Università degli Studi di Salerno

Dipartimento di Informatica

Dottorato di Ricerca in Informatica – XXXIV Ciclo

Tesi di Dottorato/Ph.D. Thesis

Multi-biometric systems integrating fixed cameras, mobile devices, and drones

Luigi De Maio

Supervisor: **Prof. Michele Nappi**

Ph.D. Program Director: **Prof. Andrea De Lucia**

AA 2020/2021

Curriculum Computer Science and Information Technology

University of Salerno

Department of Computer Science

Dottorato di Ricerca in Informatica
XXXIV Ciclo

Tesi di Dottorato / Ph.D. Thesis

Multi-biometric systems integrating fixed cameras, mobile devices, and drones

Luigi DE MAIO

Supervisor: **Prof. Michele NAPPI**

PhD Program Director: **Prof. Andrea DE LUCIA**

A.A 2020/2021

I would like to dedicate this thesis to those who love me, and to my children born during this course.

Each individual is unique, but perhaps it is more so than you think.
Not only from the physical or behavioral point of view
but also in the soul, and all three must be protected.
Use biometrics just to do that.

— Luigi De Maio

ACKNOWLEDGMENTS

Of the many who provided us with constructive comments, we are especially grateful to Prof. Michele Nappi[1], to the friends who advised and referred me, to many friends who have served as role models during the pandemic, to my PhD colleagues, to the students I followed, to the reviewers for their fruitful comments and corrections, and last but not least to my family who knew how to wait for me.

1 Full professor - Computer Science Department- University of Salerno

ABSTRACT

To identify a person by means of fully automatic systems is still an open problem and a matter of social concern. Many of the approaches coping with this problem are based on the combination of biometrics and computer vision techniques. In particular, *biometrics* was born in the wake of the wider field of pattern recognition, as that discipline analyses the characteristics of specific individuals with the aim to link their identity to what they are, rather than to what they know or possess. Although much progress has been made in this area, there are still many open problems, which limit its application in daily life at a very large scale. As with many problems in the field of pattern recognition, also for biometrics, the availability of annotated and structured data necessary for designing and validating systems represents a crucial aspect. This theme becomes even more central when systems under consideration show a complex architecture and involve advanced technology such as drones. An in-depth study of the state of the art in this direction has allowed us to identify the most interesting datasets that are currently available. The first goal of this work was to design, acquire, and annotate a large collection of data from different fixed and mobile devices. Data were collected by means of mobile devices such as commercial drones and smartphones, in combination with fixed cameras usually adopted in controlled environments. This type of architecture allows for greater versatility in capturing subjects such as shooting from multiple angles, extreme framing, and using different devices at the same time. Moreover, the characteristics and potential use of this new dataset are drawn. Secondly, we proceeded to design and develop biometric solutions that could demonstrate new integrated approaches of people face trait acquisition. In more details to demonstrate the applicability of drones in a real environment, an acquisition system via a monocular camera installed on-board of a drone has been proposed. This system shows the peculiarity that the drone moves autonomously without a pilot around a cooperative subject (autonomous un-

manned drone). The fusion of data acquired by the camera from various perspectives allows us to obtain high-quality aggregate data, useful to be compared with other data obtained from acquisitions made with other devices and protocols. At the end of the acquisition, a 3D face model is obtained by a completely automatic data processing pipeline. A further example of the effectiveness of drones in the biometric field is provided. It is explained the architecture and the results that can be obtained by a drone in building a 3D face model showing a quality comparable to that obtained with a smartphone. Current trends, implications, solutions, and main shortcomings of biometric data protection are discusses. Additionally, a sample study conducted on the combined use of biometrics and cryptography to secure biometric entities is explained and demonstrated. The potential use of these results is addressed, and discusses new advanced methods and applications of biometrics in virtual environments. Finally, the potential uses of these results are addressed, and new advanced methods and applications of biometrics in virtual environments are discussed. Conclusions are dealt with and summarized in the main contributions of the work and provides an insight on future trends in the use of drones in the field of biometrics and in the new era.

ABSTRACT IN ITALIANO

L'identificazione di una persona mediante sistemi completamente automatici rappresenta ancora un problema aperto ed è questione di interesse sociale. Molti degli approcci per affrontare questo problema si basano sulla combinazione di biometria e tecniche di visione artificiale. In particolare, la *Biometria* nasce sulla scia dell'ampio campo del riconoscimento di pattern, ed è una disciplina che analizza le caratteristiche di un individuo con l'obiettivo di legare la sua identità a ciò che è, piuttosto che a ciò che conosce o possiede. Sebbene siano stati fatti, molti progressi in questo settore, ci sono ancora molti problemi da risolvere, che limitano le sue applicazioni nella vita quotidiana su più ampia scala. Come per molti problemi nel campo del riconoscimento di pattern,

anche per la biometria, la disponibilità di dati annotati e strutturati necessari per progettare e validare i sistemi rappresenta un aspetto cruciale. Questo tema diventa ancora più centrale quando i sistemi in esame mostrano un'architettura complessa e coinvolgono tecnologie avanzate come i droni. Uno studio approfondito dello stato dell'arte in questa direzione ha permesso di individuare i dataset più interessanti attualmente disponibili. Il primo obiettivo di questo lavoro è stato progettare, acquisire e annotare un ampio corpo di dati provenienti da diversi dispositivi fissi e mobili. I dati sono stati raccolti tramite dispositivi mobili come droni commerciali e smartphone, in combinazione con telecamere fisse solitamente adottate in ambienti controllati. Questo tipo di architettura consente una maggiore versatilità nell'acquisizione di soggetti come riprese da più angolazioni, inquadrature estreme e utilizzo simultaneo di dispositivi diversi. Inoltre, vengono tracciate le caratteristiche e il potenziale utilizzo di questo nuovo set di dati. In secondo luogo, abbiamo proceduto alla progettazione e allo sviluppo di soluzioni biometriche in grado di dimostrare nuovi approcci integrati per l'acquisizione dei tratti del volto delle persone. Più in dettaglio per dimostrare l'applicabilità dei droni in un ambiente reale, è stato proposto un sistema di acquisizione tramite una telecamera monoculare installata a bordo di un drone. Questo sistema mostra la particolarità che il drone si muove autonomamente senza pilota attorno a un soggetto cooperativo. La fusione dei dati acquisiti dalla telecamera da diverse prospettive permette di ottenere dati aggregati di alta qualità, utili per essere confrontati con altri dati ottenuti da acquisizioni effettuate con altri dispositivi e protocolli. Al termine dell'acquisizione, si ottiene un modello di volto 3D tramite un processo di elaborazione dati completamente automatico. Viene fornito un ulteriore esempio dell'efficacia dei droni in campo biometrico. Viene spiegata l'architettura e i risultati che si possono ottenere da un drone nella costruzione di un modello di volto 3D che mostri una qualità paragonabile a quella ottenuta con uno smartphone. Vengono discusse le tendenze attuali, le implicazioni, le soluzioni e le principali carenze della protezione dei dati biometrici. Inoltre, viene spiegato e dimostrato un esempio di studio condotto sull'uso combinato della biometria e della crittografia per proteggere le entità biometriche. Viene esa-

minato il potenziale utilizzo di questi risultati e vengono discussi nuovi metodi avanzati e applicazioni della biometria in ambienti virtuali. Le conclusioni sono trattate e riassunte nei principali contributi del lavoro e forniscono uno spaccato sulle tendenze future nell'uso dei droni nel campo della biometria e nella nuova era.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

ACRONYMS

EDPS European Data Protection Supervisor

GDPR General Data Protection Regulation

RSA Rivest–Shamir–Adleman

NIST National Institute of Standards and Technology

MUBIDUS-I MUlti-BIometric and multipurpose Dataset developed at University of Salerno

GPUs Graphics Processing Units

PTZ Pan, Tilt and Zoom

SDK Software Development Kit

CUDA Compute Unified Device Architecture

FPS Frame Per Second.

ICP Iterative Closest Point

RMSE Root-Mean-Square Error

DA Data Augmentation

ANN Artificial Neural Networks

DNN Deep Neural Networks

GANNs Generative Neural Networks

DA-GAN Dual-Agent Generative Adversarial Network

FRR False Rejection Rate

FAR False Acceptance Rate

ERR Equal Error Rate

DNA Deoxyribonucleic Acid

FRVT Face Recognition Vendor Test

CNN Convolutional Neural Network

ROI Region of interest

SVM Support Vector Machine

ROSR Rigid-area Orthogonal Spectral Regression

HoG Histogram of the mesh Gradient

HoS Histogram of the Shape index

HoGS Histogram of the Gradient of the Shape index

GPGPU General-Purpose computing on Graphics Processing Units

OpenCL Open Computing Language

CB Cancelable Biometrics

ROS Robot Operating System

CAGR Compound annual growth rate

CMC Cumulative Match Curve

ROC Receiver Operating Characteristic

Part I

GUIDELINES

1

INTRODUCTION

When we come into the world, to recognize our mother is among the first things we do. Unconsciously, a recognition operation is performed but, perhaps, it already starts in the belly. Human being has an innate ability to distinguish faces, sounds, and flavours. Biometrics is a branch of science that analyzes individual differences and the techniques that can be used to measure them.

Day by day, biometrics increasingly fit into various facets of daily life depending on cultural, economic, and social influences [108]. This has happened since ancient times when the first evidence can be found, handprints in the prehistoric period on the wall of a cave 31,000 years ago, fingerprints on clay tablets from Babylonia 500 B.C., and so on until today.

Even if its primordial nature is owing to guarantee the identity of a person, at present, the biometric component is found in activities linked to free time, security, mobility, and recently in medicine. It is also found in more specific scopes such as law enforcement and public security, military activities, investigations, driving, marketing, etc. Iris scanning prevents people from accessing a restricted area if not authorized (security). Fingerprints found on the weapon certify who used it (investigation). A facial expression indicates the degree of satisfaction with the sight of a commercial product (marketing). Innovative applications available in complex biometric systems as well as in commercial-grade mobile devices leverage data from all possible sensors. Sensors, pre-installed on devices or externally connected, continuously measure what a person, or a group of them is or is doing at that moment. New opportunities for shooting from different angles and heights are made available using drones and robots equipped with on board sensors.

The measurements are grouped and classified in a broader context that sees their analysis also developed in the long term, both for physical and behavioral aspects. New devices and new

problems introduce new challenges in this research field that require new application methodologies. This is what has been studied and implemented during the three years of Ph.D. and is presented in this work.

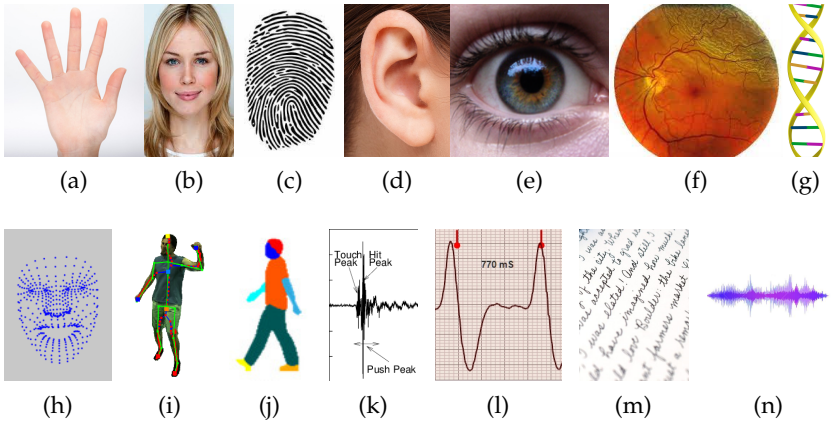
1.1 BIOMETRICS

Not everyone knew the meaning of the term *Biometrics* until ten years ago. Nowadays, its use is more widespread in everyday contexts, even in movies that often show its use. It is a very old term, derived from the Greek words *"bìos"* and *"métron"*, life and measure. It is the term to indicate a measurable physical aspect or psychological condition of a person, in addition to several purposes, usually to detect, identify, and recognize subjects. Biometrics, in general, is everything that can be measured of human beings [65].

Precisely, we can take the so-called hard biometrics, citing some such as face, iris, and measure their appearance, or soft biometrics such as emotion, voice, handwriting, keystroke, and walk and define their status. Some of these are complex traits and are believed to the uniqueness be a affected by environmental factors during pregnancy (e.g. iris and fingerprints) [71, 72].

Three types of origin for biometric traits are distinguished [19]: *genotypic* traits are those defined by an individual's genetic constitution, *randotypic* traits are those generated early in the embryo's development, and *behavioral* characteristics that a person acquires as a result of learning in the environment while growing up. These characteristics are spread randomly throughout the population and randotypic traits are usually considered the most valuable features for biometric applications due to the requirement of absolute uniqueness feature sets per subject. Table 1.1 show all three factors contribute to a biometric trait. The uniqueness of biometrics is often inversely related to the ease of acquisition. Fig.1.1 shows some biometrics from left to right in ascending order according to uniqueness.

Biometrics traits are used to measure what a person is at a given instant of time in appearance or emotional state. Recently, attention has also been given to the short-term dynamics of the measurements. Depending on how they evolve during an



(a) (b) (c) (d) (e) (f) (g)

(h) (i) (j) (k) (l) (m) (n)

Figure 1.1: The most widely studied biometrics:(a) Hand, (b) Face, (c) Fingerprint, (d) Hear, (e) Iris,(f) Retina, (g) DNA, (h) Emotion, (i) Body Pose, (j) Gait, (k) Keystroke frequency, (l) ECG signal, (m) Handwriting, (n) Voice signal. In first row are the physiological and in the second the behavioral.

Table 1.1: Relative importance of each factor (■ is low, ■■■ is high).

	Origin factors		
Biometric trait	**Genotypic**	**Randotypic**	**Behavioral**
Fingerprint (only minutia)	■	■■	■
Signature (dynamic)	■■	■	■■■
Facial geometry	■■■	■	■
Iris pattern	■	■■■	■
Retina (Vein structure)	■	■■■	■
Hand geometry	■■■	■	■
Finger geometry	■■■	■	■
Hand (Vein structure)	■	■■■	■
Ear form	■■■	■	■
Voice (Tone)	■■■	■	■■
DNA	■■■	■	■
Odor	■■■	■	■
Keyboard Strokes	■	■	■■■
Gaze (movements)	■■	■	■■■

individual's growth or aging, some biometrics are persistent, and others are not. They can also be defined as deep and non-deep based on their informative details and classified uniqueness with respect to the entire population, such as the DNA is deep,

due to the pathology indiscretions it can be contains. However, all of them fall under the protection law of privacy and their measurement requires the owner's consent or else privacy is violated. A broader discussion of biometrics can be found in the book *"Moderne tecniche di elaborazione diimmagini e biometria"* [95]. With that being said, we can give an overview of biometrics.

1.1.1 *Biometric and biometric traits*

Deoxyribonucleic Acid (DNA) may be considered the representative par excellence of biometrics. It is a very intimate photograph of an individual. It retains accurate information on how an individual is physically like, and some research investigates something about behavior [52, 61]. It is present in blood traces, hair, skin pieces, saliva, in short in all what is a biological piece of an individual. It is also unique among homozygous twins. It is immutable during growth, and it does not wear out over life unless smoking and alcohol abuse. On the other hand, it is difficult to pick up and process quickly, the amount of analysis on a sample is limited, and laboratory equipment is needed. Therefore, it limits its use to cases where a high level of accuracy is required, such as in forensics field.

Citing biometrics *Iris*, has a good level of uniqueness, they are different even in the eyes of the same individual [34]. Iris traits are less hard to obtain and to handle then the DNA, with good results even via commercial devices like smartphones. However, acquiring an iris is not easy, the muscles around the iris change shape continuously, it is full of reflections, and from afar sophisticated high-resolution cameras are needed. Capturing a photo of an iris from an uncooperative subject could be invasive and unsafe. It is not widely used because its use is limited by privacy laws which require higher levels of data protection.

The *Retina* trait has a high level of uniqueness in each individual and each eye. Its measurement is the image of the vein pattern beneath its surface in eye. The retinal vasculature is stable, unique, and not easy to change or replicate, thus being claimed to be the most secure biometric. A factor deterring public acceptance is that the image acquisition entails contact with the eyepiece. It can reveal some health conditions (e.g., hypertension

or diabetes), which is deterring the public acceptance. It is difficult to measure if the subject is non-cooperative, and the users perceive the technology as high intrusive.

The *Face* is the more familiar biometrics. It is a good compromise for easy capture and computation, and subjects undergoing measurements often are agreed. It can be captured using professional cameras or mobile devices. The level of uniqueness is not extremely great; there are people that share facial features, such as relatives or homozygous twins, and it can be difficult to tell their apart. In many circumstances, however, it is still a valuable biometric feature. As for the challenges given by the *"recognition at your fingertips"* in market for financial transactions, see smartphone applications, they pulled the best methodologies applied for the face to an accuracy close to perfection, in ideal condition [56].

The *Ear* trait is in many ways comparable with face trait [124], and often used in the absence of valid fingerprint for identification [76]. It has a uniform colour distribution, low changes over the lifetime, and low variability with expression as advantages. The pinna has a structure of cartilaginous tissue distinguishable and the ear shape is distinctive. The measurement includes taking a photo, taking a thermogram picture, and taking earmarks where are found sufficient distinction to be used for biometric purposes. For acceptable measurements, the requirement of users' cooperation is implied due it could be completely or partially covered by hair or ear muffles, as discussed by [5].

The most common biometric for identification purposes is *Fingerprint*. The same person has a different one for each finger, and it is so for the same finger of identical twins [68]. A fingerprint is a pattern on the surface of a fingertip consisting of alternating skin ridges and grooves. Achieved accuracy, proved to be very high [86], is adequate for authentication of a few hundred users. Using all fingers of the hand, additional information is provided, and a large-scale identification is possible, millions of identities can be involved. Undergoes change over time due to wear, frequently there are significant cuts and abrasions that keep adding due to manual work or accidents. Genetic factors, aging, and the environment can also be problems.

Hand Geometry biometrics is most commonly used in military field. It is based on measurements of traits of the human hand, such as the shape, size of the palm, lengths, and widths of the fingers. It has been scientifically studied for some time. Studies on the geometry of the hand have given good results on classification and verification. Results indicate that it can be considered useful for identification purposes in medium / high security environments [112]. It cannot help in systems that require identification of an individual on a large scale as in a population.

Analysis of the *Facial expressions*, *body poses*, and *gait* provide information about a person's physiology structure, even if more interesting about the emotionally state and intentions. Knowing the emotion of an individual is useful in many contexts such as marketing, security, medicine, human-computer interaction, and automotive [14]. Initially, seven are the basic emotions on which studies have focused: *anger, contempt, disgust, fear, sadness, happy, surprise* [43], and later *neutral* was added. Body pose estimation methodologies have achieved excellent levels of precision, even without using specific sensors, simply using a photo (e.g. OpenPose) [26]. Gait analysis studies the systematic human motion style and pace during walking to help athletes in sports, to identify posture disease in medicine, to locate people on the run and to recognize person for forensic use [17].

Different dynamic styles of *keystroke* to send dash and dot signals of Morse code, identify different telegraph operators. This feature allowed to authenticate messages received during the second World War [21]. It is even more accurate if a keyboard with letters and numbers is used [88]. The definitions of a precise personality attributable to a specific identity are in *handwriting* and *signature* behavioral biometrics. Handwriting and signature are behavioral biometrics and are influenced by the physical and emotional conditions of writers and signatories. They are used in the courts for calligraphic studies and investigations even for texts written in Chinese [132]. Further, professional counterfeiters may be able to reproduce them to deceive the system.

The *Voice* is behavioral biometrics due to physiological factors. It is based on physical structure of the nasal cavities, vocal tracts, mouth, and lips. necessary for the synthesis of sounds. Recognizing a voice mean recognizing a speaker [23]. The Voice timbre

changes over time in the growth phase, due to emotional state, medical condition, etc. Is not appropriate for a large-scale identification of a speaker in text-independent systems, less difficult for a text-dependent system.

The *Gaze* is never really fixed. It is characterized by micro movements of the eye, called saccadic movements, which are imperceptible to the human eye. The dynamics of these action have origin factor both physical and behavioural. In many cases, they draw out the psychology trait or the pathological state of a person. Often, they are often distinctive to the individual or relative to what one is looking at. The study on gaze analysis carried out in [24] showed that a person's way of observing a known face is different in the case of an unknown face. Recent studies have shown that it can be a distinctive trait for contactless authentication methods, [100].

Table 1.2 lists the basic features and characteristics of the biometrics traits discussed [33].

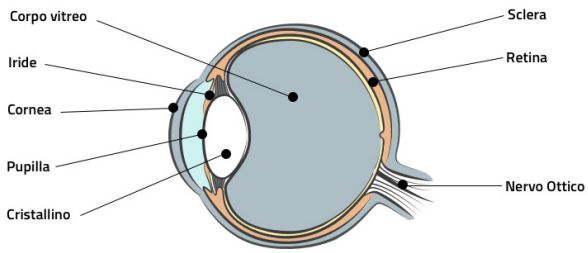


Figure 1.2: The details of eye's anatomy.

Clearly, each biometric trait can be considered as a different component to give rise to new types of measurements. In hand biometric, vein lines can be observed, for example. It is possible to consider one biometric as part of another or as a subgroup, generally. To explain, reflect on the face and its components, eyes, nose, mouth, skin, and so on for the eyes, sclera, iris, and retina Fig. 1.2. The vastness of the elements in the set of biometrics as a whole, sensors useful for the acquisition, scientific methodologies for their treatment, and fields of application can be easily understood. There is an entire category of biometrics that arises from signals transmitted by the human body. Recently, some of these have also been made acquirable by wearable devices. These

Table 1.2: Summary of the biometrics mentioned and their basic features and characteristics.

	Features and Characteristics		
Biometrics	**Principal details**	**Trait**	**Applications**
DNA	High accuracy, Low invasive, unfalsifiable, immutable (in limits)	Physical	Law Enforcement Forensics, Medicine
Retina	Low errors, unfalsifiable, more invasive, immutable (in limits)	Physical	Law Enforcement Forensics, Medicine
Iris	High accuracy and predictiveness, immutable, high randomness	Physical	Identification as Aadhaar card in India, national border, restricted access areas, airports and seaports, Medicine, Banking systems
Face	Contactless, easy storage,convenient,process fast,light and illumination conditioned, mutable over the time	Physical	Access Control Verification, Human Computer Interaction, Criminal Identification, Surveillance, Gaming
Fingerprint	Secure, reliable, high accuracy,process fast, low memory consumption, mutable by, by contact wear	Physical	Driver Authentication, Law Enforcement Forensics, License and Visa Authentication, Access control, Passport
Hand Geometry	by contact, perishable, small template, fast processing	Physical	Military access control, Commercial authentication
Facial expression	emotion conditions, non-invasive, contactless, fast processing	Both	Marketing, Medicine, Human Computer Interaction, Forensics
Body Pose	non-invasive, computational expensive, low accuracy	Behavioral	Law Enforcement Forensics Medicine
Gait	non-invasive, computational expensive, low accuracy	Both	Medical diagnose, Osteopathic and Chiropractic, Comparative biomechanics
Keystroke	non-invasive, emotional condition, small template, no interface, by contact	Behavioral	Law Enforcement Forensics Medicine
Handwriting	high accuracy, by contact	Behavioral	Law Enforcement Forensics Medicine
Signature	high accuracy, by contact	Behavioral	Banking system, Online authentication
Voice	Contactless, non-invasive,easy capture, no interface,	Both	Web based transactions, Voice Response based health and banking systems.
Gaze movement	Medium accuracy, Low invasive, unfalsifiable, mutable due to age	Both	Medicine, Recognition, Psychology

signals have long been studied from a clinical point of view but have been found to be useful for identification purposes alone or in combination with other biometric data. Electrocardiography (EKG), electroencephalogram (EEG), heart rate, dental prints are just few examples. Others come from the use of medical devices such as X-rays, tomography, and ultrasound scans. Others are strange things like skin dermis, forehead wrinkles, and hand hair. Therefore, the decision to use one instead of another should be made according to the intended and expected result. A marketing study projected to 2027 makes predictions on the most used biometric traits, as shown in Fig. 1.3.

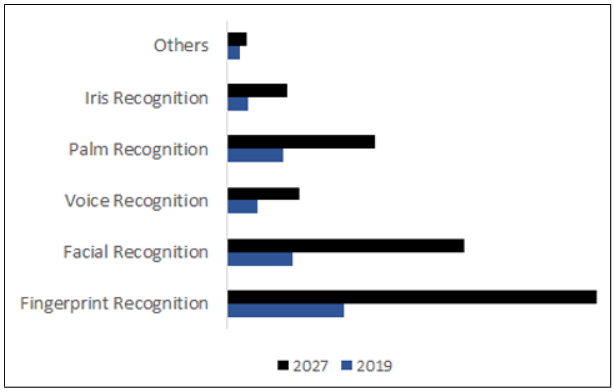


Figure 1.3: Use of biometric traits marketing study by emergenresearch.com.

1.1.2 *Biometrics classes*

The first possible subdivision of the biometric set is the physiological and behavioral one:

- *Physiological* Biometrics are those retaining biological characteristics of an individual. Among these are found DNA, retina, iris, hear, fingerprint, facial patterns, finger, and hand and palm geometry, voice, finger and palm vein, odor, ECG. and others. Look at the first row of Fig. 1.1 to see some. Some are stiff, others are elastic because the muscles of the body can change shape or trim, see the lips [36].

- *Behavioral* traits describe the behavior and psychological state of an individual, e.g., emotion, keystroke, signature, gait, body pose, handwriting, speech, and still others. Many of these are physiologically elastic, and are analyzed from the point of view of dynamics, you see the dynamics of the lips in speaking. An example of behavioral biometrics is the pattern by which a person looks at a face, as studied in [24]. Fixation points over time allow one to discriminate a subject from another. Some of behavioral biometrics are shown in Fig. 1.1 at second row.

Physical and behavioral biometrics can be combined to build multiple biometric systems able to prevent access to restricted areas or to protect private data in a more secure way. Ideal biometric trait characteristics are: the *Universality*, that is the trait should occur in every person or in many people as possible (Do all people have it?); the *Distinctiveness/Uniqueness*, that measures the degree of the trait uniqueness and ensure the dissimilarity between individuals (Can people be distinguished based on it?); the *Permanence*, meaning the feature should be reasonably immutable over lifetime, in order to remain meaningful (How permanent is it); the *Collectability/Measurability* intended as the easiness degree in acquisition or measurement of the trait (How well can it be captured and quantified?); the *Performance*, as the trait must provide adequate precision, speed and robustness with the right technology (Is the matching fast and accurate?); the *Acceptability/Intrusiveness*, ensuring that the relevant population and law well accept capture and storage of that trait (Do people accept it?); finally, the *Circumvention/Resistance*, that measures how it hard to imitate the trait with an artifact or substitute (Is it fool proof?). Usually, no argument is found but we consider it useful for this discussion to add factors: *Elasticity*: the trait changes shape through the muscles and at the and we have a new trait, such as iris and face expression (Is it mutable?); *Dynamicity*: the trait is a dynamic mutation in limited time frame, such as lip motion and saccadic eye movements. The trait is the way the transformation took place (Does it have a dynamic?).

Table 1.3 summarizes the qualitative analysis of various biometric traits considering the seven basic factors on our experience and perception. The attributes considered are acceptability, per-

Table 1.3: Summary of the biometrics mentioned and their basic features and characteristics. Absence, Low, Medium, and H are denoted by A, L, M, and H, respectively.

Biometric traits	**Universality**	**Distinctiveness**	**Permanence**	**Collectability**	**Performance**	**Acceptability**	**Circumvention**	**Elasticity**	**Dynamicity**
DNA	H	H	H	L	H	L	L	A	A
Retina	H	H	M	L	H	L	L	A	A
Iris	H	H	H	M	L	M	M	M	A
Face	H	L	M	H	L	H	H	H	M
Ear	M	M	H	M	M	H	M	A	A
Fingerprint	M	H	H	M	H	M	M	A	A
Hand Geometry	M	M	M	H	M	M	L	M	H
Facial expression	H	H	L	H	M	H	L	H	L
Body Pose	M	L	L	H	L	H	H	H	H
Gait	M	L	L	H	L	H	M	H	H
Keystroke	L	L	L	M	L	M	M	H	H
Handwriting	L	L	L	H	L	H	H	H	H
Signature	L	L	L	H	L	H	H	H	H
Voice	M	L	L	M	L	H	H	H	H

formance, universal, distinct, permanence and collectability. H, M, and L denote High, Medium, and Low values, respectively, while A denotes absence. The deep analysis performed by Jain (Jain et al., 2000) and Proença (Proença, 2007) is in line with that presented here. Table 1.3 also shows the most challenging biometrics in terms of attacks, security, complexity, distinction, and so on.

1.1.3 *Biometric traits processing techniques*

Whatever the type of biometric trait used, it must be captured, acquired, processed, and converted into a mathematical model to be then registered and protected in turn. And it is precisely the *template*, together with the real fingerprint, that is all context the main player in the recognition process of an individual. The generated template becomes the data on which operating through two possible procedures: *verification* and *identification*. In the verification, the person declares their identity by typing a personal reference code or by presenting a card in which the template was previously registered while the system compares the bleed print with the corresponding model stored in the database or in the card in a previous *enrollment* procedure. However, for the identification, the comparison takes place between the live template and all templates in the system database. More details of these procedures are in Section 1.2. The treatment of a biometric trait can be schematized in two main operations: *detection* and *features extraction*, regardless of the purpose, as shown in Fig. 1.4.

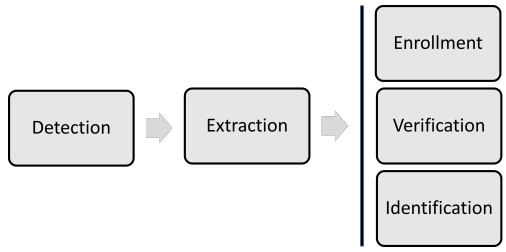


Figure 1.4: Trait processing schema.

Detection is the operation in which the information relating to a biometric trait is selected with respect to all information re-

turned by a sensor. In a photo, Region of interest (ROI) containing the face, for example, is selected. The most common methodologies used for detection are: Support Vector Machine (SVM) [97], Cascade [125], Scale-invariant Features (SIFT) [84], Random Decision Forest [59], and Convolutional Neural Network (CNN) [80]. More insights can be found in [81].

The extraction of the characteristics is operation that extracts salient and comparable information, for the purpose of recognition or verification, among those belonging to the biometric trait identified in the detection. For the treatment of signals and, in particular, the images, we find: Local Binary Pattern (LBP) [115], Gabor wavelets [12], Principal Component's analysis (PCA)[6, 69], Local Directional Number pattern (LDN)[103], Histograms of Oriented Gradients (HOG) [32], or in the world of Deep Neural Networks the CNN [39]. Some methods use the resulting new texture and leave the total interpretation of the data to the classifier. Others stack the features of salient areas or points in a vector to classify them. In both operations there is always a preprocessing pre-phase aimed at improving information and results (denoising, sharpness, and others).

1.2 BIOMETRIC SYSTEMS

Biometric systems are hardware and software architectures that use biometrics to verify and identify people in order to guarantee identity. Basically, it compares a pattern presented with one declared *Verification* or associates it with one present in a larger set *Identification*. More clearly verification is a one-to-one matching of a given biometrics to a known identity, for answering the question: *"is this the claimed person?"*. The verification aims to prevent multiple people from presenting the same identity. Examples are bank applications, computer logon, private data security, ATMs, physical access control, cellular phones, etc. Otherwise, identification is a one-to-many matching of a given biometrics against a database of known identities, for answering the question: *"who is this person?"*. The identification purpose is to prevent a single person from using multiple identities. Typical examples are criminal investigations, missing people, driver's license, etc.

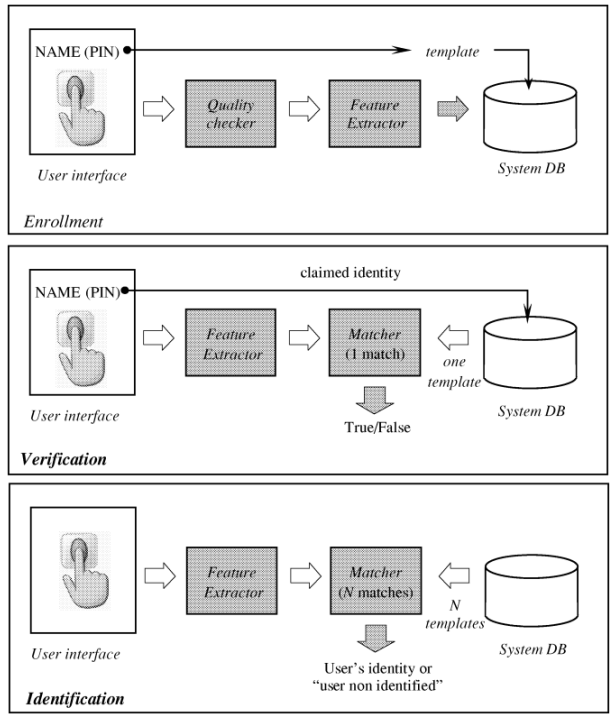


Figure 1.5: Operating modes basic diagram of a biometric system.

There is another modality, in which the biometric system can operate the *screening*: *"is this a wanted person?"*. The screening software determines whether a person is on a person watch list. Airport security, access to public events, and other surveillance solutions are examples of screening applications. A moderate number of identities are on the screening watchlist (e.g., a few hundred). In Fig. 1.5 the basic modalities schema of a system is shown graphically, with respect to biometric fingerprints.

The identities are represented through a template that preserves the distinctive characteristics of the biometrics used. The template is obtained in a preacquisition phase called *enrollment* and is encoded in a digital sequence named *biometric key*. Enrollment is the phase where the acquisition and processing of the user's biometric data is carried out. These data will be used by the system in subsequent authentication/identification operations.

The verification criterion may be expressed as in Equation 1.1. Given an input vector X_Q of biometrics features and I a claimed

identity, assess if the claim identity is True (I, X_Q) is in the class w_1 (a genuine user), else is in w_2 (an impostor). X_I is the vector corresponding to user I and is compared with (I, X_Q) to determine its category.

$$(I, X_Q) \in \begin{cases} . & w_1, \quad if\ S(X_Q, X_I) \geq t \\ & w_2, \qquad otherwise \end{cases} \tag{1.1}$$

The similarity or matching score $S(X_Q, X_I)$ is termed between feature vectors X_Q and X_I and measured by the function S, and t is the set threshold. The threshold t is necessary because the biometric measurements of the same individual are almost never identical, taken at different times.

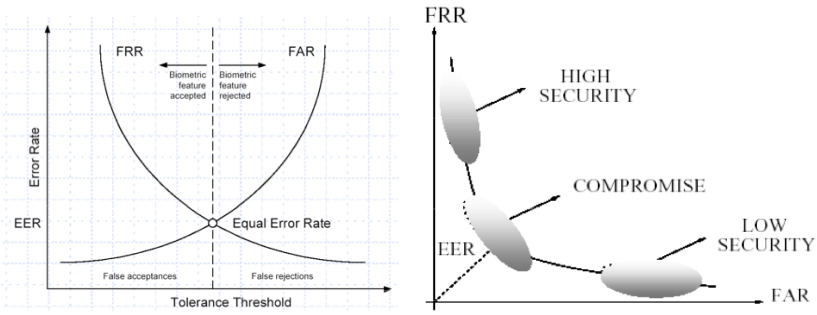
$$X_Q \in \begin{cases} I_k, & if\ \max\limits_k\ S(X_Q, X_{Ik}) \geq t, k = 1, 2, ..., N \\ I_{n+1}, & otherwise \end{cases} \tag{1.2}$$

Otherwise, the identification problem is formalized by Equation1.2. Given the biometric template X_Q, determine the identity I_k,$k \quad \in \{= 1, 2, 3..., N, N+1\}$. Here, $I_1, I_2, ..., I_N$ are the identities enrolled in the system and I_{N+1} indicates the reject case. The vector X_{I_k} corresponding to identity I_k, and t is a predefined threshold.

A biometric system is susceptible to many types of attack that can undermine system security, resulting in system failure. All attacks depicted can be divided into two categories. i) *Zero-effort attacks*: An opportunistic intruder's biometric features may be sufficiently similar to those of a properly enrolled individual, resulting in a False Match and a system security breach. This occurrence is linked to the likelihood of discovering a degree of similarity between templates originating from various sources through coincidence. i) *Adversary attacks*: This refers to the prospect that a determined impostor could impersonate a registered user by employing a lawfully enrolled user's physical or digital artifacts. An individual can also intentionally modify a biometric feature to avoid being detected by an automated biometric system [66] [65].

In the below Section 1.2.1 are discussed the metrics to evaluate the biometric system performances (FAR/FRR/EER/ROC).

1.2.1 *Performance of a biometric system*



(a) (b)

Figure 1.6: Performances curves of biometric systems.

The metrics for evaluating the performance of a biometric system are due to two factors:

1. *True Recognition* happens when a person is recognized correctly or is actually an impostor.

2. *False Recognition* represents the case in which the system fails, so recognize a false identity or not recognizing a true identity.

It followed the evaluation of two errors in false recognition:

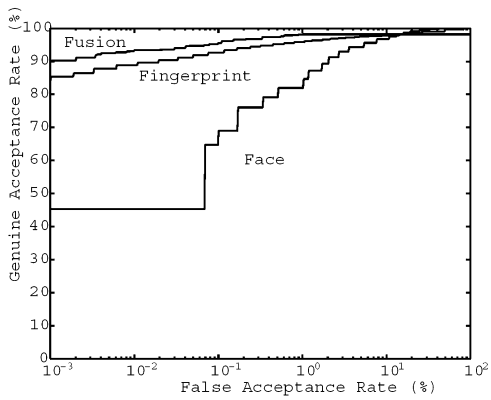
- *False Rejection Rate (FRR)* is the number of false rejects in percentage. Authorized users are rejected incorrectly.

- *False Acceptance Rate (FAR)* is the number of false acceptances. Users unauthorized are accepted erroneously.

FAR and FRR are two strictly inversely correlated quantities. The sensitivity of the system is established by regulating the relationship FRR/FAR.

Denoting by $FRR(t)$ a decreasing monotonic function and by $FAR(t)$ an increasing monotonic function, where t is the degree of tolerance and quality of the system, Fig. 1.6a. If t is low, we have a high number of false rejects, if t is high, we have a low number of false acceptances. At the intersection of the

two curves, we find the Equal Error Rate (ERR), point of balance, where $FRR(t^*)/FAR(t^*) = 1$ with t^* is the tolerance at that point. Therefore, if $t > t^*$ the ERR decreasing for $t < t^*$ ERR increasing. Through t the response of the system can be regulated, as shown in Fig. 1.6b. For some types of application, it can be convenient to have false acceptances or other false refusals, such as a turnstile in the first case to speed up the entrance or for a bank in the second case to be sure of the person entering.



(a)

Figure 1.7: ROC curve for different biometric techniques [60]

Receiver Operating Characteristic (ROC), showed in Fig. 1.7, is the curve that expresses the trade off between the FAR and FRR relating in various thresholds. The ROC curve is a widely used metric in expressing 1:1 matcher performance.

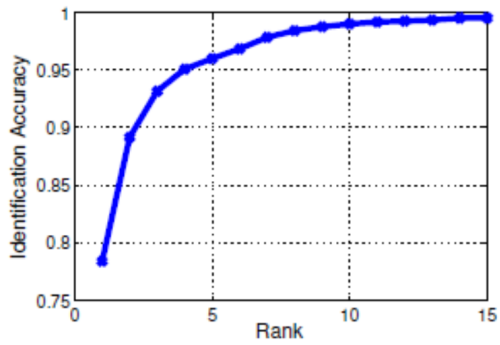


Figure 1.8: Cumulative matching characteristics (CMC) curve for subject retrieval.

The Cumulative Match Curve (CMC) is a metric that represents the performance of biometric identification systems that provide ranked lists of candidates in 1:m search engines, see Fig. 1.8.. It assesses an identifying system's ranking skills. To evaluate the CMC, the match scores between the request template and their biometric samples in the database are sorted. Lower is the rank of the genuine matching biometric in the enrollment database, better is the 1:m identification system.

Moreover, in [105] the author demonstrated a contradiction to previous concepts, that there is a relationship between the ROC and the CMC. They show that the CMC is also related to the FAR and FRR of a 1:1 matcher, i.e., the matcher that is used to rank the candidates. Therefore, the CMC provides no additional information other than the FAR and FRR curves. CMC is another way to display data and can be calculated from FAR and FRR.

1.2.2 *Open-set and closed-set identification*

Biometric recognition can be tested in closed-set or open-set scenarios, depending on the testing technique. All testing identities are predefined in the training set for the closed-set protocol. It is only reasonable to assign identities to test biometric templates. The verification is the same as conducting identification for a pair of the biometric trait in this circumstance. As a result, closed-set FR can be effectively addressed as a classification problem with separable features. The testing identities are frequently disconnected from the training set in open-set protocols, making FR more demanding while remaining near to practice. We must map the biometric trait to a discriminative feature space because it is impossible to classify biometric traits to known identities in the training set. The verification between the probing biometric trait and every identity in the gallery might be seen as biometric trait identification in this case. The key to learning discriminative large-margin features in open-set FR is to think of it as a metric learning problem. Under a given metric space, desired features for open-set FR should satisfy the requirement that the greatest intra-class distance is smaller than the minimal inter-class distance. If we want to obtain perfect accuracy using the nearest neighbor, this condition is required. Due to the intrin-

sically great intra-class variance and high interclass similarity that faces exhibit, learning features with this criterion is often difficult ([83]).

1.2.3 *Limits of Unimodal Biometric Systems*

A biometric system that operates using a single biometric feature is called a unimodal system and has the following limitations:

i) *Noise in sensed data*. The recorded data may be noisy or distorted. Examples of noisy data include scarred fingerprints and cold-changed voices. Noisy data can also be the result of sensor defects or improper maintenance (such as the accumulation of dirt on the fingerprint sensor) or poor environmental conditions (such as poor lighting of the user's face in a facial recognition system). Noisy biometric data may be mistakenly matched to templates in the database (see Noisy biometric data can be mistakenly compared to the database template, resulting in the user being mistakenly rejected.

ii) *Intra-class variations*. The biometric data collected from an individual during authentication can be significantly different from the data used to create the template during registration and will affect the matching process. This variation usually occurs when the user changes the sensor or changes settings improperly during the validation phase. As another example, different psychological conditions of individuals at different times can lead to very different behavioral traits.

ii) *Inter-class similarity*. Biometric traits are believed to vary widely from person to person, but there can be significant similarities between the classes of features used to represent those traits. This limitation limits the identity provided by the biometrics feature, due to low distinctiveness. Therefore, all biometric functions have a theoretical upper limit on their ability to distinguish themselves.

iV) *Non universality*. While each user is expected to own the biometric trait being acquired, in reality, they may not have it. Fingerprints can be deteriorated or absent after the enrollment phase occurred a long time before. It is possible of a subset of users to not own a particular biometric.

v) *Spoof attacks*. Fraudsters may attempt to forge the biometric properties of legitimate registered users to bypass the system. A relevant example is when we make use of behavioral traits such as voice and signature. However, physical properties are also susceptible to spoofing. For example, it is possible to generate an artificial finger or fingerprint.

1.2.4 *Multi-modal and Multi-biometric Systems.*

In a complex biometric system, different biometric traits, behavioral or physiological, can be used in a combined way to improve the system performance, but at the expense of computational weight. This category of systems is called Multimodal Systems. Using multiple biometric modalities, some of the limitations imposed by unimodal biometric systems can be overcome. For example, using multiple fingers of the same person or face and a single fingerprint. Such a system, known as a multimodal biometric system, is expected to be more reliable due to the presence of multiple independent evidence. These systems can also meet the high-performance requirements of various applications. The multimodal biometric system addresses the issue of non-universality as several features ensure proper population coverage. In addition, the multimodal biometric system provides anti-spoofing measures by making it difficult for an intruder to forge multiple biometric properties of a legitimate user at the same time (see the five fingerprints). The system ensures that the "living" user actually exists at the time of data acquisition by asking the user to present a random subset of biometric entities (such as the index on the right hand and face). A multimodal biometric system can operate in one of three different modes: *serial mode*, *parallel mode*, or *hierarchical mode*.

Serial mode operations typically use the biometric output to narrow down the number of possible identities before using the next feature. It acts as an index scheme for the identification system. For example, a multimodal biometric system that uses the face and fingerprint can first use face information to get some matches and then use fingerprint information to converge to a single identity, a hierarchic sequence. This contrasts with the parallel operation mode, which uses information from mul-

tiple features at the same time to perform the discovery. This difference is essential, in the parallel mode of operation, it is not necessary to process different biometric functions at the same time. You can also make decisions without getting all the features. This reduces the overall recognition time. Hierarchical schemes combine individual classifiers into a tree-like structure [66] [65].

The fusion of biometrics data can be done at different levels in the process of verification/identification: *Feature level*, *Match score level*, *Rank level*, *Decision level* in-depth is dedicated in the next Sub-Section 1.2.5.

1.2.5 *Data fusion in biometrics*

Usage of multiple biometrics does not always imply high system performance of using one. A poorly designed multiple biometrics system can reduce performance compared to a single biometrics system, increase operating costs, and cause inconvenience in management. To use more biometric traits, typically no more than two or three, in the same biometric system, architecture and methodologies of the single trait are integrated, Fig. 1.9. The resulting system can be operating in three modes: serial, parallel or hierarchical. The resulting information will be filtered, merged, or both, respectively, by mode, to produce a unique result.

In a multi-biometric system, it is necessary to decide at what level of the process the fusion of data results is to be carried out [110]. The levels of fusion can be distinguished in: *Sensor level, Feature level, Match score level, Rank level, Decision level*.

Sensor level: raw data collected by multiple sensors can be processed and integrated to generate new data from which features can be extracted. For example, in the case of face biometrics, multiple 2D textures obtained from different sensor are fused to generate 3D depth for feature extraction and matching.

Feature level: feature sets extracted from multiple data sources can be merged to create a new feature set that represents an individual. For example, you can augment the hand geometry with facial eigen-coefficients to build a new high-dimensional feature vector. You can use feature selection / transformation procedures to get a minimal set of features from a high-dimensional feature vector.

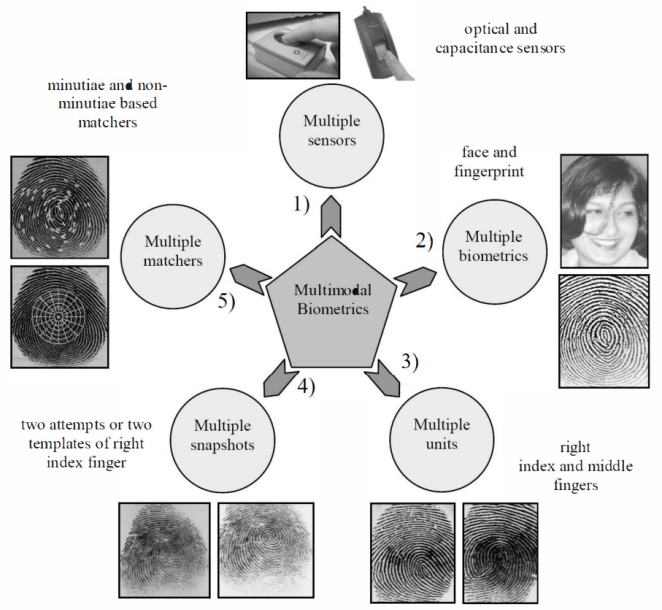


Figure 1.9: Multi-modal biometric system [101].

Match score level: In this case, multiple classifiers output a set of matching scores, and they are merged to produce a single scalar score. For example, you can combine the match scores generated by the user's face and hand modality through a simple sum rule to get a new match score and use it to make the final decision.

Rank level: this type of fusion is related to an identification system where each classifier assigns a rank to each registered ID (the higher the rank, the better the match). Therefore, a merger involves integrating multiple ranks associated with an identity and determining a new rank that will help make the final decision. You can make the final decision using techniques such as Borda count [45].

Decision level: if each matcher issues its own class label (i.e., approve or reject in the validation system, or the user's ID in the identification system), simply use a technique such as majority voting. You can get one class label. Decision making, action, knowledge space, etc

At which stage of the biometric system to use information fusion, depends on many factors and some considerations need to be made. Merging the information coming from different

sensors, *Sensor level*, is not always simple, as they represent different phenomena, they are organized differently and merged can not always help, on the other hand, when possible, the amount of information obtained is consistent. The compromise of fusion at the *Feature level* is that if on the one hand we obtain a lot of information, on the other hand it is difficult to carry out, practically impossible if very different characteristics are used. At *Match score level*, we have scalar values that are simple to manage but sufficiently rich in information, while on the other hand we need to normalize the distributions of scores from different systems. The merging of multiple ranks for the same identity is simple *Rank level*, but if one of the systems is not performing, it can greatly compromise the final result. At *Decision level*, the fusion is simple to implement (answers 0/1), but little information is available.

Part II

THE SHOWCASES

2

A NEW MULTI-BIOMETRICS DATASET, MUBIDUS-I

The widest possible and realistic database is the basic element for designing and perfecting biometric measurement and comparison methods. The creation of new sensors and sensing devices enables to open new research fields and consequently rises needs for new datasets specifically designed for the applications they give birth to. Collecting and annotating new datasets is an important contribution in this regard. In this case, the first goal of this work is to collect multi-biometric/multimodal data aimed to reduce the gap between indoor and outdoor acquisition, biometric matching of features acquired by different devices like cameras, smartphones, and a drone. With the availability of these new data, a beneficial contribution to the state of the art is given by making new experiments feasible under more realistic conditions in terms of subject posture, light exposure, and camera framing. In this chapter, we explain in detail how the new dataset was structured and what equipment was used to obtain it, MUlti-BIometric and multipurpose Dataset developed at University of Salerno (MUBIDUS-I) [35].

2.1 DATASETS AVAILABLE IN THE LITERATURE

In the literature, there are many datasets with the most disparate characteristics and purposes, collected over the years. It is impossible to list them and discuss them all. Below are some collections with similarities or something in common with the one presented in this study.

Many datasets are on individual biometrics, some of which have analogies with the proposed one. One of these datasets is MUCT [89]. It contains 3755 images of human faces, of which 76 landmarks have been located. The dataset provides information such as lighting, age, and ethnicity. The protocol used by MUCT is as follows: Each subject was photographed with five front-

facing webcams using ten different lighting systems. 2/3 of the resulting photographs were used for each subject.

Since our dataset contains iris details for many of the subjects, we decided to analyze some public eye imaging dataset (iris). VISOB is a public data set of eye images taken by 550 volunteers, [106]. Three different mobile devices were used as capture tools: iPhone 5, Samsung Note 4, and Oppo N1. The subjects made their own images autonomously. In VISOB, multiple images were recorded under four lighting conditions for each recording session: normal light, office light, low light, and daylight.

MICHE-I is another data collection consisting of 3732 eye images, [37]. Subjects were asked to take self-photographs of their own irises. Four photos were taken for each device and each recording mode (indoor and outdoor). Indoor shooting mode uses artificial light sources, sometimes combined with other daylight, while natural sunlight is used for shooting outdoors. Three different device types were used in MICHE-I: iPhone 5, Samsung Galaxy IV, and Galaxy Tablet IV.

COMPACT Is a database for facial recognition studies of subjects less collaborative [127]. The dataset consists of high-resolution images of 108 subjects taken while passing through a fully automatic detection lock. This ensures that the data collected is consistent with traditional real-life problems in terms of different distances, representations, occlusions, pose changes, and motion blur. To capture multiple positions on the subject's face, the author used a rotating platform.

The data set proposed in [107], that is EGA, aims to overcome some of the main limitations of the previous data set of faces. In this dataset, facial images are organized according to certain categories, such as ethnicity, gender, and age. In particular, the faces were grouped into homogeneous and balanced categories. Another purpose of the dataset is to support studies on the benefits that soft biometric information, such as gender, can provide to a facial recognition system.

A data record for ear recognition is UBEAR. Photographs are taken with moving objects, in dynamic lighting conditions, and without special attention from subjects concerning the ear closure and pose [104]. A sample of 126 subjects was obtained, 44.62%

are men and 55.38% are women: a total of 252 ears by a single camera with a resolution of 1280×960.

Multi-modal/Multi-biometrics datasets are a form of biometric dataset that has gotten a lot of interest throughout the research phase. This type of dataset contains different biometrics of the same individual gathered at various periods. These datasets have the potential to increase the precision and accuracy and reduce time complexity in biometric systems using data fusion techniques at the score level to be applied and deployed. For example, QUIS-CAMPI is a dataset that comprises both full body video sequences and high-resolution head samples of people in a parking lot with a biometric recognition system that operates outside in completely unconstrained and covert conditions [96].

One of the first multimodal databases was BIOMET [50]. It was created by the Multimodal Biometric Identity Verification project and compiled with five modalities and temporal variability. The database was created at three different sessions three and five months apart and contains samples of faces, voices, fingerprints, hand shapes, and handwritten signatures.

Collecting a database that had characteristics in terms of number of subjects, number of biometric traits, and number of temporarily separate acquisitions was the goal of BiosecureID [47]. The database was collected at six different locations in an uncontrolled environment that simulated a realistic scenario in which 8 different biometrics traits of 400 subjects such as speech, fingerprints, iris, hand, face, writing, signature, keystroking, were recorded in 4 sessions.

A multi-modal database containing biometrics of subjects from various countries is MobBIO [114]. It contains biometric data from 105 volunteers. Each person provided samples of their face, iris, and voice. The nationalities of the volunteers were mainly Portuguese, but volunteers from Great Britain, Romania, and Iran also attended.

A database that belongs to this category, but with the particularity of the type of uniqueness of the subjects under consideration, is the Multi-modal Biometric Recognition for Toddlers and Pre-School Children [13]. Here of over 100 children (age range of 18 months to 4 years), face, fingerprint, and iris modalities were collected during six months in two separate sessions. A Cross-

Match scanner[1] for the iris, a Slap Crossmatch LScan scanners[2] for the fingerprint and a Nikon D90 DSLR camera[3] for the face were used.

Regarding the use of drones in a biometrics context, there are a variety of datasets available, some with specific purposes, others with a more general one. Through the use of a drone, the goal is to collect data from the above, at various angles, without causing any discomfort to the person being observed.

An example is the MiniDrone dataset, which provides flights worth of data recorded in outdoor environment during a parking area monitoring [16]. The objective is to keep an eye on the area, assisting in managing of parking spaces, crowd control and reporting relevant information such as mis-parked cars, dangerous manoeuvres, number of free parking spots, suspicious behaviors, etc. The information is categorized into three categories: normal, suspicious, and illicit behaviors.

For re-identifying people purposes, the MRP Drone dataset was collected [78]. Its aim is to maintain the consistent identity of human detection recordings during flight. The data is acquired from multiple flights in both an indoor and an outdoor environment that is both unrestricted and densely populated.

In order to simulate in the wild scenarios for face recognition, the DroneSURF dataset contains 200 videos of 58 participants caught by the camera aboard a drone [74]. Each video contains a group of individuals with differences across use case, geography, and acquisition time.

In the year 2018 Kalka et al. [73] collected the IJB-S dataset which features a component for face recognition of 10 UAV based videos. The dataset includes a wide range of campus scenarios for simulating real-world use cases.

The DroneFace dataset contains a series of facial images of 11 subjects, 7 males and 4 females, collected in an uncontrolled outdoor environment with varying fixed distances and heights [62]. The UAV used is equipped with a fixed sports camera, which is pointed parallel to the aircraft's motion. A fixed sports

1 http://www.crossmatch.com/i-scan-2/
2 http://www.crossmatch.com/Guardian-USB/
3 https://www.nikon.it/it_IT/product/discontinued/digital-cameras/2015/d90

camera is mounted on the UAV, which is pointed parallel to the aircraft's motion.

Acquiring either top-view or oblique data are also the characteristics of a recent dataset P-DESTRE, collected in 2020, purely for Pedestrian Detection, the Tracking and Re-Identification [75]. It helps research efforts in these subareas to search new methods able to work in real-world conditions with crowded scenes based on very low resolution and partially obscured data.

The dataset presented in Section 2.3, was created with the intention of contributing to the state of the art in this field and providing researchers with new challenges to problems using modern detection devices precisely MUBIDUS-I.

2.2 LIMITS OF THE EXISTING DATASETS

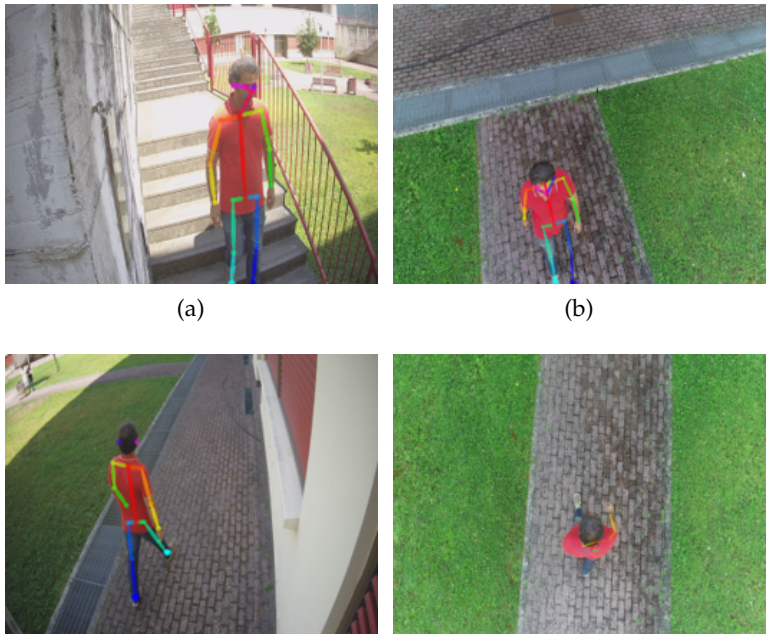
Being able to build an ideal dataset is very difficult and the evolution of devices and sensors would always lead to new needs and additions. However, it is possible from time to time to integrate with new acquisition methods and collect new datasets.

It is easy to understand that using a drone, it is likely that the quality of the video is not optimal because it is subject to factors such as height and speed of flight, viewing distance, unfavorable observation points, and environmental conditions. Otherwise, the new quality parameters in the construction of these aircraft, see flight controllers, gimbals for cameras, and SDK available, allow to partially overcome these types of problem.

The identification of a subject by means of drones is very demanding and it is more and more if the system operators are influenced by the operating environment, as analysed in [48].

Algorithms trained on datasets present in the literature or on specific ones collected for the occasion do not work well in some situations present in the presented dataset. As a demonstration of what has been said, we show in the following Fig. 2.1 some frames of the MUBIDUS-I processed with the OpenPose method [26]. Given a frame that contains the image of a person, this method is able to calculate the pose of the body parts, such as torso, head, limbs and others. As we can see, while for frame a) the calculation returns precise results, in the case of frames b), c), and d), there are some inaccuracies. For frame b), a pose is attributed to the

left leg of the subject, even if it is not present in the image. For frame c), the left arm of the subject is indicated, even if covered by the torso. In case d), viewed from above by the drone, the algorithm fails completely without being able to predict any pose. This is a proof of the problems introduced by the new dataset that will be present in many real world cases.



(a) (b)

(c) (d)

Figure 2.1: Some frames were processed by openPose algorithm.

2.3 MUBIDUS-I

MUBIDUS-I is a MUlti-BIometric and multipurpose Dataset developed at University of Salerno, as contributed to the state of the art. The dataset is very heterogeneous in terms of acquired biometric traits, acquisition protocol, devices used, pose of subjects, and lighting of the environments. It is Multi-biometrics, as it collects various captured and measurable biometric data, such as eyes, nose, mouth, ears, and periocular area. It is Multi-sensor, as different devices were used to capture the experimental data,

such as TLC, Smartphone, Drone, Macro Lens. The data was collected at different times and therefore is also Multi-session. It is Multi-protocol because there are different types of acquisitions exactly *Up/Down*, *Face Details*, *Hallway*, *Drone* protocols. The participants are all collaborative but acquired in different pose angles and in a controlled and uncontrolled way, precisely Multi-pose and Multi-mode. It is Multi-environment, the data acquisitions were made both indoors and outdoors. In the indoor case, data was recorded with cameras and mobile devices. In the outdoor, camera, phone, and included a drone all together recording. Another important feature that differentiates it is that, in certain acquisition sessions, the biometrics are acquired simultaneously from different devices, as in the case of the *Drone* session detailed below.

The original idea was to create a dataset containing video sequences of moving objects, which was made with the drone in an external environment and previously recorded in an internal environment. This idea later evolved into creating indoor photos and video sequences created with cameras and two mobile devices. The dataset contains biometric data from 80 people who work or study at the University of Salerno, 13 women, and 67 men. The age ranged from 22 to 28 years, and all subjects were of Caucasian ethnicity.

Within the dataset, there are video sequences that contain moving subjects. These recordings are indoors and under controlled conditions. Each subject has an average of 135 shots between indoors and outdoors. Indoor shots include those taken in the BipLab laboratory and those taken in the hallway with three bullet-type cameras. The outdoor acquisitions were made in the courtyard using cameras, drones, and smartphones.

Not all subjects have data collected through all protocols. Table 2.1 shows the number of acquired test subjects, categorized according to the protocol.

Unlike the data sets discussed above, MUBIDUS-I consists of images and videos in controlled and uncontrolled environments. In addition, recordings were made with the drone. Differently from MUCT, the facial images were recorded by 3 cameras that were oriented at different angles, even slightly from behind and from above as in a real world scenario, to record biometric char-

Table 2.1: Number of subjects acquired per protocol.

Up/Down	*Face Details*	*Hallway*	*Drone*	**Different subjects**
60	52	14	36	80

acteristics such as the ear, which were later also recorded by a drone and a stationary external camera. For each spectacle wearer, recordings were made with and without glasses. The laboratory artificial light was used as the lighting system.

In contrast to VISOB and MICHE-I, the latest generation of mobile devices were used to record iris. IPhone-8 and Samsung Galaxy-9 with additional macro lenses. To improve the accuracy of iris recording, subjects did not autonomously record iris images. Additionally, all rooms were under the same indoor lighting conditions. The COMPACT dataset does not include top-down capture of faces while there are in MUBIDUS-I. Table 2.2 summarizes the comparison between the datasets mentioned.

Table 2.2: Dataset comparison.

Dataset	**Subjects**	**Biometrics**	**Environ.**	**Device**
MUCT	76	face	indoor	camera
VISOB	550	multi	indoor	mobile
MICHE-I	92	iris	multi	mobile
COMPACT	108	face	indoor	camera
EGA	469	face	indoor	camera
UBEAR	126	ear	indoor	camera
QUIS-CAMPI	320	full body	outdoor	camera
DroneFace	11	face	outdoor	camera
MUBIDUS-I	80	multi	multi	multi

2.3.1 *Methods and Tools*

The dataset was created considering the following protocols: *Up/-Down*, *Face Details*, *Hallway*, *Drone*. The first three were carried out indoors, while the third protocol was limited outdoors for practical reasons. All operate under uncontrolled lighting conditions typical of real environments. The *Up/Down* and *Face Details* were captured in an ideal and controlled context, while the others were captured in real and uncontrolled contexts. Tables 2.3 and 2.4 summarize the characteristics of the protocol and the devices used, respectively. In all protocols, the participants were aware that they were being recorded and cooperated.

Table 2.3: Protocol characteristics

	Mode	**Environment**	**Data**
Up/Down	controlled	indoor	frames
Hallway	uncontrolled	indoor	videos
Face details	controlled	indoor	frames
Drone	uncontrolled	outdoor	videos

Table 2.4: Devices used

	TLC	**Smartphone**	**Drone**	**Macro**
Up/Down	yes	no	no	no
Hallway	yes	no	no	no
Face details	yes	yes	no	yes
Drone	yes	yes	yes	no

Follow the technical and functional specifications about the respective equipment. The Fig. 2.2 showing the devices and Table 2.5 their specifications.

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

Figure 2.2: The devices used in the protocols. From left to right Bullet camera, iPhone 8, Samsung Galaxy Edge 8, Aukey 3-in-1, DJI Phantom 4 Pro.

Table 2.5: Devices specifications.

Device/add-on	**n**	**Specification**
Mini Bullet Network Camera	3	8Mp, res. 3840 × 2160
iPhone 8	1	12 Mp, focal F 1.8
Samsung Galaxy Edge 8	1	12 Mp, focal F 1.7, res. 4290 × 2800
DJI Phantom 4 Pro	1	res. full HD, 30fps
Aukey 3-in-1 Phone Lens	1	Macro/Fisheye, 150° Wide Angle

2.3.2 *Up/Down*

The *Up/Down* protocol consists of two acquisition sessions that differ in the inclination and height of the cameras. In the center of the scene there is a chair, around which three video cameras are placed, on which the subject to be recorded sits. Two cameras are placed on the sides of the tripods, one at the front. The distance between the center of the side camera tripods and the center of the chair is 150 cm for all sessions. During the *Down* sessions, the cameras are "down": the lateral cameras are at a height of 116 cm and 130 cm from the face, as in Fig. 2.3.

During the *Up* session, the cameras are at 188 cm tall, precisely are "up". In these sessions, the camera distances from the face are approximately 180 cm, and they are directed downward at

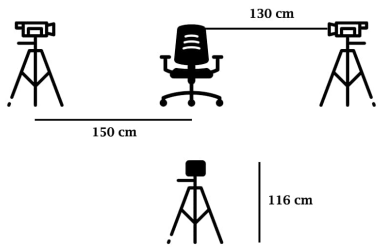


Figure 2.3: Cameras location of the *Down* session.

an angle of 45° to allow the subject to be captured, as depicted in Fig. 2.4.

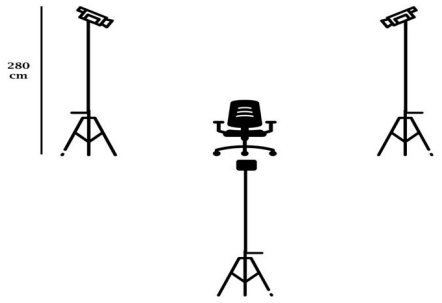
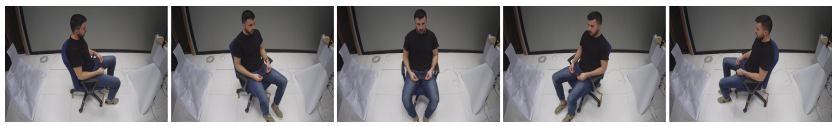


Figure 2.4: Cameras location of the *Up* session.

The protocol begins with the subject sitting in a chair where, if they have long hair, they are asked to cover their ears. During the recording, in the case of the Up sessions, they were asked to look at the camera and stare at it. During the *Down* sessions, they were told to look straight ahead. In each session, a sequence of acquisition actions is performed. First, the subject is asked to turn the body and head from the left camera and stop at an intermediate angle of about 45° to the right camera and in front of each camera. This is a controlled mode. After that, they will be asked to perform the same actions, but to stop freely at random angles at will. This is an uncontrolled mode. The test person repeats the same actions with and without glasses.

As mentioned above, software aids were used to capture simultaneous frames images from all three cameras. Therefore, in the case of controlled detection, the camcorders make simultaneous

recordings of a total of 15 images for each controlled mode for each rotation made by the subject. In uncontrolled mode, shots range from 60 to 80, depending on the speed of the subject turning. In Fig. 2.5 some sample frames recorded during the session are shown.



(a) Frames of controlled *Up* session without glasses.

(b) Frames of *Down* session without glasses in controlled mode.

Figure 2.5: Sample frames of *Up* and *Down* sessions

2.3.3 *Face Details*

Capturing biometrics traits at close range is the objective of the *Face Details* mode. Consequently, this modality aims to capture facial details, such as the nose and mouth, as well as the periocular area and the iris, at different positions of the eyeball. The task requires more explains in detail. The main theme of this modality is the periocular area. Particular attention is paid to the rotation of the iris. The recording equipment contains a chair in which the person sits. There are three cameras around the chair. They measure 116 cm in height and 40 cm from the person's face. The protocol is as follows. The subject sits with cameras close to his face and performs five rotations of the iris, taking three pictures with cameras at each rotation. When taking the first picture, the subject looks directly at the center camera. Then, keeping the face still, the person performs the following iris rotations: up, down, left, and right. This creates images of the periocular area viewed from different angles and with different iris directions, as shown in Fig. 2.6.

The same process is repeated with two mobile devices: an iPhone 8 and a Samsung Galaxy 8, placed on an easel 30 cm

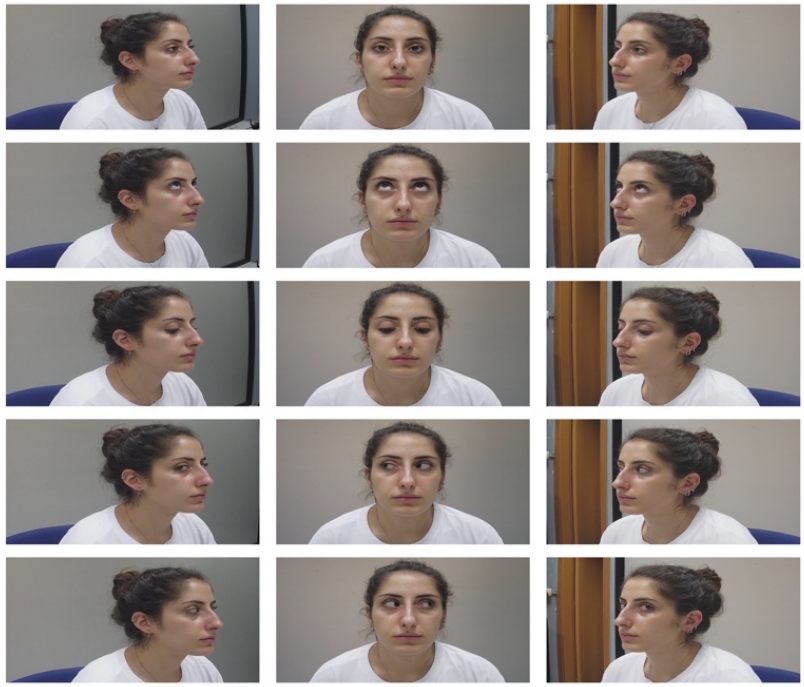


Figure 2.6: Zooming of face image from the three cameras.

from the person's face, with the camera in a frontal position. This design decision allowed us to compare the recording quality at different resolutions and to take recordings from different devices. An example of such a collection of recordings is shown in Fig. 2.7.

(a) Samsung Galaxy 8.

(b) iPhone 8.

Figure 2.7: Periocular images obtained via smartphones.

Once the recordings are obtained from these mobile devices, the iris is recorded using an iPhone 8 and a macro lens mount installed on the rear view camera. During this phase, some lighting problems were identified: when using a macro lens, the mobile device was placed too close to the iris, which affected the lighting: as in Fig. 2.8 you can see that the iris contains a partial reflection.

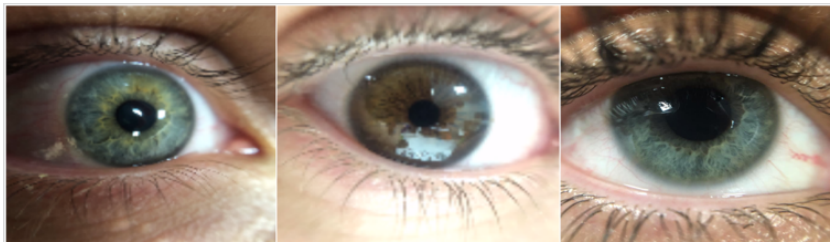


Figure 2.8: Images of iPhone 8 with on-board macro lens.

2.3.4 *Hallway*

This technique is carried out in a real-life indoor setting. Fourteen volunteers, all of whom have participated in prior studies and are aware that they are being tracked in an unrestricted manner, walk down a corridor following a preset round trip path. Three cameras are recording them at the same time throughout this walk. In the hallway, about halfway down the path, there is a turn. Cameras are positioned in such a way that the subject will be continuously captured during the trip. The walk is resumed both on the way out and on the way back in an L-shaped route. The first camera is set at the beginning, 12 meters distant from the others, pointing at the midway hallway corner. The other two cameras are placed at the corner: one is oriented towards the path start (and the first camera), and the last is positioned such that it may capture both the route back to the corner and the arrival (Fig. 2.9).

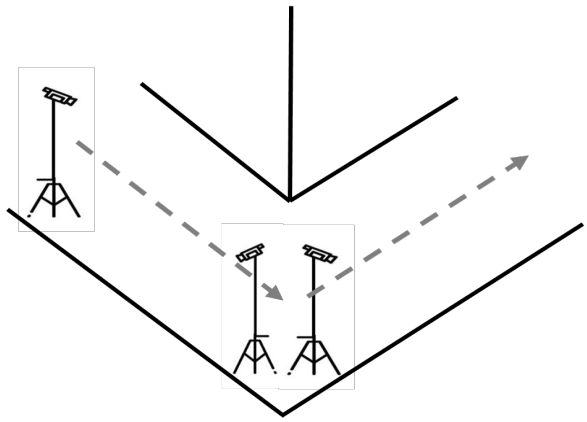


Figure 2.9: *Hallway* protocol path way scheme.

All cameras are placed at the same 273 cm high and the same tilted by 45° degrees downward with respect to the horizontal line, to record the entire body of the subject. The subjects appear in only one view in some situations, but in others they are visible back and forth, as seen in Fig. 2.10. The subject's face is plainly visible in the full-body photos, which are taken from various points of view. The artificial light is the hallway's true lighting. Simultaneous random frames are taken by three cameras with a

resolution of 3840×2160 for the acquisition. Additionally, each view's video is recorded at 10 frames per second with the same resolution as the still frames.

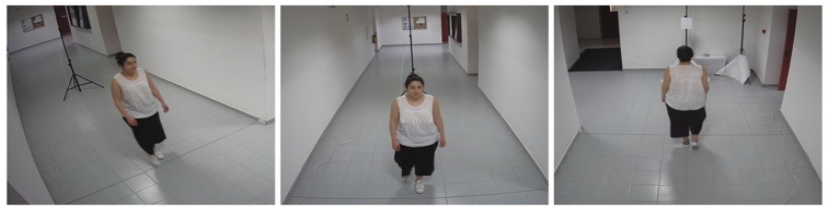


Figure 2.10: *Hallway* Protocol cameras frames from points of view.

2.3.5 *Drone*

This protocol can only be carried out in the open air. The experiment is carried out in natural light. The acquisitions are designed to mimic a hypothetical path made by free-roaming individuals. In this case, there was also a staircase on the route. Three cameras are set up in an attempt to duplicate the *Hallway* protocol's conditions: they are angled to point at the subject laterally. The cameras are 273 cm tall and arranged in a triangle arrangement to cover as much of the area where the subject would travel as feasible. A drone and a smartphone, in addition to the cameras, record the scene, the drone rising from the ground until it reaches the subject (Fig. 2.11).

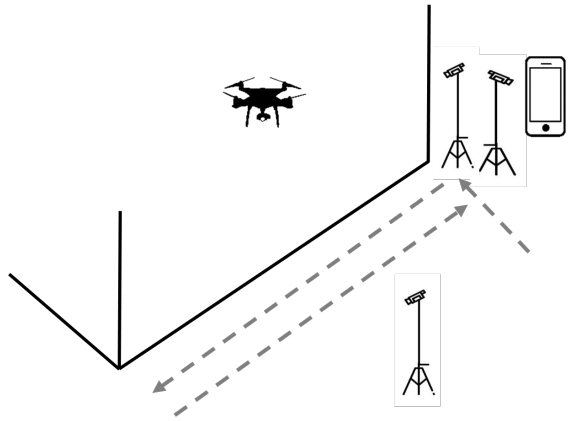


Figure 2.11: *Drone* protocol path way scheme.

Participants walk a predetermined path from a starting point to a destination location. They begin to walk down a set of steps, where the first camera captures them from the side. They turn left at the bottom of the stairs and walk straight to the location of the second camera. The second camera first takes a side view of the subject and then of the one from behind. When the journey is completed, the subjects begin to turn around. They turn right in the midst of the journey, and the drone takes off and begins recording. Another static camera photographs the subject laterally when the drone reaches the top height.

Subjects continue walking until they reach the end of the route and return back to the starting place. There are two types of acquisitions that can be made with this modality: single subject and multiple subjects. A single subject walks a predefined route in the first condition; numerous subjects are identified in the second condition to replicate a genuine situation in which more subjects wander autonomously on their own. In Fig. 2.12, you can see an example of some frames taken from the drone's video streaming. A few frames from a camera video stream recorded simultaneously with the drone are shown in Fig. 2.13.

2.3.6 *Data annotation and organization*

Each subject has voluntarily participated in the experiment and without any compensation. Before carrying out the acquisition experiments, each participant has been well informed on: protocols followed, research purposes, which would have been the acquired biometrics and from which device, who would have stored data, how data would have been made available. Participants have been informed that they could have been deleted from the database at any time if requested. Each participant has signed a consent form on the use of personal data. In the first experiment, each subject has been provided with an identification number that has not been registered anywhere. The assigned identification number has been given by the participant each time. The identification number has only been used to annotate the recordings, but no connection between the number and the signed documentation is possible unless expressly required by the subject itself. Each footage and photogram have been orga-

Figure 2.12: Some frames of the drone video recording.

Figure 2.13: The frames of the three TLC and of drone recorded simultaneously.

nized into folders named with the protocol number to which they belong. In these folders, there are others named with the

identification number of the subject to which they belong, those containing the recordings. The names of the files are made up of an initial part that contains the identification number. Each file can be downloaded from the BipLab Unisa website, with a prior application for authorization from the scientific director manager.

3

3D FACE BIOMETRICS BASED ON MOBILE DEVICES AND DRONES

New frontiers of computer vision techniques and the computational speed evolution of computing systems allow addressing the search for solutions through three-dimensional models also in the field of biometrics. 3D techniques can support the study of both soft biometrics, walking, body-pose, and hard biometrics, face, hands, and others. Classical face recognition issues such as intraclass variations like facial expressions, illumination, and pose can be simplified using three-dimensional information. Having a 3d model available allows you to check the acquisition of faces in non frontal poses more accurately. On the other hand, processing this amount of information is more computationally expansive. The algorithms that compute the data must be designed for parallel computation[1] to have acceptable response times, although this problem is gradually being simplified due to the cost of greater accessibility to the use of Graphics Processing Units (GPUs)[2]. The study presented in this chapter has focused on this particular subfield of investigation. An application example of 3D face model reconstruction by video recorded via smartphone versus drone, Section 3.2, and a research approach for the fast massive elaboration of these models, Section 3.3, are presented in this chapter.

3.1 3D FACE BACKGROUND

People's ability to recognize faces is well known. However, when recognition is iterated on a large database, fatigue occurs, the level of attention drops, and the speeds and accuracy perfor-

1 Parallel Computing means the simultaneous execution of one or more programs on multiple microprocessors or on multiple cores of the same processor for increasing the computational performance of the processing system.

2 GPU Is an electronic circuit designed to accelerate data elaboration during manipulation and alteration of graphic memory.

mances are compromised. Under these conditions, it is necessary to automate the facial biometric recognition process. Several methods have been developed for two-dimensional facial recognition as reported in the in-depth discussion Face Recognition Vendor Test (FRVT) [119]. However, several difficulties arise primarily from the significant variability and changes in facial expression, head pose, lighting conditions, and occlusions [131]. The 3D facial recognition techniques is an alternative solution to overcome the aforementioned problems above, for more details take a look in [4]. However, under facial expressions, the 3D shape of the face will be severely deformed. This distortion could affect various elastic areas, reducing the similarity between people's faces. Some attempts have been made to solve this problem. Different matching strategies can be adopted to neutralize the effect of expressions: *rigid matching*, *non-rigid matching*, *geometric form matching*, and *keypoints detection matching*.

Rigid: this category of strategies overcomes the problem considering only the rigid areas of the face (e.g., eyes, forehead, and nose). Comparison of the models is made only with the Iterative Closest Point (ICP) alignment obtained with the non-deformable points. In Ming [90] the authors, utilize Rigid-area Orthogonal Spectral Regression (ROSR) to describe and discriminate facial rigid areas as features

Non-rigid: applying morphing techniques, the facial characteristics are preserved and, at the same time, eliminate the expression information. [8] presents a strategy of non-rigid approach capable to morphed out expression deformations using PCA pattern of similar facial features.

Geometric: curves are considered the geometric features in the shape analysis. The basic idea is that a finite and indexed collection of radial curves approximate the facial surface. A radial curve of face, with the open mouth emanates from the nose, matching with a radial curve from the same face with closed mouth, as a combination of stretching and shrinking of similar points (upper lips, lower lips, etc.) [42].

Keypoints : consider salient facial landmarks of the face scan. These points are detected in 3D faces by the maximum and minimum curvatures estimated in the 3D Gaussian scale space, Lei et al. [79] detected s. Then the three quantities: Histogram

of the mesh Gradient (HoG), Histogram of the Shape index (HoS), and Histogram of the Gradient of the Shape index (HoGS) describe the local region around each prominent landmark.

Rich information from 3D measurements makes it possible to reconstruct 3D facial shapes. Therefore, they add accuracy to identification and recognition, but increase the computational workload [53]. 3D representation of the face is invariant to changes in lighting and pose. Nevertheless, 3D facial recognition has not gained popularity in real-world applications. One reason is that the scanners employed for 3D face acquisition in earlier studies are mostly bulky and expensive, and therefore not feasible in real-world circumstances. Recently, research has increasingly focused on in-depth facial images. This growth is due to advances in 3D sensing technology. As the sensor costs decline rapidly, it is possible to have very affordable depth cameras like Microsoft Kinect, ifm O3D303, Nerian Scarlet 3D Depth Camera, Intel RealSense, or even depth cameras in phones (e.g. iPhone X, Samsung Galaxy S20 Ultra, Huawei P30 Pro, Nokia 7.2 and Xiaomi Poco X2).

With the progress of 3D data acquisition technology, the number of open 3D databases is increasing day by day. Recent research on 3D face recognition has shown the use of depth or distance images for 3D face recognition tasks. The most widely applied to face depth images include the dimensionality reduction method, the local method, and the deep learning method. 3D face models have depth information absent in the typical 2D models discussed in the face recognition literature. 2D and 3D models can be used in combined approaches, such as the so-called multimodal algorithms [102]. An example method of face recognition fusing 2D and 3D models is shown in [82]. Acquisition of a 3D face model usually occurs in controlled systems designed for this specific purpose, mostly through the use of a 3D scanner. A good example is the system used to collect the FLORENCE 3D FACE dataset [11].

3.2 A NEW 3D FACE RECONSTRUCTION SYSTEM

This section presents an implementation of a biometric trait adaptive acquisition system that was part of the research pathway [2]. Furthermore, it provides an example of how to use the automatic

acquisition of 3D facial patterns that can be used for biometric recognition. The solution is accomplished using a drone and the results are compared to those achieved with the use of a smartphone. The system proposed can be exploited as an acquisition method for the collection of biometric face datasets with the added value of 3D information. It can also be used as part of a gate entry system to large areas with low population density, or for places with higher population density if fleet management is integrated. Another possible application used for this technology could be surveillance in places where environmental or human security is at risk and / or fixed cameras are vulnerable to vandalism or theft.

3.2.1 *Description of the objective*

The objective of this study is to use a 3D representation of the biometric feature to recreate the subject's face. This places our approach in the category of biometric collection systems, and more specifically, the class of subsystems that deal with the representation and manipulation of biometric traits. During a biometric acquisition protocol, the interaction between an individual and the system should be as noninvasive as possible. The greater the distance to the acquisition device, the lower the pressure felt by the subject undergoing an acquisition protocol. Wherever possible, contactless and unrestricted interfaces are preferred [40]. Minimizing the intervention of a system operator allows for the massive capture of people and it becomes crucial to automate this process. That is where the use of a drone becomes useful. The issue addressed is obtaining a stable and reliable 3D representation of a face from a monocular camera onboard the drone. The higher the reconstruction quality, the better the performance of a face recognizer based on it is projected to be.

3.2.1.1 *Drones in biometric system*

Unmanned Aerial Vehicles (UAVs), commonly known as *drones*, have recently emerged as a potential biometric purpose tool. UAVs are able to replace camera grid systems, allowing large areas to be monitored without the usage of a large number of

sensors [22]. Aerial vehicles, are provided with cameras and possibly other sensors, some of which are built into mobile devices such as smartphones.

Technological advances have changed the way in which UAVs interact with the environment and humans. In particular, the classical Pan, Tilt and Zoom (PTZ) mode, where only camera is moving parts, can be extended to Throttle, Pitch, Roll, Yaw, allowing more advanced mobile framing. The aircraft can move freely in the working environment with more degrees of freedom, which allows for a greater selection of shooting options and therefore the best possible detection of biometric targets.

The flight of these vehicles is usually managed by a pilot via an RC (remote control). In some cases, however, the pilot cannot react quickly enough in critical situations; hence, the need to automate processes such as object tracking, landing, take-off, return home, etc. Autonomous UAV flight is also desirable to reach certain destinations or avoid obstacles [29].

Drones can be used for a variety of purposes, such as surveillance, hobbies, rescue, photography, interactive social context, and more [121]. Therefore, day to day, they are becoming increasingly ubiquitous. Sometimes, the probability of success of a mission using drones can be compromised by the human factor. Because of this, appropriate interfaces exist to maximize automation and minimize the potential impact of human error, such interface as in [85]. In the present study, however, the drone interface is reduced to a minimum and the control is totally by the software. Traditional user interfaces for human-drone interaction are remote controls, phones, and simple gestures, which allow people to engage with a drone in a natural way [27].

In some circumstances, biometrics could be employed in an unconventional way as an interface for human-drone interaction. Aiming and launching flying robots with user-defined trajectories is one example. The authors of [20] present a user interface based on facial data. The approach is simple to understand and does not necessitate any user instrumentation. Beyond line-of-sight, the drone is deployed on a desired 3D trajectory. A study of integrated interface gestures and face pose estimation can be found in [94], which takes a different approach to the topic.

Combining the use of behavioral biometrics and drone flight is also possible. The trajectories of the drone connected to the emotion of the pilot are tracked in [28]. In [129] can be found as another human-robot interface tool that employs face information for control.

Drones have also been utilized to gather photographs overhead in several recent biometric datasets of facial images. Examples include the DroneFace dataset [62] and the multimodal dataset MUBIDUS-I [35]. Images were taken in both controlled and uncontrolled situations in the latter case.

3.2.1.2 *The Flight controller of the UAV*

The UAV when equipped with a ground-based controller, and a communication system between them is knowns as UAS (Unmanned Aircraft System), and drones are part of they. These aircraft are classified differently by defense or civil agencies, with categories constantly evolving. They can be classified according to their size, range, flight duration, purpose of use, and in some cases cost, as explained in [126]. Commercial models are so advanced in terms of performance and sensors that they can be compared to professional ones. The model chosen for this study is part of the first case. In origin, these vehicles were in use for missions considered "dull, dirty, or dangerous" for humans, often military [120]. Quickly, the use of commercial drones has expanded into many daily life human activities such as leisure and scientific research. Even because, they can be equipped with various sensors and modern technologies to overcome architectural barriers ensuring the safety of people. Many advantages have emerged for various critical situations in the field of security. UAVs can be flown by a human pilot under remote control (RPAS - Remote Piloted Aircraft System), autonomously using on-board computers, or as in our case, through a remote application. The flight of many of the new generation drones can be controlled via programming. This makes it possible to create intelligent flight systems like those with the DJI drone, one of the most widely used commercial systems.

3.2.2 *Method and Tools*

The drone it is programmed to adapt to the position of the subject's face in the frame and to fly autonomously to take pictures from a sufficient distance and a convenient perspective. The operator only needs to perform simple system startup and shutdown commands. The pilot-drone interface to the system should be simple since it only requires starting and stopping the application. Furthermore, the only requirement is the physical presence of a human subject at the scene. The contactless mode offers complete freedom of movement: test subjects can take control of the drone by attracting their attention. The drone then recognizes the user's face and begins hovering to take pictures. The captured images are processed to produce the 3D model. The drone model used in the experiments is a DJI Phantom 4 Pro+ [31], and is equipped with the DJI Mobile Software Development Kit (SDK)[3] [30] a proprietary development libraries.

Detection is a prerequisite for recognition. We used the Google Mobile Vision API [55], which is now part of the Machine Learning Kit, for detection. Because of its portability and scalability, the Mobile Vision API library was chosen over the standard OpenCV [18] library. The Google API and the DJI SDK can be utilized on any DJI remote flight control system without additional computing resources required. There would have been additional limits if OpenCV had been used for the detection stage.

The API's detection technique is based on a machine learning approach to object localization that ensures long-term support. To automate the flight of Phantom 4 Pro, it is used the DJI SDK, specifically the Mobile SDK, to construct a custom mobile app that fully exploits the DJI aerial platform's capabilities. With the Android Studio IDE, Android DJI SDK full compatibility [54], the Android app was built.

On the ground, the drone is initially in stand-by mode. The application instructs the drone to lift the camera and wait for face detection inside the scene in the first stage. When a face is spotted, the UAV takes flight and aligns the camera with it both horizontally and vertically before starting the preprogrammed procedure. At this point, the flight follows a geometric figure

3 Set of libraries that group functionality for multi-platform software developing

pattern around the topic, such as a rhomboid, while capturing a picture of the image, as in Fig. 3.2.

Other flying trajectories are possible, such as circles. What matters is that the subject is always framed from different perspectives. Only rhomboids were chosen for the sake of simplicity.

Finally, the 3DF Zephyr software form 3DF Flow [1] processes the video recorded by the drone and creates an accurate 3D reconstruction of the subject under acquisition. A repository containing the source code of the proposed solution is available online at `https://github.com/ldema/Remote3D`.

3.2.2.1 *Hardware*

The acquisition devices used are not particularly high-end. The drone model is a DJI Phantom 4 Pro+ with the following specifications: a flight autonomy time of 30 min, as far as 7 km control distance, 72 km/h max flight speed, up to 4K 60fps video resolution, up to 30 m sensor distance, obstacle sensing for 5 directions, 6 camera navigation system, main camera specification (FOV) 84° 8.8 mm/24 mm (35 mm format equivalent) f/2.8 - f/11 auto focus at 1 m $-\infty$. In addition, a smartphone Samsung S8 with 12 MP, f/1.7, 26mm (wide), 1/2.55", 1.4µm, dual pixel PDAF, OIS, back camera specification. The data for reconstruction and co-registration was processed on a computation system, an Asus laptop with Intel® Core™ i7 6700HQ Processor, Intel® HM170 Chipset, NVIDIA® GeForce® GTX 1070 with 8GB VRAM, 16GB DDR4 2133MHz SDRAM.

3.2.2.2 *Software*

Google Mobile Vision API: The Mobile Vision API framework allows the detection of contents present in photos or videos like specific objects, bar-codes, faces, and text. The tool is part of the ML Kit (Machine Learning for Mobile Developers Kit), a very powerful and easy-to-use package based on Google's expertise in the field of machine learning, offered to mobile developers. It includes very fast pre-trained object detectors that locate and describe the object in any orientation within images or video frames. In this application a face detector was used. It returns the region coordinates of interest containing the human face, and

it is able to extract facial landmarks positions referring to the contour of the eyes, ears, nose, cheeks, mouth, and visage. The very responsive response of this detector allows processing of the video stream from the drone in real-time .

DJI Mobile SDK software: The DJI Mobile SDK is designed to allow developers access via software to the capabilities of DJI's aircraft. It includes a set of libraries that group software calls for simplifying application development. High-level calls for low-level functionality such as flight stabilization, battery management, and signal transmission and communication, are provided from the SDK and can be imported into an Android or iOS app by its interfaces. The developers through the SDK can have easy access to many product features and capabilities such as can automate the flight, receive real-time video and sensor data, control the camera and gimbal, download saved media from the product, and monitor the state of onboard sensors. Many of these features have been used in this application work to reach the desired outcome.

Fig. 3.1 illustrates the way the DJI Mobile SDK interfaces with a mobile application and how to connect to a DJI aircraft. Furthermore, Fig. 3.2 shows an application testing by means of the flight simulator. A mobile application is developed using the DJI Mobile SDK and runs on a mobile device such as an Apple iPhone, iPad, Nexus phone, Nexus tablet, or other device that supports the applicable platform (iOS or Android).

3*DF ZEPHYR software:* It is possible to reconstruct 3D models of objects from multiple photos, by means of the 3DF Zephyr software. The software has no specific hardware requirements, however, it can take advantage of GPU computing capability if one is available. It is based on proprietary reconstruction technology, which is essentially photogrammetry. The generated model has good accuracy if the framed object is stable and the images are in focus. The software selects the most focused frames and discards those that are not sufficiently similar. Details on this step are given and best discussed in Section 3.2.4.2. Standard parameter settings were used for our experiments, with the focus set at a high level. During the render step, the software selects the best frames for reconstruction from the extracted frames.

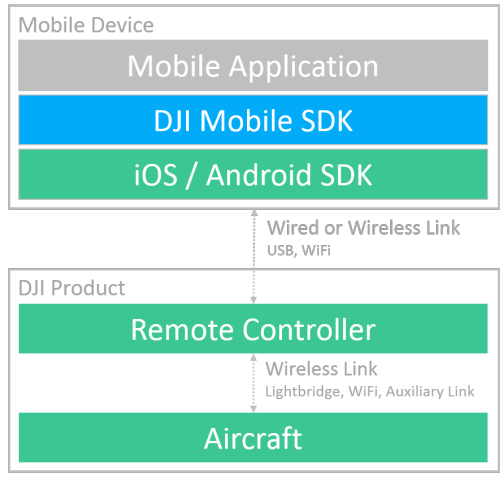


Figure 3.1: Connection diagram between the mobile device and the DJI system.

3DF Zephyr can create and export meshes and point clouds in the most popular 3D formats to generate video animations, digital elevation models, cross sections, and contour lines. You can derive the angle, area, and volume. The software application needs computationally demanding, even more by high resolution images, although multiple CPU cores and Nvidia Compute Unified Device Architecture (CUDA) technology, if available, can speed up the computation. 3DF Zephyr can import not only images, but also videos.

3.2.3 *Use of the system with safety*

Before use the SDK to automate the drone flight, some of following issues need to be addressed. Aircraft moves in spaces shared by people, structures, animals, plants, and possibly other drones. The aircraft can move at speeds of up to 20 m/s and can have a mass of up to 2.80 kilograms (kinetic energy). While the ability to change position by software programming is quite powerful, and safety systems are present on board, attention must be paid to the things and people around it.

However, developers and users should continue to monitor aircraft movements to avoid dangerous and unintentional collisions

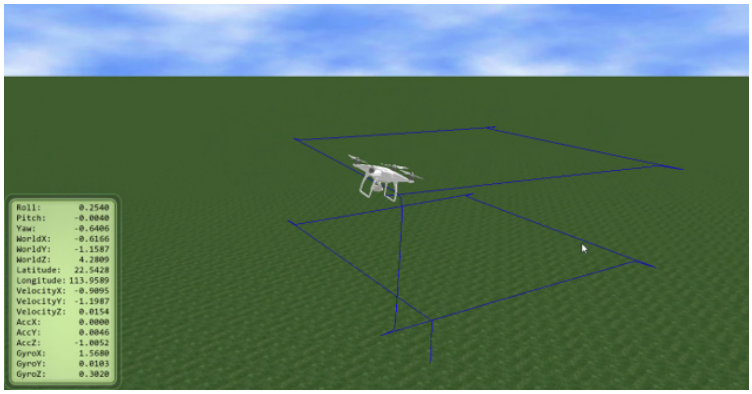


Figure 3.2: Flight simulator in action.

(share space). DJI offers a geofencing system to keep airplanes out of sensitive areas.

The UAV can cross challenging wireless areas during the flight, therefore wireless connectivity can be unpredictable. There, hundreds of milliseconds may be required to transfer a command, provided it happens, and unforeseeable situations may occur in the physical world (highly asynchronous process).

Moreover, to solve the collision problems it is possible to set the flight to a less sensitive function of the remote control, not abrupt movements, and stopping in the vicinity of obstacles

3.2.4 *Performance evaluation*

The system consists mainly of three software modules. The first module automates the subject recording process and records video images as soon as a human face is recognized in the video sequence. The drone flies autonomously and corrects its path to allow different perspective shots of the subject. The 3D reconstruction is implemented by second module. Discard useless frames: those with excessive blur, sub-optimal angles, and other flaws. Some examples of half-body models are shown in Fig. 3.5. The co-registration of the models pipeline are obtained in this way. The third module performs the core registration of the model obtained in the pipeline in the following way: the 3D model of an object reconstructed by the drone in flight is aligned together with the model of the same object captured under ideal

and cooperative conditions by the smartphone. As expected, the accuracy of the reconstructed model from the drone recordings is as high as the co-registration error is low.

The following sections 3.2.4.1 introduce the system user interface, the benchmark, and the software product used to manage point clouds. Experimental results of the section 3.2.4.3 describe the level of performance achieved and makes some observations derived from the behavior of the system.

3.2.4.1 *Human interaction and interface*

By collecting the raw data, the drone goes through a sequence of various phases and states in a certain order, similar to a finite automaton. A summary of the phases and phase transitions is shown in Fig. 3.3.

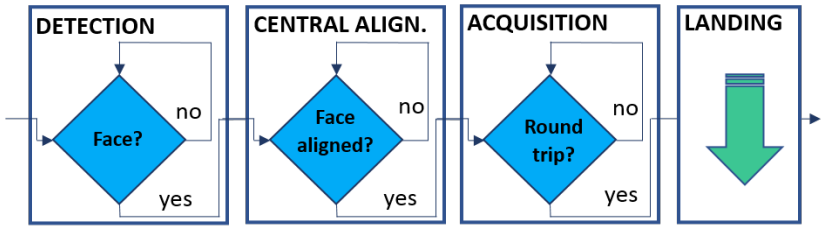


Figure 3.3: Flow chart of the drone phases in flight for data collection.

The first phase is named `DETECTION`. In the initial state named `landed` the Phantom 4 drone is at ground level. The user presses the start button on the RC to start the automation and a command is sent to the drone to tilt the gimbal upwards by 29.5°. Next, the application will then start streaming from the aircraft and the transmitted data will be processed by the facial recognition of the Google Mobile Vision API real-time. The UAV takes off and tilts the gimbal to 0°, which is the default stand by camera position, when a face is detected in the scene. This flight mode is called `take-off`. Fig. 3.4 shows an example of the real execution of this mode.

After taking off, the drone tries to align the face, both horizontally and vertically, to the center of the camera frame. This phase is called `CENTRAL ALIGNMENT` and completed the alignment, the

Figure 3.4: (Left) The drone stands by until a face is detected in the camera. (Right) The drone starting to take off when is detected a face.

aircraft starts the video recording, switching to the state `stalled flight` to perform phase `ACQUISITION`.

The drone flying around the subject actually acquires the face details, phase name `ACQUISITION`. The height should ideally be at eye level. The parameters Yaw and Pitch are modulated in rapid succession to achieve height stabilization. By pressing a designated button, the drone can be stopped from continuing its trajectory at any time for safety concerns.

Once the flight path has been completed, all data have been acquired, the aircraft goes back to the ground level in the `landed`, and the camera returns to the original 29.5° orientation, phase `LANDING`.

3.2.4.2 *3D Reconstruction phase*

A video recording of the subject has been captured after the automated drone flies and then the 3DF Zephyr program can begin 3D reconstruction. First of all, the video is loaded into the software and then is separated into frames, which are treated as separate single images. In the software, it is possible to set parameters for the frame rate, for extraction Frame Per Second. (FPS)[4], for the automatic analysis of the blur artifacts in each frame, and for the threshold to reject outlier frames that do not contain a component similar enough to the others. It is also possible to intervene manually and cut some parts of the video

4 The number of consecutive full-screen images that are displayed each second

that are not needed for extraction, if necessary - e.g., those not including the subject.

The next phase called `STRUCTURE FROM MOTION` processes all images that were originally loaded into the software, normalizes their orientation, and creates an initial representation of the photographed object, which is defined as a scattering point cloud[5].

An example of the results from the last two phases is shows in Fig. 3.5.

Figure 3.5: 3D reconstruction of half body left and center with FHD and C4K resolution from drone and right with FHD resolution from mobile device.

3.2.4.3 *Results and discussion*

Based on the system described above, 3D models of 20 participants were obtained in the distance range of between 2.5 and 3 meters at different resolutions. The group of people subjected to the experiments were men between the ages of 40 and 50, 1.7 to 1.8 meters tall, and of Caucasian descent with no occlusion in the facial area, except for one person who wore a mustache. The environment consists of an open space outside and without obstruction in the flight zone between the drone and the subject to be acquired, and behind the drone for at least 2 meters.

The recordings were made with minimal wind and in light conditions either sunny, but with no shadows in the recording area. Four different resolutions were considered: FHD, 2.7K, 4K, and C4K. 3D models contain a large amount of information. In this thesis, we focus on the alignment and 3D face reconstruction of the obtained models.

5 Set of three-dimensional vertices for 3D models.

Quantitative data from the experiments are collected in Table 3.1. The 'Processing time' value refers to the 640×480 of each cropped video frame, coincident with the bounding box enclosing the half body of the subjects.

The main contribution of these quantitative results refers to the influence of the video resolution on the calculation time and the extraction of point clouds, which in turn should affect the precision of the reconstructed model.

Table 3.1: Quantitative data from the experimental session.

Data information	**Drone**	**Drone**	**Drone**	**Drone**	**Mobile**
Video resolution	FHD	2.7K	4K	C4K	FHD
Video duration (sec)	[90,120]	[90,120]	[90,120]	[90,120]	[25,35]
Video-frames used	[80,130]	[80,120]	[75,110]	[80,110]	[50,75]
Processing time (min)	[90,120]	[90,120]	[120,180]	[120,180]	[50,75]
Point cloud size	[1k,2k]	[2k,3k]	[3k,4k]	[3k,3.5k]	[4.5k,6k]

From the table, it can be noted that for the drone data, the variations of the data processing time at different resolutions are fairly limited and that for videos in 4K, the point cloud size is higher than under other conditions. This suggests that the information captured at resolutions below 4K are not enough to achieve quality to produce accurate 3D models, and the C4K resolution may include noise. The experimental results presented in Table 3.2 and in Figs. 3.7, 3.8 endorse further this observation.

To assess the accuracy of the drone's 3D reconstruction, the 3D face mesh was aligned with the ideal model acquired via a mobile device under laboratory-controlled conditions. The ICP method was used to align the models [15] and Root-Mean-Square Error (RMSE) as a co-registration error metric to estimate how the 3D models fit together, according to the following equation:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\text{Drone}_i - \text{Mobile}_i)^2} \tag{3.1}$$

where N is the point cloud size in the 3D model saved by the co-registration process, while Drone_i and Mobile_i are the corresponding points of the two clouds found closest during the co-registration through the ICP algorithm. The mean RMSE achieved in the experimental session for all participants acquired

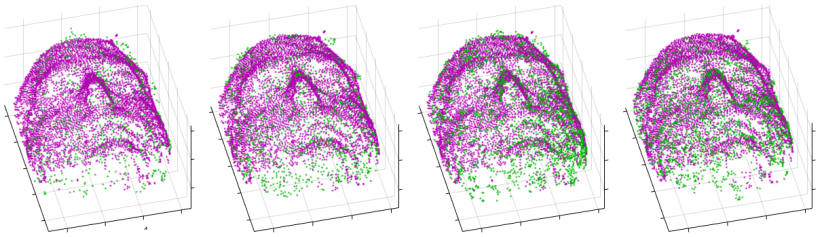
and reconstructed is shown in Table 3.2. The table also reports the mean variance of the co-registration of the point clouds. The most promising results are achieved with video resolution at 4K where the error is the lowest, as expected from the preliminary point cloud size considerations. Visually the Fig. 3.8c highlights the impact of this result.

Table 3.2: RMSE and variance of the 3D pint cloud alignment.

	Drone Video Resolution			
	FHD	**2.7K**	**4K**	**C4K**
RMSE	0.0356	0.0270	0.0264	0.0275
variance	0.0007	0.0005	0.0004	0.0005

Taking half-body 3D models of a subject, from a drone at different video resolutions and that from the mobile, the region of interest of the face has been extracted and compared. In Fig. 3.7a are shown the extraction visual results.

The overlap between the segmented point clouds, from the drone and mobile, can be observed in Fig. 3.6. Different colors show the drone model versus smartphone model, and how the two points clouds are co-registered for each comparison. When both point clouds contribute to the fully coloring plot, they are well co-registered.



(a) FHD (b) 2.7K (c) 4K (d) C4K

Figure 3.6: The mobile vs. drone point cloud registration at different resolutions.

Similar occurs in Fig. 3.7b, where the two textures are interleaved, the overlapping error is low. The overlap is greatest at 4K drone resolution, as it can be seen. The comparison results

in Table 3.2 show that the 4K resolution produces the lowest Root-Mean-Square Error (RMSE) and mean variance.

(a) (left to right) *MobileFHD*, *DroneFHD*, 2.7*K*, 4*K*, C4*K*.

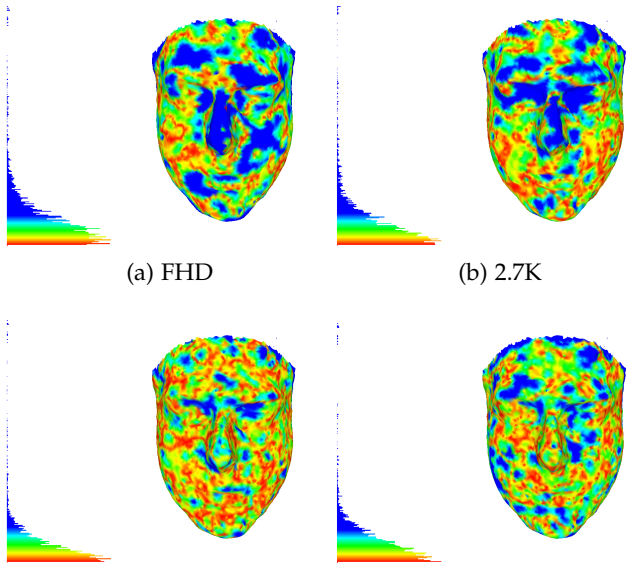
(b) (left to right) *MobileFHDvsDroneFHD*, 2.7*K*, 4*K*, C4*K*.

Figure 3.7: Crops of the 3D faces. On top portions extracted from the 3D models acquired by mobile device and by drone in controlled conditions at different resolutions. On bottom visual results of the overlap between the 3D model acquisition from mobile and by the drone. The overlap is not satisfactory when a model prevails on the other. The more accurate the overlapping, the more interleaved the textures.

Fig. 3.8 shows the distance maps obtained by the comparison of the ideal 3D model from the mobile acquisitions with those generated by images acquired from the drone during flight. The figure, in particular, shows how the map differs depending on the video resolution used in the comparison. The Hausdorff distance is used as the distance measure and the distributions from the nearest to farthest pixels are depicted on the left side of each map. As can be analyzed in the histograms, at 4K resolution, the best similar reconstruction is achieved.

When precise overlapping of the two models occurs, a high concentration of red and yellow points is accumulated at the bottom of the histogram (red/yellow color), while the blue one is the farthest. With FHD and 2.7K, you can see a wide area of non-ideal overlap. The resolutions of 4K and C4K are about the

same, but upon a thorough inspection, the lowest value cards that correspond to the 4K resolution are more compact and provide slightly better overlap than C4K.



(a) FHD (b) 2.7K

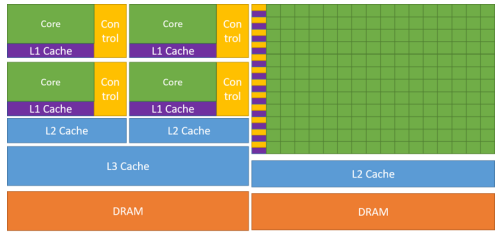
(c) 4K (d) C4K

Figure 3.8: Maps of Hausdorff distance at different video resolutions, Mobile vs Drone.

3.3 MASSIVE 3D FACE MATCHING

3D reconstruction of human faces can provide a lot of information to biometric recognition algorithms, which can help them to perform better. Making biometric data computation faster is a critical step in real-time scenarios. It is even more so when dealing with three-dimensional face models, where the amount of data and preprocessing complexity is significant. Classics programming methodologies or those based on neural networks are no longer computational sufficient for the purpose and therefore we move on to others. One is getting General-Purpose computing on Graphics Processing Units (GPGPU) programming into play. The GPGPU is used for processing that is extremely demanding in terms of processing power, and for which traditional CPU

architectures do not have sufficient processing capacity. By their nature, these processes are highly parallel and therefore able to greatly benefit from the typical architecture of GPUs (Fig. 3.9). CPUs used for fast sequential operations (10 thrs), GPUs for many parallel operations (1000 thrs). GPUs Compared to CPUs have a lot of memory bandwidth, very low latency, and higher machine instruction speed. Applications are typically developed by a mix of parallel parts (GPU) and sequential parts (CPU) to maximize the overall performance. Furthermore, GPU and CPU work independently and asynchronously on separate memory spaces. The available GPGPU programming technologies are: CUDA is Nvidia's proprietary technology for GPU computing; ROCm and HIP are similar technologies to CUDA, but they are open source and developed by AMD; Open Computing Language (OpenCL) is a library based on ANSI C and C ++ 14 languages.



(a) CPU (b) GPU

Figure 3.9: CPU and GPU basic architectures.

Preliminary experiments have been conducted in order to verify if what has been said has a basis. The experiments carried out foresaw a one-to-many comparison on large sets of 3D face images to perform their recognition in computationally competitive times. The study has been conducted with the aim of massively processing a face data set (1 million subjects) in search of a specific face in time equal to real-time, Fig. 3.10.

For the first experimental phase, the problem has been moved from 3D space to 2D space to valuate a 3D face identification method but with a computational complexity equal to 2D. In short, it was tried to consider a 3-channel 2D face image with R, G, B channels, as the representation of a 3D model. This is possible using normal maps [3]. In a 2D image representation of

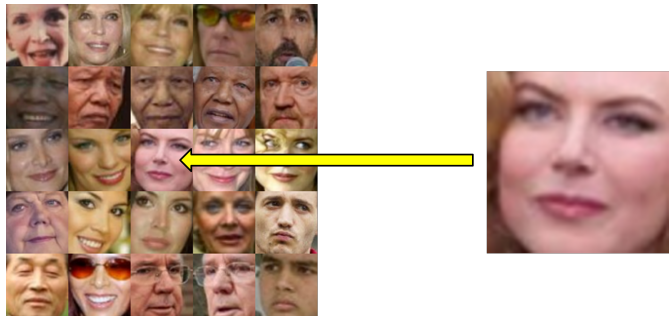


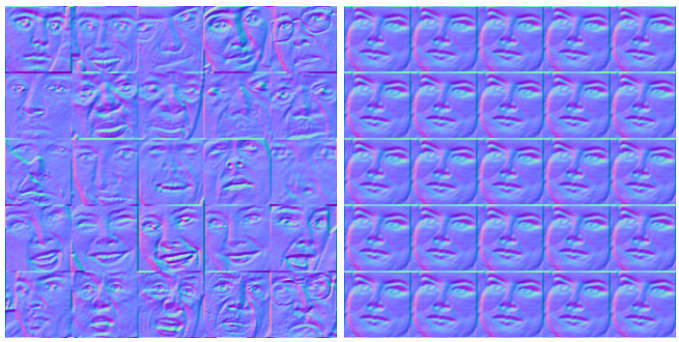
Figure 3.10: Objective: search a face identity in a huge dataset in real-time.

a normal map, the three channels' values of a point (pixel) are the intensities of the vectors x, y, z called normal of the 3D model at that point respectively of a view, Fig. 3.11.

Having a very large dataset of 3D face models, equal to 1 million, even better if different subjects is not easy. Then the 2D LFWcrop face dataset (64x64 face image dimension) of approximately 13,000 subjects have been used [113], and the normal maps for each image have been generated and duplicated for useful numbers of times. This is because the first experimental part is more interesting for verifying the speed of massive calculation rather than accuracy.

Figure 3.11: 2D Normal map image of the sample face.

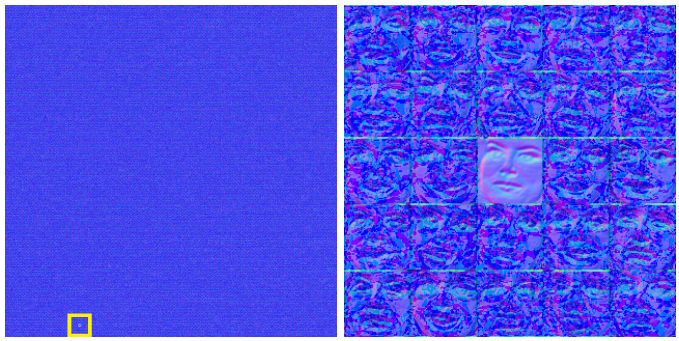
For the valuation, the basic idea is to prepare two matrices, called mega-matrix containing a predetermined number of normal maps of face images. One of the mega matrices includes the faces of the dataset and the other, the face to be sampled, duplicated as many times as there are faces in the first, Fig. 3.12).



(a) Dataset faces (b) Sample faces

Figure 3.12: The two mega-matrices normal map portions. (a) the Dataset mega-matrix, and (b) the Sample mega-matrix.

Once obtained, the two matrices have been transferred to the GPU memory and processed in bulk by means of an atomic operation. Atomic parallel operations are as fast as they can be as a basic computation within a GPU. The Boolean bitwise AND was computed between the two mega matrices as showed in Fig. 3.13. The figure shows the case of perfect matching when the face image is present in the dataset.



(a) mega-matrix (b) zoom

Figure 3.13: The mega-matrix bitwise computation result (a), with the relative region of interest zooming portion (b)

If the {Face Dataset AND (NOT Face Sample)} bitwise operation is made on the mega-matrices the better the matching of face images is than minor are different bits. Fig. 3.14 shows the result

of a perfect matching in the black region area. The difference computational time with the previous operation is imperceptible.

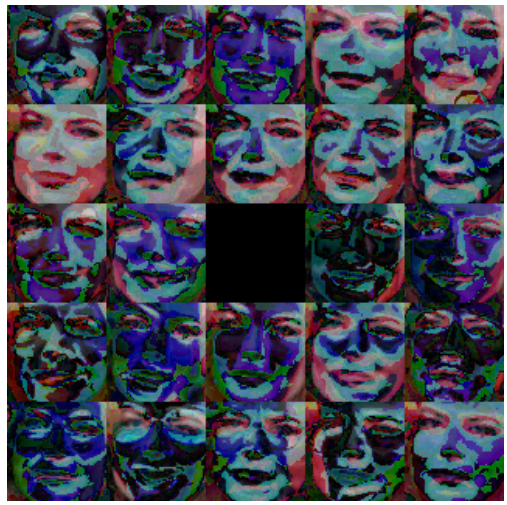


Figure 3.14: The Difference mega-matrices detail in Dataset AND (NOT Sample Face) bit-wise operation

The experiment was iterated to test different models of GPU video card performance using different dimensions of the mega-matrices about the number of face images. Table 3.3 shows the specifications of the tested GPU model. For any iteration, initially a 10K step of face images from the dataset has been used between 0 and 100K, and after a 50K step of face images from 100K to 1M. The size of a face image, 64x64 pixel in RGB format, is equal to 12.288 bytes. The following Table 3.4 shows the size in bytes of any mega-matrices.

Table 3.3: GPU specifications models used in the experiments.

GPU model	**VRAM**	**TxB**	**GPU Clk**	**Mem Clk**
Geforce GTX 970M	3Gb	1024	1038 MHz	5 Gbps
Quadro P4000	8Gb	1024	1227 MHz	6 Gbps
GeForce RTX 2080 Ti	12Gb	1024	1750 MHz	14 Gbps

To bitwise compute the two maga-matrices, first they would be transferred to the guest memory. These take time comparable

Table 3.4: Byte size of mega-matrices.

Faces	Size
10K	118 Mb
20K	235 Mb
30K	352 Mb
40K	469 Mb
50K	586 Mb
60K	704 Mb
70K	820 Mb
80K	936 Mb
90K	1 Gb

Faces	Size
100K	1,2 Gb
150K	1,7 Gb
200K	2,3 Gb
250K	2,9 Gb
300K	3,4 Gb
350K	4 Gb
400K	4,6 Gb
450K	5,2 Gb
500K	5,7 Gb

Faces	Size
550K	6,3 Gb
600K	6,8 Gb
650K	7,4 Gb
700K	8 Gb
750K	8,6 Gb
800K	9,2 Gb
850K	9,7 Gb
900K	10,3 Gb
950K	10,9 Gb
1 M	11,5 Gb

to the computational time. Fig. 3.15 show the GPU computation times. As you can see, the graph data is truncated due to the memory capacity of the GPUs model and the grow of time is directly proportional to the amount of data processed. Up to 100K, the growth time is 1, double data needs double time over the trend factor decrease.

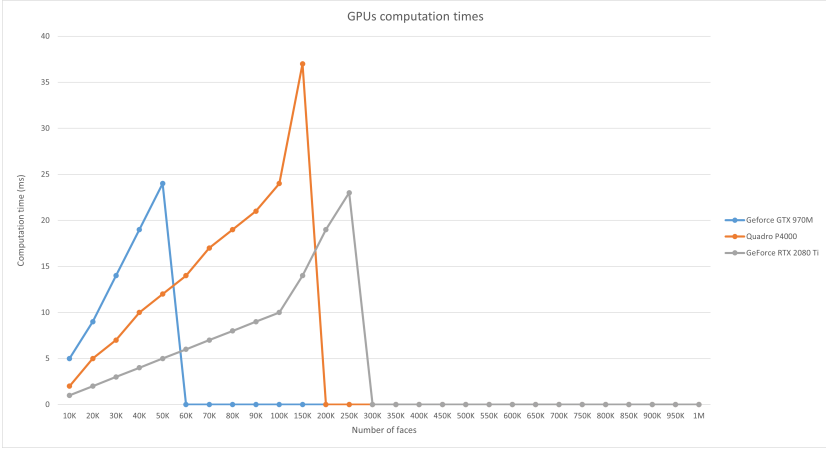


Figure 3.15: The benchmark of GPU computation

Memory allocation on the guest (GPU) takes more consuming time of computational time and therefore it cannot be overlooked.

The GPU memory allocation times related to the GPUs models and to the amount of memory are shown in in Fig. 3.16. The growth factor is less than 0.5. The allocation of double memory does not imply a double allocation time. This is even more true as the memory performance of the video card increases.

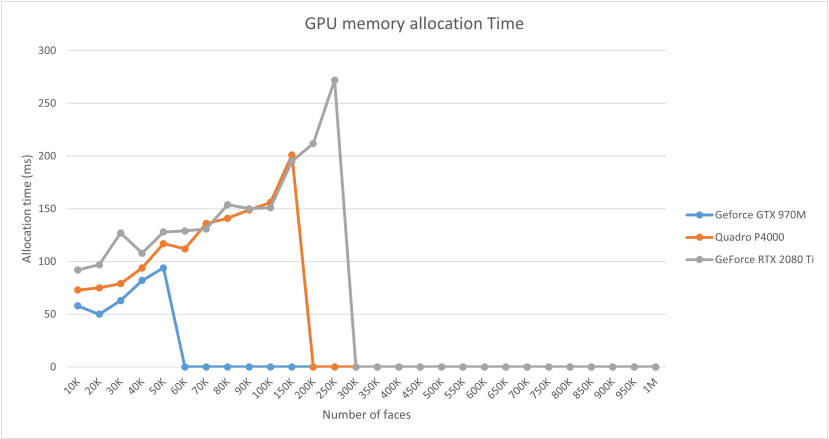


Figure 3.16: The benchmark of memory GPU allocation

Once obtained distance/difference value for each face image which a label or index has been assigned, these values have been to be sorted in order to identify the face images similar to the one it has been looking for. In many real-time scenarios, and this is one of them, efficient sorting algorithms are a key requirement. Parallel computing solutions on CPU and GPU are implemented to grow up the performance. In [57] the author has investigated some implementations of the fastest GPU and CPU sorting algorithms. An exhaustive survey of GPU based sorting algorithms can be found in [116]. The sorting time is very fast on GPUs, therefore is negligible related to the whole operation, as shown in Fig. 3.17. In addition to the fact that GPU times are substantially lower than CPU times, the computing time for CPUs climbs linearly as the number of elements increases, whereas the GPUs growth trend is more attenuated. If the number of elements to be sorted is lower to 100K, CPU and GPU have the same computation time, under 50K elements the CPU is better, as shown in the chart of Fig. 3.17.

It can infer that the tested method for massive 3D recognition of the face is a viable potential solution in real-time scenarios,

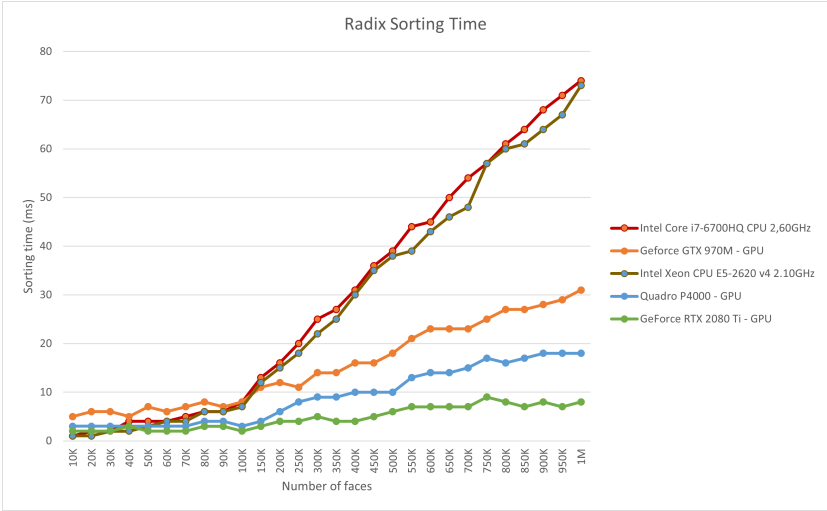


Figure 3.17: The benchmark of parallel sorting CPUs vs GPUs

thanks also to the possibility of scaling up to meet increased workloads. It is critical to choose characteristics to employ for the generation of the templates in this type of approach, since they must allow comparison with bitwise operations. In a future experimental phase, it is planned to identify the most suitable preprocessing function to uniform the model to a frontal pose as suggested in the work [4] and select a feature extraction function more suitable for the massive bitwise comparison. The experiments will focus on the accuracy of the identification system considering that it can be usefully compensated by the speed of performance.

4

TRENDS, LACKS AND CONTROVERSIES

The biometrics market is growing rapidly. Let us summarize a study by a marketing agency, Mordor Intelligence: The next generation biometrics market posted a Compound annual growth rate (CAGR) of 35.53% during the forecast period (2021-2026), Fig. 4.1. A paradigm shift toward more data protection and less security threats in the business of the next-generation biometrics market is one of the most important trends. End users prefer to rely on integrated solutions rather than traditional methods.

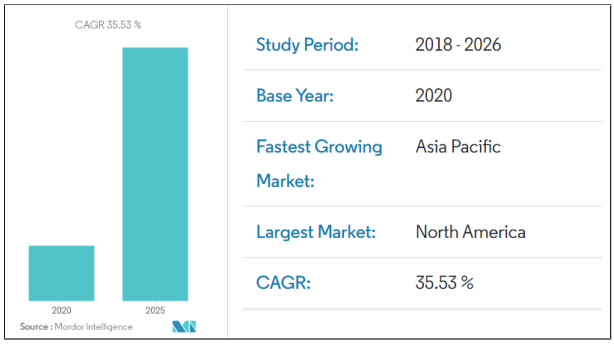


Figure 4.1: Market Summary, CAGR of 35.53% during the forecast period (2021 -2026).

The next-generation biometrics market is expected to grow significantly due to an increase in the number of terrorist activities and an increase in theft of important data and information that raises national security concerns. Key factors such as the growth of electronic passport programs, government support, and their widespread use in criminal identification are the major drivers behind the market. With the rise of airport security initiatives and attempts to reduce crime rates, investment in biometric systems around the world is increasing.

Various government initiatives such as electronic passports, electronic driver's licenses, border controls, national ID cards, etc. are being implemented in developed countries with advanced

biometrics, Fig. 4.2. Iris recognition is the fastest of this type of solution. It is one of the growing segments. Some of the advantages of this technology are that it is easy to use, difficult to counterfeit, and accurate. Applications of iris recognition in the consumer electronics field are expected to show the highest growth rates during the forecast period, primarily due to the commercialization of various electronic devices based on iris scanning, such as smartphones, tablets, smartwatches, and notebooks. However, factors such as high deployment costs and the threat of privacy breaches are expected to impede market growth.

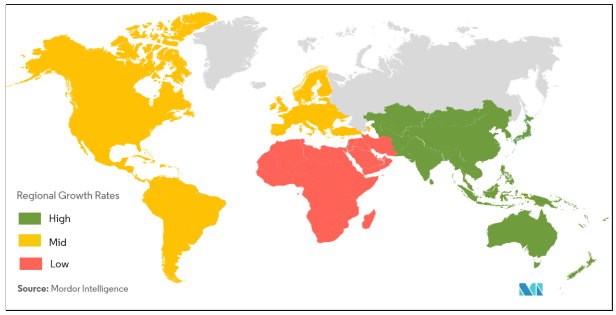


Figure 4.2: Next Generation Biometrics Market - Grow Rate by Region (2019-2024)

4.1 BIOMETRIC LIFE, IMPLICATIONS

Today, we try to measure everything about a person, from how he looks at a screen to how the mouse moves, to alteration of pulsations, microexpressions, or micro movements of the eyes [25]. Many new generation sensors allow for all this and can be installed on devices such as smartphones or smartwatches (wearable devices). Health monitoring, why not profile individuals? This question shifts the discourse on biometrics to a fine line of privacy. The regulations on biometric measurements are specific to each country and generally follow a political direction. Some privacy guarantors believe that one biometrics trait rather than another is more invasive from the point of view of the content (see DNA and Retina). Capturing a photo without consent is a different violation than taking measurements of it, which must be authorized by the guarantor. By measuring the facial features

of a photo, sensitive information is extracted that simple storage does not. To meet these requirements and to protect sensitive data extracted from the measurements, various methodologies have been developed. Sensible data protection techniques, such as encryption, are useful, as discussed in Section 4.3. In addition, there are laws that regulate the practices of acquiring, sharing, and storing such data. The independent European supervisory authority is the European Data Protection Supervisor (EDPS) by means of General Data Protection Regulation (GDPR) laws.

Another aspect, often overlooked in favor of the insistent request for protection of the individual, is the moral and philosophical question concerning biometrics, as argued in [92]. Epstein, claims that Biometrics can be seen as a tool for the development of state authority [44]. To extremes, by transforming the human subject into a series of biometric characteristics, biometrics dehumanizes the person, would violate physical integrity, and, in the end, offend human dignity. Agamben foresees the reduction to naked bodies for the whole of humanity, and Biometrics will usher in this new world [7].

As mentioned above, it is difficult to design an objective biometric technology if the system is subjective and error-prone. The proliferation of biometric technology in the public and private sectors raises these concerns. The increasing commercialization of biometric data by the private sector increases the risk of losing human value. Businesses value biometrics more than individuals.

Thus, modern society should mature a "biometric consciousness" that stimulates an informed public debate about these technologies and their applications, and accountability on the part of the state and the private sector, ownership, and access to your physical data and other intellectual property generated by your physical data should be understood as rights.

Meanwhile, other researchers have shown that the globalized world is home to many people with weak or no citizen identities. Most developing countries have weak and untrustworthy documentation, while for the poorest countries, it is nonexistent at all. Without the identity of a certified individual, there is no legal security or civil freedom. Any person can claim their rights, including the right to refuse to verify their identity, only if they can identify themselves. In this sense, biometrics can play

a fundamental role in supporting and promoting human dignity and respect for fundamental rights as a science that enables an individual's unique identification.

4.2 CANCELLABLE BIOMETRICS.

Biometric recognition solutions suffer from privacy and security concerns. Allowing the acquisition of your biometrics means exposing you to improper use. So, to protect your identity, we have to protect your biometrics, a paradox that exists in all protection systems, but that is the way it is. An example is very trivial; it is how to use an alarm system to protect yourself from intruders, and the alarm system must provide protection against tampering. Furthermore, an essential aspect of the question is, if a biometrics is stolen and used for illicit purposes, that biometrics no longer guarantees the identity if the systems that use it are not able to repel spoofing attacks by distinguishing the original one from the one stolen. The best strategy to preserve biometrics is to not expose them in their full, or to expose them in a changed form, with the goal of keeping their core features for identification and verification. An alternative option is to convert or distort sensitive data in such a way that recovering the original data is challenging. Another key feature is that if the previous instance of biometrics is compromised, a new instance must always be possible.

The Cancelable Biometrics (CB) are one of the solutions to address these concerns. Introduced first by Soutar et al. [117] in 1998 and then defined by Patel et al. [99] in 2015 as: *"Cancelable Biometrics consist of intentional, repeatable distortions of Biometric signals based on transforms which provide a comparison of Biometric templates in the transformed domain"*. The CB must meet four important requirements that are: *Diversity*, *Reusability*or *Revocability*, Non-invertibility, and *Performance*.

- *Diversity*: The new CB of the same biometrics must be more different than previous.
- *Reusability/Revocability*: The CB must be regenerated if compromised.
- *Non-invertibility*: The original biometrics cannot be reconstructed from the CB.

- *Performance*: The generated CB does not degenerate the recognition performance.

To date, various techniques for generating the template have been pioneered, such as non-invertible geometric transformation, hashing, random projections and permutations, filtering, fuzzy vault (Sub-Section 4.2.1), cryptography (Sub-Section 4.2.2), etc. argued in [87, 98, 99].

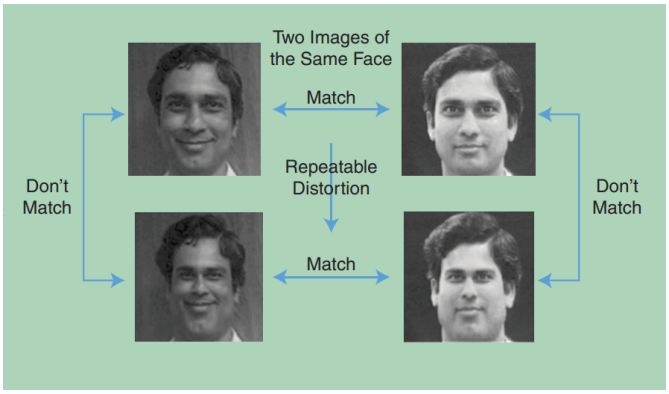


Figure 4.3: The generation basic scheme with non-linear geometric transformation.

An example of cancelable biometrics for face recognition is illustrated in Fig. 4.3. Prior to feature extraction, the face image is morphed into the original pixel signal domain. The morphed version does not match the original face, whereas the two instances of the morphed face match [98].

4.2.1 *The Fuzzy vault technique*

The first technique proposed for the generation of erasable biometric keys was more cumbersome, but for a long time it was the only technique available. Again, the idea was to hide the information so that it could only be reconstructed by those authorized. Many times, these techniques depend on the type of biometrics for which they were designed; for example, biohashing was designed for fingerprints and iris, but it works for any biometric trait capable of giving a numerical vector as characteristics. Instead, the fuzzy vault technique [70], was designed primarily for

fingerprints, but it should be borne in mind that this technique also works for other characteristics derived from biometrics other than the fingerprint. For fingerprints, in biometric systems, the representative points of the trait, called minutiae, are considered. Examples of minutiae are, for example, the points where the lines join or fork or others where the noose is formed. The basic idea is to add false points to the template so that, in comparison, only those who coded the erasable template know how it was built. When the subject needs to be recognized, his details are acquired, and the intersection is made between them and the points previously saved during registration. If the subject is the same: his minutiae on his finger will catch most of the real minutiae in the container, discarding the fakes. If the subject is an impostor: his minutiae do not match those of the container, then he will take real minutiae, but most of the minutiae he will take will be false ones, so the reconstruction fails [128].

One of the limitations of this technique is that if more sets were stolen, making the intersection between all stolen sets would reveal the true points (always present points). Another question is whether it is possible to use a biometric key to encrypt data, for which a contribution to the state of the art is made with the solution presented in Subsection 4.3.

4.2.2 *Cryptography basic concepts*

Cryptology is a science that studies speech and secret writing and is divided into two major branches: *Cryptography*: the science of writing messages that no one beyond the true recipient will be able to read, from the Greek kryptós = hidden and from the Greek theme, gráphò that is, to write. *Cryptoanalysis*: the science that deals with reading encrypted information through the breaking of encrypted systems, therefore, studies the vulnerabilities and improvements that can be applied to the former.

Cryptography provides a suitable tool to keep secret all information that is not intended to be disclosed publicly, so that the possibility of accessing it is given only to authorized persons.
Two basic operations can be performed: *Encryption*: this is the operation by which the information is hidden and is carried out using a special algorithm called a cipher; the information

to be encrypted is known as clear text. Encryption uses a key as a fundamental means to convert clear text into ciphertext or cryptograms.
Decryption: it is the reverse operation with respect to encryption, that is, the conversion from cipher text to clear text; it too uses the cipher key.

Encryption algorithms can encrypt data using biometrics. We must also keep an eye on cryptanalysis because from the moment a new scheme is pulled out, we must prove that it is safe, that is, we must analyze if and what are the vulnerabilities of the code and how they can be solved. An encryption algorithm is a technique whose purpose is to hide information by encrypting the data (placing it in a form that is not readable by the person in possession of it), but the authorized party must have a way to decrypt the data.

In order to encrypt and decrypt data, the schemes used today are based on a key, that is, they take a set of data and a key that governs the encryption process and apply transformations that depend on the key to change the data and make it unreadable. In the decryption process, using the same key, you are able to reverse this process. Techniques that use the same key to encrypt and decrypt are called *symmetric or secret key*. If instead of using one key, two are used, a public one available to everyone and a private one (known only to the user), we are talking about *asymmetric keys*. This scheme works in such a way that if the user encrypts the data with their private key, someone can decrypt them using the public key and vice versa. This versatility then makes it possible to use these schemes not only to encrypt and decrypt data, but also to implement signature algorithms, for example, if I want to certify a signature of a document, if we want to guarantee who sent a message, therefore, you can use these protocols to do more operations than symmetric ones that only do encryption and decryption. Since the asymmetric ones are particularly heavy and slow, in general, what you do is encrypt a set of data with a symmetrical pattern and then encrypt the key of the symmetrical pattern with an asymmetrical pattern (security problem).

4.2.3 *Bio-cryptosystems*

Systems that manage to extract a cryptographic key[1] from biometrics are called *bio-cryptosystems*. These systems perform a binding operation between the information coming from a biometric key[2] and what should be the cryptographic key I want to use to be able to encrypt. There are two possibilities: either I make the biometric key stable (I make the system always extract the same sequence of bits for the same subject implies stabilizing the biometric key) or I can come up with ad hoc schemes for encryption and decryption that are robust with respect to small and possible errors of the key (the fuzzy vault is an example implies robust for the minutiae). There are systems (key-generation) that directly generate a key, taking out from the biometric key a smaller set of additional data that allows it to be stabilized. These additional data are called help-data[3]. There are key-binding systems: Find the biometric key and the cryptographic key from the supplier, generating data that represent how the two keys are joined [67].

Only recently has the question of the variability of biometric data of a single individual has been studied for the generation of biometric keys in the context of the cryptographic system [122]. The problem is more complex, for the same biometric entity the extracted biometric data is significantly different due to acquisition characteristics, at different times. Fig. 4.4a shows a simplistic biometrics-based key release method [122], where a correct biometric template match releases a cryptographic key. This method is vulnerable to attacks on the biometric template database, the cryptographic key database, and the biometric matcher. The second method (Fig. 4.4b), has the advantage of the so-called biometrics-based key generation methods [122], the system is not vulnerable to template information database attacks because secrets and biometric templates are securely stored in cryptographic biometric templates. An example of the combination of

1 Is a string of data that is used to lock or unlock cryptographic functions,
2 Numerical expression of the salient measurements of a biometric trait. Usually organized in a vector or matrix, they are used in matching operations.
3 Data in support of biometric data whose purpose is to make them stabilize for the purposes of the encryption

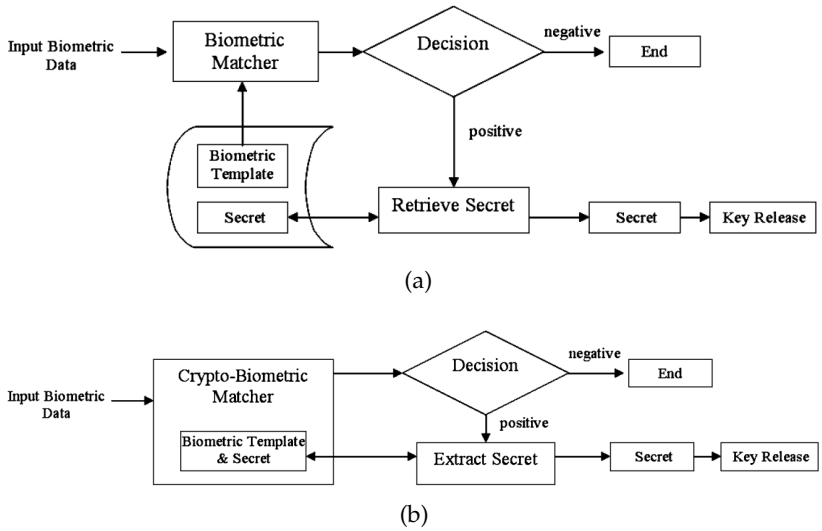


Figure 4.4: Two modes of combining biometrics with cryptography: (a) key release and (b) key generation [123]

biometric key and cryptography is presented in the following Section 4.3.

4.3 A NEW BIO-CRYPTOSYSTEM, FACE BIOMETRIC & RSA ENCRYPTION

An original example of biometric data protection techniques with key-binding is one of the works published *"An Encryption Approach Using Information Fusion Techniques Involving Prime Numbers and Face Biometrics"* [64]. Two methodologies of two unconnected research areas have been joined, Biometrics and Cryptography. Specifically, the data on face biometric traits were combined with Public-key Cryptography, with the purpose of generating a new type of Rivest–Shamir–Adleman (RSA)[4] key [109], precisely a hybrid key.

The method presented is based on a general requirement: it can be applied to any biometric trait considered (iris, face biometrics, etc.) from which you can extract a Biometric Code. The new technique devised to merge the biometric code and the public

4 RSA is a public-key cryptosystem used for secure data transmission.

Figure 4.5: The hybrid key generation basic scheme.

key is named FIF (Face Information Fusion). Fig. 4.5 shows the process diagram of the basic idea for generating the hybrid key. The main phases of the system that carries out the data fusion are listed below. The generation process goes through three main stages which accomplish the data fusion:

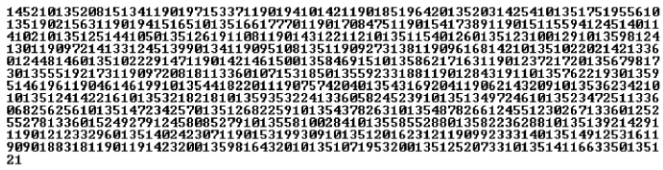
- *Face Algorithm*: it is the stage for generating the biometric code. Two different approaches are examined to extract the face features. The first use SIFT (Self Invariant Feature Transform) as described in [51]. SIFT features have many interesting properties, one of which is highly discriminating, and the other is the ability to automatically extract stable interest points in the image. SFA (Split Face Architecture) was used in the second approach, as stated in [38]. The authors' goal in this last article was to address face recognition following plastic surgery.

- *RSA*: it is the stage for creating the number code and the private key; the basic requirement here is that the key must be as much random as possible to avoid an external intruder. This requirement refers to the process of combining the elements of two arrays into a single matrix that will serve as the key. We choose to employ the primality and secrecy provided by the private key of the RSA algorithm 4 [109]. The value of the same key is then used to determine how many elements of the biometric component array and of n are in the union matrix.

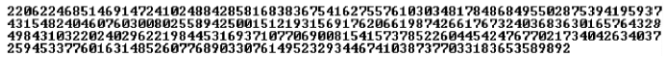
- *FIF Algorithm*: it is the data fusion stage. The FIF technique attempts to generate a Fusion Key using Face Biometrics and Numerical Data. The transformation of two vectors into one matrix is a vital step in this technique. In the current implementation, it is critical that this matrix is square and

that its order is determined by the number of components of the two vectors.

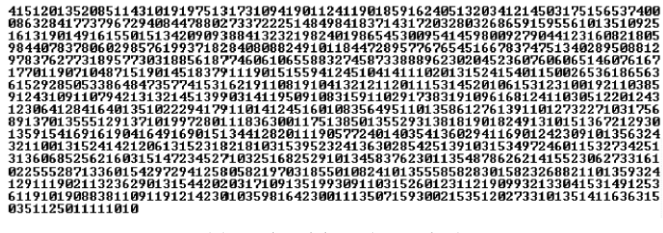
In short, for the FIF algorithm, the RSA private key indicates the order (indices) in which the blocks in the final matrix (hybrid key) are built. The algorithm is reversible only if you know the biometric component and the private key. Fig. 4.6 shows an example of the keys involved.



(a) Biometric Code



(b) Private Key (1024 bit)



(c) Hybrid key (2048 bit)

Figure 4.6: Example of involved key code.

In an efficient cyber-protection solution, the first requirement is the randomness of the key. The randomness of a sequence is determinate by the P-value, with a range value between 0 and 1. A number sequence is accepted as random if P-value is greater than 0,01. In the tests, we used a private key of RSA with 1024 as the bits dimension, and the Hybrid Face Codes of dimension 2048 bits, calculated from 100 faces. The P-Value indicators resulting from National Institute of Standards and Technology (NIST) [5] [111] are shown in Table 4.1, with respect to the sequence generated by the FIF system of 100 faces.

5 NIST Statistic tests that verify the randomness of a binary sequence of a fixed dimension.

Table 4.1: P-value Test NIST.

Test NIST	**P-value**
Frequency	0,489429
Block Frequency	0,821653
Cumulative Sums1	0,484992
Cumulative Sums2	0,373882
Runs	0,182665
Longest Run	0,326421

The FIF, Hybrid Information Fusion algorithm, has a general philosophy that is completely independent of the biometric component we've looked at and is completely reversible, but only if the biometric component and the private key generated by the public-key cryptography algorithm are both available. An example of a possible application of this encryption technique is to ensure the security of blockchain and electronic currency transactions, such as access to very private areas, classified and confidential documents, privileges to activate critical, military, or defensive infrastructures, and so on.

The results of NIST statistic testing show that the codes generated by the FIF algorithm are truly random. As a result, these findings highlight the potential prospect of applying them to high-security or high-fraud-risk operations, such as online or data-transfer transactions.

The problem of support or token can be solved using a hash technique (SHA-2, 512 bits, [1]) and a conversion of the digest code to Base 64 (to decrease memory waste and obtain more compact information), according to the literature. This code would be associated to the Hybrid Face Code, allowing you to use the second code in the event that the first is lost.

Part III

FINAL DISCUSSIONS

5

3D FACE BIOMETRICS IN VIRTUAL ENVIRONMENTS

The use of 3D in the biometric field is not a recent issue, but continues to be studied and new perspectives are coming. Having a 3D model available of a face make it possible to achieve higher levels of accuracy than 2D models for biometric functionality. This is due to the high level of detail of the biometric properties and the potential to evaluate them from different perspective when transferred to a 2D reference system. Obtaining large numbers of 3D models of biometrics is not easy because either the acquisition operations are very long and expensive, or specific devices are required, but in any case, finding many subjects is always a complicated thing. One solution is to think virtual, both for the generation of models and for their field experimentation.

5.1 FROM REAL TO VIRTUAL ROUND-TRIP

In the field of biometrics, moving from the real world to the virtual is not always possible, but it is worth trying. Often it is a round trip because you switch to the virtual world and then bring the results to the real world. Therefore, even if the separation between the two worlds must be advantageous, there must never be a cleavage between them to obtain applicable solutions in reality. The steps are not painless from a computational point of view, but this is a problem of future architectures and partly solved. Virtual data and semi-synthetic data can be merged to generate a large number of models, with features near to real, useful for the training needs of next generation neural networks. Many publications in the multimodal fusion literature present results based on approximately 100 genuine participants, with little consideration of the reality that such results could be highly biased. The authors of [41] first address this issue and present a novel technique for evaluating multibiometric systems on standard size databases of real subjects. Furthermore, it is logical to consider the idea of em-

ploying virtual subject databases, which are individuals created by merging distinct biometric traits (modalities) from different people or obtained like from the process presented in chapter 3. If this approach is valid, it will make creating multimodal data easier because it will just be necessary to integrate two or more databases with roughly the same number of individuals, each containing each distinct modality.

Synthetic datasets in combination with virtual environments are tools for this purpose. Some examples of synthetic biometric traits are covered in the following sections.

5.1.1 *Synthetic data*

Synthetic data is information that has been created artificially using algorithms based on features of the real world. The constraints that these generations must follow are that the set obtained must keep the fundamental properties of the real world and must not alter the problem unless that is the goal. As stated below, there are various advantages to creating this type of dataset, particularly in the biometric, financial services, and healthcare industry. They increasingly replace real data in tests to validate mathematical models and train machine learning models.

The advantages of their use include the reproducibility of the test in a precise way, the lack of need to comply with regulations, the achievement of the quality of the real test in subsequent steps massive increase in the number of tests and the variety of simulated conditions, the introduction of rare, unexpected, and unreal conditions.

The disadvantages include inconsistencies when trying to replicate complex real systems or conditions if the synthetic model is not accurate, in-depth knowledge of the nature of the original data, and the risk of having models that do not produce the expected results due to the unknown variability of the data in the real world.

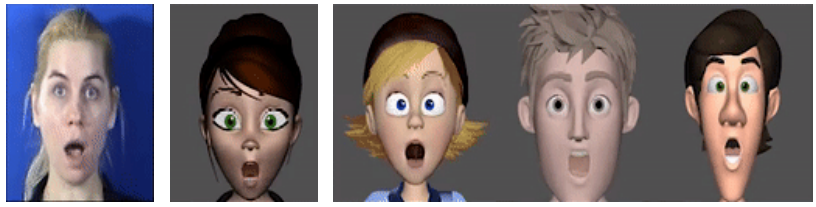
5.1.1.1 *Big Extension*

The first and most important benefit is the ability to gather a large amount of data, referring to Data Augmentation (DA) in technical terms, which is particularly beneficial for training Artificial Neural Networks (ANN), which requires sampling of many tens of thousands, if not hundreds of thousands, of samples. DA is frequently obtained by altering the properties of a source data set. In the case of image classification, alterations such as distortion, rotation, noise insertion, flipping, blurring, color changes, and illumination changes are used. In a 3D face model, something more advanced could be generated. In [63] the authors have demonstrated the performers Deep Neural Networks (DNN) techniques applied to 3D data augmentation. They synthesize images of 3D face shape estimates with new conditions of viewpoint and lighting conditions. This augmented model can then be applied for face recognition, face classification, and face landmarks. Major properties are preserved diversity to cover the variability in real-world scenarios, and fidelity to not insert unrealistic artifacts. Visual results are shown in Fig.5.1.

Figure 5.1: Examples of morphed emotional faces.

5.1.1.2 *Privacy guarantee*

If synthetic data are not attributable to an identity, they do not require authorization to be used or stored because they respect the privacy of personal data. A very useful example is the generation of a synthetic dataset of facial expressions suitable for testing emotion identification, the dataset FERG in Fig. 5.2. This dataset is built starting from a group of people where each subject makes expressions Fig. 5.2a. For each subject and for each expression, a character series is generated with different aspects, but which retain the expression of the subject [10], Figures 5.2b and 5.2c. This method is very useful because also in this field it is not easy to find many subjects. In this case, the identity and



(a) Source (b) Primary Character (c) Secondary Character

Figure 5.2: FERG dataset emotion example.

image are totally protected because the dataset can be distributed even without the face image of the original subjects and without compromising its purpose.

5.1.1.3 *Features injection*

Database features in the real world can be enriched with new information not originally present. An example is an experiment conducted in order to realize a high-resolution synthetic model of an iris. Other synthetic models already exist in the literature, but in this case, it differs in terms of purpose. The intention was to generate a series of light reflections inside the iris such as to make extremes about the difficulty of its segmentation. The virtual results can be seen in the Fig. 5.3.

Figure 5.3: Synthetic Iris biometrics trait examples.

5.1.1.4 *From 2D to 3D*

Starting from the consideration that there are many datasets of faces in 2D images in the literature, some authors have focused on the possibility of generating the respective 3D models. In [77] the authors set themselves the goal of the 3D reconstruction of a face from a single in-the wild image with an increasing level of detail and high fidelity characteristics. If on the one hand, the availability of many 2d images with faces allows the generation of innumerable 3D models, on the other hand as many starting 3D models are necessary. Furthermore, even if the visual effects are considerable, in essence, the 3D models are not very different as a point cloud shape. In any case, the result is a realistic rendering and the carried out models are useful and good as can be seen in Fig. 5.4.

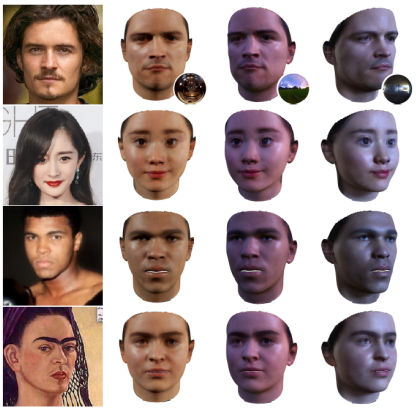


Figure 5.4: From 2D (left) to 3D face reconstructions under different environment maps with added spot lights.

5.1.1.5 *Generative models*

Non-existent data generation consistent with existing data is a prerogative of Generative Neural Networks (GANNs). GANNs learning by samples how to transfer predetermines features form sample to one other preserving some characteristics of the destination sample. The authors of [130] propose a novel Dual-Agent Generative Adversarial Network (DA-GAN) for synthesis of profile views to generate high-quality faces that are really useful for unconstrained face recognition. The biggest obstacle to learning a well-performing pose-invariant model for unconstrained face recognition is the high variety and few profile face photos for each participant. To solve this issue, they trained the DA-GAN on face photos with various predefined poses (i.e., yaw angles), and explicitly augmented the existing training data while balancing the pose distribution without further human annotation efforts. In Fig. 5.5 are shown some DA-GAN results.

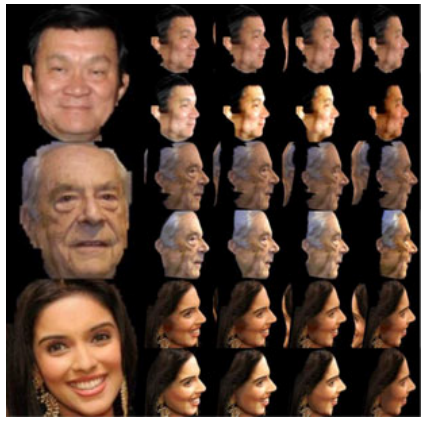


Figure 5.5: Refined results of DA-GAN

5.1.2 *Synthetic Environment*

Virtual Environments mirror the visual and dynamics characteristics of the real world to obtain new synthetic worlds in which to immerse human-machine interaction or design interactions between systems and environments. There are many fields of application, automotive, medicine, military, and gaming, in which

there has been the greatest development, and many others. A virtual environment can be thought of as a set of synthetic datasets between static and animated elements with which it is possible to manage interactions.

Suffice it to say that in the automotive field, virtual environments built in doc, with myriads of sensors and functions, sell modeled and tested future mechanisms for autonomous driving which will then be placed on the market. The level of reality, understood not only from a visual point of view, but also from an interactive point of view, is already very high and in some cases the cognitive sensation perceived by the user bordered on reality.

Building an adequate visual environment requires a lot of design and development efforts, and computationally powerful tools such as modern GPUs are available. The most widely used tools are Unity, Blender, Unreal3D, and Robot Operating System (ROS). While the first were born for gaming, ROS is a tool that has evolved especially for the management of robots in the industrial sector.

The purposes are important because, in cases where prototyping is dangerous both for equipment and people, they solve the problem. An example is the prototyping of the project presented in the 3.2 session, in which the use of a simultaneous driver for the drone is shown. Although without high-resolution features, the drone simulator allowed us to prototype the flight around a point in which a subject would then be placed for video recording.

5.2 WHAT COULD BE DONE

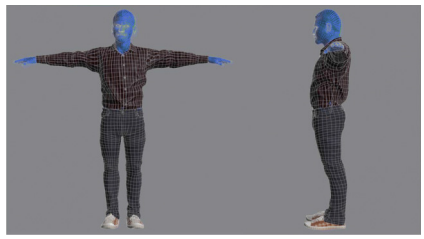


Figure 5.6: An illustration of the avatar construction.

The results presented in Section 3.2 can be further integrated in a broader experimentation involving the use of 3D face models in a virtual environment. The face region of the 3D models obtained can be nested in animated characters named avatars, inserted into the virtual environment to create an architecture suitable for biometric purposes and collect new synthetic datasets [49], Fig. 5.6. The solutions carried out can then be verified in real circumstances by involving the subjects used in the acquisition by means of a drone. An example of a first prototype implemented with the intention of displaying features is shown in Fig. 5.7.

(a) (b)

Figure 5.7: Prototypical environment of acquisition by means of drone. Different viewpoints of the scene: a) from the observer, b) from the drone.

6

CONCLUSIONS AND FUTURE WORKS

From the study carried out so far in the work of PhD, it is possible to draw the following conclusions. Biometrics still represents an open and constantly evolving field of research. New sensors and software technologies broaden the field of applications, while also improving the precision level and reliability of biometric systems. The evolution of mobile devices is increasing the demand for this approach in everyday life to ensure security and reliable services related to personal identity. The results and contributions of this work are described below. Further discussion of future scenarios that may occur and which would require further investigation in this area.

6.1 CONTRIBUTIONS, RESULTS, AND DISCUSSION

The evolution of surveillance systems always requires new solutions, and therefore leads to the creation of more complete and multipurpose datasets suitable for biometric recognition and identification. Following an in-depth study, we realized the lack of a dataset that gathered more features present even in datasets already existing but completely unrelated to each other and not applied to the same subjects.

Due to the variety of acquisitions available, such as biometric traits, mobile devices, pose, lighting, control conditions, and the use of a drone, MUBIDUS-I can simplify the study and execution of identification experiments in a variety of real-world scenarios involving surveillance systems equipped with a variety of acquisition devices.

Mobile devices, drones, and new categories of devices may be added to these systems in the near future. A broader version of MUBIDUS-I is currently in work in progress. The new version will increase the number of biometrics recorded, the number of subjects, and the type of recording devices.

Although several multimodal biometric databases are already available for research purposes, but none of them can be used like MUBIDUS-I: acquisitions of different biometric traits on different devices are simultaneous from different views; not all datasets use the same subjects acquired in different sessions at a distance of time and under different conditions; multimodal datasets use specific devices while this uses easily available and everyday ones; the variety of protocols, and biometrics in the dataset allow you to simulate real-world environments and conditions, and can be combined in heterogeneous ways to create different experimental conditions; most of the datasets are collected for general-purpose monitoring or activity recognition, with limited focus on person identification or face recognition, this not; we plan to expand it to include a further 3D model dataset of the face generated by the acquisition data of the first; the coming dataset show that it was a good intuition (others worked independently like PDESTRE [75]) All of this in one dataset.

Due to drones are rapidly populating human spaces and curiosity, the number of studies on facial recognition using drones is growing as [9], and some nice ones like what explores the use of facial expressions to represent emotions on social drones [58]. Therefore, having a system available that automatically captures faces is important, more and more for 3D face model. Furthermore, the proposed system can be remodelled and integrated to carry out the recognition method.

The 3D representation of the human face can add important information to improve the performance of the biometric recognition approach. In the Section3.2 proposes an automated method for 3D reconstruction of faces with a single dynamic source.

By means of a commercial UAV that autonomously flies along a dynamic trajectory to capture pictures of the face, the 3D face model is computed. Images are processed into a true 3D reconstruction suitable for both biometric purposes and all application areas where these models can be used. Obstacle avoidance sensors allow objects to be approached within 1.5 m, so obtaining a geometric representation that is accurate enough to provide adequate 3D reconstruction quality is not an easy task. Proximity sensors can be disabled via DJI SDK, but this is not recommended for user safety as the entire process is automated. By keeping

this constraint active, we have demonstrated that in-flight drones can achieve accurate 3D reconstruction.

The 3D models of faces obtained from controlled acquisition in a laboratory environment that were collected using a standard mobile device with low camera noise, were compared with those of the wild environment. The co-recording technique of the Iterative Closest Point algorithm revealed that the models acquired by the drone at various resolutions are mostly equivalent to the model obtained in the absence of environmental noise. Higher resolutions of the camera embedded in the aircraft produced the most promising results as expected. However, based on the results of an experiment with 20 people, 4K video resolution has greater accuracy compared to C4K, and low computing time.

The next step in this direction will be to solve some Remote Controller performance issues and move most of the process directly into it.

Another useful development would be to further specialize the presented system for biometric applications. Other physical and behavioral traits could be used, e.g., iris or ear. and gait.

The results obtained show that UAVs can be used to automate the collection and storage of facial biometric information. They also suggest that the proposed approach may allow monitoring large areas of open access without human intervention, at least during the detection phase.

We hope that technological developments related to computing power and battery consumption will enable all computing parts to be installed on board in the near future, to achieve the real-time computation. Therefore, new sensors on future UAVs will make it easier and faster to capture 3D models of the face.

6.2 LOOKING TO THE FUTURE

The arrival of 5G has enhanced the sensor connection network for the smart cities to become. Information from countless sensors or large data streams from more complex devices, such as drones and robots, and super cameras, will transmit at super real-time speeds [91].

This technological condition will allow implementations in the planning stage to be able to become operative. Projects like the

Figure 6.1: 5G solutions of remote 3D Reconstruction.

one presented can be inserted in many contexts of cities and easily remotely controlled, Figure 6.1, even if some problems still need to be solved. Connections to 5G base stations cannot be maintained during flight of the drone, but handovers to 4G will occur. Furthermore, the higher the flight altitude, the lower the throughput and the more frequent handovers, as highlighted in [93]. Nevertheless, the tests continue: Ferro et al. experimented aerial drones and exploited the 5G mobile network communication for face recognition [46].

Furthermore, the transition to a more complex solution, such as that of using swarms of drones, robots, or various sensors will be more possible, without great effort [118] (Figure 6.2).

Figure 6.2: Advanced solution with swarms of drones.

BIBLIOGRAPHY

[1] 3Dflow. *3DF Zephyr Sofware*. 2020. URL: https://www.3dflow.net/it//.

[2] Andrea F. Abate, Luigi De Maio, Riccardo Distasi, and Fabio Narducci. "Remote 3D face reconstruction by means of autonomous unmanned aerial vehicles." In: *Pattern Recognition Letters* 147 (2021), pp. 48–54. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2021.04.006. URL: https://www.sciencedirect.com/science/article/pii/S0167865521001367.

[3] Andrea F. Abate, Michele Nappi, Stefano Ricciardi, and Gabriele Sabatino. "Fast 3D face recognition based on normal map." In: *IEEE International Conference on Image Processing 2005* 2 (2005), pp. II–946.

[4] Andrea F. Abate, Michele Nappi, Daniel Riccio, and Gabriele Sabatino. "2D and 3D face recognition: A survey." In: *Pattern Recognition Letters* 28.14 (2007). Image: Information and Control, pp. 1885 –1906. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2006.12.018. URL: http://www.sciencedirect.com/science/article/pii/S0167865507000189.

[5] Ayman Abaza, Arun Ross, Christina Hebert, Mary Ann F Harrison, and Mark S Nixon. "A survey on ear biometrics." In: *ACM computing surveys (CSUR)* 45.2 (2013), pp. 1–35.

[6] Muzammil Abdulrahman, Tajuddeen R. Gwadabe, Fahad J. Abdu, and Alaa Eleyan. "Gabor wavelet transform based facial expression recognition using PCA and LBP." In: *2014 22nd Signal Processing and Communications Applications Conference (SIU)*. 2014, pp. 2265–2268. DOI: 10.1109/SIU.2014.6830717.

[7] Giorgio Agamben. *HOMO SACER: Sovereign Power and Bare Life*. stanford university Press, 2020.

[8] F. Al-Osaimi, M. Bennamoun, and A. Mian. "An Expression Deformation Approach to Non-rigid 3D Face Recognition." In: *International Journal of Computer Vision* 81.3 (2009), pp. 302–316. ISSN: 1573-1405. DOI: 10.1007/s11263-008-0174-0. URL: https://doi.org/10.1007/s11263-008-0174-0.

[9] Giuseppe Amato, Fabrizio Falchi, Claudio Gennaro, Fabio Valerio Massoli, and Claudio Vairo. "Multi-Resolution Face Recognition with Drones." In: *2020 3rd International Conference on Sensors, Signal and Image Processing*. SSIP 2020. Prague, Czech Republic: Association for Computing Machinery, 2020. ISBN: 9781450388283. DOI: 10.1145/3441233.3441237. URL: https://doi.org/10.1145/3441233.3441237.

[10] D. Aneja, B. Chaudhuri, A. Colburn, G. Faigin, L. Shapiro, and B. Mones. "Learning to Generate 3D Stylized Character Expressions from Humans." In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2018, pp. 160–169.

[11] Andrew D. Bagdanov, Alberto Del Bimbo, and Iacopo Masi. "The Florence 2D/3D Hybrid Face Dataset." In: *Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding*. J-HGBU '11. ACM, 2011. ISBN: 978-1-4503-0998-1. DOI: 10.1145/2072572.2072597. URL: http://doi.acm.org/10.1145/2072572.2072597.

[12] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. "Recognizing facial expression: machine learning and application to spontaneous behavior." In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 2. 2005, 568–573 vol. 2. DOI: 10.1109/CVPR.2005.297.

[13] Pratichi Basak, Saurabh De, Mallika Agarwal, Aakarsh Malhotra, Mayank Vatsa, and Richa Singh. "Multimodal biometric recognition for toddlers and pre-school children." In: *2017 IEEE International Joint Conference on Biometrics (IJCB)*. 2017, pp. 627–633. DOI: 10.1109/BTAS.2017.8272750.

[14] Massimo Bertozzi, Alberto Broggi, Massimo Cellario, Alessandra Fascioli, Paolo Lombardi, and Marco Porta. "Artificial vision in road vehicles." In: *Proceedings of the IEEE* 90.7 (2002), pp. 1258–1271.

[15] P. J. Besl and N. D. McKay. "A method for registration of 3-D shapes." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2 (1992), pp. 239–256.

[16] Margherita Bonetto, Pavel Korshunov, Giovanni Ramponi, and Touradj Ebrahimi. "Privacy in mini-drone based video surveillance." In: *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. Vol. 04. 2015, pp. 1–6. DOI: 10.1109/FG.2015.7285023.

[17] Imed Bouchrika, Michaela Goffredo, John Carter, and Mark Nixon. "On Using Gait in Forensic Biometrics." In: *Journal of Forensic Sciences* 56.4 (2011), pp. 882–889. DOI: https://doi.org/10.1111/j.1556-4029.2011.01793.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1556-4029.2011.01793.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1556-4029.2011.01793.x.

[18] G. Bradski. *The OpenCV Library*. 2000.

[19] Manfred Bromba. *Biometrics FAQ's*. 2006. URL: http://www.bromba.com/faq/biofaqe.htm. (accessed: 01.09.2016).

[20] J. Bruce, J. Perron, and R. Vaughan. "Ready—Aim—Fly! Hands-Free Face-Based HRI for 3D Trajectory Control of UAVs." In: *2017 14th Conference on Computer and Robot Vision (CRV)*. 2017, pp. 307–313. DOI: 10.1109/CRV.2017.39.

[21] William Lowe Bryan and Noble Harter. "Studies in the physiology and psychology of the telegraphic language." In: *Psychological Review* 4.1 (1897), p. 27.

[22] K. Bálint. "UAVs with Biometric Facial Recognition Capabilities in the Combat Against Terrorism." In: *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*. 2018, pp. 000185–000190. DOI: 10.1109/SISY.2018.8524800.

[23] J.P. Campbell. "Speaker recognition: a tutorial." In: *Proceedings of the IEEE* 85.9 (1997), pp. 1437–1462. DOI: 10.1109/5.628714.

[24] V. Cantoni, M. Porta, L. De Maio, R. Distasi, and M. Nappi. "Towards a novel technique for identification based on eye tracking." In: *2012 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings*. 2012, pp. 1–4. DOI: 10.1109/BIOMS.2012.6345780.

[25] Virginio Cantoni, Chiara Galdi, Michele Nappi, Marco Porta, and Daniel Riccio. "GANT: Gaze analysis technique for human identification." In: *Pattern Recognition* 48.4 (2015), pp. 1027–1038.

[26] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. *OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*. 2019. arXiv: 1812.08008 [cs.CV].

[27] J. R. Cauchard, A. Tamkin, C. Y. Wang, L. Vink, M. Park, T. Fang, and J. A. Landay. "Drone.io: A Gestural and Visual Interface for Human-Drone Interaction." In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2019, pp. 153–162. DOI: 10.1109/HRI.2019.8673011.

[28] J. R. Cauchard, K. Y. Zhai, M. Spadafora, and J. A. Landay. "Emotion encoding in Human-Drone Interaction." In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2016, pp. 263–270. DOI: 10.1109/HRI.2016.7451761.

[29] P. Chen and C. Lee. "UAVNet: An Efficient Obstacel Detection Model for UAV with Autonomous Flight." In: *2018 International Conference on Intelligent Autonomous Systems (ICoIAS)*. 2018, pp. 217–220. DOI: 10.1109/ICoIAS.2018.8494201.

[30] DJI. *DJI Mobile SDK*. 2018. URL: https://developer.dji.com/.

[31] DJI. *DJI Phantom 4 Pro+*. 2020. URL: https://www.dji.com/it/phantom-4-pro.

[32] Mohamed Dahmane and Jean Meunier. "Emotion recognition using dynamic grid-based HoG features." In: *2011 IEEE International Conference on Automatic Face Gesture Recognition (FG)*. 2011, pp. 884–888. DOI: 10.1109/FG.2011.5771368.

[33] Shaveta Dargan and Munish Kumar. "A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities." In: *Expert Systems with Applications* 143 (2020), p. 113114. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2019.113114. URL: https://www.sciencedirect.com/science/article/pii/S0957417419308310.

[34] John Daugman. "Probing the Uniqueness and Randomness of IrisCodes: Results From 200 Billion Iris Pair Comparisons." In: *Proceedings of the IEEE* 94.11 (2006), pp. 1927–1935. DOI: 10.1109/JPROC.2006.884092.

[35] L. De Maio, R. Distasi, and M. Nappi. "MUBIDUS I - Multibiometric and Multipurpose Dataset." In: *Proc. SITIS 2019 – The 15th International Conference on Signal Image Technology & Internet based Systems*. 2019, pp. 748–753. DOI: 10.1109/SITIS.2019.00124.

[36] Maria De Marsico, Michele Nappi, and Daniel Riccio. "FARO: FAce Recognition Against Occlusions and Expression Variations." In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 40.1 (2010), pp. 121–132. DOI: 10.1109/TSMCA.2009.2033031.

[37] Maria De Marsico, Michele Nappi, Daniel Riccio, and Harry Wechsler. "Mobile Iris Challenge Evaluation (MICHE)-I, biometric iris dataset and protocols." In: *Pattern Recognition Letters* 57 (2015). Mobile Iris CHallenge Evaluation part I (MICHE I), pp. 17–23. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2015.02.009. URL: https://www.sciencedirect.com/science/article/pii/S0167865515000574.

[38] Maria De Marsico, Michele Nappi, Daniel Riccio, and Harry Wechsler. "Robust face recognition after plastic surgery using region-based approaches." In: *Pattern Recognition* 48.4 (2015), pp. 1261–1276.

[39] Maria De Marsico, Alfredo Petrosino, and Stefano Ricciardi. "Iris recognition through machine learning techniques: A survey." In: *Pattern Recognition Letters* 82 (2016). An insight on eye biometrics, pp. 106–115. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2016.02.001. URL: https://www.sciencedirect.com/science/article/pii/S0167865516000477.

[40] R. Donida Labati, A. Genovese, V. Piuri, and F. Scotti. "Toward Unconstrained Fingerprint Recognition: A Fully Touchless 3-D System Based on Two Views on the Move." In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 46.2 (2016), pp. 202–219. ISSN: 2168-2232. DOI: 10.1109/TSMC.2015.2423252.

[41] Bernadette Dorizzi, Sonia Garcia-Salicetti, and Lorene Allano. "Multimodality In Biosecure: Evaluation On Real Vs. Virtual Subjects." In: vol. 5. June 2006, pp. V –V. DOI: 10.1109/ICASSP.2006.1661469.

[42] Hassen Drira, Boulbaba Ben Amor, Anuj Srivastava, Mohamed Daoudi, and Rim Slama. "3D Face Recognition under Expressions, Occlusions, and Pose Variations." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.9 (2013), pp. 2270–2283. DOI: 10.1109/TPAMI.2013.48.

[43] Paul Ekman. "Universals and cultural differences in facial expressions of emotion." In: *Nebraska Symposium on Motivation* 19 (1971), pp. 207–283. ISSN: 0146-7875(Print).

[44] Charlotte Epstein. "Guilty Bodies, Productive Bodies, Destructive Bodies: Crossing the Biometric Borders." In: *International Political Sociology* 1 (2007), pp. 149 –164. DOI: 10.1111/j.1749-5687.2007.00010.x.

[45] Scott L Feld and Bernard Grofman. "The Borda count in n-dimensional issue space." In: *Public Choice* 59.2 (1988), pp. 167–176.

[46] Erina Ferro, Claudio Gennaro, Alessandro Nordio, Fabio Paonessa, Claudio Vairo, Giuseppe Virone, Arturo Argentieri, Andrea Berton, and Andrea Bragagnini. "5G-enabled

security scenarios for unmanned aircraft: Experimentation in urban environment." In: *Drones* 4.2 (2020), p. 22.

[47] Julian Fierrez, Javier Galbally, Javier Ortega-Garcia, Manuel R. Freire, Fernando Alonso-Fernandez, Daniel Ramos, Doroteo Torre Toledano, Joaquín González-Rodríguez, Juan A. Sigüenza, and Javier Garrido Salas. "BiosecurID: a multimodal biometric database." In: *Pattern Analysis and Applications* 13 (2009), pp. 235–246.

[48] Matthew C. Fysh and Markus Bindemann. "Person Identification from Drones by Humans: Insights from Cognitive Psychology." In: *Drones* 2.4 (2018). ISSN: 2504-446X. DOI: 10.3390/drones2040032. URL: https://www.mdpi.com/2504-446X/2/4/32.

[49] Matthew C. Fysh, Iliyana V. Trifonova, John Allen, Cade McCall, A. Mike Burton, and Markus Bindemann. "Avatars with faces of real people: A construction method for scientific experiments in virtual reality." In: *Behavior Research Methods* (2021). ISSN: 1554-3528. DOI: 10.3758/s13428-021-01676-5. URL: https://doi.org/10.3758/s13428-021-01676-5.

[50] Sonia Garcia-Salicetti, Charles Beumier, Gérard Chollet, Bernadette Dorizzi, Jean Leroux les Jardins, Jan Lunter, Yang Ni, and Dijana Petrovska-Delacrétaz. "BIOMET: A Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities." In: *Audio- and Video-Based Biometric Person Authentication*. Ed. by Josef Kittler and Mark S. Nixon. Springer Berlin Heidelberg, 2003, pp. 845–853. ISBN: 978-3-540-44887-7.

[51] Cong Geng and Xudong Jiang. "Face recognition using sift features." In: *2009 16th IEEE International Conference on Image Processing (ICIP)*. 2009, pp. 3313–3316. DOI: 10.1109/ICIP.2009.5413956.

[52] Agnieszka Gidziela et al. "Using DNA to predict behaviour problems from preschool to adulthood." In: *medRxiv* (2021). DOI: 10.1101/2021.02.15.21251308.

[53] D. Giorgi et al. "A Critical Assessment of 2D and 3D Face Recognition Algorithms." In: *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. 2009, pp. 79–84.

[54] Google. *Android Studio Developers*. 2020. URL: https://developer.android.com/studio.

[55] Google. *Mobile Vision*. 2020. URL: https://developers.google.com/vision.

[56] Patrick Grother, Mei Ngan, and Kayee Hanaoka. *Ongoing Face Recognition Vendor Test (FRVT) Part 2: Identification*. en. 2018. DOI: https://doi.org/10.6028/NIST.IR.8238.

[57] Linh Ha, Jens Krüger, and Cláudio T. Silva. "Fast four-way parallel radix sorting on GPUs." English (US). In: *Computer Graphics Forum* 28.8 (2009), pp. 2368–2378. ISSN: 0167-7055. DOI: 10.1111/j.1467-8659.2009.01542.x.

[58] Viviane Herdel, Anastasia Kuzminykh, Andrea Hildebrandt, and Jessica R. Cauchard. "Drone in Love: Emotional Perception of Facial Expressions on Flying Robots." In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2021. ISBN: 9781450380966. URL: https://doi.org/10.1145/3411764.3445495.

[59] Tin Kam Ho. "Random decision forests." In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol. 1. 1995, 278–282 vol.1. DOI: 10.1109/ICDAR.1995.598994.

[60] Lin Hong and Anil Jain. "Integrating faces and fingerprints for personal identification." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.12 (1998), pp. 1295–1307. DOI: 10.1109/34.735803.

[61] Leroy Hood and David Galas. "The digital code of DNA." In: *Nature* 421.6921 (2003), pp. 444–448. ISSN: 1476-4687. DOI: 10.1038/nature01410. URL: https://doi.org/10.1038/nature01410.

[62] Hwai-Jung Hsu and Kuan-Ta Chen. "DroneFace: An Open Dataset for Drone Research." In: *Proceedings of the 8th ACM on Multimedia Systems Conference*. MMSys'17. Taipei, Taiwan: Association for Computing Machinery, 2017. ISBN: 9781450350020. DOI: 10.1145/3083187.3083214. URL: https://doi.org/10.1145/3083187.3083214.

[63] T. Huang, S. Hsu, and L. Fu. "Data Augmentation via Face Morphing for Recognizing Intensities of Facial Emotions." In: *IEEE Transactions on Affective Computing* 01 (5555), pp. 1–1. ISSN: 1949-3045. DOI: 10.1109/TAFFC.2021.3096922.

[64] Gerardo Iovane, Carmen Bisogni, Luigi De Maio, and Michele Nappi. "An Encryption Approach Using Information Fusion Techniques Involving Prime Numbers and Face Biometrics." In: *IEEE Transactions on Sustainable Computing* 5.2 (2020), pp. 260–267. DOI: 10.1109/TSUSC.2018.2793466.

[65] A. K. Jain, A. Ross, and S. Prabhakar. "An introduction to biometric recognition." In: *IEEE Transactions on Circuits and Systems for Video Technology* 14.1 (2004), pp. 4–20. ISSN: 1558-2205. DOI: 10.1109/TCSVT.2003.818349.

[66] A.K. Jain, A. Ross, and S. Pankanti. "Biometrics: a tool for information security." In: *IEEE Transactions on Information Forensics and Security* 1.2 (2006), pp. 125–143. DOI: 10.1109/TIFS.2006.873653.

[67] A.K. Jain, A. Ross, and S. Pankanti. "Biometrics: a tool for information security." In: *IEEE Transactions on Information Forensics and Security* 1.2 (2006), pp. 125–143. DOI: 10.1109/TIFS.2006.873653.

[68] Anil K. Jain, Salil Prabhakar, and Sharath Pankanti. "On the similarity of identical twin fingerprints." In: *Pattern Recognition* 35.11 (2002), pp. 2653–2663. ISSN: 0031-3203. DOI: https://doi.org/10.1016/S0031-3203(01)00218-7. URL: https://www.sciencedirect.com/science/article/pii/S0031320301002187.

[69] Chen Juanjuan, Zhao Zheng, Sun Han, and Zhang Gang. "Facial expression recognition based on PCA reconstruction." In: *2010 5th International Conference on Computer Science Education*. 2010, pp. 195–198. DOI: 10.1109/ICCSE.2010.5593658.

[70] Ari Juels and Madhu Sudan. "A fuzzy vault scheme." In: *Designs, Codes and Cryptography* 38.2 (2006), pp. 237–257.

[71] PETER C. KRONFELD. "CHAPTER 1 - The Gross Anatomy and Embryology of the Eye." In: *Vegetative Physiology and Biochemistry*. Ed. by HUGH DAVSON. Academic Press, 1962, pp. 1–62. ISBN: 978-1-4832-3090-0. DOI: https://doi.org/10.1016/B978-1-4832-3090-0.50007-1. URL: https://www.sciencedirect.com/science/article/pii/B9781483230900500071.

[72] Henry S. Kahn, Mariaelisa Graff, Aryeh D. Stein, Patricia A. Zybert, Ian W. McKeague, and L. H. Lumey. "A fingerprint characteristic associated with the early prenatal environment." In: *American Journal of Human Biology* 20.1 (2008), pp. 59–65. DOI: https://doi.org/10.1002/ajhb.20672. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ajhb.20672. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ajhb.20672.

[73] Nathan D. Kalka, Brianna Maze, James A. Duncan, Kevin O'Connor, Stephen Elliott, Kaleb Hebert, Julia Bryan, and Anil K. Jain. "IJB–S: IARPA Janus Surveillance Video Benchmark." In: *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. 2018, pp. 1–9. DOI: 10.1109/BTAS.2018.8698584.

[74] Isha Kalra, Maneet Singh, Shruti Nagpal, Richa Singh, Mayank Vatsa, and P. B. Sujit. "DroneSURF: Benchmark Dataset for Drone-based Face Recognition." In: *2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019)*. 2019, pp. 1–7. DOI: 10.1109/FG.2019.8756593.

[75] S. V. Aruna Kumar, Ehsan Yaghoubi, Abhijit Das, B. S. Harish, and Hugo Proença. "The P-DESTRE: A Fully Annotated Dataset for Pedestrian Detection, Tracking, and

Short/Long-Term Re-Identification From Aerial Devices." In: *IEEE Transactions on Information Forensics and Security* 16 (2021), pp. 1696–1708. DOI: 10.1109/TIFS.2020.3040881.

[76] Hanna-Kaisa Lammi. "Ear biometrics." In: *Lappeenranta University of Technology* (2004).

[77] Alexandros Lattas, Stylianos Moschoglou, Baris Gecer, Stylianos Ploumpis, Vasileios Triantafyllou, and Stefanos Zafeiriou. "AvatarMe: Realistically Renderable 3D Facial Reconstruction "in-the-wild"." In: (Mar. 2020).

[78] Ryan Layne, Timothy M. Hospedales, and Shaogang Gong. "Investigating Open-World Person Re-identification Using a Drone." In: *Computer Vision - ECCV 2014 Workshops*. Ed. by Lourdes Agapito, Michael M. Bronstein, and Carsten Rother. Springer International Publishing, 2015, pp. 225–240. ISBN: 978-3-319-16199-0.

[79] Yinjie Lei, Yulan Guo, Munawar Hayat, Mohammed Bennamoun, and Xinzhi Zhou. "A Two-Phase Weighted Collaborative Representation for 3D partial face recognition with single sample." In: *Pattern Recognition* 52 (2016), pp. 218–237. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2015.09.035. URL: https://www.sciencedirect.com/science/article/pii/S0031320315003660.

[80] Wei Li, Kai Liu, Lin Yan, Fei Cheng, YunQiu Lv, and LiZhe Zhang. "FRD-CNN: Object detection based on small-scale convolutional neural networks and feature reuse." In: *Scientific Reports* 9.1 (Nov. 2019), p. 16294. ISSN: 2045-2322. DOI: 10.1038/s41598-019-52580-0. URL: https://doi.org/10.1038/s41598-019-52580-0.

[81] Yali Li, Shengjin Wang, Qi Tian, and Xiaoqing Ding. "A survey of recent advances in visual feature detection." In: *Neurocomputing* 149 (2015), pp. 736–751. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2014.08.003. URL: https://www.sciencedirect.com/science/article/pii/S0925231214010121.

[82] Shisong Lin, Mengchao Bai, Feng Liu, Linlin Shen, and Yicong Zhou. "Orthogonalization-Guided Feature Fusion Network for Multimodal 2D+3D Facial Expression Recog-

nition." In: *IEEE Transactions on Multimedia* PP (2020), pp. 1–1. DOI: 10.1109/TMM.2020.3001497.

[83] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. "SphereFace: Deep Hypersphere Embedding for Face Recognition." In: Apr. 2017. DOI: 10.1109/CVPR.2017.713.

[84] D.G. Lowe. "Object recognition from local scale-invariant features." In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 2. 1999, 1150–1157 vol.2. DOI: 10.1109/ICCV.1999.790410.

[85] Salvatore Luongo, Marianna Di Gregorio, Giuliana Vitiello, and Angela Vozella. "Human Machine Interface Issues for Drone Fleet Management." In: *Human Systems Engineering and Design*. Ed. by Tareq Ahram, Waldemar Karwowski, and Redha Taiar. Springer International Publishing, 2019, pp. 791–796. ISBN: 978-3-030-02053-8.

[86] Davide Maltoni, Dario Maio, Anil K. Jain, and Salil Prabhakar. "Handbook of Fingerprint Recognition." In: *Springer Professional Computing*. 2003.

[87] Manisha and Nitin Kumar. "Cancelable Biometrics: a comprehensive survey." In: *Artificial Intelligence Review* 53.5 (2020), pp. 3403–3446. ISSN: 1573-7462. DOI: 10.1007/s10462-019-09767-8. URL: https://doi.org/10.1007/s10462-019-09767-8.

[88] Roy A Maxion and Kevin S Killourhy. "Keystroke biometrics with number-pad input." In: *2010 IEEE/IFIP International Conference on Dependable Systems & Networks (DSN)*. IEEE. 2010, pp. 201–210.

[89] S. Milborrow, J. Morkel, and F. Nicolls. "The MUCT Landmarked Face Database." In: *Pattern Recognition Association of South Africa* (2010). http://www.milbo.org/muct.

[90] Yue Ming. "Rigid-area orthogonal spectral regression for efficient 3D face recognition." In: *Neurocomputing* 129 (2014), pp. 445–457. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2013.09.014. URL: https://www.sciencedirect.com/science/article/pii/S0925231213009235.

[91] Rupendra Nath Mitra and Dharma P. Agrawal. "5G mobile technology: A survey." In: *ICT Express* 1.3 (2015). Special Issue on Next Generation (5G/6G) Mobile Communications, pp. 132–137. ISSN: 2405-9595. DOI: https://doi.org/10.1016/j.icte.2016.01.003. URL: https://www.sciencedirect.com/science/article/pii/S2405959515300503.

[92] Emilio Mordini and Sonia Massari. "Body, biometrics and identity." In: *Bioethics* 22.9 (2008), pp. 488–498.

[93] Raheeb Muzaffar, Christian Raffelsberger, Aymen Fakhreddine, José López Luque, Driton Emini, and Christian Bettstetter. "First Experiments with a 5G-Connected Drone." In: *Proceedings of the 6th ACM Workshop on Micro Aerial Vehicle Networks, Systems, and Applications*. DroNet '20. Toronto, Ontario, Canada: Association for Computing Machinery, 2020. ISBN: 9781450380102. DOI: 10.1145/3396864.3400304. URL: https://doi.org/10.1145/3396864.3400304.

[94] J. Nagi, A. Giusti, G. A. D. Caro, and L. M. Gambardella. "Human Control of UAVs using Face Pose Estimates and Hand Gestures." In: *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2014, pp. 1–2.

[95] MIchele Nappi and Daniel Riccio. *Moderne tecniche di elaborazione di immagini e biometria*. CUA - Coop. Univ. Athen, 2008. ISBN: 8890306122.

[96] Joao Neves, Juan Moreno, and Hugo Proença. "QUIS-CAMPI: an annotated multi-biometrics data feed from surveillance scenarios." In: *IET Biometrics* 7 (2017). DOI: 10.1049/iet-bmt.2016.0178.

[97] E. Osuna, R. Freund, and F. Girosit. "Training support vector machines: an application to face detection." In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1997, pp. 130–136. DOI: 10.1109/CVPR.1997.609310.

[98] Vishal M. Patel, Nalini K. Ratha, and Rama Chellappa. "Cancelable Biometrics: A review." In: *IEEE Signal Processing Magazine* 32.5 (2015). DOI: 10.1109/MSP.2015.2434151.

[99] Vishal M Patel, Nalini K Ratha, and Rama Chellappa. "Cancelable biometrics: A review." In: *IEEE Signal Processing Magazine* 32.5 (2015), pp. 54–65.

[100] Marco Porta, Piercarlo Dondi, Nicola Zangrandi, and Luca Lombardi. "Gaze-Based Biometrics From Free Observation of Moving Elements." In: *IEEE Transactions on Biometrics, Behavior, and Identity Science* (2021), pp. 1–1. DOI: 10.1109/TBIOM.2021.3130798.

[101] Salil Prabhakar and Anil K. Jain. "Decision-level fusion in fingerprint verification." In: *Pattern Recognition* 35.4 (2002), pp. 861–874. ISSN: 0031-3203. DOI: https://doi.org/10.1016/S0031-3203(01)00103-0. URL: https://www.sciencedirect.com/science/article/pii/S0031320301001030.

[102] Antonio Rama, Francesc Tarrés, Jürgen Rurainsky, and Peter Eisert. "2D-3D Mixed Face Recognition Schemes." In: *Recent advances in face recognition* (2008). DOI: 10.5772/6398.

[103] Adin Ramirez Rivera, Jorge Rojas Castillo, and Oksam Oksam Chae. "Local Directional Number Pattern for Face Analysis: Face and Expression Recognition." In: *IEEE Transactions on Image Processing* 22.5 (2013), pp. 1740–1752. DOI: 10.1109/TIP.2012.2235848.

[104] Rui Raposo, Edmundo Hoyle, Adolfo Peixinho, and Hugo Proença. "UBEAR: A dataset of ear images captured on-the-move in uncontrolled conditions." In: *2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM)*. 2011, pp. 84–90. DOI: 10.1109/CIBIM.2011.5949208.

[105] N. K. Ratha, A. W. Senior, R. M. Bolle, S. Pankanti, and J. H. Connell. "The Relation between the ROC Curve and the CMC." In: *Proceedings. Fourth IEEE Workshop on Automatic Identification Advanced Technologies*. Los Alamitos, CA, USA: IEEE Computer Society, 2005, pp. 15–20. DOI: 10.1109/AUTOID.2005.48. URL: https://doi.ieeecomputersociety.org/10.1109/AUTOID.2005.48.

[106] Ajita Rattani, Reza Derakhshani, Sashi K. Saripalle, and Vikas Gottemukkula. "ICIP 2016 Competition on Mobile Ocular Biometric Recognition." In: *IEEE International Conference on Image Processing (ICIP) 2016, Challenge Session on Mobile Ocular Biometric Recognition*. 2016.

[107] Daniel Riccio, Genny Tortora, Maria De Marsico, and Harry Wechsler. "EGA — Ethnicity, gender and age, a pre-annotated face database." In: *2012 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings*. 2012, pp. 1–8. DOI: 10.1109/BIOMS.2012.6345776.

[108] Chris Riley, Kathy Buckner, Graham Johnson, and David Benyon. "Culture & biometrics: regional differences in the perception of biometric authentication technologies." In: *AI & SOCIETY* 24.3 (2009), pp. 295–306. ISSN: 1435-5655. DOI: 10.1007/s00146-009-0218-1. URL: https://doi.org/10.1007/s00146-009-0218-1.

[109] R. L. Rivest, A. Shamir, and L. Adleman. "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems." In: *Commun. ACM* 21.2 (1978). ISSN: 0001-0782. DOI: 10.1145/359340.359342. URL: https://doi.org/10.1145/359340.359342.

[110] Arun Ross and Anil K. Jain. "Multimodal biometrics: An overview." In: *2004 12th European Signal Processing Conference*. 2004, pp. 1221–1224.

[111] Andrew Rukhin, Juan Soto, James Nechvatal, Miles Smid, and Elaine Barker. *A statistical test suite for random and pseudorandom number generators for cryptographic applications*. Tech. rep. Booz-allen and hamilton inc mclean va, 2001.

[112] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos. "Biometric identification through hand geometry measurements." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.10 (2000), pp. 1168–1171. DOI: 10.1109/34.879796.

[113] Conrad Sanderson and Brian C. Lovell. "Multi-Region Probabilistic Histograms for Robust and Scalable Identity Inference." In: *Advances in Biometrics*. Ed. by Massimo Tistarelli and Mark S. Nixon. Springer Berlin Heidelberg, 2009, pp. 199–208. ISBN: 978-3-642-01793-3.

[114] Ana F. Sequeira, João C. Monteiro, Ana Rebelo, and Hélder P. Oliveira. "MobBIO: A Multimodal Database Captured with a Portable Handheld Device." In: *Proceedings of the 9th International Conference on Computer Vision Theory and Applications - Volume 3: VISAPP, (VISIGRAPP 2014)*. INSTICC. SciTePress, 2014, pp. 133–139. ISBN: 978-989-758-009-3. DOI: 10.5220/0004679601330139.

[115] Caifeng Shan, Shaogang Gong, and Peter W. McOwan. "Facial expression recognition based on Local Binary Patterns: A comprehensive study." In: *Image and Vision Computing* 27.6 (2009), pp. 803–816. ISSN: 0262-8856. DOI: 10.1016/j.imavis.2008.08.005.

[116] Dhirendra Pratap Singh, Ishan Joshi, and Jaytrilok Choudhary. "Survey of GPU Based Sorting Algorithms." In: *Int. J. Parallel Program.* 46.6 (2018). ISSN: 0885-7458. DOI: 10.1007/s10766-017-0502-5. URL: https://doi.org/10.1007/s10766-017-0502-5.

[117] Colin Soutar, Danny Roberge, Alex Stoianov, Rene Gilroy, and Bhagavatula Vijaya Kumar. "Biometric encryption using image processing." In: *Optical Security and Counterfeit Deterrence Techniques II*. Vol. 3314. International Society for Optics and Photonics. 1998, pp. 178–188.

[118] Konstantina Spanaki, Erisa Karafili, Uthayasankar Sivarajah, Stella Despoudi, and Zahir Irani. "Artificial intelligence and food security: swarm intelligence of AgriTech drones for smart AgriFood operations." In: *Production Planning & Control* 0.0 (2021), pp. 1–19. DOI: 10.1080/09537287.2021.1882688. eprint: https://doi.org/10.1080/09537287.2021.1882688. URL: https://doi.org/10.1080/09537287.2021.1882688.

[119] National Institute of Standards and Commerce Department Technology (NIST). *Face Recognition Vendor Test (FRVT)*

Part 21: Identification. Government. Commerce Department. 2019. URL: https://www.govinfo.gov/app/details/GOVPUB-C13-a4a30985dc259996c42d67593fec5166.

[120] L. Takayama, W. Ju, and C. Nass. "Beyond dirty, dangerous and dull: What everyday people think robots should do." In: *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2008, pp. 25–32.

[121] R. Tariq, M. Rahim, N. Aslam, N. Bawany, and U. Faseeha. "DronAID : A Smart Human Detection Drone for Rescue." In: *2018 15th International Conference on Smart Cities: Improving Quality of Life Using ICT IoT (HONET-ICT)*. 2018, pp. 33–37. DOI: 10.1109/HONET.2018.8551326.

[122] U. Uludag, S. Pankanti, S. Prabhakar, and A.K. Jain. "Biometric cryptosystems: issues and challenges." In: *Proceedings of the IEEE* 92.6 (2004), pp. 948–960. DOI: 10.1109/JPROC.2004.827372.

[123] Umut Uludag, Sharath Pankanti, and Anil K Jain. "Fuzzy vault for fingerprints." In: *International Conference on Audio- and Video-Based Biometric Person Authentication*. Springer. 2005, pp. 310–319.

[124] B. Victor, K. Bowyer, and S. Sarkar. "An evaluation of face and ear biometrics." In: *2002 International Conference on Pattern Recognition*. Vol. 1. 2002, 429–432 vol.1. DOI: 10.1109/ICPR.2002.1044746.

[125] P. Viola and M. Jones. "Rapid object detection using a boosted cascade of simple features." In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Vol. 1. 2001, pp. I–I. DOI: 10.1109/CVPR.2001.990517.

[126] Adam Watts, Vincent Ambrosia, and Everett Hinkley. "Unmanned Aircraft Systems in Remote Sensing and Scientific Research: Classification and Considerations of Use." In: *RS* 4 (2012), pp. 1671–1692. DOI: 10.3390/rs4061671.

[127] Michal Wlodarczyk, Damian Kacperski, Wojciech Sankowski, and Kamil Grabowski. "COMPACT: Biometric Dataset of Face Images Acquired in Uncontrolled Indoor Environ-

ment." In: *Computer Science* 20 (2019). DOI: 10.7494/csci.2019.20.1.3020.

[128] Dacheng Xu and Bailiang Li. "A pseudo-random sequence fingerprint key algorithm based on fuzzy vault." In: *2009 International Conference on Mechatronics and Automation*. 2009, pp. 2421–2425. DOI: 10.1109/ICMA.2009.5246046.

[129] N. Yao, E. Anaya, Q. Tao, S. Cho, H. Zheng, and F. Zhang. "Monocular vision-based human following on miniature robotic blimp." In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 2017, pp. 3244–3249. DOI: 10.1109/ICRA.2017.7989369.

[130] Jian Zhao, Lin Xiong, Jianshu Li, Junliang Xing, Shuicheng Yan, and Jiashi Feng. "3D-Aided Dual-Agent GANs for Unconstrained Face Recognition." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.10 (2019), pp. 2380–2394. DOI: 10.1109/TPAMI.2018.2858819.

[131] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. "Face Recognition: A Literature Survey." In: *ACM Comput. Surv.* 35.4 (2003). ISSN: 0360-0300. DOI: 10.1145/954339.954342. URL: https://doi.org/10.1145/954339.954342.

[132] Yong Zhu, Tieniu Tan, and Yunhong Wang. "Biometric personal identification based on handwriting." In: *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*. Vol. 2. 2000, 797–800 vol.2. DOI: 10.1109/ICPR.2000.906196.