M. Alkoot and J. Kittler. Modified product fusion. *Pattern Recogn. Lett.*, 23(8):957–965, 2002.

[169] G. Ananthakrishnan, H. Dibeklioglu, M. Lojka, A. Lopez, S. Perdikis, U. Saeed, A. Salah, D. Tzovaras, and A. Vogiannou. Activity-related biometric authentication. In *eNTERFACE 2008, 4th Summer Workshop on Multimodal Interfaces, August 4th-29th, 2008, Paris, France*, pages 56–72, 08 2008.

[170] P. Bajcsy, W. W. Feng, and P. Kumar. Relevance assignment and fusion of multiple learning methods applied to remote sensing image analysis. In *SMC-IT '06: Proceedings of the 2nd IEEE International Conference on Space Mission Challenges for*

Information Technology, pages 291–298, Washington, DC, USA, 2006. IEEE Computer Society.

[171] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners (Studies in Fuzziness and Soft Computing).* Springer, 1 edition, February 2008.

[172] I. Bloch. Information combination operators for data fusion: a comparative review with classification. *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, 26(1):52–67, 1996.

[173] A. V. Bogdanov. Neuroinspired architecture for robust classifier fusion of multisensor imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(5):1467–1487, 2008.

[174] P. Bullen. *Handbook of Means and their Inequalities.* Kluwer, 2003.

[175] A. Del Amo, J. Montero, and E. Molina. Representation of consistent recursive rules. *European Journal of Operational Research*, 130(1):29–53, April 2001.

[176] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition).* Wiley-Interscience, 2 edition, November 2000.

[177] J. Esteban, A. Starr, R. Willetts, P. Hannah, and P. B. Cross. A review of data fusion models and architectures: towards engineering guidelines. *Neural Comput. Appl.*, 14(4):273–281, 2005.

[178] M. Ferro, G. Pioggia, A. Tognetti, N. Carbonaro, and D. De Rossi. A Sensing Seat for Human Authentication. *IEEE Transactions on Information Forensics and Security*, 4(3):451–459, September 2009.

[179] L. Hong, A. K. Jain, and S. Pankanti. Can multibiometrics improve performance? In *Proc. IEEE Workshop on Automatic Identification Advanced Technologies (WAIAT-99)*, pages 59–64, 1999.

[180] X. Huang and N. V. Boulgouris. Human Gait Recognition Based on Multiview Gait Sequences. *EURASIP Journal on Advances in Signal Processing*, 2008(Article ID 629102):8 p.+, 2008.

[181] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):226–239, 1998.

[182] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms (Trends in Logic, Vol. 8).* Springer, 1 edition, July 2000.

[183] L. I. Kuncheva. "Fuzzy" versus "nonfuzzy" in combining classifiers designed by boosting. *IEEE Trans. Fuzzy Systems*, 11(6):729–741, 2003.

[184] R. C. Luo and M. G. Kay, editors. *Multisensor Integration and Fusion for Intelligent Machines and Systems.* Ablex Publishing Corporation, Norwood, NJ, 1995.

[185] F. Matta and J.-L. Dugelay. Physiological and behavioural approaches for person recognition using videos. In *IPTA 2008, 1st workshop on Image Processing Theory, Tools and Applications, 23-26 November, 2008, Sousse, Tunisia*, page 6, 11 2008.

[186] H. B. Mitchell. *Multi-Sensor Data Fusion: An Introduction.* Springer, 1 edition, September 2007.

[187] A. C. Morris, S. Jassims, H. Sellahewa, and J. Koreman. Multimodal person authentication on a smartphone under realistic conditions. In *Proc. SPIE Conference on Mobile Multimedia/Image Processing for Military and Security Applications, Florida*, 2006.

[188] A. Patra and S. Das. Enhancing decision combination of face and fingerprint by exploitation of individual classifier space: An approach to multi-modal biometry. *Pattern Recognition*, 41(7):2298–2308, 2008.

[189] N. Poh and S. Bengio. Can chimeric persons be used in multimodal biometric authentication experiments. In *Proc. 2nd Joint AMI/PASCAL/IM2/M4 Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI). Volume LNCS 3869. (2005) 87100*, pages 11–13, 2005.

[190] N. Poh and S. Bengio. Using Chimeric Users to Construct Fusion Classifiers in Biometric Authentication Tasks: An Investigation. In *Proc. 2006 IEEE International Conference onAcoustics, Speech and Signal Processing, ICASSP.*, pages V–1077–V–1080, 2006.

[191] A. Riera. TN00250 - Electrophysiological Biometric Databases Available. Technical report, Starlab Barcelona S.L., 02 2011.

[192] A. Riera, S. Dunne, I. Cester, and G. Ruffini. STARFAST: a wireless wearable EEG/ECG biometric system based on the ENOBIO sensor. In *Proc. 5th Workshop on Wearable Micro and Nanosystems for Personalised Health (pHealth 08)*, pages 56–72, 05 2008.

[193] A. Ross and A. K. Jain. Multimodal biometrics: An overview. In *Proceedings of 12th European Signal Processing Conference (EUSIPCO'04)*, pages 1221–1224, 2004.

[194] I. J. Rudas and J. Fodor. Information aggregation in intelligent systems using generalized operators. *International Journal of Computers, Communications & Control*, 1(1):47–57, 2006.

[195] S.Argyropoulos, D.Tzovaras, D.Ioannidis, and M.G.Strintzis. A channel coding approach for human authentication from gait sequences. *IEEE Transactions on Information Forensics and Security*, 4(3):428–440, 2009.

[196] A. Soria-Frisch. *Soft Data Fusion for Computer Vision*. PhD thesis, TU Berlin, ISBN: 3-8167-6689-7, 2005.

[197] A. Soria-Frisch, A. Riera, and S. Dunne. Fusion operators for multi-modal biometric authentication based on physiological signals. In *Proc. 2010 IEEE International Conference on Fuzzy Systems (FUZZ'IEEE)*, pages 1–7, July 2010.

[198] S. Stramondo, C. Bignami, M. Chini, N. Pierdicca, and A. Tertulliani. Satellite radar and optical remote sensing for earthquake damage detection: results from different case studies. *International Journal of Remote Sensing*, 27(20):4433–4447, 2006.

[199] B. Waske and J. A. Benediktsson. Fusion of Support Vector Machines for Classification of Multisensor Data. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(12):3858–3866, November 2007.

[200] B. Waske and S. van der Linden. Classifying multilevel imagery from sar and optical sensors by decision fusion. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(5):1457–1466, 2008.

[201] R. R. Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst. Man Cybern.*, 18(1):183–190, 1988.

[202] A. Soria-Frisch. TN00251 - Multi-Modal Biometric Databases Description. Technical report, Starlab Barcelona S.L., 02 2011.