MobBIO: A Multimodal Database Captured with a Portable Handheld Device

Ana F. Sequeira[1,2], João C. Monteiro[1,2], Ana Rebelo[1] and Hélder P. Oliveira[1]
[1]*INESC TEC, Porto, Portugal*
[2]*Faculdade de Engenharia, Universidade do Porto, Porto, Portugal*
{*ana.filipa.sequeira, joao.carlos.monteiro*}*@fe.up.pt*, {*arebelo, helder.f.oliveira*}*@inescporto.pt*

Keywords: Biometrics, Multimodal, Database, Portable Handheld Devices.

Abstract: Biometrics represents a return to a natural way of identification: testing someone by what (s)he is, instead of relying on something (s)he owns or knows seems likely to be the way forward. Biometric systems that include multiple sources of information are known as multimodal. Such systems are generally regarded as an alternative to fight a variety of problems all unimodal systems stumble upon. One of the main challenges found in the development of biometric recognition systems is the shortage of publicly available databases acquired under real unconstrained working conditions. Motivated by such need the *MobBIO* database was created using an Asus EeePad Transformer tablet, with mobile biometric systems in mind. The proposed database is composed by three modalities: iris, face and voice.

1 INTRODUCTION

In almost everyone's daily activities, personal identification plays an important role. The most traditional techniques to achieve this goal are knowledge-based and token-based automatic personal identifications. Token-based approaches take advantage of a personal item, such as a passport, driver's license, ID card, credit card or a simple set of keys to distinguish between individuals. Knowledge-based approaches, on the other hand, are based on something the user knows that, theoretically, nobody else has access to, for example passwords or personal identification numbers (Prabhakar et al., 2003). Both of these approaches present obvious disadvantages: tokens may be lost, stolen, forgotten or misplaced, while passwords can easily be forgotten by a valid user or guessed by an unauthorized one. In fact, all of these approaches stumble upon an obvious problem: any piece of material or knowledge can be fraudulently acquired (Jain et al., 2000).

Biometrics represents a return to a more natural way of identification: many physiological or behavioural characteristics are unique between different persons. Testing someone by what this someone is, instead of relying on something he owns or knows seems likely to be the way forward (Monteiro et al., 2013).

Several biological traits in humans show a considerable inter-individual variability: fingerprints and palmprints, the shape of the ears, the pattern of the iris, among others, as depicted on Figure 1. Biometrics works by recognizing patterns within these biological traits, unique to each individual, to increase the reliability of recognition. The growing need for reliability and robustness, raised some expectations and became the focal point of attention for research works on biometrics.

Most biometric systems deployed in real-world applications rely on a single source of information

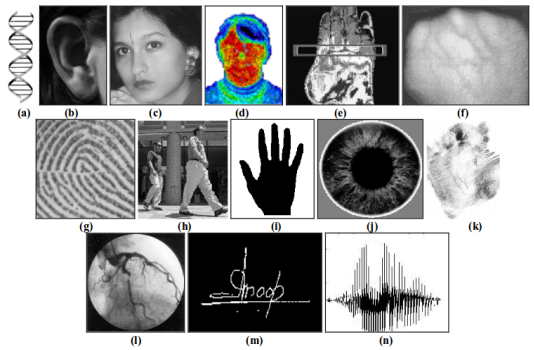


Figure 1: Examples of some of the most widely studied biometric traits: (a) DNA, (b) Ear shape, (c) Face, (d) Facial Thermogram, (e) Hand Thermogram, (f) Hand veins, (g) Fingerprint, (h) Gait, (i) Hand geometry, (j) Iris, (k) Palm print, (l) Retina, (m) Keystroke and (n) Voice. Extracted from (Jain et al., 2002).

Table 1: Comparative data analysis of some common biometric traits. Adapted from (Jain et al., 2000) and (Proença, 2007).

Requirements				
Traits	**Universality**	**Uniqueness**	**Collectability**	**Permanence**
DNA	High	High	Low	High
Ear	Medium	Medium	Medium	High
Face	High	Low	High	Medium
Facial Thermogram	High	High	High	Low
Hand Veins	Medium	High	High	Medium
Fingerprint	Medium	High	High	Medium
Gait	Low	Low	High	Low
Hand Geometry	Medium	Medium	High	Medium
Iris	High	High	Medium	High
Palm Print	Medium	High	Medium	High
Retina	High	High	Low	Medium
Signature	Medium	Low	High	Low
Voice	Medium	Low	Medium	Low

to perform recognition, thus being dubbed *unimodal*. Extensive studies have been performed on several biological traits, regarding their capacity to be used for unimodal biometric recognition. Table 1 summarizes the analysis performed by Jain (Jain et al., 2000) and Proença (Proença, 2007), regarding the qualitative analysis of individual biometric traits, considering the four factors laid out in the previous section. Careful analysis of the advantages and disadvantages laid out in the previously referred table seems to indicate a couple of general conclusions: (1) there is no "gold-standard" biometric trait, i.e. the choice of the best biometric trait will always be conditioned by the means at our disposal and the specific application of the recognition process; (2) some biometric traits seem to present advantages that counterbalance other trait's disadvantages. For example, while voice's permanence is highly variable, due to external factors, the iris patterns represent a much more stable and hard to modify trait. However, iris acquisition in conditions that allow accurate recognition requires specialized NIR illumination and user cooperation, while voice only requires a standard sound recorder and even no need for direct cooperation of the individual.

This line of thought seems to indicate an alternative way of stating the two conclusions outlined in the previous paragraph: even though there is no "best" biometric trait *per se*, marked advantages might be found by exploring the synergistic effect of multiple statistically independent biometric traits, so that each other's pros and cons counterbalance resulting in an improved performance over each other's individual accuracy. Biometric systems that include multiple sources of information for establishing an identity are known as *multimodal biometric systems* (Ross and Jain, 2004). It is generally regarded, in many reference works of the area, that multimodal biometric systems might help cope with a variety of generic problems all unimodal systems generally stumble upon, regardless of their intrinsic pros and cons (Jain et al., 1999). These problems can be classified as:

1. Noisy data: when external factors corrupt the original information of a biometric trait. A fingerprint with a scar and a voice altered by a cold are examples of noisy inputs. Improperly maintained sensors and unconstrained ambient conditions also account for some sources of noisy data. As an unimodal system is tuned to detect and recognize specific features in the original data, the addition of stochastic noise will boost the probabilities of false identifications (Jain and Ross, 2004).

2. Intra-class variations: when the biometric data acquired from an individual during authentication is different from the data used to generate the template during enrolment (Jain and Ross, 2004). This may be observed when a user incorrectly interacts with a sensor (e.g. variable facial pose) or when a different sensor is used in two identification approaches (Ross and Jain, 2004).

3. Inter-class similarities: when a database is built on a large pool of users, the probability of different users presenting similarities in the feature space of the chosen trait naturally increases (Ross and Jain, 2004). It can, therefore, be considered that every biometric trait presents an asymptotic behaviour towards a theoretical upper bound in terms of its discrimination, for a growing number of users enrolled in a database (Jain and Ross, 2004).

4. Non-universality: when the biometric system fails to acquire meaningful biometric data from the

user, in a process known as failure to enrol (FTE) (Jain and Ross, 2004).

5. Spoof attacks: when an impostor attempts to spoof the biometric trait of a legitimately enrolled user in order to circumvent the system (Jain et al., 2002).

It is intuitive to note that taking advantage of the evidence obtained from multiple sources of information will result in an improved capability of tackling some of the aforementioned problems. These sources might be more than just a set of distinct biometric traits. Other options, such as *multiple sensors*, *multiple instances*, *multiple snapshots* or *multiple feature space representations* of the same biometric are also valid options, as depicted on Figure 2 (Jain and Ross, 2004).

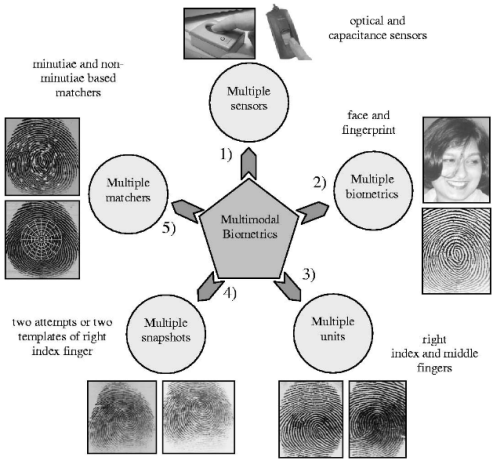


Figure 2: Scenarios in a multimodal biometric system. From (Ross and Jain, 2004).

The development of biometric recognition systems is generally limited by the shortage of large public databases acquired under real unconstrained working conditions. Database collection represents a complicated process, in which a high degree of cooperation from a large number of participants is needed (Oliveira and Magalhães, 2012). For that reason, nowadays, the number of existing public databases that can be used to evaluate the performance of multimodal biometric recognition systems is quite limited.

Motivated by such need we present a new database, named *MobBIO*, acquired using a portable handheld device, namely an Asus EeePad Transformer tablet. With this approach we aim to tackle not only the ever growing need for data, but also to provide a database whose acquisition environment follows the rapid evolution of our networked society from simple communication devices to mobile personal computers. The proposed database is composed by three modalities: iris, face and voice. A possible schematics of a multimodal system trained for the MobBIO database is presented on Figure 3.

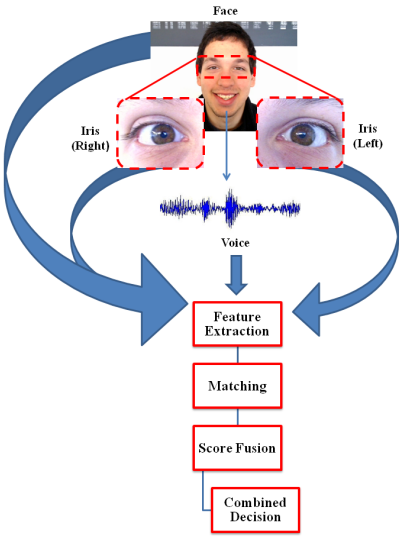


Figure 3: Flowchart of a generic multimodal system working on the modalities present in the MobBIO database.

The remainder of this paper is organized as follows: Section 2 summarizes the state-of-the-art concerning available multimodal biometric databases; Section 3 presents the MobBIO database and its specifications; and finally the conclusions and future work prospects regarding possible improvement to the database are summarized in Section 4.

2 MULTIMODAL DATABASES

A strong trend observed as of lately is the appearance of multimodal databases. As already referred, it seems obvious that the complementarity of some biometric traits will bring advantages and, consequently, a more accurate biometric recognition. When it comes to the choice of a biometric trait a vast list of possibilities is found, as shown in previous section. This diversity gives rise, in existing multimodal databases, to many possible combinations of traits.

The first multimodal database with 5 modalities and time variability, launched by the *Multimodal Biometric Identity Verification* project, was the BIOMET (Garcia-Salicetti et al., 2003). The database was constructed in three different sessions, with three and five months spacing between them and contains samples of face, voice, fingerprint, hand shape and handwritten signature.

On 2003, the Biometric Recognition Group - ATVS made public and freely available the MCYT-Bimodal Biometric Database (Fierrez-Aguilar et al., 2003). This database includes fingerprint and handwritten signature, in two versions containing data from 75 and 100 users, respectively offline and online signature acquisition.

Within the *M2VTS project (Multi Modal Verification for Teleservices and Security applications)* the database XM2VTS (Poh and Bengio, 2006) was launched, comprising several datasets including face images and speech samples. According to its authors, the goal of using a multimodal recognition scheme is to improve the recognition efficiency by combining single modalities, namely face and voice features. At cost price, sets of data taken from this database are available including high quality color images, 32 KHz 16-bit sound files, video sequences and a 3d Model of each subjects head.

In the aforementioned databases there are several limitations, such as the absence of important traits (e.g., iris), limitations at sensors level (e.g., sweeping fingerprint sensors), and informed forgery simulations (e.g., voice utterances pronouncing the PIN of another user) (Ortega-Garcia et al., 2010). The BioSec Multimodal Biometric Database Baseline (Fierrez-Aguilar et al., 2007) was an attempt to overcome some of these limitations. This database included real multimodal data from 200 individuals in two acquisition sessions including fingerprint, iris, voice and face. However the two releases of this database are now under construction and are not available at the moment. An enlarged version of the previous database is The Multiscenario Multienvironment BioSecure Multimodal Database (BMDB) (Ortega-Garcia et al., 2010) which comprises signature, fingerprint, hand and iris acquired in three different scenarios. This database is not freely accessed.

The *WVU/CLARKSON: JAMBDC - Joint Multimodal Biometric Dataset Collection* project gave rise to a series of biometric datasets, available under request and with costs. Integrated within the aforementioned project, the West Virginia University constructed two releases of biometric data containing six distinct biometric modalities: iris, face, voice, fingerprint, hand geometry and palmprint. The two releases differ only in the number of subjects. Within the same initiative, the Clarkson University created another dataset which contains image and video files for the same modalities except for hand geometry (Crihalmeanu et al., 2007).

The MOBIO database (McCool et al., 2012) consists of bi-modal audio and video data taken from 152 people. The speech samples and the face videos were recorded using two mobile devices: a mobile phone and a laptop computer.

The emergence of portable handheld devices, for multiple everyday activities, has created a necessity for the development of mobile identity verification applications. The objective of research is to create a reliable, portable way of identifying and authenticating individuals. To pursue this goal, the availability of testing databases is crucial, so that results obtained by different methods may be compared. It is noted that the existing databases do not completely fulfill the requirements of this line of research. On one hand, there are limitations in the variety and combination of biometric traits, and on the other hand some of the databases are not public accessible limiting their usability.

3 MobBIO: DATABASE OVERVIEW

The reasons to create the MobBIO multimodal database are related, on one hand, with the raising interest in mobile biometrics applications and, on the other hand, with the increasing interest in multimodal biometrics. These two perspectives motivated the creation of a database comprising face, iris and voice samples acquired in unconstrained conditions using a mobile device, whose specifications will be detailed in further sections. We also stress the fact that there is no multimodal database with similar characteristics, regarding both the traits and the unique acquisition conditions.

As voice is the only acoustic-based biometric trait and the facial traits - face and iris - are the most instinctive regions for a mobile device wielder to photograph, we chose these three traits for the MobBIO database. In the choice of such traits it was also taken into account that the design of consumer mobile devices is extremely sensitive to cost, size, and power efficiency and that the integration of dedicated biometric devices is, thus, rendered less attractive (Shi et al., 2011). However, the majority of the developed iris recognition systems rely on near-infrared (NIR) imaging rather than visible light (VL). This is due to the fact that fewer reflections from the cornea in NIR imaging result in maximized signal-to-noise ratio (SNR) in the sensor, thus improving the contrast of iris images and the robustness of the system (Monteiro et al., 2013). As NIR illumination is not an acceptable alternative we obtain iris images with simple VL illumination, even though this results in considerably noisier images.

Mobile device cameras are known to present lim-

itations due to their increasingly thin form factor. Therefore, these devices inherently lack high quality optics like zoom lenses and larger image sensors. Nevertheless, for most daily uses, the quality is considered good enough by most consumers (Tufegdzic, 2013). Regarding acoustic measurements, no hardware improvements can solve the problems that harm the performance of voice recognition: environmental noise and voice alterations by external noises, such as emotional state or illness, need to be accounted for, by the algorithm (Khitrov, 2013). Multimodal approach may help counter image-based difficulties, like low illumination or rotated images, with voice-based features or vice-versa. By exploring multiple sensors the intrinsic hardware-based limitations of each one can be balanced by the other, resulting in a synergistic effect in terms of biometric data quality.

The creation of this database seems a valuable resource for future research and its purpose goes far beyond its immediate application in the "MobBIO 2013: 1st Biometric Recognition with Portable Devices Competition"[1] that was launched in January of 2013. This competition is embraced by ICIAR2013[2].

3.1 Description of the Database

The MobBIO Multimodal Database comprises the biometric data from 105 volunteers. Each individual provided samples of face, iris and voice. The nationalities of the volunteers were mainly portuguese but also participated volunteers from U.K., Romania and Iran. The average of ages was approximately 34, being the minimum age 18 and the maximum age 69. The gender distribution was 29% females and 71% males.

The volunteers were asked to sit, in two different spots of a room with natural and artificial sources of light, and then the face and eye region images were captured by sequential shots. The distance to the camera was variable (10-50 cm) depending on the type of image acquired: closer for the eye region images and farther away for face images. For the speech samples, the volunteers were asked to get close to the integrated microphone of the device and the recorder was activated and deactivated by the collector. The equipment used for the samples acquisition was an Asus Transformer Pad TF 300T, with Android version 4.1.1. The device has two cameras one frontal and one back camera. The camera we used was the back camera, version TF300T-000128, with 8 MP of resolution and autofocus.

[1]http://www.fe.up.pt/∼mobbio2013/

[2]http://www.iciar.uwaterloo.ca/iciar13/

For the voice samples, the volunteers were asked to read 16 sentences in Portuguese. The collected samples had an average duration of 10 seconds. Half of the read sentences presented the same content for every volunteer, while the remaining half were randomly chosen among a fixed number of possibilities. This was done to allow both the application of text-dependent and text-independent methods, which comprise the majority of the most common speaker recognition methodologies (Fazel and Chakrabartty, 2011).

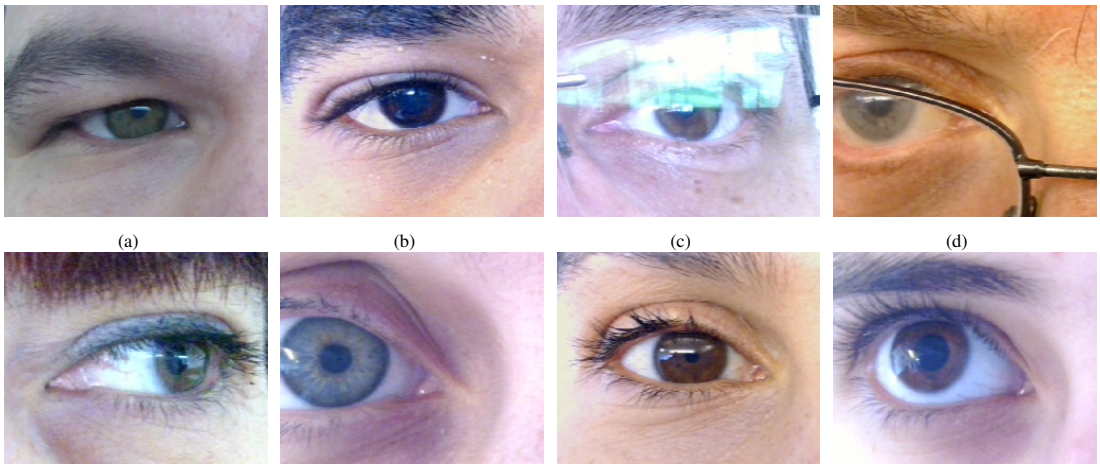
The iris images were captured in two different lighting conditions, with variable eye orientations and occlusion levels, so as to comprise a larger variability of unconstrained scenarios. For each volunteer 16 images (8 of each eye) were acquired. These images were obtained by cropping a single image comprising both eyes. Each cropped image was set to a 300×200 resolution. Some examples of iris images are depicted in Figure 4.

The iris images can, by themselves, constitute an important tool of work concerning iris recognition in mobile devices environment. This dataset is provided with manual annotation of both the limbic and pupillary contours, so that the segmentation methods applied to its images can be evaluated. An example of such annotation is shown in Figure 5.

Face images were captured in similar conditions as iris images, in two different lighting conditions. A total of 16 images were acquired from each volunteer, with a resolution of 640×480. Some examples are illustrated in Figure 6.

4 CONCLUSIONS

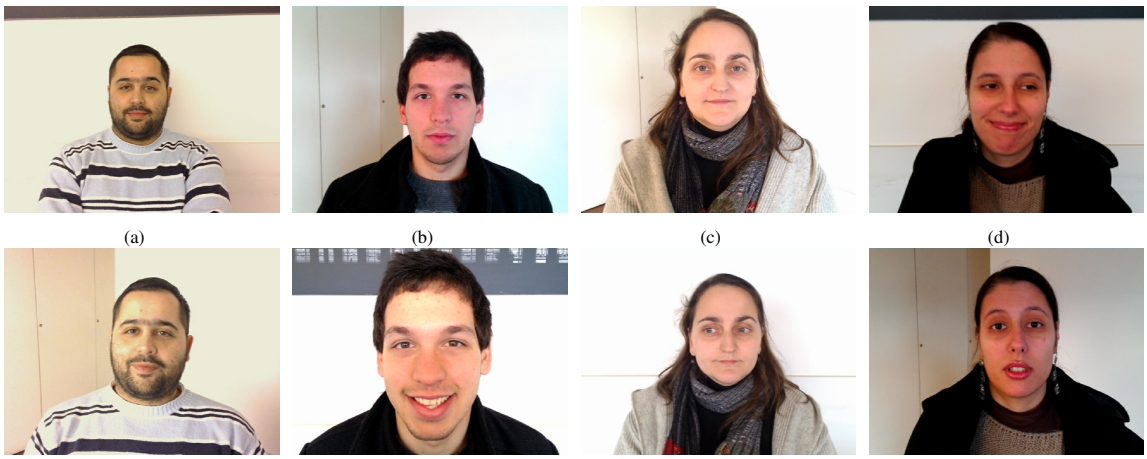
The increased use of handheld devices in everyday activities, which incorporate high performance cameras and sound recording components, has created the possibility for implementing image and sound processing applications for identity verification. The aim to produce reliable methods of identifying and authenticating individuals in portable devices is of utterly importance nowadays. The research in this field require the availability of databases that resemble the unconstrained conditions of this scenarios. We aim to contribute to the research in this area by deploying a multimodal database whose characteristics are valuable to the development of state-of-the-art methods in multimodal recognition. The manual annotation of iris images is a strong point of this database as it allows the evaluation of developed methods of segmentation with this noisy images. For the future, the other samples will also be annotated manually: the



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4: Examples of iris images from MobBIO database: a) Heavily occluded; b) Heavily pigmented; c) Glasses reflection; d) Glasses occlusion; e) Off-angle; f) Partial eye; g) Reflection occlusion and h) Normal.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 6: Examples of face images from MobBIO database.

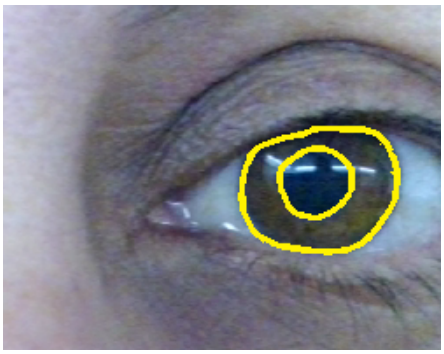


Figure 5: Example of a manually annotated iris image.

face will be identified in face images and the silence and speech will be identified in the sound recordings.

It might be argued that the use of this particular one in the research community may be limited. It would be better if the voice samples were recorded both in English as well as Portuguese, and the images stored in several resolutions and more challenging real-life conditions, such as variable illuminations. This set of suggestions will surely be taken into consideration for future improvements over the present dataset.

This database has already been tested in another work (Monteiro et al., 2014) concerning iris segmentation. Also, the iris image collection has allowed the construction of a dataset of fake images (MobBIO*fake*), composed by printed copies and their respective originals. This database was developed for the purpose of iris liveness detection research, and was already tested in the scope of a different work (Sequeira et al., 2014).

ACKNOWLEDGEMENTS

The authors author would like to thank Fundação para a Ciência e Tecnologia (FCT) - Portugal the financial support for the PhD grants with references SFRH/BD/74263/2010 and SFRH/BD/87392/2012.

REFERENCES

Crihalmeanu, S., Ross, A., Schuckers, S., and Hornak, L. (2007). A protocol for multibiometric data acquisition, storage and dissemination. Technical report, WVU, Lane Department of Computer Science and Electrical Engineering.

Fazel, A. and Chakrabartty, S. (2011). An overview of statistical pattern recognition techniques for speaker verification. *IEEE Circuits and Systems Magazine*, 11(2):62–81.

Fierrez-Aguilar, J., Ortega-garcia, J., Torre-toledano, D., and Gonzalez-rodriguez, J. (2003). Mcyt baseline corpus: A bimodal biometric database. *IEE Proc.Vis. Image Signal Process.*, 150:395–401.

Fierrez-Aguilar, J., Ortega-Garcia, J., Torre-Toledano, D., and Gonzalez-Rodriguez, J. (2007). Biosec baseline corpus: A multimodal biometric database. *Pattern Recognition*, pages 1389–1392.

Garcia-Salicetti, S., Beumier, C., Chollet, G., Dorizzi, B., les Jardins, J. L., Lunter, J., Ni, Y., and Petrovska-Delacrétaz, D. (2003). Biomet: a multimodal person authentication database including face, voice, fingerprint, hand and signature modalities. In *Audio-and Video-Based Biometric Person Authentication*, pages 845–853. Springer.

Jain, A., Bolle, R., and Pankanti, S. (2002). Introduction to biometrics. In *Biometrics*, pages 1–41.

Jain, A., Hong, L., and Kulkarni, Y. (1999). A multimodal biometric system using fingerprint, face and speech. In *Proceedings of 2nd International Conference on Audio-and Video-based Biometric Person Authentication, Washington DC*, pages 182–187.

Jain, A., Hong, L., and Pankanti, S. (2000). Biometric identification. *Communications of the ACM*, 43(2):90–98.

Jain, A. K. and Ross, A. (2004). Multibiometric systems. *Communications of the ACM*, 47(1):34–40.

Khitrov, M. (2013). Talking passwords: voice biometrics for data access and security. *Biometric Technology Today*, 2013(2):9 – 11.

McCool, C., Marcel, S., Hadid, A., Pietikainen, M., Matejka, P., Poh, N., Kittler, J., Larcher, A., Levy, C., Matrouf, D., et al. (2012). Bi-modal person recognition on a mobile phone: using mobile phone data. In *IEEE International Conference on Multimedia and Expo Workshops*, pages 635–640. IEEE.

Monteiro, J. C., Oliveira, H. P., Sequeira, A. F., and Cardoso, J. S. (2013). Robust iris segmentation under unconstrained settings. In *Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP)*, pages 180–190.

Monteiro, J. C., Sequeira, A. F., Oliveira, H. P., and Cardoso, J. S. (2014). Robust iris localisation in challenging scenarios. In *CCIS Communications in Computer and Information Science*. Springer-Verlag.

Oliveira, H. P. and Magalhães, F. (2012). Two unconstrained biometric databases. In *Image Analysis and Recognition*, pages 11–19. Springer.

Ortega-Garcia, J., Fierrez, J., Alonso-Fernandez, F., Galbally, J., Freire, M. R., Gonzalez-Rodriguez, J., Garcia-Mateo, C., Alba-Castro, J.-L., Gonzalez-Agulla, E., Otero-Muras, E., et al. (2010). The multiscenario multienvironment biosecure multimodal database (bmdb). *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(6):1097–1111.

Poh, N. and Bengio, S. (2006). Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication. *Pattern Recognition*, 39(2):223–233.

Prabhakar, S., Pankanti, S., and Jain, A. K. (2003). Biometric recognition: Security and privacy concerns. *Security & Privacy, IEEE*, 1(2):33–42.

Proença, H. (2007). *Towards Non-Cooperative Biometric Iris Recognition*. PhD thesis.

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

Sequeira, A. F., Murari, J., and Cardoso, J. S. (2014). Iris liveness detection methods in mobile applications. In *Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP)*.

Shi, W., Yang, J., Jiang, Y., Yang, F., and Xiong, Y. (2011). Senguard: Passive user identification on smartphones using multiple sensors. In *IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications*, pages 141–148.

Tufegdzic, P. (2013). iSuppli: Smartphone cameras are getting smarter with computational photography; Last check: 06.06.2013.