BEAT

Biometrics Evaluation and Testing

http://www.beat-eu.org/

Funded under the 7th FP (Seventh Framework Programme)

Theme SEC-2011.5.1-1
[Evaluation of identification technologies, including Biometrics]

D3.6: Description of Biometric Databases and Protocols

Due date: 30/11/2012
Project start date: 01/03/2012
WP Manager: Julian Fierrez
Submission date: November 23, 2012
Duration: 48 months
Revision: 1

Author(s): N. Poh (UNIS), C-H. Chan (UNIS), J. Kittler (UNIS), Javier Gabbaly (UAM), Julian Fierrez (UAM), André Anjos (IDIAP), Sébastien Marcel (IDIAP), Cagatay Karabat (TUBITAK)

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

D3.6: Description of Biometric Databases and Protocols

Abstract:

The role of databases and evaluation protocols in evaluating biometric performance cannot be over-emphasized. While a good database should allow testing of a biometric system in situations where the system is intended to be used, a good protocol should ensure that the performance results so obtained be repeatable.

The objective of this deliverable is to provide a survey of different databases and evaluation protocols of different biometric modalities, including both real and synthetic ones, as well as fusion. Since a few databases will be selected to be used on a prototype of BEAT platform, synthetic databases can be advantageous as they can be used without any privacy restriction. Furthermore, some noise factors can be better controlled during data generation. Despite these advantages, databases with real biometric data still should be the ultimate benchmark. Therefore, both types of databases will be discussed.

Contents

1 Introduction

An increasing number of biometric databases have been published and made available for the research community. This is mainly due to efforts in evaluating and benchmarking the biometric technology as well as challenge programmes that aim to improve the technology. This deliverable describes a number of biometric databases in existence as well as highlight some experimental protocols. This deliverable directly provides input to D3.7 which is about the databases and evaluation protocols that will be included into the BEAT platform.

While it is not possible to include all possible biometric databases in existence, we will highlight major databases reported for the three most commonly used biometric modalities, i.e., face, fingerprint, and iris. Apart from these modalities, real multi-modal biometric databases which contain more than one biometric modalities of same subjects will also be included.

While making databases available is essential for research, it is not a sufficient condition to guarantee repeatability of experimental results. It is equally important to establish how enrolment, validation and test data are divided. Documents and file lists that dictates these data partitions are called *experimental protocols* or simply “protocols”.

This document is organised as follow: Section 2 presents some commonly used databases, including multi-modal databases and those that provide processed data such as scores that are used for benchmark fusion algorithms. Section 3 underpins some general principles for defining experimental protocols. Some examples of commonly used protocols are also described. Section 4 then presents some selected databases in details. This is followed by conclusions in Section 5.

2 Databases

2.1 Biometric Repositories

A number of major biometric research centres worldwide, standards bodies, and research consortiums often collect and publish biometric database repositories. A repository is understood to contain a number of databases, possibly of several biometric modalities. The following are some examples of popular biometric database repositories. They are not necessary complete but do serve the purpose of illustrating the existence of a variety of biometric databases made available for research purposes.

- The CASIA database (`http://www.cbsr.ia.ac.cn/english/Databases.asp`) from the Centre for Biometrics and Security Research, China contains iris, gait, fingerprint, handwriting, action, palmprint, multi-spectral palmprint and various forms of face biometrics, including visual, near infra-red, thermal infra-red and 3D face images.

- The Biometric Consortium, a US-based (`http://www.biometrics.org/html/research.html`) has a repository of databases and research groups working on various biometric modalities

- The Computer Vision Research Lab, University of Notre Dame, has a dedicated webpage (`http://www.nd.edu/~cvrl/CVRL/Data_Sets.html`) that provides a number of biometric databases. The Multiple Biometric Grand Challenge (MBGC) database, Face Recognition Grand Challenge (FRGC) and Iris Challenge Evaluation (ICE2005) are becoming de-facto data for experiments as they have been used by NIST for performance evaluation. Other databases have been collected for evaluating biometric performance over time where ageing may become an important factor, on twin subjects, and in cross-device matching.

- The Biosecure database was collected by an EU project bearing the same name [1]. An important feature of this database is its breadth coverage of scenarios, which contain Internet, desktop, and mobile scenarios. Apart from the multi-modal aspect of the data, two devices have been used. This allows the use of cross-device matching.

- NIST series of challenges have produced a number of face, fingerprint, iris as well as multi-modal databases. The link http://www.nist.gov/itl/iad/ig/special_dbases.cfm is a good starting point, summarising software resources and databases, as well other dedicated sub-domain webpages such as

  - fingerprint.nist.gov
  - face.nist.gov
  - iris.nist.gov
  - mbark.nist.gov (for multi-modal)

  Other resources include biometric standards, biometric evaluation tool kit, and biometric web services (for mobile and remote authentication).

- Fingerprint vendor tests have also produced number of versions of fingerprint databases, including synthetic ones.

- The Face recognition homepage (`http://www.face-rec.org/databases`) is an independent initiative that publishes an ever-growing list of face databases. This website is regularly updated and so the latest face databases are likely to be found there.

Having presented some online repositories, the following sections describe some of the popular databases and how they can be used.

2.2 Face Databases

Face is one of the most popular biometrics and it is commonly used in biometric recognition systems. Face Databases are imagery data that are collected and used to evaluate the performance of face recognition systems. In the literature, a number of face databases exist in order to evaluate the performance of face recognition system. They are often designed to evaluate the following factors known to influence the system performance [2]:

- face image resolution,
- facial image quality,
- head orientation,
- facial expression,
- lighting conditions,
- occlusion,
- ageing and facial surgery.

In this section, we give a brief survey of popular face databases. Since there are over thirty publicly available face databases, we only select the most popular ones that are reported in the literature.

2.2.1 AR Database

The AR database was collected at the Computer Vision Centre in Barcelona, Spain in 1998. It contains images of 116 individuals (63 men and 53 women). The imaging and recording conditions (camera parameters, illumination setting, camera distance) were carefully controlled and constantly recalibrated to ensure that settings are identical across subjects. The resulting RGB colour images are 768×576 pixels in size. The subjects were recorded twice at a 2week interval. During each session 13 conditions with varying facial expressions, illumination and occlusion were captured.

2.2.2 AT&T Face Database

One of the first and most used databases is AT&T (formerly Olivetti ORL) database [3] that contains 10 different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, with varying the lighting conditions, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All images were taken against dark homogeneous background with the subjects in an upright, frontal position. The more information about this database can be found at `http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html`.

2.2.3 BANCA Database

The BANCA database [4] is a realistic and challenging multi-modal database intended for training and testing multi-modal verification systems. The BANCA database was captured in four European languages in two modalities (face and voice). For recording, both high and low quality microphones and cameras were used. The subjects were recorded in three different scenarios, controlled, degraded and adverse over 12 different sessions spanning three months. In total 208 people were captured, half men and half women.

The information of this database can be found at `http://www.ee.surrey.ac.uk/CVSSP/banca/`.

2.2.4 BiosecurID database

The BiosecurID Multi-modal Biometric database was acquired by a consortium of six Spanish Universities in the framework of the BiosecurID project [5]. The database includes eight uni-modal biometric traits, namely: speech, iris, face (still images, videos of talking faces), handwritten signature and handwritten text (on-line dynamic signals, off-line scanned images), fingerprints, hand (palmprint, contour-geometry) and key-stroking. The database comprises 400 subjects captured in 4 different sessions in a 8 month time span.

The face subset contains for each of the 4 sessions: 4 frontal images (BMP not compressed), with no specific background conditions (except that no moving objects are permitted); 1 video sequence of five seconds saying an eight digit PIN assigned to the subject. Both the audio (PCM 8 bit) and video (29 frames per seconds) are captured with the webcam Philips ToUcam Pro II. No movement in the background is permitted. Thus, there are a total of $400 \times 4 \times 4 = 6,400$ still images, and $400 \times 4 = 1,600$ face videos. This dataset will soon be made available through the Biometric Recognition Group-ATVS webpage[1].

2.2.5 CAS-PEAL Database

The CAS-PEAL (pose, expression, accessory, lighting) Chinese face database [6] was collected at the Chinese Academy of Sciences (CAS) between August 2002 and April 2003. It contains images of 66 to 1040 subjects (595 men, 445 women) in seven categories: pose, expression, accessory, lighting, background, distance, and time. For the pose subset, nine cameras distributed in a semicircle around the subject were used. Images were recorded sequentially within a short time period (2 seconds). In addition, subjects were asked to look up and down (each time by roughly 30) for additional recordings resulting in 27 pose images. The current database release includes 21 of the 27 different poses. It may also be used for eye detection, face pose estimation, and facial expression recognition. The more information on this database can be found at `http://www.jdl.ac.cn/peal/index.html`. The release agreement can also be found at `http://www.jdl.ac.cn/peal/index.html`.

2.2.6 CMU Pose, Illumination, and Expression (PIE) Database

The CMU (Carnegie Mellon University) PIE database [7] was collected between October and December 2000. CMU PIE Database is one of the largest datasets developed to investigate the affect of Pose, Illumination and Expression. The CMU PIE database contains 41,368 images obtained from 68 individuals. The subjects were imaged in the CMU 3D Room using a set of 13 synchronized high-quality colour cameras and 21 flashes. The resulting RGB colour images are 640×480 in size. The more information about this

[1] `http://atvs.ii.uam.es/`

database can be found at http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261.

2.2.7 The Colour FERET Database

The FERET program set out to establish a large database of facial images that was gathered independently from the algorithm developers [8]. Dr. Harry Wechsler at George Mason University was selected to direct the collection of this database. The database collection was a collaborative effort between Dr. Wechsler and Dr. Phillips of NIST. The images were collected in a semi-controlled environment. To maintain a degree of consistency throughout the database, the same physical set-up was used in each photography session. Because the equipment had to be reassembled for each session, there are some minor variations in images collected on different dates. The FERET database was collected in 15 sessions between August 1993 and July 1996. The database contains 1564 sets of images for a total of 14,126 images that includes 1199 individuals and 365 duplicate sets of images. A duplicate set is a second set of images of a person already in the database and was usually taken on a different day. For some individuals, over two years had elapsed between their first and last sittings, with some subjects being photographed multiple times. This time lapse was important because it enabled researchers to study, for the first time, changes in a subject's appearance that occur over a year.

2.2.8 Face Recognition Grand Challenge (FRGC) Database

The FRGC database [9] is jointly sponsored by several government agencies interested in improving the capabilities of face recognition technology. The primary goal of the FRGC database is to evaluate face recognition technology. It may also be used for eye detection.

The Face Recognition Grand Challenge (FRGC) project has officially ended, however, researchers can still obtain FRGC data for experimentation in their organization. Data and software licenses will need to be signed by legal authorities who are approved to sign licenses on behalf of your organization. If you are interested in obtaining the FRGC data set, please contact the FRGC Liaison at frgc@nist.gov. The request must come from a full-time employee or faculty member of his/her organization/university.

2.2.9 Japanese Female Facial Expression (JAFFE) Database

The JAFFE database [10] contains 213 images of 10 Japanese female models obtained in front of a semi-reflective mirror. Each subject was recorded three or four times while displaying the six basic emotions and a neutral face. The camera trigger was controlled by the subjects. The resulting images have been rated by 60 Japanese women on a 5-point scale for each of the six adjectives. The rating results are distributed along with the images. The JAFFE database is available free of charge for use in non-commercial research. The more information about this database can be found at http://www.kasrl.org/jaffe.html.

2.2.10 Korean Face Database (KFDB)

The Korean Face Database (KFDB) contains facial imagery of a large number of Korean subjects collected under carefully controlled conditions [11]. Similar to the CMU PIE database, this database has images with varying pose, illumination, and facial expressions. There are 1000 subjects in the database. Similar to the CMU PIE database, the images with varying pose, illumination, and facial expressions were recorded. Pose images were collected in three styles: natural (no glasses, no hair band to hold back hair from the forehead), hair band, and glasses. The subjects were also asked to display five facial expressions neutral, happy, surprise, anger, and blink which were recorded with two different coloured lights, resulting in 10 images per subject. In total, 52 images were obtained per subject. The more information about this database can be found at `http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261`.

2.2.11 MOBIO database

The MOBIO database [12] 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 at the various institutes. However, the same office was not always used. This meant that the recordings do not have a controlled background, nor are 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. The MOBIO database[2] is publicly available for research purposes and can be downloaded after completing and signing an End User License Agreement (EULA).

2.2.12 Multiple Biometric Grand Challenge (MBGC) Database

The goal of the Multiple Biometrics Grand Challenge (MBGC) [13] is to improve the performance of face and iris recognition technology from biometric samples acquired under unconstrained conditions. The MBGC is organized into three challenge problems. Each challenge problem relaxes the acquisition constraints in different directions. In the Portal Challenge Problem, the goal is to recognize people from near-infra-red (NIR) and high definition (HD) video as they walk through a portal. Iris recognition can be performed from the NIR video and face recognition from the HD video. The availability of NIR and HD modalities allows for the development of fusion algorithms. The Still Face Challenge

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

Problem has two primary goals. The first is to improve recognition performance from frontal and off angle still face images taken under uncontrolled indoor and outdoor lighting. The second is to improve recognition performance on still frontal face images that have been resized and compressed, as is required for electronic passports. In the Video Challenge Problem, the goal is to recognize people from video in unconstrained environments. The video is unconstrained in pose, illumination, and camera angle. All three challenge problems include a large data set, experiment descriptions, ground truth, and scoring code.

The Multiple Biometric Grand Challenge (MBGC) project has officially ended, however, researchers can still obtain MBGC data for experimentation in their organization. Data and software licenses will need to be signed by legal authorities who are approved to sign licenses on behalf of your organization. Anybody interested in obtaining the MBGC data set, should contact the MBGC Liaison at mbgc@nist.gov. The request must come from a full-time employee or faculty member of his/her organization/university.

2.2.13 XM2VTS Database

The Extended M2VTS (XM2VTS) multi-modal face database [14] includes still colour images, audio data, video sequences and 3D Model. It contains four recordings of 295 subjects taken over a period of four months. Each recording contains a speaking head shot and a rotating head shot. Videos/Image captured at resolution 720 by 576 pixels in a controlled environment. The testing for face verification is performed using the Lausanne protocol which splits the database into training, development and evaluation sets. The information of this database can be found at `http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/`.

2.2.14 Yale Face Database

The Yale Face Database [15] contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: centre-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink. The database is publicly available for non-commercial use. The more information about this database can be found at `http://cvc.yale.edu/projects/yalefaces/yalefaces.html`.

2.3 Fingerprint Databases

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 which are the result of collaborative efforts in recent research projects. Examples of these joint efforts include previous European projects such as BioSec [16] or the BioSecure Network of Excellence [1] in addition to national projects such as the Spanish Ministerio de Ciencia y Tecnología (MCYT) [17] or BioSecureID [5] databases.

Apart from the fingerprint datasets included in multi-modal databases, other efforts have been directed 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 [18].

Over the last years have also been dedicated to the generation of synthetic fingerprint data, the most relevant being the Synthetic Fingerprint Generator (SFingGe) developed by the Biometric System Laboratory-DEIS from the University of Bologna [19].

2.3.1 MCYT-DB

The Multi-modal Biometric MCYT Database was acquired thanks to the joint efforts of four Spanish academic institutions [17]. It contains fingerprint and signature of 330 subjects captured in one single acquisition session. The fingerprint subset comprises 12 samples of the eight fingers and two thumbs of each subject captured with two different sensors (optical and capacitive), thus leading to a total of 79,200 fingerprint images. One of the unique characteristics of this database is the different control level with which the different samples of each finger were acquired: three present a high control level, three a medium control level, and the remaining six a low control level.

2.3.2 Biosec-DB

This database was acquired in the framework of the FP6 EU BioSec Integrated Project [16]. The corpus consist of fingerprint images acquired with three different sensors, frontal face images from a webcam, iris images from an iris sensor, and voice utterances acquired both with a close-talk headset and a distant webcam microphone. The BioSec baseline corpus includes real multi-modal data from 200 individuals in 2 acquisition sessions. The fingerprint sub-corpus comprises 4 samples captured with 3 different sensors (thermal, optical and capacitive) of the index and middle fingers of both hands, thus completing a total of 19,200 fingerprint images. Part of this fingerprint dataset was used in the Fingerprint Verification Competition 2006 (see short description below) and is publicly available at the Biometric Recognition Group-ATVS webpage[3].

2.3.3 BiosecurID-DB

The BiosecurID Multi-modal Biometric database was acquired by a consortium of six Spanish Universities in the framework of the BiosecurID project [5]. The database includes eight uni-modal biometric traits, namely: speech, iris, face (still images, videos of talking faces), handwritten signature and handwritten text (on-line dynamic signals, off-line scanned images), fingerprints, hand (palmprint, contour-geometry) and keystroking. The database

[3] `http://atvs.ii.uam.es/`

comprises 400 subjects captured in 4 different sessions in a 8 month time span. The fingerprint subset contains four samples per session (four sessions) captured with 2 different sensors (thermal and optical) of the index and middle fingers of both hands, thus completing a total of 51,200 fingerprint images. This dataset will soon be made available through the Biometric Recognition Group-ATVS webpage[4].

2.3.4 Biosecure-DB

As was mentioned already in the description of the iris-related databases, the acquisition of the Multi-modal Biosecure-DB was conducted by 11 European institutions participating in the BioSecure Network of Excellence [1]. The fingerprint sub-corpus comprised in the dataset 2 (captured under controlled conditions) contains the thumb, index and middle fingers of both hands of 650 users, captured in two separate acquisition sessions, with two different sensors (thermal and capacitive), with 2 samples per finger and per session (a total of 16,008 fingerprint images). Moreover, it is compatible in terms of sensors used and protocols followed with Biosec and BiosecurID. This database is publicly available through the Biosecure Foundation[5].

2.3.5 The series of FVC-DBs

The Biometric System Laboratory-DEIS from the University of Bologna has been leading since the year 2000, on a biannual basis, the series of Fingerprint Verification Competitions (FVC 2000, FVC 2002, FVC 2004 and FVC 2006). The organization of these independent evaluation campaigns has required the acquisition of different very valuable fingerprint datasets which have become a standard for the assessment of fingerprint recognition systems performance [18]. These datasets include:

- FVC 2000. Three real datasets captured respectively with two optical and a capacitive sensor, and a dataset of synthetic fingerprints generated with the tool SFinGe. All the four datasets comprise 110 different fingers, with 8 samples per finger.
- FVC 2002. As in the previous edition three real datasets and one synthetic (generated with the SFinGe tool) were used in the competition. The real datasets were captured with two optical and a capacitive sensors (different from those used in FVC 2000). Again all the four datasets contain 110 different fingers, with 8 samples per finger.
- FVC 2004. Again, four datasets were used in the competition, three real (captured with two optical sensors and a thermal sweeping sensor) and one synthetic. The sample distribution is also the same as in the previous editions: 110 classes, 8 samples per class. In this case low quality samples were intentionally acquired (very high skin distortion, very dry or wet skin, dirty acquisition sensors...) in order to test the systems under very challenging conditions.

[4] http://atvs.ii.uam.es/
[5] http://biosecure.it-sudparis.eu/AB/

- FVC 2006. Following the protocol used in the previous editions three real datasets (captured with an optical, capacitive and thermal sweeping sensors) and one synthetic (generated with version 3.0 of SFinGe) were released for FVC 2006. In this case there were 150 users, with 12 samples per user. No deliberate distortions were introduced in the images.

2.3.6 SFinGe-DB

Up to date, the most efficient method for the generation of synthetic fingerprint samples was proposed in [19] and was further developed and described in [20]. This fingerprint synthetic generation method was embedded in an automatic software tool named SFinGe which may be obtained from its developer's webpage, the Biometric System Laboratory-DEIS from the University of Bologna [6].

This automatic application was used for the generation of the synthetic datasets which were part of the series of FVC competitions. The software permits to have control over the number of classes (i.e., fingers), number of samples per class and quality of the samples which will be generated.

2.4 Iris Databases

Currently, several real iris databases have been captured and some synthetic iris datasets are also available. The most relevant of each type, real and synthetic, are presented next.

2.4.1 MBGC-DB

The Multiple Biometric Grand Challenge is a project with the goal to investigate, test and improve performance of face and iris recognition technology [21]. Among the technology development areas within the MBGC, the MBGC portal challenge database provides Near Infra-red (NIR) iris still images and videos.

2.4.2 ICE-DB

The iris image datasets used in the Iris Challenge Evaluations (ICE) in 2005 and 2006 [22] 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.

2.4.3 MMU-DB

The Multimedia University has released two iris databases. The MMU1 iris database [23] comprises a total number of 450 iris images which were collected using a LG IrisAccess2200

[6] http://bias.csr.unibo.it/research/biolab

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 belong to 100 volunteers of different age and nationality. Each of them gives 5 iris images for each eye.

2.4.4 UBIRIS-DB

The UBIRIS.v1 database (2004) [24] comprises 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) [25], 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.

2.4.5 BATH-DB

The BATH iris database [26] 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 formed by 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 infra-red illuminator and a consistent image capture setup.

2.4.6 BiosecurID-DB

The BiosecurID Multi-modal Biometric database was acquired by a consortium of six Spanish Universities in the framework of the BiosecurID project [5]. The database includes eight uni-modal biometric traits, namely: speech, iris, face (still images, videos of talking faces), handwritten signature and handwritten text (on-line dynamic signals, off-line scanned images), fingerprints, hand (palmprint, contour-geometry) and keystroking. The database comprises 400 subjects captured in 4 different sessions in a 8 month time span. The iris subset contains 4 samples per session and per iris, thus completing a total of $400 \times 2 \times 4 \times 4 = 12,800$ iris images captured with the LG Iris Access EOU 3000 camera. The samples were acquired removing the glasses, while the use of contact lenses is saved in a non-biometric data file. This dataset will soon be made available through the Biometric Recognition Group-ATVS webpage[7].

[7] `http://atvs.ii.uam.es/`

2.4.7 BioSecure-DB

The BioSecure database [1] 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. It is publicly available and can be obtained from the BioSecure Foundation website [8].

2.4.8 CASIA-IrisV3-DB

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 [27]. It is publicly available and can be obtained from the Centre for Biometrics and Security Research website [9].

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 infra-red illuminator. 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 a close-up iris camera which comprises a circular NIR LED array, with suitable luminous flux for iris imaging that 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 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.

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

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

2.4.9 CiTER Texture-Based Synthetic Iris Database

In this database the synthetic irises are generated in two stages following the algorithm described in [28]. In the first stage, a Markov Random Field model is used to generate a background texture representing the global iris appearance. In the next stage, a variety of iris features, viz., radial and concentric furrows, collaret and crypts, are generated and embedded in the texture field. It contains 1000 different synthetic subjects with 7 samples per subject and it is publicly available at the CiTER website [10].

2.4.10 CiTER Model-Based Synthetic Iris Database

The gallery of synthetic iris images are generated in five steps using the model-based, anatomy-based approach proposed in [29], with 40 controllable random parameters such as fibre size, pupil size, iris thickness, top layer thickness, fibre cluster degree, iris root blur range, the location of the collaret, the amplitude of the collaret, top layer transparency parameter, eye angle, eye size etc.. It contains 10,000 different irises, with 16 images per subject, 1 good quality image, 15 degraded images, with combination effects of noise, rotation, blur, motion blur, low contrast and specular reflection. For each image segmentation results (unwrapped template, enhanced template and occlusion mask) are provided. It is also publicly available at the CiTER website.

2.5 Vein Databases

Vein recognition systems are amongst the newest biometric technologies to have emerged in recent years. Vein authentication uses the vascular patterns of an individual's palm / finger / back of the hand as personal identification data. Veins and other subcutaneous features in the human hand present large, robust, stable and largely hidden patterns. The de-oxidized haemoglobin in the vein vessels absorbs light having a wavelength in the near-infra-red area. When an infra-red ray image is captured only the blood vessel pattern containing the de-oxidized haemoglobin are visible as a series of dark lines. Based on this feature, the vein authentication device translates the black lines of the infra-red ray image, and then matches it with the previously registered pattern of the individual.

Vein authentication technology consists of a small vein scanner - the users simply need to hold the palm / finger / back of hand a few centimetres over the scanner and the scanner reads the unique vein pattern. Vein recognition works by recording subcutaneous Infra Red (IR) absorption patterns to produce unique and private identification templates for users. Subcutaneous features can be conveniently imaged within the wrist, palm, and dorsal surfaces of the hand.

There are number of vein (hand vein, finger vein, palm vein) databases available in the literature. In this section, we provide a survey of popular vein databases.

[10] http://www.citer.wvu.edu/

2.5.1 Bosphorus Hand Vein Database

The hand vein data is acquired using NIR imaging technology with a monochrome NIR CCD camera (WAT-902H2 ULTIMATE) equipped with an infra-red lens. The back of the hand is irradiated by two IR light sources. The images have 300×240 pixel size with a gray-scale resolution of 8-bit.

There are overall 1575 images of the left hands of 100 subjects distributed as:

- Three left-hand images per subject taken under normal conditions (N: Normal),
- Three left-hand images per subject after having carried a bag weighing 3 kg. for one minute (B: Bag),
- Three left-hand images per subject after having squeezed an elastic ball repetitively (closing and opening) for one minute (Activity: A),
- Three left-hand images per subject after having cooled the hand by holding an ice pack on the surface of the back of the hand (Ice: I).
- Three right-hand images per subjects under normal conditions.
- Images of the left hands of 25 subjects after a time lapse ranging from two months to five months.

Here are the website addresses that you can find more information on Bosphorus hand vein database:

- The licence agreement can be found at `http://bosphorus.ee.boun.edu.tr/hand/HowtoObtain.aspx`.
- The privacy statement can be found at `http://bosphorus.ee.boun.edu.tr/hand/privacy.aspx`.
- The terms of use can be found at `http://bosphorus.ee.boun.edu.tr/hand/terms.aspx`.

2.5.2 CASIA Multi-Spectral Palmprint Database

CASIA Multi-Spectral Palmprint Image Database contains $7,200$ palm images captured from 100 different people using a self-designed multiple spectral imaging device. All palm images are 8 bit gray-level JPEG files. For each hand, we capture two sessions of palm images. The time interval between the two sessions is more than one month. In each session, there are three samples. Each sample contains six palm images which are captured at the same time with six different electromagnetic spectrum. Wavelengths of the illumination corresponding to the six spectrum are $460nm, 630nm, 700nm, 850nm, 940nm$ and white light respectively. Between two samples, a certain degree of variations of hand postures are allowed. The aim is to increase diversity of intra-class samples and simulate practical use.

There are no pegs to restrict postures and positions of palms in the device. Subjects are required to put his palm into the device and lay it before a uniform-coloured background. The device supplies an evenly distributed illumination and captures palm images using a CCD camera fixed on the bottom of the device. A control circuit is designed in order to adjust spectrum automatically.

The database is released for research and educational purposes. They accept no liability for any undesirable consequences of using the database. All rights of the CASIA database are reserved. No person or organization is permitted to distribute, publish, copy, or disseminate this database. In all documents and papers that report experimental results based on this database, their efforts in constructing the database should be acknowledged as: Portions of the research in this paper use the CASIA-MS-Palmprint V1 collected by the Chinese Academy of Sciences' Institute of Automation (CASIA) and a reference to "CASIA-MS-Palmprint V1, http://biometrics.idealtest.org/" should be included. A copy of all reports and papers that are for public or general release that use the CASIA-MS-Palmprint V1 should be forwarded upon release or publication to Professor Tieniu Tan whose contact details can be found at `http://biometrics.idealtest.org/dbDetailForUser.do?id=6`.

2.5.3 PKU Finger Vein Database

The PKU Finger Vein Database (V2) contains 4574 gray-scale images corresponding to 431 different fingers, which were collected on the first semester of 08-09 school year. This database was founded earlier than that of D1016, however, this database had not been hand-picked when it was created on December 1st 2008.

The PKU Finger Vein Database (V3) - Test Version contains 5379 gray-scale images corresponding to 398 different fingers, which was founded later than that of D1015, also, this database had been hand-picked when it was created on May 22nd 2009.

The PKU Finger Vein Database (V4) contains 1597 gray-scale images corresponding to 200 different fingers, each having 8 samples. This database is picked by A.I. program from a large database first, and then picked by hand to further ensure that the images are in good quality. So it is a good choice for those who are new to this area, and also a good choice for more advanced researchers.

All rights of the PKU Finger Vein Database (V2, V3, and V4) are reserved by the Artificial Intelligence Lab of Peking University. The database is only available for research and non-commercial purposes. Commercial distribution or any act related to commercial use of this database is strictly prohibited.

The more information on this database can be found at `http://rate.pku.edu.cn/RATE/listdatabases`.

2.6 Multi-modal Databases

A multi-modal biometric database implies the inclusion of more than one biometric modalities. In many audio-visual related applications, ace and speech biometric modalities are arguably the most common ones, early examples are M2VTS which includes face, speech, and talking lip modalities, its extended version called XM2VTS, and the BANCA database. Table 1 lists some existing multi-modal biometric databases.

- **BM** databases [1]. The Biosecure Multi-modal database was jointly conducted by 11 European institutions participating in the BioSecure Network of Excellence. BMDB comprised of three different datasets which are DS1 is captured on Internet scenario, DS2 is on Desktop scenario and DS3 is on Mobile scenario. The three datasets of BMDB include a common part of audio and video data, which comprises still face images and talking face videos. Also, signature and fingerprint data has been acquired both in DS2 and DS3. Additionally, hand and iris data was acquired in DS2.

- **BiosecurID** database [5]. The BiosecurID database was collected in 6 Spanish institutions in the framework of the BiosecurID project funded by the Spanish Ministry of Education and Science. It has been collected in an office-like uncontrolled environment (in order to simulate a realistic scenario), and was designed to comply with the following characteristics: 400 subjects, 8 different traits (speech, iris, face still and talking face, signature, handwriting, fingerprint, hand and keystroking) and 4 acquisition sessions distributed in a 4 month time span.

- **BioSec** database [16]. BioSec was an Integrated Project (IP) of the 6th European Framework Programme which involved over 20 partners from 9 European countries. One of the activities within BioSec was the acquisition of a multi-modal database. This database was acquired at four different European sites and includes face, speech (both with a webcam and a headset), fingerprint (with three different sensors) and iris recordings. The baseline corpus [16] comprises 200 subjects with 2 acquisition sessions per subject. The extended version of the BioSec database comprises 250 subjects with 4 sessions per subject (about 1 month between sessions). A subset of this database was used in the last International Fingerprint Verification Competition [30] held in 2006.

- **MyIDEA** database [31], which includes face, audio, fingerprint, signature, handwriting and hand geometry of 104 subjects. Synchronized face-voice and handwriting-voice were also acquired. Sensors of different quality and various scenarios with different levels of control were considered in the acquisition.

- **BIOMET** database [32], which offers 5 different modalities: audio, face images (2D and 3D), hand images, fingerprint (with an optical and a capacitive sensor) and signature. The database consists of three different acquisition sessions (with 8 months between the first and the third) and comprises 91 subjects who completed the three sessions.

Table 1: Multi-modal databases taken from [1].

Database	Year	Users	Sessions	Traits	2Fa	3Fa	Fp	Ha	Hw	Ir	Ks	Sg	Sp
MOBIO	2010	150	12	2	x								x
BM	2008	971 (DS1)	2	2	x								x
		667 (DS2)	$2^{\dagger_2}$	3	x					x			x
		713 (DS3)	2	4	x								x
BiosecurID	2007	400	4	8	x		x	x	x	x	x	x	x
BioSec	2007	250	4		x		x			x			x
MyIDEA	2005	104	3	6	x		x	x	x			x	x
BIOMET	2003	91	3	6	x	x	x	x				x	x
MBioID	2007	120	2	6	x	x	x			x		x	x
M3	2006	32	3	3	x		x						x
FRGC	2006	741	variable	2	x	x							
MCYT	2003	330	1	2			x					x	
BANCA	2003	208	$4^{\dagger_1}$	2	x								x
Smartkom	2002	96	variable	4			x	x				x	x
XM2VTS	1999	295	4	2	x								x
M2VTS	1998	37	5	2	x								x
BT-DAVID	1999	124	5	2	x								x

$\dagger_1$: In each of the 4 sessions, data were collected in three acquisition conditions (controlled, adverse, and degraded).
$\dagger_2$: In each of the 2 sessions, 2 devices were used.

2.7 Score-level Databases for Fusion Algorithms

Apart from various uni-modal and multi-modal biometric databases, for the purpose of evaluating score calibration procedures and fusion algorithms, a common set of matching scores are often required. By keeping the same score data set and changing only a score calibration or fusion module, one can systematically examine the effectiveness of the latter module, independent of the feature extraction and matching modules. This approach can clarify the gain brought by score calibration or fusion which would have otherwise conflated by modifications introduced to the feature extraction and matching modules of its constituent uni-modal biometric sub-systems.

The following characteristics are desirable for a biometric score database:

- Scores are made publicly available for download. This ensures that different methods can be compared on the same baseline and hence, the results of which are comparable.
- The process in generating the scores is well documented. This includes proper documentation of the underlying database, its characteristics and context of application, and the underlying matching algorithms.

- Protocols specifying the training and test data sets are published.
- Real multi-modal biometrics rather than chimeric databases.

For the purpose of benchmarking score calibration and fusion, we argue that biometric matching scores should be made publicly available. This is in order to guarantee the repeatability of experiments. Furthermore, since it is impossible to reverse the process, i.e., generating the biometric sample directly from a matching score, there is literally no concern with respect to the subject's privacy.

There have been a number of score-level benchmark databases available. The earliest published one is possibly the XM2VTS score-level benchmark database [33]. Other databases include the NIST heterogeneous speaker verification systems score [34], the BANCA score data set[11], the NIST BSSR1 score data set[12], and the Biosecure DS2 score+quality database [35][13].

3 Protocols

An experimental protocol partitions a database into number of dataset in order to demonstrate the merit of a biometric system. For an instance, to show the merit of a pose-invariant face recognition system, a database is divided into number of data sets based on the head pose. The use of experimental protocols allow researchers to directly compare state-of-the-art systems with their own developed system without the need of rerunning the former systems which can be time consuming. More importantly, because of the complexity involved in a piece of face recognition algorithm, it is often not possible to replicate an experiment exactly if the source codes are not provided. Clearly, with a common and well designed experimental protocol, different systems can be benchmarked without the need of rerunning and re-implementing various systems.

When partitioning a database, there are at least two types of partitions which play very different roles: development and evaluation data sets. The former is used for algorithm development and tuning whereas the latter is used for reporting performance in an unbiased way as much as possible.

The discussion in the following section draws from our experience in participating and organising various biometric performance competition and evaluation campaigns, as well as design and collection of various databases. The campaigns organised by members of the consortium, include MOBIO Face and Speaker Verification Evaluation [12] organised in conjunction with ICPR2010, Face Video Competition organised in conjunction with ICB2009 [36], and a benchmark of fusion algorithms organised in conjunction with the Biosecure Workshop 2007 [37].

[11] http://personal.ee.surrey.ac.uk/Personal/Norman.Poh/web/banca_multi
[12] http://www.nist.gov/itl/iad/ig/biometricscores.cfm
[13] http://personal.ee.surrey.ac.uk/Personal/Norman.Poh/web/fusionq

3.1 Development and Evaluation Data Sets

A *development* set is used for developing a biometric system. This includes tuning any hyper-parameters of the system, including determining the decision threshold. Examples of hyper-parameters are number of hidden units of a multilayer perceptron, the number of Gaussians in Gaussian Mixture Models, the number of orientations and frequencies for Gabor filters, the number of filter banks to be used when representing a speech signal, and the parameters needed when segmenting an iris image. An *evaluation* set is used uniquely to report the system performance using various metrics discussed in Deliverable D3.3.

For each of the two data set, FNMR and FMR have to be computed. One of the rules is to give sets of matching and non-matching pairs in order to address the problem whether a sample pair belongs to the same individual. An alternative rule is further to make distinction between two important data partitions: enrolment and test sets. An *enrolment* data set serves uniquely to enrol users into a biometric system. The biometric samples in this data set constitute typically biometric *templates* in a *gallery*.

A test set, on the other hand, contains data that are used to test the system. A biometric sample in this data set is also called a *probe* or test sample. This data set includes samples of imposters (external users) and gallery users. Imposter samples are used to measure the distribution of non-matching scores, while gallery samples are for the distribution of matching scores. Therefore, the gallery samples in the test set only share the subjects but not the samples with the enrolment data set. In contrast to the former rule, it is possible to address a problem whether a sample is an example of a particular individual.

In FRGC V2.0 database [38], its protocol does not separate the database into development and evaluation sets. A verification rate at 0.1% FAR (False Acceptance Rage) is used to report the system accuracy and therefore the decision threshold is unnecessary. The shortcoming of this protocol is that the quoted performance will be biased, and overly optimistic on this database, as it will not be representative of an independent test result. On the other hand, in Lausanne protocol [39], although it partitions the database into development and evaluation sets, it uses the same enrolment samples. Thus, the quoted performance will still be biased or over-fitted as the system hyper-parameters are normally tuned by using the enrolment samples. In general, the gallery and imposter subjects between development and evaluation sets are desired to be non-overlapping to reduce the degree of performance bias. Certainly, the degree of bias can be further verified by testing the same system along with all the hyper-parameters fixed on several independent databases [40].

3.2 Background/cohort data set

In addition to the above data partitions, we also optionally reserve another data partition called a *background* data set. It is so called because many algorithms require an external data set to initialise the system parameters. For instance, face recognition systems that rely on PCA and LDA algorithms often require this background data in order to build their respective projection matrix. Another example is the state-of-the-art GMM classifier used for speaker verification. This classifier requires a background data set to represent non-target users, i.e., users who are not part of the development or evaluation data sets.

The advantage of including a background data set is to remove any variation in performance caused by the choice of this data set. In the competition use-case, if two participants use two LDA-based algorithm, any observed variation in performance cannot be attributed to the choice of background data set if a common one has been used. This will help us to understand why one system is better than the other. In this example, it is clear that from a scientific perspective, it is important to restrict the choice of background data set.

The cohort data set can also be used for experiments or competitions where *cohort-based score normalisation* is permitted. In this setting, the biometric samples of subjects in the cohort set (or "cohort subjects") participate in each biometric comparison as if they are templates in the gallery. Therefore, if there are C cohort samples, a query involving one probe sample actually entails $C + 1$ comparisons. An example of cohort-based score normalisation technique is T-norm [41]. Another example of cohort-based classifier is One-Shot similarity (OSS) measure [42]. It builds the binary classifier using the cohort samples and test pair. Although being a simple technique, it can often boost the system performance by a significant margin, often decreasing false match rate (FMR) (see D3.3) without increasing false non-match rate (FNMR).

In summary, setting aside a background data set that is dedicated to algorithm development can remove variation to this choice. This is an advantage when comparing algorithms because the choice of background data set has been controlled for. Consequently, any variation in performance between two algorithms cannot be attributed to this factor. This allows a better interpretation of experimental results.

3.3 Some Properties

Table 2 illustrates the different partitions of data whereas Table 3 illustrates the different partitions of users in a cross-tabulated format of development versus evaluation set and enrolment (gallery) versus test (probe) set. We also introduce the variables that represent them, reusing the same notation in D3.3: $\mathcal{U}$ for user index set and $\{\mathcal{G}, \mathcal{B}, \mathcal{C}, \mathcal{E}\} \subset \mathcal{X}$ for denoting biometric samples in the enrolment set (gallery), test (probe) set, and background (cohort) set and imposter set respectively, recalling that $\mathcal{X}$ is the domain of biometric features. We use the superscript *dev* to denote the development set and similarly *eva* to denote the evaluation set This notation is extremely flexible as we can describe the users in $dset \in \{dev, eva\}$ of partition $\{\mathcal{G}, \mathcal{B}, \mathcal{C}, \mathcal{E}\} \subset \mathcal{X}$ as $\mathcal{U}_{\mathcal{X}}^{dset}$.

The second row of Table 3 states that users who are present in the gallery should be present in the probe set. However, in addition, the probe set should contain out-of-sample (external) users whom the system has never seen. This is consistent with the open-set identification scenario or a typical authentication scenario. The second row of Table 2 reflects that of Table 3. It states that a probe set should contain samples of enrolled users as well as that of out-of-sample (external) users, represented by $\mathcal{B}^{1,dset}$ and $\mathcal{B}^{0,dset}$, respectively for $dset \in \{dev, eva\}$.

Table 2: Data partitions

	Background (Cohort)	Development	Evaluation
Enrolment (Gallery)	$\mathcal{C}$	$\mathcal{G}^{dev}$	$\mathcal{G}^{eva}$
Test (Probe)		$\mathcal{B}^{dev} = \mathcal{B}^{1,dev} \cup \mathcal{B}^{0,dev}$	$\mathcal{B}^{eva} = \mathcal{B}^{1,eva} \cup \mathcal{B}^{0,eva}$

Table 3: User partitions

	Background (Cohort)	Development	Evaluation
Enrolment (Gallery)	$\mathcal{U}_{\mathcal{C}}$	$\mathcal{U}_{\mathcal{G}}^{dev}$	$\mathcal{U}_{\mathcal{G}}^{eva}$
Test (Probe)		$\mathcal{U}_{\mathcal{B}}^{dev} = \mathcal{U}_{\mathcal{G}}^{dev} \cup \mathcal{U}_{\mathcal{E}}^{dev}$	$\mathcal{U}_{\mathcal{B}}^{eva} = \mathcal{U}_{\mathcal{G}}^{eva} \cup \mathcal{U}_{\mathcal{E}}^{eva}$

It is important to ensure that the background, development, and evaluation data sets are *mutually exclusive*, i.e, $\mathcal{X}' \cap \mathcal{X}'' = \emptyset$ for all

$$\mathcal{X}', \mathcal{X}'' \in \{\mathcal{C}, \mathcal{G}^{dev}, \mathcal{G}^{eva}, \mathcal{B}^{1,dev}, \mathcal{B}^{1,eva}\}.$$

In the table, for the background data set, it is not necessary to make the distinction between the enrolment and the test data set because such distinction is often not required.

The development set should contain samples that are representative of the evaluation data set. However, by the mutual exclusivity criterion, both of them must contain different subjects.

The concepts of development and evaluation sets are common for biometric authentication/verification. For instance, they are used for NIST speaker evaluation campaigns [43]. However, for biometric identification, especially closed-set identification, the two data sets are often not distinguished. This is fine as long as an identification system has no adjustable hyper-parameter. In reality, this is not always the case. Therefore, even for biometric identification, a competition team should be given some development set in order to allow them to tune their system.

Regarding the size between the development and the evaluation set, we recommend two settings, depending on the purpose of an experiment. A ratio that favours a larger evaluation set is recommended if the objective is to measure generalisation performance to a precision as high as possible. This ratio could be, for instance, 1:10.

Another practical choice proposed in [44] to avoid the use of large database is to divide the database to 10 subsets and then apply leave-one-out cross validation scheme. In each experiment, nine of the subsets should be combined to form a development set, with the tenth subset used for evaluation.

3.4 Evaluating Fusion Algorithms

In the case of a two-fold cross-validation stated above, it is possible to assess the performance of a fusion algorithm by using one fold of data to train a fusion classifier and another fold as a test data. The role of the two data sets can be reversed. The resultant performance metrics can be combined to report the final performance.

To do so, we can reuse the notation established in D3.3. Let $s(g, b)$ be a similarity score by combining a gallery sample (template), $g \in \mathcal{G}^{dset}$ with a probe $b \in \mathcal{B}^{dset}$, for each of $dset \in \{dev, eva\}$. Let the resultant set of scores be $\mathcal{S}_*^{dset}$ which contains both match and non-match samples. The two-fold experiment for fusion works as follow:

- Train a fusion classifier on $\mathcal{S}_*^{dev}$ and test it on $\mathcal{S}_*^{eva}$ to get a metric m_1.
- Train another fusion classifier on $\mathcal{S}_*^{eva}$ and test it on $\mathcal{S}_*^{dev}$ to get a metric m_2.
- Report the final metric which is a combination of m_1 and m_2.

A two-fold fusion experimental protocol can be found in [33]. The above idea can be extended easily to k-fold cross validation.

Metric combination has not been discussed but the idea can be explained intuitive. If a metric consists of a pair of FMR and FNMR, it is straightforward to combine two such metrics by working out the total false matches and false non-matches of the two metrics. By normalising both quantities with their respect to the total non-match and match comparisons, one obtains the combined FMR and FNMR. This idea can be found in [33]. Moreover, the idea can be generalized to combining several DET curves where each point on a DET curve (which is a pair of FMR and FNMR) across several experiments can be grouped together. One way of grouping these points is by a common decision threshold criterion, for instance, minimizing weighted error rate (WER) at a given weight β:

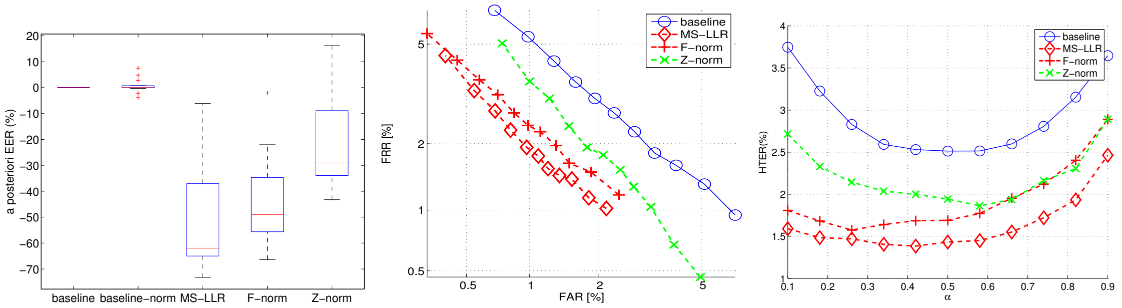
$$\text{WER}_\beta(\Delta) = (1 - \beta)\text{FNMR}(\Delta) + (\beta)\text{FMR}(\Delta) \tag{1}$$

where $\beta \in [0, 1]$ adjusts the contribution between FMR and FNMR (see Deliverable D3.3 for details).

Therefore, each point on a DET curve has an associated β value. If there are N DET curves, N metrics of (FMR,FNMR) can be combined to form a final combined metrics. This approach was reported in [45]. An example of combined EER, DET and EPC is shown in Figure 1.

4 Recommended databases and protocols for inclusion on the BEAT platform

In this section, we selected a few databases that will be considered for inclusion on the BEAT platform. The databases should satisfy the requirements of use-cases as described in Deliverable D2.1 as well as the legal requirements as set out in Deliverable D9.1.



(a) relative change of EER (b) DET (c) EPC

Figure 1: Performance of several methods, namely, baseline, Z-norm, F-norm and MS-LLR score normalization procedures, on 14 speaker verification experiments carried out on the XM2VTS database (a) in box-plots showing the distribution of relative change of EERs; (b) in *pooled* DET curves; and (c) in *pooled* EPC curves. The box in a box-plot contains the first and the third quantiles of EERs. The dashed lines ended with horizontal lines show the 95% confidence of the data. Outliers are plotted with "+". Diagram taken from [45].

4.1 Face modality

The MOBIO database

The MOBIO database [46] is unique because it is a bi-modal dataset including challenging face and speaker data that was captured almost exclusively on mobile phones. It consists of over 61 hours of audio-visual data with 12 distinct sessions usually separated by several weeks. In total there are 192 unique audio-video samples for each of the 150 participants, this is almost twice the size of Phase I of the MOBIO database. This data was captured at 6 different sites over one and a half years with people speaking English (see more details on Table 4).

Capturing the data on mobile phones makes this database unique because the acquisition device is given to the user, rather than being in a fixed position. This means that the microphone and video camera are no longer fixed and are now being used in an interactive and uncontrolled manner.

Site	Phase I			Phase II		
	subjects (female/male)	sessions	shots	subjects (female/male)	sessions	shots
BUT	**33** (15/18)	6	21	**32** (15/17)	6	11
IDIAP	**28** (7/21)	6	21	**26** (5/21)	6	11
LIA	**27** (9/18)	6	21	**26** (8/18)	6	11
UMAN	**27** (12/15)	6	21	**25** (11/14)	6	11
UNIS	**26** (6/20)	6	21	**24** (5/19)	6	11
UOULU	**20** (8/12)	6	21	**17** (7/10)	6	11

Table 4: A summary of the number of: subjects per site, sessions per subject and shots (questions) per session.

Unique characteristics. The MOBIO database presents new unique challenges, including:

- High variability of pose and illumination conditions, even during recordings;
- High variability in the quality of speech;
- Variability in the acquisition environments in terms of acoustics as well as illumination and background.

Protocols The database is split into three non-overlapping partitions for: training, development and evaluation. The data is split so that two of the six sites, used for data collection, are used in totality for one partition. This ensures that no identity in one partition will be present in the other partitions. For clarity of use, protocol files are shipped with the database at the address indicated below and should be respected when possible to ease algorithm comparison. Details and full specifications can also be found in [46].

Use-cases. The MOBIO database is specially suited for the following use-cases defined in Deliverable D2.1:

- Use-case 2.1: Benchmarking of biometric modalities (face and speaker);
- Use-case 2.3: Evaluation of biometric (face and speaker) system components;
- Use-case 2.5: Comparative evaluation;
- Use-case 2.6: Biometric (face and speaker) algorithm optimization;
- Use-case 2.7: Biometric (face and speaker) system optimization;
- Use-case 2.10: Educational resource.

Licensing and availability. The MOBIO database is freely available through IDIAP's website[14].

4.2 Fingerprint modality

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 [5]. This database is perfectly suited for an eventual implementation of the BEAT platform for the evaluation of fingerprint-related algorithms.

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

[14] http://www.idiap.ch/dataset/mobio/

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 5: A summary of the fingerprint data captured in BMDB-DS2 in each session and for each of the two sensors used in the acquisition (optical and thermal).

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.

The fingerprint sub-corpus comprised within DS2 was captured was captured in two separate acquisition sessions with the optical sensor Biometrika FX2000 and with the Thermal Sweeping sensor Yubee by Atmel. The data available for each session is summarised in Table 5. The data is stored as bmp images of 296×560 pixels captured at a resolution of 569 dpi.

Unique characteristics. The DS2 dataset presents several characteristics which do not possess the other fingerprint datasets available nowadays, and which make it specially suited for the performance and testing objectives defined within the BEAT project:

- 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 [16] or BiosecurID [5].
- 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 BEAT).
- 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.

Protocol. Each of the 6 fingers (index, middle and ring of both hands) of each subject is considered as a separate user. Thus, the database comprises $667 \times 3 \times 2 = 4002$ fingerprint users.

A possible evaluation protocol for the fingerprint data captured with each of the two sensors (optical and thermal) is given in Fig. 2.

Sensor interoperability may be evaluated by using the training and development data of one of the sensors, and the test data of the other.

Use-cases. Due to the aforementioned characteristics, the BioSecure-DS2 Fingerprint Subcorpus is specially suited for the next use-cases defined in D2.1 for the BEAT platform:

		BioSecure DS2 Fingerprint (667subjects x 3fingers x 2hands = 4002 users)		
Session	**Sample**	**2000 Users**	**602 Users**	**1400 Users**
1	1	Training (enrolled templates)	Development (impostors)	Test (impostors)
	2	Development (clients)		
2	1	Test (clients)		
	2			

Figure 2: Evaluation protocol for the fingerprint data captured with each of the two sensors in BioSecure-DS2 Fingerprint Subcorpus.

- Use-case 2.1: Benchmarking of biometric modalities (fingerprint).
- Use-case 2.3: Evaluation of biometric (fingerprint) system components.
- Use-case 2.5: Comparative evaluation.
- Use-case 2.6: Biometric (fingerprint) algorithm optimization.
- Use-case 2.7: Biometric (fingerprint) system optimization.

Licensing.The BioSecure Multi-modal DB is publicly available through the BioSecure Foundation[15].

4.3 Iris modality

The BiosecurID Iris Database

The BiosecurID database was collected at 6 different Spanish academic institutions in an office-like uncontrolled environment simulating a realistic scenario [5].

The iris sub-corpus comprised within the BiosecurID database was captured in 4 separate acquisition sessions with the LG Iris Access EOU 3000 sensor. The data available for each session is summarised in Table 6. The data is stored as bmp images of 640×480 pixels.

Unique characteristics. The database presents several remarkable characteristics which make it unique for the purposes set in the BEAT project:

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

Users	Eyes	Samples/finger	Total Samples/Session
400	Rigth/Left (2)	4	$400 \times 4 \times 2 = 3200$

Table 6: A summary of the iris data captured in the BiosecurID DB in each of the four sessions.

- Number of biometric traits: speech, iris, face (photographs and talking faces videos), signature and handwriting (on-line and off-line), fingerprints, hand (palmprint and contour-geometry), and keystroking.
- Number of subjects: a total of 400 subjects.
- Number of sessions: 4 sessions distributed in a 4 month time span. Thus, three different levels of temporal variability are taken into account: 1) within the same session (the samples of a given biometric trait are not acquired consecutively), 2) within weeks (between two consecutive sessions), and 3) within months (between non-consecutive sessions). This is specially relevant in traits such as face, handwriting or signature which present a significant variation through time.

Another design principle in the BiosecurID database was to have balanced demographic groups. Thus, all sites were asked to acquire 30% of the subjects between 18 and 25 years of age, 20% between 25 and 35, 20% between 35 and 45, and the remaining 30% of the subjects above 45 years of age. The gender distribution was forced to be balanced and only a 10% difference was permitted between male and female sets. The ethnicity of most subjects correspond to white/caucasian Spaniards.

All relevant non-biometric data of each subject was stored in an independent file (available with the biometric samples) so that experiments regarding specific demographic groups can be easily carried out. The available information in these files includes: age, gender, handedness, manual worker (yes/no), and vision aids (glasses, contact lenses, none).

Protocol. Each of the 2 irises (right and left) of each subject is considered as a separate user. Thus, the database comprises $400 \times 2 = 800$ iris users.

A possible evaluation protocol for the iris data in BiosecurID DB is given in Fig. 3.

Use-cases. Due to the aforementioned characteristics, the BiosecurID Iris Subcorpus is specially suited for the next use-cases defined in D2.1 for the BEAT platform:

- Use-case 2.1: Benchmarking of biometric modalities (iris).
- Use-case 2.3: Evaluation of biometric (iris) system components.
- Use-case 2.4: Supporting user specific studies of biometric modality (iris) sensitivity to degradation phenomena such as: ageing.
- Use-case 2.5: Comparative evaluation.
- Use-case 2.6: Biometric (iris) algorithm optimization.

<table>
<tr><th></th><th></th><th colspan="3">BiosecurID Iris
(400subjects x 2iris = 800users)</th></tr>
<tr><th>Session</th><th>Sample</th><th>200 Users</th><th>50 Users</th><th>150 Users</th></tr>
<tr><td rowspan="2">1</td><td>1</td><td rowspan="2">Training
(enrolled templates)</td><td rowspan="8">Development
(impostors)</td><td rowspan="8">Test
(impostors)</td></tr>
<tr><td>2</td></tr>
<tr><td rowspan="2">2</td><td>1</td><td rowspan="2">Development (clients)</td></tr>
<tr><td>2</td></tr>
<tr><td rowspan="2">3</td><td>1</td><td rowspan="4">Test (clients)</td></tr>
<tr><td>2</td></tr>
<tr><td rowspan="2">4</td><td>1</td></tr>
<tr><td>2</td></tr>
</table>

Figure 3: Evaluation protocol for the iris data comprised in the BiosecurID Iris Subcorpus.

- Use-case 2.7: Biometric (iris) system optimization.

Licensing.The BiosecurID DB does not have yet a set licensing procedure but is announced to be made publicly available through the Biometric Recognition Group - ATVS [16].

4.4 Vein modality

CASIA Multi-Spectral Palmprint Database

The CASIA databases were collected for number of biometric traits (iris, face, fingerprint, handwriting, palmprint, gait etc.). The CASIA Multi-Spectral Palmprint Database consists of palm print images of 100 individuals (six samples per individual), captured under six different NIR illuminators. Wavelengths of the illuminator corresponding to the six spectrum are 460nm, 630nm, 700nm, 850nm, 940nm and white light respectively. Palm veins are most visible under the illuminator at 940 nm wavelength.

Unique characteristics. The CASIA database has several remarkable properties which make it unique for the purposes set in the BEAT project:

- Number of biometric traits: iris, face, fingerprint, handwriting, palmprint, gait.
- Number of images and subjects: 7,200 palm images captured from 100 different people.
- Number of sessions: For each hand, palm images are captured at two sessions. The time interval between the two sessions is more than one month.

Protocol. The database has 3,600 images for each hand (7,200 in total). The protocol can be built for each hand separately. The database is divided into three disjoint sets of palm vein images. The first set contains training set of 1985 palm images and evaluation set contains 966 palm images while query set contains 649 palm images. The training set is used to build genuine models, the evaluation set is used to obtain the genuine and imposter scores for verification thresholds and the query set of palm images is used for obtain the verification rates. Alternatively, Leave-one-out technique can be used in verification and the nearest neighbour search method can be used in identification.

Use-cases. By taking account the aforementioned properties, the CASIA Multi-Spectral Palmprint Database is especially suitable for the below uses-cases defined in D2.1 for the BEAT platform:

- Use-case 2.1: Benchmarking of biometric modalities (vein).
- Use-case 2.3: Evaluation of biometric (vein) system components.

[16] `http://atvs.ii.uam.es/`

- Use-case 2.5: Comparative evaluation.
- Use-case 2.6: Biometric (vein) algorithm optimization.
- Use-case 2.7: Biometric (vein) system optimization.

Licensing. The database is released for research and educational purposes. They accept no liability for any undesirable consequences of using the database. All rights of the CASIA database are reserved. Any person or organization is not permitted to distribute, publish, copy, or disseminate this database. In all documents and papers that report experimental results based on this database, their efforts in constructing the database should be acknowledged as: Portions of the research in this paper use the CASIA-MS-Palmprint V1 collected by the Chinese Academy of Sciences' Institute of Automation (CASIA) and a reference to "CASIA-MS-Palmprint V1, http://biometrics.idealtest.org/" should be included. A copy of all reports and papers that are for public or general release that use the CASIA-MS-Palmprint V1 should be forwarded upon release or publication to Professor Tieniu Tan whose contact details can be found at `http://biometrics.idealtest.org/dbDetailForUser.do?id=6`.

4.5 Multi-modal score-level fusion

Multi-modal score-level fusion Databases

For the purpose of benchmarking fusion algorithms, we consider three data sets:

- Biometric Scores Set - Release 1 (BSSR1)[17]: It contains three data sets with the following suggested usage:
    1. multi-modal fusion of face and fingerprint involving 517 individuals.
    2. two finger fusion involving 6000 individuals.
    3. two-sample or two-algorithm fusion involving 3000 individuals.
- XM2VTS score-level fusion benchmark (XM2VTS fusion) database[18]: It contains the outcome of 7 face and 6 speaker recognition systems of 200 clients. Configurations for multi-modal and intra-modal are published. However, it is conceivable to realize multi-algorithmic and multi-sample fusion as well.
- BANCA score-level fusion data set (BANCA fusion) [19]: It contains 1186 face and speech experimental outcomes.

[17] `http://www.nist.gov/itl/iad/ig/biometricscores.cfm`
[18] `http://personal.ee.surrey.ac.uk/Personal/Norman.Poh/web/fusion`
[19] `http://personal.ee.surrey.ac.uk/Personal/Norman.Poh/web/banca_multi`

Table 7: The Lausanne and fusion protocols of the XM2VTS database. Numbers quoted below are the number of samples.

Score Data sets	Lausanne Protocols	
	LP1	LP2
client development set	600 (3×200)	400 (2×200)
impostor development set	40,000 ($25 \times 8 \times 200$)	
client evaluation set	400 (2×200)	
impostor evaluation set	112,000 ($70 \times 8 \times 200$)	

While BSSR1 is primary used for identification, the XM2VTS and BANCA fusion data set is used for authentication/verification. Therefore, both data sets can complement each other.

Unique characteristics. A unique feature about the XM2VTS fusion data set is that both the development and evaluation contains the same 200 clients, it is often used for benchmarking client-specific score normalization and client-specific fusion procedures. In this case, the development set is used for training the above procedures and the evaluation set is used uniquely for testing.

As for the BANCA fusion data set, it has some interesting features, as listed below:

- It arguably offers the largest number of score sets; it contains 1186 baseline face and speech experiments
- The data contains controlled (clean), adverse (under challenging conditions) and degraded conditions (with different devices), hence allowing algorithms to be evaluated in these specific conditions, e.g., [47].
- The baseline classifiers are very diverse, containing both generative and discriminative classifiers: MLP, GMM and SVM, as well as template-based (with correlation measures). This allows for multi-modal scalability studies, e.g., [48].

Protocol. The BSSR1 does not have a defined protocol. Therefore, there is no development or evaluation set. This means that fusion algorithms that need training (and the parameters of which need tuning) may be trained or tested on different data sets. As a consequence, different fusion algorithms reported on this data set cannot be compared with each other.

In comparison, the XM2VTS fusion data set has two defined protocols, namely Lausanne *fusion* protocols I and II which follow closely the original Lausanne protocols I and II. They are listed in Table 7. In this table, development and evaluation sets are clearly distinguished.

Use-cases. The proposed multi-modal score-level fusion Databases are specially suited for the following use-cases defined in Deliverable D2.1:

- Use-case 2.1: Benchmarking of biometric modalities (Multi-modal);
- Use-case 2.3: Evaluation of biometric (Multi-modal) system components;
- Use-case 2.5: Comparative evaluation;
- Use-case 2.6: Biometric (Multi-modal) algorithm optimization;
- Use-case 2.8: Multi-modal biometric system brokerage
- Use-case 2.10: Educational resource.

Licensing and availability. Because no biometric images are involved, the above data sets are available for download and can be freely used for experimentation without restriction

5 Conclusions

This deliverable includes a brief literature survey of databases for four selected biometric modalities, namely, face, fingerprint, iris, and vein modalities. Several desirable characteristics of biometric experimental protocols are also discussed. The deliverable also lists some examples of candidate database to be considered for the BEAT platform.

References

[1] Javier Ortega-Garcia, Julian Fierrez, Fernando Alonso-Fernandez, Javier Galbally, Manuel R. Freire, Joaquin Gonzalez-Rodriguez, Carmen Garcia-Mateo, Jose-Luis Alba-Castro, Elisardo Gonzalez-Agulla, Enrique Otero-Muras, Sonia Garcia-Salicetti, Lorene Allano, Bao Ly-Van, Bernadette Dorizzi, Josef Kittler, Thirimachos Bourlai, Norman Poh, Farzin Deravi, Ming W.R. Ng, Michael Fairhurst, Jean Hennebert, Andreas Humm, Massimo Tistarelli, Linda Brodo, Jonas Richiardi, Andrzej Drygajlo, Harald Ganster, Federico M. Sukno, Sri-Kaushik Pavani, Alejandro Frangi, Lale Akarun, and Arman Savran, "The multiscenario multienvironment biosecure multimodal database (bmdb)," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 32, pp. 1097–1111, 2010.

[2] Ralph Gross, "Face databases," in *Handbook of Face Recognition*, A.Jain S.Li, Ed. Springer, New York, February 2005.

[3] F. S. Samaria, , and A.C. Harter, "Parameterisation of a stochastic model for human face identification," 1994.

[4] Enrique Bailly-Bailliére, Samy Bengio, Frédéric Bimbot, Miroslav Hamouz, Josef Kittler, Johnny Mariéthoz, Jiri Matas, Kieron Messer, Vlad Popovici, Fabienne Porée, Belén Ruíz, and Jean-Philippe Thiran, "The banca database and evaluation protocol," in *AVBPA*, 2003, pp. 625–638.

[5] 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. Cardeoso-Payo, A. Viloria, C. E. Vivaracho, Q. I. Moro, J. J. Igarza, J. Sanchez, I. Hernaez, C. Orrite-Uruuela, F. Martinez-Contreras, and J. J. Gracia-Roche, "BiosecurID: a multimodal biometric database," *Pattern Analysis and Applications*, vol. 13, pp. 235–246, 2009, To appear.

[6] Wen Gao, Bo Cao, Shiguang Shan, Delong Zhou, Xiaohua Zhang, and Debin Zhao, "The cas-peal large-scale chinese face database and evaluation protocols," Tech. Rep., Joint Research and Development Laboratory, CAS, 2004.

[7] Terence Sim, Simon Baker, and Maan Bsat, "The cmu pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1615–1618, 2003.

[8] P. Jonathon Phillips, Hyeonjoon Moon, Patrick J. Rauss, and Syed A. Rizvi, "The feret evaluation methodology for face-recognition algorithms," in *CVPR*, 1997, pp. 137–143.

[9] P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek, "Overview of the

face recognition grand challenge," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, Washington, DC, USA, 2005, CVPR '05, pp. 947–954, IEEE Computer Society.

[10] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, Washington, DC, USA, 1998, FG '98, pp. 200–, IEEE Computer Society.

[11] Bon-Woo Hwang, Hyeran Byun, Myoung-Cheol Roh, and Seong-Whan Lee, "Performance evaluation of face recognition algorithms on the asian face database, kfdb," in *Proceedings of the 4th international conference on Audio- and video-based biometric person authentication*, Berlin, Heidelberg, 2003, AVBPA'03, pp. 557–565, Springer-Verlag.

[12] Chris McCool and Sébastien Marcel, "Mobio database for the icpr 2010 face and speech competition," Idiap-Com Idiap-Com-02-2009, Idiap, 11 2009.

[13] P. Jonathon Phillips, Patrick J. Flynn, J. Ross Beveridge, W. Todd Scruggs, Alice J. O'Toole, David Bolme, Kevin W. Bowyer, Bruce A. Draper, Geof H. Givens, Yui Man Lui, Hassan Sahibzada, Joseph A. Scallan, Iii, and Samuel Weimer, "Overview of the multiple biometrics grand challenge," in *Proceedings of the Third International Conference on Advances in Biometrics*, Berlin, Heidelberg, 2009, ICB '09, pp. 705–714, Springer-Verlag.

[14] K Messer, J Matas, J Kittler, J Luettin, and G Maitre, "Xm2vtsdb: The extended m2vts database," in *Second International Conference on Audio and Video-based Biometric Person Authentication*, 1999.

[15] Peter N. Belhumeur Athinodoros S. Georghiades and David J. Kriegman, "From few to many: Generative models for recognition under variable pose and illumination," 2000, pp. 277–284, IEEE Computer Society.

[16] J. Fierrez, J. Ortega-Garcia, D. Torre-Toledano, and J. Gonzalez-Rodriguez, "BioSec baseline corpus: A multimodal biometric database," *Pattern Recognition*, vol. 40, pp. 1389–1392, 2007.

[17] 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*, vol. 150, no. 6, pp. 391–401, 2003.

[18] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, "Performance evaluation of fingerprint verification systems," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 3–18, 2006.

[19] R. Cappelli, D. Maio, and D. Maltoni, "Synthetic fingerprint-database generation," in *Proc. IEEE Int. Conf. on Pattern Recognition (ICPR)*, 2002, vol. 3, pp. 744–747.

[20] Raffaele Cappelli, *Handbook of fingerprint recogniion*, chapter Synthetic fingerprint generation, pp. 271–302, Springer, 2009.

[21] NIST, "The multiple biometric grand challenge db," http://www.nist.gov/itl/iad/ig/mbgc.cfm.

[22] NIST, "National insitute of standards and technology: Iris challenge evaluation data," http://iris.nist.gov/ice/.

[23] MMU, "Multimedia university iris database," http://pesona.mmu.edu.my/c̃cteo/.

[24] H. Proenca and L. A. Alexandre, "The ubiris iris image database," 2004, http://iris.di.ubi.pt.

[25] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and 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*, vol. 32, pp. 1529–1535, 2010.

[26] University of Bath, "University of bath iris image database," http://www.bath.ac.uk/elec-eng/research/sipg/irisweb/.

[27] CASIA, "The chinese academy of sciences (casia) iris image database v3.0," http://www.cbsr.ia.ac.cn/IrisDatabase.htm.

[28] S. Shah and A. Ross, "Generating synthetic irises by feature agglomeration," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, 2006, pp. 317–320.

[29] J. Zuo, N. A. Schmid, and X. Chen, "On generation and analysis of synthetic iris images," *IEEE Trans. on Information Forensics and Security*, vol. 2, pp. 77–90, 2007.

[30] FVC2006, "Fingerprint Verification Competition - http://bias.csr.unibo.it/fvc2006/default.asp," 2006.

[31] B. Dumas, C. Pugin, J. Hennebert, D. Petrovska-Delacretaz, A. Humm, F. Evequoz, R. Ingold, and D. Von Rotz, "MyIdea - multimodal biometrics database, description of acquisition protocols," Tech. Rep., Proc. COST-275 Workshop on Biometrics on the Internet, 2005.

[32] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J.L. les Jardins, J. Lunter, Y. Ni, and D. Petrovska-Delacretaz, "BIOMET: A multimodal person authentication database including face, voice, fingerprint, hand and signature modalities," *Proc. International Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA*, pp. 845–853, 2003.

[33] N. Poh and S. Bengio, "Database, protocol and tools for evaluating score-level fusion algorithms in biometric authentication," *Pattern Recognition*, vol. 39, no. 2, pp. 223–233, February 2005.

[34] Niko Brummer, Lukas Burget, Jan Cernock, Ondrej Glembek, Frantisek Grzl, Martin Karafit, D. A. van Leeuwen, Pavel Matejka, Petr Schwarz, and A. Strasheim, "Fusion of heterogeneous speaker recognition systems in the stbu submission for the nist speaker recognition evaluation 2006," *IEEE Transactions on Audio, Speech & Language Processing*, pp. 2072–2084, 2007.

[35] N. Poh, T. Bourlai, and J. Kittler, "A multimodal biometric test bed for quality-dependent, cost-sensitive and client-specific score-level fusion algorithms," *Pattern Recognition*, vol. 43, no. 3, pp. 1094–1105, 2010.

[36] N. Poh, J Kittler, C. H. Chan, S. Marcel, C. Mc Cool, E.A.Rua L. A. Castro, M. Villegas, R. Paredes, V. Struc, N. Pavesic, A.A. Salah, H. Fang, and N. Costen, "An evaluation of video-to-video face verification," *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 5, no. 4, pp. 781–801, 2011.

[37] N. Poh, T. Bourlai, J. Kittler, and al, "Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms," *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 4, no. 4, pp. 849–866, 10 2009.

[38] P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, and William Worek, "Preliminary face recognition grand challenge results," in *in Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, 2006, pp. 15–24.

[39] Jiri Matas, Miroslav Hamouz, Kenneth Jonsson, Josef Kittler, Yongping Li, Constantine Kotropoulos, Anastasios Tefas, Ioannis Pitas, Teewoon Tan, Hong Yan, Fabrizio Smeraldi, N. Capdevielle, Wulfram Gerstner, Yousri Abdeljaoued, Josef Bigün, S. Ben-Yacoub, and E. Mayoraz, "Comparison of face verification results on the xm2vts database," in *ICPR*, 2000, pp. 4858–4863.

[40] Patrick J. Grother, George W. Quinn, and P J. Phillip, "Report on the evaluation of 2d still-image face recognition algorithms," Tech. Rep. 7709, NIST, June 2010.

[41] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score Normalization for Text-Independant Speaker Verification Systems," *Digital Signal Processing (DSP) Journal*, vol. 10, pp. 42–54, 2000.

[42] L. Wolf, T. Hassner, and Y. Taigman, "The one-shot similarity kernel," in *IEEE International Conference on Computer Vision (ICCV)*, Sept. 2009.

[43] A. Martin, M. Przybocki, and J. P. Campbell, *The NIST Speaker Recognition Evaluation Program*, chapter 8, Springer, 2005.

[44] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

[45] N. Poh and J. Kittler, "Incorporating Variation of Model-specific Score Distribution in Speaker Verification Systems," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 3, pp. 594–606, 2008.

[46] Chris McCool, Sébastien Marcel, Abdenour Hadid, Matti Pietikainen, Pavel Matejka, Jan Cernocky, Norman Poh, J. Kittler, Anthony Larcher, Christophe Levy, Driss Matrouf, Jean-François Bonastre, Phil Tresadern, and Timothy Cootes, "Bi-modal person recognition on a mobile phone: using mobile phone data," in *IEEE ICME Workshop on Hot Topics in Mobile Multimedia*, July 2012.

[47] N. Poh, J. Kittler, S. Marcel, D. Matrouf, and J.-F. Bonastre, "Model and score adaptation for biometric systems: Coping with device interoperability and changing acquisition conditions," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, aug. 2010, pp. 1229–1232.

[48] N. Poh and J. Kittler, "On Using Error Bounds to Optimize Cost-sensitive Multimodal Biometric Authentication," in *Proc. 19th Int'l Conf. Pattern Recognition (ICPR)*, 2008, pp. 1–4.