UNIVERSIDADE DA BEIRA INTERIOR
Faculdade de Engenharia

Biometric Recognition Systems

Course Report

Hugo Pedro Martins Carriço Proença

In the scope of the application for the degree of *Agregado*, submitted at the University of Beira Interior, Covilhã, Portugal, in accordance to the Decree-Law n. 239/2007.

May, 2015

Preliminary Remark

This report was prepared in the scope of the application for the degree of *Agregado*, submitted at the University of Beira Interior, Covilhã, Portugal, in accordance to the Decree-Law n. 239/2007.

The ***Biometric Recognition Systems*** curricular unit is described, which was designed as a **new unit** to be included in the curricular structure of the 2^{nd} cycle of studies in Informatics Engineering of the University of Beira Interior.

Contents

1 Introduction **9**
- 1.1 Goals and Syllabus . . . 9
- 1.2 Biometrics Foundations and Emerging Relevance . . . 10

2 Course Context **13**
- 2.1 Context . . . 13
- 2.2 University of Beira Interior . . . 13
- 2.3 Masters in Informatics Engineering . . . 14
- 2.4 Similar Courses . . . 14
  - 2.4.1 National Level . . . 14
  - 2.4.2 International Level . . . 16
  - 2.4.3 Research Groups and Relevant Conferences / Journals . . . 19

3 Course Description **23**
- 3.1 Intended Recipients and Learning Goals . . . 23
  - 3.1.1 General Goals . . . 23
  - 3.1.2 Specific Goals . . . 24
- 3.2 Prerequisites and Corequisites . . . 24
- 3.3 Detailed Program . . . 26
  - 3.3.1 Calendar . . . 27
- 3.4 Teaching / Learning Activities . . . 28
  - 3.4.1 Backward Design . . . 29
  - 3.4.2 Theoretical Classes . . . 30
  - 3.4.3 Practical Classes . . . 31
  - 3.4.4 Classes Scheduling . . . 32
  - 3.4.5 Evaluation / Grading System . . . 33
- 3.5 Feedback Plan . . . 35
  - 3.5.1 Educative Assessment . . . 35
  - 3.5.2 Auditive Assessment . . . 35
- 3.6 Course Bibliography and Information Sources . . . 36
  - 3.6.1 Textbooks . . . 36
  - 3.6.2 Research Papers . . . 36
  - 3.6.3 Software . . . 37
  - 3.6.4 Web Sources . . . 37

4 Conclusion **39**

A Support Materials **41**
A.1 Announcements 41
A.2 Theoretical Lectures 41
A.3 Laboratorial Sheets 250
A.3.1 Additional Exercises 268
A.3.2 Hands-on Assignment 271
A.4 Exams 274

Document Structure and Notation

This document is organised as follows: Chapter 1 gives a summary of the proposed course, along with the main rationale behind it. Chapter 2 gives a global perspective of the course and its context, in terms of the institution and of the cycle of studies that hosts it, along with a description of the most relevant similar courses found at the national and international levels. Chapter 3 gives a detailed description of the course, its intended recipients, general and specific goals, plus the detailed program and classes schedule. Finally, Chapter 4 concludes this document, giving some final remarks and pointing for further directions for the course.

For comprehensibility, the **particularly important** parts of the document include **Summary Tables**, that give a cohesive / condensed perspective of the information in the corresponding part:

Key statement 1

Key statement 2

Regarding the notation used in this document, the different types of elements are denoted by a unique key, in order to simplify their reference in other parts of the document. The protocol for naming such keys is given in Table 1: goals for the course are identified by a ”GG" followed by a sequential number (#SEQ). Similarly, a specific goal is identified by "SG" and a sequential number, while topics and sub-topics of the program are respectively identified by "PT" and "ST", followed by sequential numbers that turn them unique.

Table 1: Protocol used for naming the key terms used in this document.

Key	Description
GG_#SEQ	General goal of the course
SG_#SEQ	Specific goal of the course
PT_#SEQ	Topic of the course program
ST_#SEQ	Sub-topic of the course program

Chapter 1

Introduction

1.1 Goals and Syllabus

Only a few years ago, biometric recognition was considered a *shallow topic*, with applications restricted to criminal forensics, and the concern of a relatively small number of researchers. Recently, growing attention has been paid to the concepts of security and safety in modern societies, which increased the interest about biometrics and promoted its mainstream use for computer authentication, access control, document security and surveillance. A large number of novel research units was created, while many others adjusted their focus to the development of innovative solutions for biometric recognition. This was accompanied by new exciting scientific events, where the advances in the biometrics state-of-the-art have been reported, both in terms of the traits possible to use (e.g., fingerprints, face, iris, voice, retina and other novel biometrics), and of the algorithms claimed to be most suitable for each phase of the recognition processing chain.

In the scope of biometrics research, a tricky concept is the recognition **robustness**, which essentially refers to the constraints demanded in data acquisition to obtain data with enough quality. In this context, the ultimate goal is the development of fully **non-cooperative recognition systems**, which will turn the identification process completely transparent to users. This is exactly the main motivation behind the research carried out by the author of this document, who has been focusing his efforts in this topic over the last decade.

Biometrics is now accepted as one of the most vibrant topics inside the artificial intelligence, pattern recognition and computer vision domains, and is the subject of many publications and funded projects. Private and governmental agencies are investing large amounts of money in the development of less intrusive biometric recognition technologies, which are about to be deployed. Moreover, according to some of the most prominent researchers in this area, the emerging concept of **large-scale identification** will turn biometrics into one of the ***grand-challenges*** faced by the scientific community, as no automaton had ever interacted with as many people as biometric systems are likely to interact in the next few years.

Despite the relevance of biometrics, **there is not a single course about this topic lectured in any Portuguese university either at the undergraduate or master levels**, which is in evident disparity to the international scenario and was one of the **justifications** for the proposed course. Two other complementary factors were the requirement of **discriminability** for the courses lectured at each institution **with respect to competitors**, which is particularly important for small institutions located in the interior part of the country. Also, the fact that the Department of Computer Science of the University of Beira Interior (UBI) is well known by the international biometrics research community, also increases the importance of offering a specialised formation on biometrics.

According to the above motivations, this document describes the **Biometric Recognition Systems** course, designed to be included in the curricular structure of the 2^{nd} level of studies in Informatics Engineering at the University of Beira Interior.

In short, the course will provide a global perspective about the **biometrics research and technology**, highlighting its most relevant advantages over the traditional recognition solutions (e.g., passwords, smart cards or pins) in terms of effectiveness, security, user convenience and burden. During the semester, students should gain knowledge about the multi-disciplinary features of biometrics, in terms of image and signal processing, pattern recognition, computer vision, artificial intelligence, optics, information security and systems engineering. By the end of the course, students should be able to:

1. Obtain a **cohesive perspective of biometrics**: its history, current state and expected further directions;
2. Understand the **basic modules** of the biometrics **processing chain**, starting from data acquisition up to decision;
3. Gain knowledge about the **state-of-the-art algorithms** for each phase / trait of the biometrics recognition process;
4. **Compare objectively** the different **traits / technologies** available, and select the **most convenient** for a particular scenario;
5. **Design and evaluate computational systems** that include biometrics;
6. Faithfully address the **ethical / legal issues** behind the development of a biometric recognition system;
7. Objectively evaluate the **state-of-the-art** in biometrics research and point for possible improvements.

The key information about the **Biometric Recognition Systems** course is as follows:

Institution	University of Beira Interior
Organic Unit	Faculty of Engineering
Department	Informatics
Subject Areas	Engineering, Computer Science
Mode of Study	Full-time
Attendance Mode	Campus
Duration	One semester
Schedule	Theoretical classes 2h / week, Laboratorial classes 2h / week. Total 15: weeks.
Total ECTS	6

1.2 Biometrics Foundations and Emerging Relevance

The term *Biometrics* has more than one meaning: even though the *automatic recognition of individuals based on their biological / behavioural traits* is now the most popular in the literature, the first explicit reference to biometrics dates back to 1901, when Francis Galton edited the first issue of the journal *Biometrika*, as an attempt to describe the *novel contributions in terms of the application to biology of the modern methods of statistics*, with applications in biology, agriculture, medicine, public health, demography, actuarial science, and related fields. Rigorously, the correct way to discriminate both meanings of *Biometrics* and to actually refer to the automatic identification of human beings should be "biometric recognition", which can even be regarded as a particular case of the more broad concept of *Biometrics*, but focused in human recognition.

Even implicitly, the first references to biometrics date back to 3,000 years ago, with the bible itself containing several statements related to biometric recognition. As an example, in Judges 12:5-6: *"Then*

said the men of Gilead unto him, Say now Shibboleth: and he said Sibboleth: for he could not frame to pronounce it right. Then they took him, and slew him at the passages of the Jordan: and there fell at that time of the Ephraimites forty and two thousand", describes a particular case of a phenotypic biometric (voice) used to identify Ephraimites, the enemies of the Gileadites.

The pioneering semi-automated methods for human recognition were based on speech and appeared in the 1940s, while the first fully automated techniques were based on fingerprints, handwriting signature and face, having emerged in the 1960s, supported by breakthrough advances in hardware technologies [30]. During the 1970s, the first generation of biometric recognition systems was deployed commercially and became extremely popular during the 1980s, particularly for banking and government applications. Since then, new biometric traits have been proposed (e.g., the iris, retina or the palm veins structure) and the current challenges are now related with the systems' massification and the deployment at worldwide scales.

Fig. 1.1 (taken from[1]) illustrates the joint civil and military revenue expected for the biometrics market, in which the current state is regarded as the *tip of the iceberg* [34]. This is mostly due to the deployment of biometric recognition systems at extremely large scale scenarios (nations with over 1,000 millions inhabitants) and due to the incorporation of biometric technologies in devices that are massively used, such as smart phones or cars. Hence, *the future of biometric applications looks extremely bright, especially with the growing interest in smart and safe cities, that Frost & Sullivan has recognised as key areas of technology innovation and growth. As such concepts are becoming international trends, biometrics should undoubtedly become a part of that holistic solution: digitalisation and the evolution of fast data transfer wireless internet should enable the interconnection of devices (surveillance cameras, sensors, etc.) and systems (video management system, access control system, etc.) which should be scattered around the city* [34].

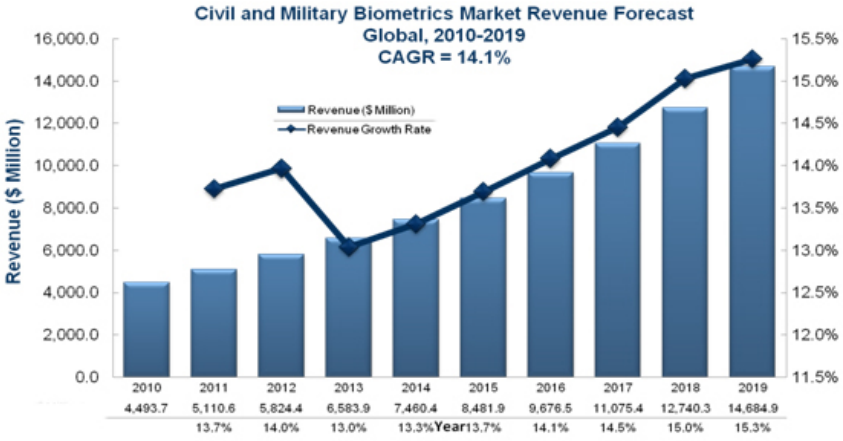


Figure 1.1: Expected joint civil and military revenue for the biometrics market. The horizontal axis is the time line, and the vertical axes gives (at left) the total revenue and (at right) the growth rate.

Biometrics is among the markets with the largest expected increases in global revenue for the next years;

The Department of Informatics of the UBI is well known by the international biometrics community, and has the scientific / academic surrounding propitious to teach a strong course on biometrics;

Portuguese universities are expected to promote differentiation, which is particularly important for institutions in the interior part of the country, with evident difficulties in attracting students from the more populated (coast) regions.

[1] http://www.frost.com/prod/servlet/market-insight-print.pag?docid=260998675

Chapter 2

Course Context

2.1 Context

The proposed course on **Biometric Recognition Systems** should be integrated in the 2^{nd} cycle of studies leading to the *Master* degree in Informatics Engineering, which aims to form engineers with a solid knowledge and the appropriate skills for the current labor market, as well to provide a basis for the frequency of the third cycles of studies (Ph.D.) in Informatics Engineering.

The contents and learning plans were designed to be faithfully integrated into the existing curricular structure of the cycle of studies, according to a multidisciplinary perspective. Also, the course is expected to profit from the research laboratories that have been active in biometrics research. Particularly, the interaction between the students and researchers the *SOCIA: Soft Computing and Image Analysis Lab.* and in the *Pia-Cv: Pattern and Image Analysis* lab. from the *IT: Instituto de Telecomunicações* is encouraged.

2.2 University of Beira Interior

The genesis of the University of Beira Interior (UBI) was in 1973, when the *Instituto Politécnico da Covilhã* (IPC) was formally created. In 1979, the institution became the *Instituto Universitário da Beira Interior* (IUBI) and was finally converted into UBI in 1986. Currently, the UBI hosts over 6,000 students, distributed over five organic unities: Arts and Humanities, Sciences, Health Sciences, Economical and Social Sciences and Engineering. The university offers a broad range of cycles of studies, from the first (B.Sc. and B.Eng.), second (M.Sc. and M.Eng.) to the third cycle (Ph.D.) of studies.

In summary, the main mission of UBI is to promote the high level qualification, the production, transmission, criticism and diffusion of knowledge, culture, science and technology, by means of study, teaching and research activities [41]. UBI is currently known by its high quality levels in terms of learning, research and knowledge transfer activities, assuming a strong partnership with its surroundings. In terms of major goals, UBI aims to be known as an institution that 1) promotes the quality of its graduated, masters and doctors; 2) is open to the surrounding society, both at regional, national and international levels; 3) promotes the social and economical development of the *Beira Interior* region and of the country itself; and 4) is able to promote technological-based start-up companies.

During its relatively short history, the UBI has been promoting human and academic principles that not only define its identity but also dictate its collective efficiency: a) **Intellectual liberty**, by promoting creativity and innovation; b) **Academic integrity**, by assuring that learning / research activities are intellectual and morally independent; c) **Diversity**, by promoting a global conscientiousness about tolerance, mutual respect and diversity; d) **Excellence**, by meeting the most demanding standards for learning / research; e) **Social responsibility**, by taking into account the social, cultural and ambient needs in its surroundings; f) **Learning for life**, by promoting the knowledge search as a way to improve the quality

of populations' life; and g) **Rationality**, by assuring that an appropriate balance between the goals and the available resources is found.

2.3 Masters in Informatics Engineering

The second cycle of studies leading to the *Master* degree in Informatics Engineering aims to form engineers with a solid knowledge and the appropriate skills for the current labor market, as well to provide them with a basis for the frequency of the third cycle of studies (Ph.D.) in Informatics Engineering or related.

All the courses in this cycle of studies were designed to articulate teaching, learning and research, being encouraged the integration and contact of the students with the research groups in the Department of Informatics. At the ultimate level, the main objective of this cycle of studies is **to develop the critical sense of students and their skills in terms of problem solving**. The idea is that any recently graduated Master is able to perform engineering acts in computer and informatics engineering, according to the premium standards of the professional boards for engineers at the international level. In result of the most recent definition of its structure, the cycle of studies is divided into two years / four semesters, requiring students to complete 120 ECTS, in accordance to the main principles of the Bologna process.

The curricular structure is divided into two academic years: during the first year students have to approve on ten curricular units (60 ECTS), of which five are mandatory and the remaining five are chosen by students, allowing them to choose their own academic path, according to the Bologna spirit. Every curricular unit corresponds to six ECTS, which facilitates the comparability and transparency within the European area of higher education, and the introduction of the **Diploma Supplement** provided to students upon the degree conclusion. Finally, the second year is fully devoted to the dissertation, which exclusivity augments the responsibilities of students / advisors to produce high quality work and tends to orient the dissertations for scientific research, usually carried out in one of the surrounding research laboratories.

2.4 Similar Courses

As stated above, courses covering biometrics topics are still hard to find in Portuguese universities. The exceptions are mainly at the third cycle of studies, integrated in Doctoral programs from the electrical, biomedical engineering and computer science domains. For most of the existing courses, the goals can be regarded as simple introductions to the fundamental concepts associated with biometrics, focusing much more in the analysis of the state-of-the-art (by analysing and summarising research papers) than the *learn how to do it* paradigm.

In opposition, the number of courses and - even - complete cycles of studies available at the undergraduate and postgraduate levels in international universities is quite large, which also points for the relevance that this topic should acquire soon also in the national context.

This section summarises and compares the contents and main focus of the courses and cycles of studies available, both at the national and international levels. The descriptions below are exclusively on the basis of the information publicly available in the web. We start by describing the existing options in terms of national universities, followed by the analysis of the international context.

2.4.1 National Level

Even though the offer in Portuguese institutions about biometrics technology is short, there are some courses that can be understood as *adjacent to biometrics*, i.e., they cover topics that are highly similar to the used in biometrics recognition. In this family, the *Digital Image Processing* course at the Faculty of Sciences from the University of Lisbon, and the *Image Analysis and Processing* course at the Department

of Physics from the University of Coimbra can be referred. However, this section exclusively compares the courses or learning programs that contain explicit references to biometrics recognition, which are summarized in Table 2.1. The most obvious conclusion is that the offer is concentrated in the 3^{rd} cycles of studies, with a unique exception consisting of a complete cycle of studies that focuses biometrics but also has more broad purposes, not specifically focusing the recognition process.

Table 2.1: Summary of the courses / learning programs explicitly related with biometric recognition that are available at the national level.

Title	**Institution**	**Scientific Area(s)**	**Head Lecturer**
Visual Surveillance and Biometrics [42]	University of Coimbra (UC)	3^{rd} cycle in Electrical and Computer Engineering	Jorge Batista
Master in Biostatistics and Biometry [39]	Open University (Universidade Aberta)	2^{nd} cycle	-
Image Analysis and Processing in Engineering [18]	Faculty of Engineering, University of Oporto (FEUP)	3^{rd} cycle in Mechanical Engineering	João Tavares
Biometric Systems [23]	Technical Superior Institute (IST)	3^{rd} cycle in Electrical and Computer Engineering, 3^{rd} cycle in Information Security	Ana Fred
Modelling and Classification of Biomedical Signals [24]	Technical Superior Institute (IST)	3^{rd} cycle in Biomedical Engineering	Ana Fred

The *Visual Surveillance and Biometrics* course from the University of Coimbra is perhaps the closest (in terms of content) to the course described in this document, exploring the underlying technologies and the potential of biometrics / visual surveillance for human identification tasks. This course also covers topics about human action and activity recognition, being organised in two different parts: 1) at first, regular lectures are given in an expositive way; and 2) the second part consists of reading and seminar presentations given by students, as well in the analysis of progress reports on students' projects. Each student is expected to read a set of research papers, and the corresponding discussion about each paper is opened to the whole class. The evaluation is carried out according to class participation (20%), paper presentation (30%) and - predominantly - the final project evaluation (50%).

The *Image Analysis and Processing in Engineering* course from the Faculty of Engineering of the University of Porto has a scope evidently broader than biometrics. The course structure and the corresponding topics covered turn it more focused in medical image analysis, modulation and diagnosis than in biometric recognition itself. Even though, the course explicitly states *biometry* and *surveillance* as topics covered during the semester, being organised in terms of tutorials given by the instructor. Near the end of the course, students are expected to present their work, along with a computational prototype, being the final mark given according to the computational work (65%), written report (25%) and public presentation (10%).

The *Biometric Systems* course from the Technical Superior Institute (IST) of the University of Lisbon is given at the 3^{rd} level of studies for two different Doctoral programs: Information Security and Electrical Engineering and Computers. The course is based on case studies and on two types of projects, according to students choices: either book-based or data centric projects. The evaluation is carried out based on the grades from the case study (30%) and -mainly - from the project (70%). Still at the IST, the course *Modelling and Classification of Biomedical Signals* is taught by the same instructor and shares a much similar structure and contents to the former course, but in this case with emphasis putted on medical diagnosis and decision.

From our viewpoint, the absence of biometric recognition courses at more preliminary levels (2^{nd} and even 1^{st} cycles of studies) generates an opportunity for differentiation: we argue that such a course will be of interest for a significant number of prospective students, due to the increasing requirements

from the industrials / commercial contexts and the expected augmenting impact of the *biometrics* topic.

The absence of biometric recognition courses at 1^{st} and 2^{nd} cycles of studies in Portuguese universities generates an opportunity of differentiation for the UBI.

2.4.2 International Level

Undoubtedly, the number and variety of courses and learning programs about biometric recognition taught in foreign universities has been increasing remarkably. Table 2.2 summarises these courses, in a way similar to the used for the description at national level. Again, note that the list is not exhaustive, and the main criteria for selection were the scientific relevance of the head lecturer, along with the reputability of the university / research institute.

Three complete cycles of studies focused exclusively on biometrics were found, all of them corresponding to 2^{nd} cycles of studies. The cycle *Information Security and Biometrics* taught at the University of Kent [49] aims to providing a thorough understanding of the theories, concepts and techniques for the design, development and effective use of secure information system. Most courses on this cycle are actually focused on information security (e.g., trust, security and privacy management, or system security), but the program also contains courses exclusively centered on biometrics (e.g., industrial context of biometrics). The cycle of studies at La Trobe University mixes the biometrics an bio-computing domains, offering courses specialised on biometric recognition (e.g., biometric security), but also several courses more focused in the processing of biological signals (e.g., techniques in biotechnology and bioinformatics technologies). Anna University offers a M.Eng. on *Biometrics and Cyber security* [2] cycle of studies, with courses spreading through four semesters and covering fundamental concepts such as probability, statistics, computer networks, and also more specific topics about biometrics, image processing, ethics, and digital forensics. This cycle contains a large set of elective courses, enabling students to define their own trajectory according to their desired skills.

In terms of individual courses, the *Introduction to Biometrics* is lectured at the University of Twente [44] in an open way, i.e., to external students from 6 different participating programs of the 2^{nd} level of studies (from the computer science, engineering and mathematics domains). The course has only two contact hours per week and mainly covers the scientific issues of biometrics: its background theory, the available traits, data covariates, and performance evaluation measures. Also, instructors stress the goal of providing sufficient background knowledge to read and understand scientific publications on this topic.

The majority of the available courses is lectured a the U.S.A., probably due to the concerns about security / safety that are particularly evident there. A course named as *Biometrics* is lectured for the undergraduate level at the University of California [47], which shares evident similarities with the structure of the course proposed in this document. They start by providing the background in machine learning and imaging, regarding the processing chain in biometrics. Next, the second half of course details individual biometrics, methods for spoofing biometrics, and policy implications. In a project spaced over the course, students implement a biometric recognition method, system, or application of their choice and appropriate to their background. Another university with a strong tradition in biometrics research is undoubtedly the West Virginia University, that offers an undergraduate course on *Introduction to Biometrics Systems* [54], containing chapters about the classical traits: face, iris and fingerprint, but also about hand geometry and multimodal biometrics. The chapter about performance evaluation is given, in a rather unusual way, at the middle of the semester, between the "hand" and "face" chapters. Similarly, the University of Louisville offers the undergraduate course *Introduction to Biometrics* [50] that is almost exclusively centered on face recognition, starting by an introduction to statistical pattern recognition, followed by skin and face detection, face alignment and recognition chapters. Michigan State University is among the institutions with a strongest focus on biometrics, due to the research group leaded by A. K. Jain. It offers two courses for the 2^{nd} and 3^{rd} levels [28]. They address biometric systems based

Table 2.2: Summary of the courses / learning programs about biometric recognition that are available at the international level.

Title	**Institution**	**Scientific Area(s)**	**Head Lecturer**
Introduction to Biometrics [44]	University of Twente	2^{nd} cycle in Mechanical, Electrical and Biomedical Engineering, Computer Science, Applied Mathematics and Security	Berk Gokberk
Biometrics [47]	University of California, San Diego	1^{st} cycle in Computer Science	David Kriegman
Introduction to Biometrics Systems [54]	West Virginia University	1^{st} cycle in Computer Science	Natalia A. Schmid
Introduction to Biometrics [50]	University of Louisville	1^{st} cycle in Computer Science	Aly A. Farag
Biometric Technologies [46]	University of Calgary	1^{st} cycle in Computer Science	Marina Gavrilova
Biometrics Institute Intensive Course [6]	The Biometrics Institute	-	James L. Wayman
Biometric technologies - Certificate IV [10]	Canberra Institute of Technology	Post-graduation	-
Biometric Systems [45]	University of Applied Sciences Salzburg	2^{nd} cycle in Computer Science	-
Information Security and Biometrics [49]	University of Kent	2^{nd} cycle	{Complete Cycle}
Biometrics and Cybersecurity [48]	University of Hertfordshire	2^{nd} cycle in Computer Science	Andrey Mokhov
Biometrics [22]	Indiana-Purdue University	1^{st} cycle in Computer and Information Science	-
Biocomputing and Biometric Security [25]	La Trobe University	2^{nd} cycle	{Complete Cycle}
Biometrics [9]	Brooklyn Polytechnic University	1^{st} cycle in Computer Science	Nalini Ratha
Biometrics Systems [36]	Denmark Technical University	2^{nd} cycle in Computer Science	Cristoph Bunch
Biometrics [32]	Northwestern University	2^{nd} cycle in Electrical Engineering	Xin Chen
Biometrics, Advanced Biometrics [28]	Michigan State University	2^{nd} and 3^{rd} cycles in Computer Science and Engineering	Arun Ross
Biometrics [21]	Indian Institute of Technology Delhi	1^{st} cycle in Electrical Engineering	Ajay Kumar
Biometrics and Image Analysis [40]	University at Buffalo	1^{st} cycle in Electrical Engineering	Venu Govindaraju
Biometrics [38]	The University of Winnipeg	1^{st} cycle in Electrical Engineering	Sergio Camorlinga
Biometric Systems [43]	Universitá of Siena	3^{rd} cycle in Information Engineering	Fabio Scotti
Biometrics [14]	Columbia University	2^{nd} cycle in Computer Science	Peter Belhumeur
Biometrics and Cyber Security [12]	Anna University	2^{nd} cycle	{Complete Cycle}
Biometrics [51]	University of Southampton	2^{rd} cycle in Electrical Engineering	Mark Nixon

on fingerprints, voice, face, hand geometry, palm print, iris, retina, and other emerging modalities. Multimodal biometric systems deserve a chapter, as do performance evaluation measures and issues related to the security and privacy aspects. The *Biometrics* course at the Indiana-Purdue university [22] offers a graduate certificate on biometrics, composed by two core courses (Statistics and Pattern Recognition) and a specialisation course on biometrics computing, with strong focus on the core of technology and on the traits that are typically used in recognition. Columbia university offers a *Biometrics* course [14], integrated at the 2^{nd} cycle of studies, with focus on two biometric traits (iris and face) and the algo-

rithms typically used in the processing chain of both kinds of systems. Fingerprint recognition is also superficially studied.

Still at the undergraduate level, the New York Polytechnic School of Engineering (at Abu Dhabi and Shangai) offers a *Biometrics* [9] course for Computer Science students that concentrates on the advantages that biometrics brings to computer security. The course also analyses the recognition rates of the different traits and the corresponding privacy concerns, compared to the alternative use of passwords and smart cards.

In India, there are several technical institutes and universities with courses on biometrics, among which the *Biometrics* at the Indian Institute of Technology [21] is highlighted. A well known researcher in the biometrics field (Ajay Kumar) has set up a course with a highly typical structure, divided into chapters exclusively devoted to the classical traits (fingerprint, iris and face), and other emerging biometrics (such as pal print). This course also contains an introductory chapter about statistical pattern recognition, and biometrics basis, and concludes with a multimodal biometrics chapter.

In Canada, the University of Winnipeg offers a *Biometrics* course [38] which programs starts in a rather unusual way, by introducing the performance evaluation measures and only then the fundamental concepts behind fingerprint, face and iris recognition. The last part of the course is about multibiometrics and novel traits (e.g., ear, gait or hand).

The quantity of courses about biometrics lectured in Europe in much smaller than in the U.S.A.. Here, the University of Calgary is one of the few that lectures a course at the undergraduate level, named *Biometric Technologies* [46], aiming to study biometrics algorithms that touch the applied sciences and computer graphics domains, with explicit references to visualisation and surface modelling technologies. In Austria, the University of Applied Sciences of Salzburg offers the *Biometric Systems* [45] course, which distinguishes from others mainly by regarding biometrics as simple applications of image and signal processing, that shares the main feature of seeking the systems optimization with respect to risk minimization. At the Denmark Technical University, the *Biometrics Systems* course [36] also starts by the description of the goals in biometrics and its performance evaluation measures, followed by chapters devoted to fingerprint, face and vein recognition, together with data security and privacy issues. This course contains a research seminar, where students are asked to address selected topics from state-of-the-art biometric research. The *Biometrics* course at the University of Southampton [51] overviews the fundamental concepts on computer vision and image analysis behind a biometric recognition system, followed by chapters devoted to gait (the main research topic of the instructor) and iris. This is also one of the few courses that includes a demonstration module. Finally, the course lectured at the *Universitá of Siena*, named *Biometric Systems* [43] also covers the basics of biometrics and chapters about face, fingerprint and iris traits. Its singularity lies in the fact of having a chapter exclusively about DNA recognition, along with a discussion about security / privacy issues and performance evaluation measures. Lastly, the *Biometrics and Cybersecurity* post-graduated course at the University of Hertfordshire [48] has a focus on biometric systems, signal processing and speech processing. The Cybersecurity component covers the digital rights management issue and the security of distributed systems.

To conclude, the *Biometrics Institute Intensive Course* at the Biometrics Institute [6] and the *Biometric technologies - Certificate IV* taught at the Canberra Institute of Technology [10] have an intensive nature, covering superficially most of the above listed topics: both courses summarise the available traits to perform biometric recognition, summarise the current international standards and conclude by illustrating practical cases studies of successful biometric recognition systems. Due to the short period for the courses, it is impossible that students *actually* design and implement a biometric recognition system, which is not even a definitive goal, due to the nature of most of the candidate for both courses (lack of prerequisites).

In short, when comparing the courses available at the national and international levels, the most obvious conclusion is the disparity in terms of the levels of studies where biometric recognition systems are taught: national programs are predominantly in the third level of studies, in opposition to the courses

taught at international level, that not only introduce biometrics topics at the undergraduate level, but also devote complete cycles of studies to the issue of biometric recognition. This obviously points for a larger relevance of biometrics in the international than in the national domain, which was one of the major reasons behind the proposal of the curricular unit described in this document.

2.4.3 Research Groups and Relevant Conferences / Journals

There is a growing number of active research groups concerned about the biometrics technology. In recent years, not only several groups fully devoted to biometric recognition were created (e.g., the Spanish Biometric Recognition Group - AVTS [3]), but also other groups redefined their strategies, having pointed a strong focus in biometrics research (e.g., the IDIAP research institute [20], in Switzerland). In result of this growing number of research groups, many grants and funded projects have naturally appeared, with a significant number of relevant publications that conducted to actual advances in the state-of-the-art. Table 2.3 gives a short list of the most relevant research groups with interests on biometrics, particularly those that over the last few years have published most of the relevant works. Note that this list is far from exhaustive, as many other groups could have been also included here.

Table 2.3: Summary of some of the most active research groups in the biometrics domain of knowledge.

Group	**Institution**	**Head Lecturer**
Forensics and Biometrics Group [37]	University of Auckland (New Zealand)	Catherine Watson
Biometrics [48]	University of Surrey (United Kingdom)	Joseph Kittler
Biometric Recognition Group - AVTS [3]	Universidad Autonoma de Madrid (Spain)	Javier Ortega-Garcia and Joaquin Gonzalez-Rodriguez
Center for Biometric Research [11]	Chinese Academy of Sciences (China)	Tieniu Tan
IDIAP Research Institute [20]	Ecole Polytechnique Fédérale de Lausanne (Switzerland)	Sebastian Marcel
DISI - Department of Computer Science and Engineering [17]	University of Bologna (Italy)	Dario Maio and David Maltoni
National Biometric Test Center [29]	San Jose State University (U.S.A.)	Jim Wayman
CyLab. Biometrics Center [15]	Carnegie-Mellon University (U.S.A.)	Marios Savvides
Biometric Research Lab. [5]	Indian Institute of Technology Kanpur (India)	Phalguni Gupta
Biometrics Research Group [4]	Sabanci University (Turkey)	Berrin Yanikoglu
I^2SIS, Institute for the Interdisciplinary Studies in Identity Science [19]	University of North-Carolina (U.S.A.)	Karl Ricanek
da/sec - Biometrics and Internet-Security Research Group [16]	Hochschule Darmstadt University of Applied Sciences (Germany)	Harald Baier and Christoph Busch
Biosecure [8]	Paris Sud University (France)	Bernardette Dorizzi
Notre Dame ReSource: Biometrics Research [33]	University of Notre-Dame (U.S.A.)	Kevin Bowyer
Maryland Cybersecurity Center [27]	University of Maryland (U.S.A.)	Jonathan Katz
Biometrics Research Group [7]	Michigan State University (U.S.A.)	Anil K. Jain
SOCIA: Soft Computing and Image Lab.[35]	University of Beira Interior	Luís A. Alexandre

In Table 2.4 we summarise the international conferences that were found the most relevant in reporting biometrics research. The criteria used for this selection are quite subjective, and not necessarily measured in terms of the conference impact factor. Instead, we privileged events **fully devoted to biometrics**, where most of the world leading researchers on biometrics frequently give keynote speeches, while **keeping a low acceptance rate** that *a priori* guarantees the quality of the selected works. In

parallel, we also give a short list of the international journals where the most relevant algorithms on biometric recognition (e.g., measured in terms of the number of citations) are typically published.

Among the existing events, the *ICB: International Conference on Biometrics* and the *BTAS: Biometrics Technologies Applications and Systems Conference* should be highlighted. Both conferences are currently, along with *CVPR: the International Conference on Computer Vision and Pattern Recognition*, the major events where the most significant advances in the state-of-the-art have been presented.

CVPR is known as a premium conference on Computer Vision field, and has in the last editions continuously organised the *CVPRW: Computer Vision and Pattern recognition Workshop on Biometrics*, which was observed to be a success (due to the number of submissions) and one of the workshops with the lowest acceptance rates. In fact, it is interesting to observe that a large proportion of the works selected for inclusion in CVPRW have been invariable extended and published in most of the premium journals in adjacent fields (e.g., IEEE-TPAMI or IEEE-TIFS).

The ICB was established in 2006 is a biennial conference that merged AVBPA (*Audio- and Video-based Person Authentication*), ICBA (*International Conference on Biometric Authentication*) and other minor biometrics workshops. The conference has broad scope and its proceedings include papers that advance biometric technologies, sensor design, feature extraction and matching algorithms, analysis of security and privacy, and evaluation of social impact of biometrics technology. Topics of interest include all areas of biometrics research and applications.

The BTAS conference is relatively recent (started in 2007), and every year runs in Washington D.C. (U.S.A.), having rapidly evolved to the premium conference on this area. It usually invites outstanding speakers for its keynote sessions and is perhaps the event where the most vibrating discussion about biometrics take place. Every 4 years, ICB and BTAS join and form the *IJCB: International Joint Conference on Biometrics*.

In Table 2.4 we summarize the most important events and the where research about biometrics has been published. Note that the actual number of sources is much larger, which is particularly evident in international conferences that are regularly organised in China / India, fully devoted to biometric recognition, but where the relevance / quality of most of the works published there decreases significantly the significance of such events.

There is a reduced offer at the national level, which is mainly concentrated in 3^{rd} cycles of studies

At the international level, biometrics courses are rapidly spreading and appear in several of the top universities and research centers

Biometrics is one of the few topics of knowledge with fully devoted conferences and journals, published by well known entities (mostly IEEE and Elsevier)

Table 2.4: Summary of the most important conferences and journals that have been publishing the state-on-the-art research on biometrics.

Conference / Journal	**Publisher**
IEEE Transactions on Pattern Analysis and Machine Intelligence	IEEE
IEEE Transactions on Image Processing	IEEE
IEEE Transactions on Information Forensics and Security	IEEE
Computer Vision and Image Understanding	Elsevier
Pattern Recognition	Elsevier
Image and Vision Computing	Elsevier
Pattern Recognition Letters	Elsevier
IET Computer Vision	IET
IET Biometrics	IET
Biometrics Theory, Applications and Systems Conference (BTAS)	IEEE
International Conference on Biometrics (ICB)	IEEE
ICB + BTAS = IAPR / IEEE International Joint Conference on Biometrics (IJCB)	IEEE
International Conference on Computer Vision and Pattern Recognition (CVPR)	IEEE
International Conference on Pattern Recognition (ICPR)	IEEE
International Conference on Image Processing (ICIP)	IEEE
International Conference on Computer Vision (ICCV)	IEEE

Chapter 3

Course Description

3.1 Intended Recipients and Learning Goals

The **Biometric Recognition Systems** course was not designed under a content-centered perspective, i.e., that covers topics X, Y, or Z. Instead, we began by stating two fundamental questions: *"what do we want to assure that students really know, upon completing the course?"*, or even more important *"what do we want that students keep, several years after the course is over?"*. Next, we also considered *"what are the typical skills of the prospective students?"* and *"what are students' typical gaps in knowledge that will difficult the learning process?"*

According to the questions above, and considering that most of the students (over 90%) should come from the 1^{st} cycle in studies in Informatics Engineering in the UBI, we decided to create a course with a strong focus on the **engineering** behind the development of **biometrics recognition**, leading to the final designation of **Biometric Recognition Systems**. The main idea is that during the course students get an overall picture of biometrics, the available traits, and the typical phases of a biometric recognition system. Next, a strong effort should be putted in perceiving the most typical algorithms in each phase of the biometric recognition chain. Then, and perhaps the most discriminating point with respect to competing courses (mainly at the international level), the focus is putted in systems development and integration: the analysis of requisites, system planning, implementation and deployment of practical biometric recognition systems, according to the standards offered by software engineering, information security and information systems.

Biometrics Recognition Fundamentals + Secure Systems Engineering = **Biometric Recognition Systems**

3.1.1 General Goals

We considered as *general goals* those that must be achieved by students at a level immediately below the *fundamental* goal of the course, which can be informally stated as: **assure that graduated students are able to design from scratch, implement and maintain a biometric recognition system**. In order to accomplish such ambition, four general goals were identified:

ID	Description
GG_01	Broad understanding of how to provide effective biometric identity management systems
GG_02	Obtain the system development skills that allow the adaptation to dynamic and fast-changing industrial environments
GG_03	Gain knowledge about the industrial / commercial context in which biometric recognition systems are integrated
GG_04	Understand the ethical and legal issues, as well the relevant standards in the context of biometric recognition

3.1.2 Specific Goals

Each of the above described general goals can be divided into a group of specific goals, more easily measurable. These goals are given below, together with the general goal that each one regards:

ID	General	Description
SG_01	GG_01	Gain knowledge about the emerging technologies in the biometrics recognition field
SG_02	GG_01	Understanding the mathematical / conceptual models behind the development of a biometric recognition system
SG_03	GG_01	Understand the typical phases of a biometric recognition systems, and how these phases can be regarded from the pattern recognition perspective.
SG_04	GG_01	Gain knowledge about the most typical algorithms used in each phase of the biometric recognition process
SG_05	GG_01	Ability to plan, design and implement modern biometric recognition system
SG_06	GG_01	Ability to compare in an objective way the advantages / disadvantages associated with each biometric trait, in order to fit the specificities of a particular environment
SG_07	GG_02	Ability to decide about the most appropriate hardware / software platforms for the deployment of a biometric recognition system
SG_08	GG_01	Gain knowledge about the widely acceptance performance evaluation measures typically used in biometrics research.
SG_09	GG_03	Ability to design tests for a biometric recognition system, during the deployment / validation phases.
SG_10	GG_04	Meet the standards and legal regulations required in the development of a specific biometric recognition system

3.2 Prerequisites and Corequisites

As can be inferred from the the course program, the idea in **Biometric Recognition Systems** is to integrate knowledge from a broad range of topics, such as the pattern recognition, computer vision, artificial intelligence, software engineering and optics domains. As in most of the courses lectured at the 2^{nd} cycle of studies in Engineering, an appropriate background in mathematics is undoubtedly a plus. Figure 3.1 illustrates the topics of knowledge usually associated with biometrics, which at the end determine the course prerequisites.

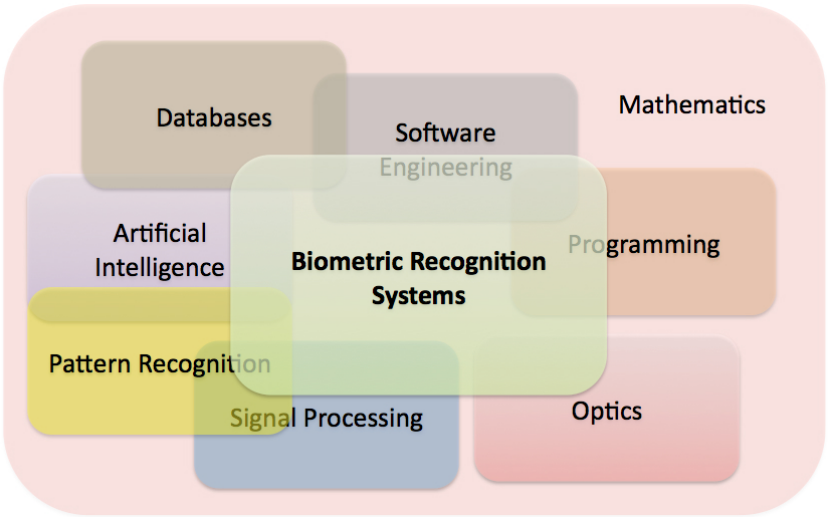


Figure 3.1: Topics of knowledge that are typically associated with the planning and development of a biometric recognition system.

The curricular structure of the 1^{st} cycle of studies in Informatics Engineering at the UBI offers a solid background in programming, data structures, software engineering and information security. Also, it contains courses on artificial intelligence, database systems, linear algebra and statistics. As most of the prospective students for the proposed course should come from the 1^{st} cycle of studies in Informatics Engineering, we assume that such requisites are fulfilled by most of them. However, there are often other candidates that come mostly from the polytechnic institutes in the Beira Interior region (e.g., *Instituto Politécnico da Guarda* and the *Instituto Politécnico de Castelo Branco*), for whose the complete fulfilment of requisites cannot be assured.

The **Biometric Recognition Systems** course was planned to be included in the *second semester* of the **first year**, along with the translation of the Computational Intelligence course into the first semester of the course, turning it a mandatory curricular unit. This creates a corequisite for students at the enrolment period, being desirable that every candidate to the **Biometric Recognition Systems** course also enrolls at the Computational Intelligence course, guaranteeing that students get a deep insight about machine learning statistical and neural-based algorithms, before the beginning of the second semester.

The fulfilment of prerequisites can be easily verified in terms of the skills / competences before the admission to the course. Topics of such skills are enumerated below, along with a small description of the desired competences:

1. Mathematics
   (a) Basic matrices operations,
   (b) Systems of linear equations, determinants;
   (c) Vector spaces: basis, linear combinations and subspaces;
   (d) Statistics: data distributions;
2. Programming and Data Structures
   (a) C, C++ programming;
   (b) MATLAB® fundamentals;
   (c) Lists, trees and graphs, search strategies, computational complexity;
3. Software Engineering
   (a) Functional decomposition, structured programming;
   (b) Agile development methodologies;
4. Pattern Recognition
   (a) Cycle of pattern recognition: acquisition, detection, segmentation, feature extraction and classification;
   (b) Basics of statistical pattern recognition;
5. Artificial Intelligence
   (a) Supervised / non-supervised model learning;
   (b) Neural networks;
   (c) Model selection
6. Information Security
   (a) Network security protocols, loss prevention;

(b) Vulnerability assessment;

7. Database Systems

   (a) Relational model;

   (b) Database Management Systems (DBMS);

8. Signal Processing

   (a) Filters: convolution, correlation;

   (b) Time / frequency domains;

9. Human-Computer Interaction

   (a) Interaction design;

   (b) information visualisation;

From the list of requisites given above, the most evident issue is the absence of previous background about signal processing, i.e., the introductory concepts on signal decomposition / superposition and the fundamental theorem of convolution. In fact, there is a complete absence of any course or module about this kind of topics in the curricular structure of the 1^{st} cycle of studies in Informatics Engineering at the UBI, which we regard as a gap. This problem can be alleviated by the existence of such kind of course (Signal Analysis) in the curricular structure of the 1^{st} cycle of studies in Electrical Engineering at the UBI. Hence, it will be desirable that students assist (at least) to some modules of this course along with the Biometric Recognition Systems classes, as both should be included in the second semester (February to July).

Problem: the engineering of a biometric recognition system requires a relatively large set of previous skills. Some students (particularly those coming from outside the 1^{st} cycle in Informatics Engineering) may not meet the prerequisites.

Solution: coordinate with teachers from B.Sc. / B.Eng. courses taught at the UBI, in order to let students from Biometric recognition systems assist to *key* classes of their courses (e.g., Signal Analysis or Linear Algebra).

3.3 Detailed Program

Below we list the program topics designed for the course (denoted by **PT** plus a sequential number) and the sub-topics that compose it (**ST** plus a sequential number), together with their title. As it will be detailed in the calendar, every program topic will be lectured during 1 to 4 weeks, according to their complexity and the relevance for the course goals. **PT_04** is the longest program topic, as it is evidently the more complex module in terms of algorithms that students need to learn and also the one with the heaviest requirements on programming. Several other program topics are expected to be lectured in a single week (e.g., **PT_01** and **PT_08**).

ID	Topic	Sub-Topic
PT_01	**Biometric recognition: cohesive perspective**	-
	ST_01	Definitions
	ST_02	Goals
	ST_03	Applications
	ST_04	Historical Overview
PT_02	**Biometric recognition essentials**	-
	ST_05	Modes of Functioning
	ST_06	Systems Classification
	ST_07	Biometric Traits
	ST_08	Multibiometrics
PT_03	**Biometrics from the pattern recognition perspective**	-
	ST_09	Typical processing chain
	ST_10	Data covariates
	ST_11	Challenges
PT_04	**Recognition methods and techniques**	-
	ST_12	Signal Processing Basics
	ST_13	Detection
	ST_14	Segmentation
	ST_15	Feature extraction
	ST_16	Feature normalization
	ST_17	Classification
	ST_18	Fusion
PT_05	**Ethical issues, legal regulations and standards**	-
	ST_19	Ethics considerations
	ST_20	National / international legal regulations
	ST_21	Standards
PT_06	**Engineering of biometric recognition systems**	-
	ST_22	Requisites analysis
	ST_23	Development, integration and maintenance
PT_07	**Performance evaluation**	-
	ST_24	Errors
	ST_25	Performance evaluation measures
PT_08	**Current trends in biometric recognition**	-
	ST_26	Recognition robustness
	ST_27	Non-cooperative recognition
	ST_28	Large-scale deployment
	ST_29	Visual surveillance and forensics

3.3.1 Calendar

The Table below gives a calendar about the topics and sub-topics that should be given to students in every week of the semester, together with the dates for the start of assignments and the corresponding due dates. The horizontal series at the top of the Table denote the index of the weeks and the topics / sub-topics appear spread across the Table rows.

	Weeks														
Topics	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
PT_01															
ST_01															
ST_02															
ST_03															
ST_04															
PT_02		**A**													
ST_05															
ST_06															
ST_07															
ST_08															
PT_03															
ST_09															
ST_10															
ST_11															
PT_04															
ST_12															
ST_13															
ST_14															
ST_15															
ST_16															
ST_17															
ST_18															
PT_05															
ST_19															
ST_20															
ST_21															
PT_06															
ST_22															
ST_23															
PT_07															
ST_24															
ST_25															
PT_08															**D**
ST_26															
ST_27															
ST_28															
ST_29															

In the Table above, the "A" in black denote the starting date for the **Hands-on** assignments course, while the "D" symbol is used to denote the due date of the corresponding assignment.

3.4 Teaching / Learning Activities

Considering the context of the course, the fact of being taught at a post-graduated level, and according to the spirit of Bologna, four different types of teaching / learning activities were planned:

1. **Theoretical classes**. The course comprises 15 theoretical sessions, two hours each. They are

planned not only according to the expositive paradigm, but also include short quizzes that should keep high levels of motivation from the students side. Also, the active participation, both in terms of questions / doubts is encouraged, and interruptions positively regarded. Also, the board will be used to illustrate the key insight concepts, as well as to solve basic exercises about each topic. The slides of each theoretical session will be available in the course web page immediately after the session, i.e., before the practical / laboratories classes.

2. **Laboratorial classes**. A total of 15 sessions, two hours each, is planned. The functioning of these classes will be clearly divided into two different parts: a) solving laboratorial exercises about fundamental topics learned in the previous theoretical session; and b) questions / doubts about the hands-on project. During the laboratories sessions, students will be asked to solve the exercises individually, during a closed time period. Then, a *brainstorming* phase among the elements in the class should occur, in order to compare and criticise alternative solutions. A list of all exercises should be available in the course web page at the beginning of each week.

3. **Hands-on project**. This project should be developed by groups of two students, mainly outside the classes schedule, and comprises the complete development of a biometric recognition system, designed to work on a practical (real or not) scenario, with constraints specific to each group. In order to accomplish the project goals, students are expected to analyse carefully the system requirements, and to use appropriate software engineering tools that guarantee that the resulting system will meet its goals. Also, the interaction between team mates will also be positively regarded, enabling students to auto-organise and to establish scheduling / priority plans, in order to accomplish a goal.

4. **Individual study**. Classes should be complemented by individual study, (estimated in 2 hours / week) order to consolidate the most important information taught during the lectures. Under this perspective, students are also expected to follow some of the additional references for further readings, according to their preferred traits or phases of the biometric recognition processing chain.

In order to promote the interaction between the instructor and students, and to clarify any issue regarding the theoretical / laboratories classes or even the development of the hands-on projects, the instructor is expected to reserve 3 hours per week to attend (at the office) students of the Biometric recognition Systems class. According to our previous experience in teaching / learning activities of M.Sc. / M.Eng. students, these sessions are an excellent opportunity to introduce challenging problems in the state-of-the-art and to discuss any ideas that might lead to advances in that domain of knowledge.

Below we summarise the most relevant information about the teaching / learning process planned for this course. As can be seen, summing all the teaching / learning activities yields a total of 160 hours of efforts related with this course, which corresponds to a total of 6 European Credit Transfer and accumulation System (ECTS).

15 theoretical classes × **2** hours = **30** hours

15 laboratories classes × **2** hours = **30** hours (22 hours solving exercises + 8 hours hands-on project)

Hands-on project, **2** hours × **15** weeks = **30** hours

Individual study, **3** hours × **17** weeks = **51** hours

Office sessions + evaluation, **1** hour × 15 weeks + 4 hours = **19** hours

160 hours; **6** ECTS

3.4.1 Backward Design

Most of the classes scheduling, both theoretical and practical, was carried out according to the **Backward Design** [55] paradigm, putting the focus on goals, and on our understanding about what we want

students to reach in the end and beyond. In the classical way, one can design the classes of a course to hit all the main topics in a set textbooks, without anticipating the major difficulties faced by students, in a *we'll know it when we see* perspective. Oppositely, the idea of backward design is to start the planning of all classes centered on goals, particularly on long-term goals that should endure over time. This paradigm can be regarded as the inversion of a pyramid of priorities, i.e., the teaching / learning activities follow from rather than lead to the goals.

As given in the sections above, we started by stating a fundamental goal for the course: *graduated students should be able to design from scratch, implement and deploy / maintain a biometric recognition system, for a practical scenario*. Next, we attempted to answer the question about *how can the instructor confirm that students learn the fundamental concepts in the course?*, which was made by planning the different pieces of evaluation. Finally, the partial goals and the scheduling of classes were conditioned to meet the goal, assuring that teaching / learning activities match its nature.

Backward Design I: *what is really essential for learners to know*?

Backward Design II: *how will the students show that they learned what they were expected to*?

Backward Design III: *what learning experiences will promote understanding, interest and goals*?

3.4.2 Theoretical Classes

The theoretical classes of this course should not only be expositive, according to the *passive learning* paradigm, but should also promote students participation and interaction with the instructor and the remaining elements of the class.

Classes will be based in oral presentations given by the instructor, making use of digital slides. Additionally, most topics should be complemented by multimedia pieces of information, which are known to be effective in transmitting the key insight behind most algorithms / techniques used in biometrics and related fields.

The instructor should introduce the fundamental concepts, principles and methods for each module of the program. Next, such concepts should be illustrated by intuitive examples, visual aids or demonstrators. Under the *active learning* paradigm, students are asked to develop their skills in mining and joining pieces of knowledge and to infer the most interesting applications under each corresponding perspective, with the instructor's guidance.

The above teaching / learning paradigm is innovative with regard to the classical way to taught theoretical classes (exclusively expositive). However, the type of topics covered in this course turns it particularly suitable for such an *open* paradigm, keeping in mind that – at the bottom level - it's up to the instructor to set the *standards* for learning and classroom environment and define the strict limits that assure the quality of the learning process.

Under an individual study perspective, students are expected to look up some important background materials before each class, assuring that they attend classes with the support knowledge required to properly understand the materials to be covered there. According to our previous experience, this is an excellent way to keep students motivated and even to encourage a healthy competition among them, which should be evident by the quantity / pertinence of the questions stated during the class. Also, it is particularly interesting to observe the - most times – the complementary perspectives of topic that different students obtain individually, enable them perceive in a more clear way the *big picture* behind a class topic.

Another interesting possibility is the idea of *daily in-class quizzes*, which should happen in a non-regular and pre-determined schedule, ensuring high levels of motivation of students during the whole semester. The idea in these quizzes is to ask students about fundamental (basic) concepts taught in the previous classes, which can be important to understand further topics.

In annex we include the complete set of slides designed for the theoretical sessions of the first edition of the Biometric Recognition Systems course.

3.4.3 Practical Classes

The practical classes planned for this course are centered on students, giving them an active role for learning by *doing*, *experimenting*, *simulating* and *discussing* with the remaining class about the specific advantages /shortcoming of every method / algorithm over competitors. In this type of classes, the role of the instructor is mainly to provide the initial step during the first 10/15 minutes of the class, assuring that a problem is framed in an objective way and that plausible solutions are considered. Next, it is important to pause for students to consolidate their understanding of the problem, find gaps, and discuss plans.

Clearly, the instructor should play the role of **tutor** here, by providing experience, encouraging students, establishing the expectations for an exercise, and assuring that rules are kept. For problems found particularly hard by a large proportion of the students, the instructor should be able to illustrate / model similar problems, and eventually give some tip that will lead students to plausible solutions.

According to our previous experience, it's not uncommon that a part of the students attend the practical classes without having consolidated appropriately the topics of the last theoretical lecture, which can dramatically reduce their effectiveness in the practical class. To alleviate this problem, the first minutes of the laboratories class can start by a **background knowledge probe**, requiring simple responses from the students side about the last learned unit. This kind of probes not only helps the instructor to determine an effective starting point for a session, helps students focusing their attention on the essential materials, and also motivates them for proper consolidation of materials before the practical classes.

Figure 3.2: Example of a normalised iris image used to illustrate an exercise about the creation of binary classification models, under the "work with real data" paradigm.

Whenever possible, we will privilege the **working with real data** paradigm, which is particularly suitable given the nature of the course. As an example, Fig. 3.2 illustrates an image of a normalised texture of the iris (i.e., invariant to scale and rotation) with portions occluded by eyelids, eyelashes and reflections due to the surrounding . Using a set of learning data, it students will be asked to create an automated model able to distinguish the unoccluded regions of the iris. If this problem is framed in the right way, it is essentially a binary classification problem at the *pixel* level. As most of the students are expected to have previous experience on binary classification models, this will increase the levels of motivation of students and the plausibility of the solutions found.

Finally, we intend to promote a **competitive learning** paradigm for particular exercises of some of the classes, in which the success of each student is measured in a relative way, with respect to the success obtained by the remaining elements of the class. The idea is to give to all class the same problem and learning set, asking to each student to generate a corresponding model. Next, the instructor receives all the models generated in the class and uses a performance evaluation set to automatically rank the models with respect to an effectiveness measure previously announced to all class. From our previous experience, this is extremely motivating for students, in result of their natural ambition of being *distinguished among peers*. On the other side, anticipating that it will be particularly important to not discourage / disappoint the classmates with the worst grades, the results announced to the class should regard exclusively the *top-n* rank, and not the complete list of results.

Theoretical classes: passive and active learning paradigms: daily in-class quizzes, knowledge mining

Practical / laboratories classes: tutoring, background knowledge probes, working with real data, competitive learning

Hands-on project: groups 2 students, self-organisation goal oriented, real scenarios, covering complete development chain.

3.4.4 Classes Scheduling

We now summarise the contents planned for each week of the semester, for the theoretical and laboratories classes.

Week 1	**Theoretical**	What's a biometric recognition system? Definitions and examples; goals of a biometric recognition system; illustration of a broad range of applications for biometrics; history of biometrics: from the Bible to the state-of-the-art, passing on Bertillon.
	Labs.	Introduction to MATLAB®; Image processing basics; image acquisition, image processing, computer vision system and statistics toolboxes.

Week 2	**Theoretical**	Biometrics modes of functioning: verification vs. identification systems; frequent observers; classification of biometrics systems: overt vs. covert, habituated vs. non-habituated, attended vs. non-attended, standard vs. non-standard environments, public vs. private, open vs. close; typical biometric traits: fingerprint, face, iris, palm vein, retina and signature; emerging traits: periocular, gait and ear. Multi-biometric recognition systems.
	Labs.	Least-squares basics; image registration exercises: Using SIFT [26] descriptors and homography matrices.

Week 3	**Theoretical**	Biometrics typical processing chain: data acquisition, detection, segmentation, normalisation, feature extraction, classification and fusion; examples of relevant algorithms for each phase of the processing chain: Adaboost-based detector, Hough transform, Gabor filters, similarity metrics and Bayesian decision rules.
	Labs.	Image filtering basics; image smoothing; edge detection: Sobel, Laplacian and Canny detectors.

Week 4	**Theoretical**	Notion of system covariate; recognition robustness; robustness against data variation factors: acquisition-noise, changes in translation, scale, rotation, resolution, focus, perspective, occlusions and environmental (lighting) conditions; challenges behind the development of robust recognition systems. Biometric menagerie.
	Labs.	Color spaces: examples; Global / local image equalisation algorithms: histogram equalisation and stretch; adaptive histogram equalization; contrast enhancement; homomorphic filtering.

Week 5	**Theoretical**	Signal processing basics: linear systems, signal decomposition / superposition. The frequency domain, Discrete Fourier Transform and Fast Fourier Transform. Convolution and filters. Edge detection
	Labs.	Signal processing on MATLAB®. Examples: image smoothing and edge detection

Week 6	**Theoretical**	The Viola and Jones [53] object detector: weak features, integral image, Adaboost-based model learning, object detection.
	Labs.	Object detection according to MATLAB® Computer Vision toolbox; learning detectors.

Week 7	**Theoretical**	Parametric vs. non-parametric segmentation methods; form fitting algorithms; introduction to active contours and deformable models: snakes; geodesic active contours.
	Labs.	Notion of parametric space: Hough transform; line, circular and elliptical Hough transform; snakes implementation and testing; active contours without edges [13].

Week 8	**Theoretical**	Texture / Shape descriptors: histograms of oriented gradients (HOGs), local binary patterns (LBPs) and basic of Fourier descriptors; basics of image filtering: convolution and correlation. examples of kernels: Gaussian, Gabor, and multi-lobe differential filters.
	Labs.	The Fourier transform; Fast Fourier Transform (FFT); signal processing fundamental theorem; Gabor and multi-lobe differential image filtering.

Week 9	**Theoretical**	Feature normalisation: min/max and mean/var rules; Typical probability distributions: uniform, normal and binomial distributions; Notion of eigenvector; analysis of principal components; PCA and LDA; non-linear classification algorithms: neural networks and support vector machines; fusion at different levels: data, features, scores and responses; Dempster-Shaffer fusion theory.
	Labs.	Analysis of eigenvectors and principal components: toy examples; image features PCA; linear classification through LDA.

Week 10	**Theoretical**	Ethics in biometrics: *bios* vs. *zoe* discrimination; right to privacy vs. right to anonymity; social / cultural impact of biometrics; (anti) spoofing techniques; security vs. privacy trade-offs; overview of official ethical reports: *RAND, BIOVISION, OECD* and *NSTC*.
	Labs.	Exercices MATLAB® Neural Network and Classification toolboxes; Bagging, Boosting and Random Forest examples.

Week 11	**Theoretical**	The Universal declaration of human rights; National regulations: the *Comissão Nacional para a Proteção de Dados*; International regulations and standards: *8 CFR 103.16 - Collection, use and storage of biometric information*; European data protection directive *95/46/EC*, European *ART 29 - Working Party on the Protection of Individuals with regard to the Processing of Personal Data.*
	Labs.	Hands-on biometric system I: choosing the trait and acquisition environment; data acquisition setup and protocol; expected data variations; data collection.

Week 12	**Theoretical**	Biometrics systems engineering: analysis of requirements; stakeholder identification and interviews; contract-style requirement lists; measurable goals; use cases; requirements issues and refinement; alternative solutions; scalability and portability. Development; high-level design; prototyping biometric recognition systems; rapid application development frameworks; system interfaces; performance optimization frameworks; debugging, testing and validation; proof-of-concept system.
	Labs.	Hands-on biometric system II: feature extraction, normalisation and classification.

Week 13	**Theoretical**	Biometrics systems engineering: integration and maintenance; data storage and interface management; security configuration; job scheduling; preventive maintenance plans; system breakdown plans. Biometric interfaces.
	Labs.	Hands-on biometric system III: system test and maintenance plan.

Week 14	**Theoretical**	Types of errors in biometric recognition: false acceptances, false rejections; Confusion matrices; acceptance thresholds; equal error rate (EER); receiver operating characteristic (ROC) curves, detection-error tradeoff (DET) curves; rank-N curves; area under curve (AUC) and decidability index.
	Labs.	Bootstrapping performance evaluation; students cross-evaluation of hands-on biometric systems: spoofing and anti-spoofing strategies.

Week 15	**Theoretical**	Challenges behind the development of robust recognition systems; hardness of non-cooperative recognition scenarios; requirements in large-scale deployment; bridging the gap between biometrics and visual surveillance or *"how close are we getting to George Orwel's Big Brother?"*
	Labs.	Ideas for biometrics research; overview of the state-of-the-art; cohesive perspective of innovations.

3.4.5 Evaluation / Grading System

The evaluation protocol established for this course considers the nature of its prospective students, their typical skills and previous background, along with the course topics and goals. Also, the evaluation moments should spread as much as possible along the semester, in order to keep the students motivated, while attempting to minimize early withdrawals, a typical problem identified in several courses of the Department. Hence, the designed evaluation scheme is based on the following points:

- **Constraint 1**: there is a previous history of early withdraws from a significant number of students, for some of the courses taught at the Department;

- **Background skills**: most students should come from the 1^{st} cycle in Informatics Engineering, and have particular skills on software development processes;

- **Constraint 2**: the curricular structure of the 1^{st} cycle in Informatics Engineering at UBI lacks on a signal processing course;

In agreement with the statements above, the evaluation is comprised by three parts: 1) a theoretical evaluation weighting 40% of the final mark; 2) a laboratorial evaluation that accounts for 30% of the final mark; and the 3) hands-on project that, along with the corresponding report and presentation, weights 30% of the final mark. All components are evaluated on the [0,20] scale, resulting in the final mark m_f:

$$m_f = [0.4\ 0.3\ 0.3]\ .[m_t\ m_l\ m_h]^T, \tag{3.1}$$

being m_t the theoretical mark, and m_l and m_h the laboratories and hands-on marks respectively.

The theoretical component mark m_t yields from a written individual exam, scheduled for a date immediately after the 15^{th} week of the semester. In this exam students will have 120 minutes to answer to all questions, counting on any paper documentation that they may find appropriate to take for the exam (i.e., open-book exam). Typically, this exam consists on:

- General questions about biometrics, emphasising topics that will demand a critical analysis from students (i.e., avoiding definitions and closed answers);

- Simple calculus required for some phase of the algorithms covered in the course (e.g., normalisation of a small data set, illustration of discriminating hyperplanes)

- Discussion about the applicability of the different traits for a specific environment (e.g., *is the iris always preferable over the face*?)

- Schematisation of development plans for a biometrics system;

- Calculus of performance evaluation measures;

- High-level software implementation of core tasks of some algorithms learned in the course.

The laboratories mark m_l will be evaluated as much as possible under a continuous evaluation paradigm, by making notes of the commitment of each student at each laboratories session, and the corresponding success in each task proposed during the semester. The instructor should give timely feedback of this component, desirably immediately after each laboratorial session, to keep students motivated and to promote their ambitions on improvements.

Finally, the hands-on mark m_h will yield from the integrated project that students should carry during the whole semester, partially developed in the practically classes. For most of the hands-on work, students should count on the help of Ph.D. students at the SOCIA Lab.: Soft Computing and Image Analysis Lab, located at room 6.12 of the UBI. They will arrange themselves in groups of two elements (at most), being asked to:

- Imagine a practical scenario where a biometric system could be useful;

- Confirm that the SOCIA Lab. has appropriate sensors for data acquisition;

- Write a requisites analysis report;

- Design an integrated solution of a biometric recognition system, in terms of: 1) the used trait; 2) acquisition setup; 3) use cases; 4) programming languages; 5) data interfaces; 6) installation and maintenance plans; 7) ethical report;

- Obtain the corresponding performance evaluation measures.

As stated above, the hands-on project should count on the valuable help of some of the Ph.D. students that are currently carrying their work in the SOCIA Lab. / Pia-Cv group of the *IT: Instituto de Telecomunicações*.

Theoretical written evaluation: 40%

Laboratorial evaluation excesses solving + class participation + *daily in-class quizzes*: **30%**

Hands-on project: Groups of two students, practical scenario **30%**

3.5 Feedback Plan

This section describes the major ways of interaction between the instructor and students, which was divided into two sub-sections: the *educative* assessment part describes how the should the students receive feedback about the learning process and their achievements in that scope. On an inverse way, the *auditive* assessment part regards the interaction mechanisms from students to the teacher, about the main difficulties and gaps that they find in some part(s) of the materials, which will be important for improving further editions of the course.

3.5.1 Educative Assessment

The educative assessment should be given to students by three major channels: 1) during the laboratories classes, immediately after the exercises solving; 2) by electronic mail, at an individual level; 3) during the weekly office opening hours. Regardless the specific channel, we consider the **FIDeLity** paradigm [1] for educative assessment, i.e., attempting to keep it **F**requent, establishing a frequency that is known by everybody in the class; **I**mmediate, not letting any relevant information be forgotten by time lapses; **D**incriminating, by stating clearly the difference between a poor, minimally acceptable, good or outstanding work and **L**ovine, keeping empathy in the way feedback is given.

3.5.2 Auditive Assessment

Auditive assessment is expected to be particular useful for further improvements on the course. Specifically, the instructor should get information about the topics that students found more / less interesting, and the particular difficulties that they faced when learning each module of the course. Auditive assessment should be carried out by means of **Time for Reflection** questionnaires, given to students at the end of each major part of the course (end of the forth, ninth and fifteenth weeks), and asking for their opinion / contributions / suggestions regarding further improvements. These questionnaires include questions such as:

- *What was the important thing you've learned about the module X*?
- *Did you understood the outcomes of this part of the course*?
- *How likely you might integrate the knowledge from module X in your further tasks*?
- *Did you find any point in this module particularly (un)appealing*?
- *How did you find the articulation between the topics given in the theoretical class and the exercises given in the laboratories classes*?
- *In what way your previous background on X augmented the difficulty in learning module Y appropriately*?

3.6 Course Bibliography and Information Sources

Here we announce the most relevant information sources about the Biometric Recognition System course, covering a much broad as possible range of sources, from those that give a global perspective about the topic, to the state-of-the-art papers that describe recognition algorithms used as baselines in current research. Also, we refer to the integrated development environments and programming languages that are typically used in the development of biometric recognition systems. Finally, we also announce several URLs that are an excellent choice for quick information access. In the rightmost column of each table we highlight some of the sources, considered the most important for the context of the course.

3.6.1 Textbooks

There is a large quantity of textbooks available on the market fully devoted to biometrics. We highlight the set below, according to the introductory level of most of them and also due to the scientific relevance of the corresponding authors.

Title	Authors	Info.	Type	Key
Handbook of Biometrics	A.K Jain, P. Flynn and A. Ross (eds.)	Springer-Verlag, ISBN: 978-0-387-71040-2, 2008	General	
Handbook of Iris Recognition	M. Burge and K. Bowyer (eds.)	Springer-Verlag, ISBN: 978-1-4471-4402-1, 2013	Iris	
Handbook of Fingerprint Recognition	D. Maltoni, D. Maio, A. K. Jain and S. Prabhakar	Springer-Verlag, ISBN: 978-1-8488-2254-2, 2009	Fingerprint	
Handbook of face Recognition	S. Li and A. K. Jain (eds.)	Springer-Verlag, ISBN: 978-0-8572-9932-1, 2011	Face	
Introduction to Biometrics	A.K Jain, A. Ross and K. Nandakumar	Springer-Verlag, ISBN: 978-0-387-77326-1, 2012	General	
Handbook of Multibiometrics	A. Ross, K. Nandakumar and A.K Jain	Springer-Verlag, ISBN: 978-0-387-33123-2, 2006	Multi-biometrics	
Signal and Image Processing for Biometrics	J. Scharcanski, H. Proença and E. Du (eds.)	Springer-Verlag, ISBN: 978-3-642-54079-0, 2014	General	

3.6.2 Research Papers

The field of biometrics is particularly rich in terms of the quantity and quality of the previous publications. Apart from existing several information sources (conferences and journals) that are exclusively devoted to biometrics research, most of the premium journals and conferences in the pattern recognition / computer vision / artificial intelligence and image processing domains also accept papers about biometrics, some times even conduct special sessions / special issues centered on biometrics research. Hence, among the large number of publications potentially interesting for students, we highlight a small set of these, written in a particularly clear way, and describing algorithms that might easily adapted to different biometric traits. Also, we privilege papers highly cited in the biometrics literature.

Title	Authors	Info.	Type	Key
Biometric recognition	A. K. Jain	Nature, 449, 38-40, 2007	General	
High Confidence Visual recognition of Persons by a Test of Statistical Independence	J. Daugman	IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(11), 1148-1161, 1993	Iris	
Robust Face Recognition via Sparse Representation	J. Wright, A. Yang, A. Ganesh and S. Sastry	IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2), 210-227, 2009	Face	
Statistical pattern recognition: A review	A.K. Jain, R.P.W. Duin, and J. Mao.	IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1), 4-37, 2000	General	
Face Verification Across Age Progression	N. Ramanathan and R. Chellapa	IEEE Transactions on Image Processing, 15(11), 3349-3361, 2006	Face	
Pores and Ridges: High-Resolution Fingerprint Matching Using Level 3 Features	A. K. Jain, Y. Chen and M. Demirkus	IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(01), 3349-3361, 2007	Fingerprint	
Soft Biometrics; Human Identification Using Comparative Descriptions	D. Reid, M. Nixon and S. Stevenage	IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(06), 1216-1228, 2007	Soft Biometrics	
Performance of biometric quality measures	P. Grother and E. Tabassi	IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4), 531-543, 2007	General	
Iris Recognition: On the Segmentation of Degraded Images Acquired in the Visible Wavelength	H. Proença	IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(8), 1502-1516, 2010	Iris	

3.6.3 Software

Most of the source code about biometrics research is publicly available by means of MATLAB® functions and toolboxes, due to the fact of being an effective rapid application development environment, with an enormous amount of previously available functions that support the implementation process. For a long time, the main disadvantage of MATLAB® was the fact of being an interpreted language (non-compiled), which significantly reduced its performance in terms of execution time. Even though performance was consistently improved in the recent editions of MATLAB®, when it comes to prototype implementation, the language C/C++ is still the main choice, due to its portability and to the effectiveness of the resulting application executables. Also, another significant gap of MATLAB® is its price, even for student individual licenses, which has been augment ion the relevance of Wolfram MATHEMATICA, as a replacing (freely available) platform.

Software	Description
MATLAB®, Mathworks	Rapid Application Deployment (RAD) environment. Availability of toolboxes. Easy of programming.
Wolfram MATHEMATICA	Free available RAD environment
Open CV: Open Source Computer Vision	BSD License, C/C++, Python and Java interfaces. Windows, Linux, MAC OS, iOS and Android development.
CImg: C++ Template Image Processing Toolkit	Open source, GPL License, C++ development
Research code on Computer Vision[1]	Binary plus source code files about a broad range of computer vision algorithms

3.6.4 Web Sources

Below we list a set of resources that are freely available at the web, which were selected according to their relevance and are organised with regard to the different biometric traits and to complementary aspects of the engineering of biometric recognition (associations, companies, standards and systems).

URL	Description
General	
http://www.cse.unr.edu/~bebis/CS790Q/Lect/IntroBiometrics.pdf	Biometrics: A *Grand* challenge, by A.K. Jain.
http://www.research.ibm.com/people/a/aws/icapr.html	Tutorial on automated biometrics, by Nalini K. Ratha and Andrew Senior
http://homepages.inf.ed.ac.uk/rbf/IAPR/researchers/PPRPAGES/pprtut.htm	IAPR Tutorials on Topics in Statistical Pattern Recognition
http://www.umiacs.umd.edu/~ramani/cmsc828d/lecture1.pdf	Fundamentals of Computer Vision: Introduction, by Larry Davis, Ramani Duraiswami, Daniel De Menthon and Yiannis Aloimonos
Iris	
http://www.cl.cam.ac.uk/~jgd1000/	Personal web page of the inventor of the most relevant iris recognition system.
Face	
http://www.face-rec.org/	Face recognition homepage
Novel Biometrics	
http://www.comm.toronto.edu/~kostas/Publications2008/pub/60.pdf	Gait recognition: A challenging signal processing technology for biometric identification
www.sersc.org/journals/IJCA/vol3_no1/3.pdf	Palm Vein Authentication System: A Review
http://www.cmeri.res.in/rnd/srlab/cvision/face\%20pericular.php	Face and periocular biometrics
Fingerprint	
http://www.cedar.buffalo.edu/~govind/CSE666/fall2007/FP_Tutorial.pdf	A Tutorial on Fingerprint recognition, by D. Maltoni
Associations / Companies / Standards	
http://www.biometricupdate.com/biometric-companies	List of biometric companies I
http://www.dmoz.org/Computers/Security/Biometrics/Companies/	List of biometric companies II
http://www.biometrics.dod.mil/References/Tutorial/	U.S.A Department of Defense, The Defense Forensics & Biometrics Agency
http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/ngi	F.B.I. next generation identification web page.
http://www.eab.org/contact/secretariat.html?ts=1410419005930	European Association for Biometrics.
http://uidai.gov.in/biometric-devices.html	Largest biometric system in the world.
http://www.safran-group.com/site-safran-en/innovation-429/areas-of-expertise/biometrics/	World leader in biometric systems deployment.
http://www.biometrics.gov/standards/	List of biometric recommended standards
Systems	
http://www.biometrics.org/html/examples/examples.html	Large list of biometric systems examples, Biometric consortium
http://iris.usc.edu/Vision-Notes/bibliography/contents.html	Computer vision bibliography. Excellent choice to get the state-of-the-art in most topics related to computer vision / image processing tasks.
https://www.linkedin.com/groups/Biometrics-Network-Forum-40210	*LinkenIn*® Biometrics Network & Forum
Datasets	
http://datasets.visionbib.com/	List of datasets related to computer vision / image processing.
http://iris.di.ubi.pt/	UBIRIS: Noisy Visible Wavelength Iris Image Databases

Chapter 4

Conclusion

This document gives a detailed description of the **Biometric Recognition Systems** course, proposed to be included in the second semester of the first year of the 2^{nd} cycle of studies in Informatics Engineering at the University of Beira Interior (UBI). We justify the requirements for this type of course, as well compare it to similar courses that are available both at the national and international levels.

We start by giving a cohesive perspective of the goals behind this course, and detail its syllabubs, program topics and sub-topics, along with a schedule of the subjects covered in each theoretical class and in each laboratorial session. The slides and exercise sheets are given in annex, as are examples of exams and of suggestions for hands-on projects.

According to our previous experience in teaching at the 2^{nd} cycle of studies in Informatics Engineering at UBI, such context determined most of the choices described in this document: 1) the prospective students have particular interest about **engineering** and **functioning** topics, in opposition to the relevance they give to more *formal* or *theoretical* concepts, which we believe to have roots in lower average marks that are required to be admitted at UBI than at the biggest Portuguese universities (e.g., universities in Lisbon or Porto). Actually, this was one of the reasons to name this course as **Biometric Recognition Systems**, keeping the focus in concepts that are more typically associated to informatics / computer science than to other domains (e.g., signal processing), which could have been naturally covered in a deepen way.

This way, the idea was to design a course that particularly fits its academic surroundings, and that is expected to motivate the interest from a large proportion of the prospective students, while guaranteeing that the most fundamental concepts about the planning, development and maintenance of a biometric recognition system are transmitted to students.

Main goal: Give an **integrated** perspective of the **whole development chain** of a biometric recognition system;

Focus on **engineering** of biometrics systems, according to the typical students skills and competences;

Attempt to enforce **partnerships** with **surrounding entities** (e.g., small companies or schools), by providing them working prototypes of biometric recognition solutions.

Appendix A

Support Materials

A.1 Announcements

- The most relevant information about the ”Biometric Recognition Systems" course is available at the **`http://www.di.ubi.pt/~hugomcp/biometrics/`** web page. Here will be announced any relevant information to students, making available the theoretical slides and practical sheets, as well the description of the practical projects and of the relevant third-party sources of information;
- Accommodations for **students with disabilities** who believe that they may need accommodations in this class are encouraged to contact the professor (`hugomcp@di.ubi.pt`) or the Department secretariat as soon as possible, in order to **ensure that proper accommodation is implemented in a timely fashion**.

A.2 Theoretical Lectures

The theoretical lectures should ideally be given at the **beginning of the week**, and **always before the practical / laboratories sessions**, in order to ensure that contents are given in an articulated / integrated way.

[0]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

University of Beira Interior, Department of Informatics
Hugo Pedro Proença, 2014/2015

Class Schedule

Time	Monday	Room	Tuesday	Room	Wednesday	Room	Thirsday	Room	Friday	Room
8-9	-	-	-	-	-	-	-	-	-	-
9-10	-	-	-	-	-	-	-	-	-	-
10-11	-	-	-	-	-	-	-	-	-	-
11-12	-	-	-	-	-	-	-	-	-	-
12-13	-	-	-	-	-	-	-	-	-	-
13-14	-	-	-	-	-	-	-	-	-	-
14-15	Theorethical	6.02	-	-	-	-	-	-	-	-
15-16			-	-	-	-	-	-	-	-
16-17	-	-	Lab.	6.13	-	-	Office	3.18	-	-
17-18	-	-			-	-			-	-
18-19	-	-	-	-	-	-	-	-	-	-

Course Web Page

http://www.di.ubi.pt/~hugomcp/biometrics/

- **Contents:**
  - Marks
    - Announcement of the marks at both the theoretical, laboratorial and hands-on components
  - Theoretical slides and laboratorial sheets
  - Announcements
    - Any information relevant to the course
    - Due date for assignments, data sets,...
  - Links
    - Tutorials, relevant research papers
  - Etc...

Evaluation Criteria

1. **Attendance**
   1. The attendance of students are recorder, both at the theorethical and practical classes. Each non-justified absence (Administrative Services of UBI) will imply a penalty of 0.25 in the final mark of the course.
2. **Laboratorial mark (30%)**
   1. Effectiveness in solving exercises; motivation; participation in classes
3. **Hands-on Project (30%)**
   1. The hands-on project should be carried out by students in groups of 2 elements
   2. The success in the course is conditioned to the obtainance of the minimal value of 8 points (8/20) in the hands-on project.
   3. The presentation of the final report will be made at the class, and is mandatory for all groups.

Evaluation Criteria

1. **Written Frequency Exam (40%)**
   1. **F1** - Thursday, January 5th 2015, 19:30, Room 6.05
2. Exams
   1. Exams **E1** and **E2.** Date and time to be defined by the Administrative Services (12 points).
3. Working-Students (enrollment at the administrative services is mandatory)
   1. The evaluation criteria are the same, with exception to the penalty due to non-attendance.

Prerequisites

Prerequisites:

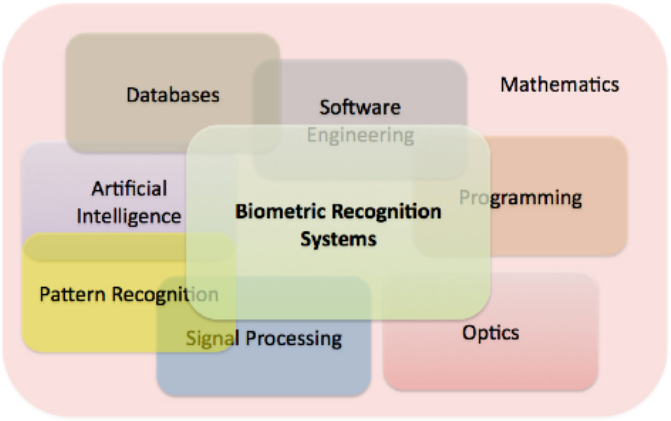
Topics that any student should previously kwow, before the starting date of the course. Even though they are not included in the course program, the success on this course strongly depends of the previous students background on these topics.

- As can be inferred from the course program, the **Biometric Recognition Systems** course integrates knowledge from a broad range of topics, such as the Pattern Recognition, Computer Vision, Artificial Intelligence, Software Engineering, Mathematics and Optics domains.
- Also, as in most of the courses lectured at the 2nd cycle of studies in Engineering, an appropriate background in Mathematics is undoubtedly a plus.

Prerequisites

Prerequisites:

Topics that any student should previously kwow, before the starting date of the course. Even though they are not included in the course program, the success on this course strongly depends of the previous students background on these topics.



Fundamental Goal

- Fundamental Goal:

"Assure that graduated students are able to design, implement and maintain a biometric recognition system from scratch"

Major Goals

- **[GG_01]** Broad understanding of how to provide effective biometric identity management systems;
- **[GG_02]** Obtain the system development skills that allow them to adapt to dynamic and fast-changing industrial environments;
- **[GG_03]** Gain knowledge about the industrial / commercial context in which biometric recognition systems are integrated;
- **[GG_04]** Understand the ethical and legal issues, as well the relevant standards in the context of biometric recognition;

Specific Goals

- **[SG_01] (GG_01)** Gain knowledge about the emerging technologies in the biometrics recognition field;
- **[SG_02] (GG_01)** Understanding the mathematical / conceptual models behind the development of a bio- metric recognition system;
- **[SG_03] (GG_01)** Understand the typical phases of a biometric recognition systems, and how these phases can be regarded from the pattern recognition perspective.
- **[SG_04] (GG_01)** Gain knowledge about the most typical algorithms used in each phase of the biometric recognition process
- **[SG_05] (GG_02)** Ability to plan, design and implement modern biometric recognition system

Specific Goals

- **[SG_06] (GG_01)** Ability to compare in an objective way the advantages / disadvantages associated with each biometric trait, in order to fit the specificities of a particular environment
- **[SG_07] (GG_02)** Ability to decide about the most appropriate hardware / software platforms for the deployment of a biometric recognition system
- **[SG_08] (GG_01)** Gain knowledge about the widely acceptance performance evaluation measures typically used in biometrics research.
- **[SG_09] (GG_03)** Ability to design tests for a biometric recognition system, during the deployment / validation phases.
- **[SG_10] (GG_04)** Meet the standards and legal regulations required in the development of a specific bio- metric recognition system

Course Program

- **PT_01:** Biometric recognition: cohesive perspective
  - **ST_01:** Definitions
  - **ST_02:** Goals
  - **ST_03:** Applications
  - **ST_04:** Historical Overview
- **PT_02:** Biometric recognition essentials
  - **ST_05:** Modes of Functioning
  - **ST_06:** Systems Classification
  - **ST_07:** Biometric Traits
  - **ST_08:** Multi-biometrics

Course Program

- **PT_03:** Biometrics from the pattern recognition perspective
  - **ST_09:** Typical processing chain
  - **ST_10:** Data covariates
  - **ST_11:** Challenges
- **PT_04:** Recognition methods and techniques
  - **ST_12:** Detection
  - **ST_13:** Segmentation
  - **ST_14:** Feature extraction
  - **ST_15:** Feature normalization
  - **ST_16:** Classification
  - **ST_17:** Fusion

Course Program

- **PT_05:** Ethical issues, legal regulations and standards
  - **ST_18:** Ethics considerations
  - **ST_19:** National / international legal regulations
  - **ST_20:** Standards
- **PT_06:** Engineering of biometric recognition systems
  - **ST_21:** Requisites analysis
  - **ST_22:** Development, integration and maintenance
- **PT_07:** Performance evaluation
  - **ST_23:** Errors
  - **ST_24:** Performance evaluation measures
- **PT_08:** Current trends in biometric recognition
  - **ST_25:** Recognition robustness
  - **ST_26:** Non-cooperative recognition
  - **ST_27:** Large-scale deployment
  - **ST_28:** Visual Surveillance and Forensics

Classes Scheduling

	Weeks														
Topics	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
PT_01															
ST_01															
ST_02															
ST_03															
ST_04															
PT_02															
ST_05		1													
ST_06															
ST_07															
ST_08															
PT_03															
ST_09															
ST_10															
ST_11															
PT_04															
ST_12															
ST_13															
ST_14															
ST_15															
ST_16															
ST_17															
PT_05															
ST_18															
ST_19															
ST_20															
PT_06															
ST_21															
ST_22															
PT_07															
ST_23															
ST_24															
PT_08															
ST_25															
ST_26															
ST_27															
ST_28															

Bibliography: Textbooks

- **Handbook of Biometrics;** A.K Jain, P. Flynn and A. Ross (eds.), Springer-Verlag, ISBN: 978-0-387- 71040-2, 2008
- **Handbook of Iris Recognition;** M. Burge and K. Bowyer (eds.), Springer-Verlag, ISBN: 978-1- 4471-4402-1, 2013
- **Handbook of Fingerprint Recognition;** D. Maltoni, D. Maio, A. K. Jain and S. Prabhakar, Springer-Verlag, ISBN: 978-1- 8488-2254-2, 2009
- **Handbook of face Recognition;** S. Li and A. K. Jain (eds.), Springer-Verlag, ISBN: 978-0- 8572-9932-1, 2011
- **Introduction to Biometrics;** A.K Jain, A. Ross and K. Nandakumar, Springer-Verlag, ISBN: 978-0-387- 77326-1, 2012
- **Handbook of Multibiometrics;** A. Ross, K. Nandakumar and A.K Jain, Springer-Verlag, ISBN: 978-0-387- 33123-2, 2006
- **Signal and Image Processing for Biometrics;** J. Scharcanski, H. Proença and E. Du (eds.), Springer-Verlag, ISBN: 978-3-642- 54079-0, 2014

Bibliography: Research Papers

- Biometric recognition; A. K. Jain, *Nature*, 449, 38-40, 2007
- High Confidence Visual recognition of Persons by a Test of Statistical Independence; J. Daugman, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11), 1148-1161, 1993
- Robust Face Recognition via Sparse Representation; J. Wright, A. Yang, A. Ganesh and S. Sastry, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2), 210-227, 2009
- Statistical pattern recognition: A review; A.K. Jain, R.P.W. Duin, and J. Mao, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4-37, 2000.
- Face Verification Across Age Progression; N. Ramanathan and R. Chellapa, *IEEE Transactions on Image Processing*, 15(11), 3349-3361, 2006

Bibliography: Research Papers

- Pores and Ridges: High-Resolution Fingerprint Matching Using Level 3 Features; A. K. Jain, Y. Chen and M. Demirkus, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(01), 3349-3361, 2007
- Soft Biometrics; Human Identification Using Comparative Descriptions; D. Reid, M. Nixon and S. Stevenage, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(06), 1216-1228, 2007
- Performance of biometric quality measures; P. Grother and E. Tabassi, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4), 531-543, 2007
- Iris Recognition: On the Segmentation of Degraded Images Acquired in the Visible Wavelength; H. Proença, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), 1502-1516, 2010

Bibliography: Aditional sources

- A set of aditional information sources can be found at the course web page.
  - "Web sources" section
- We include the URLs of relevant tutorials, systems, datasets, standards, associations and companies
- Resources are grouped according to the biological trait that they regard (iris, face, fingerprint,...)

Hugo Pedro Proença, UBI, 2014

Hugo Pedro Proença, UBI, 2014

[1]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

Syllabus

Theoretical	What's a biometric recognition system? Definitions and examples; Goals of a biometric recognition system; Illustration of a broad range of applications for biometrics; History of biometrics: from the Bible to the state-of-the-art, passing on Bertillon
Labs	Introduction to MATLAB ; Image processing basics; Image acquisition, image processing, computer vision system and statistics toolboxes.

Biometrics

- "***Biometrics*** *consists of methods for uniquely* ***recognizing humans*** *based upon one or more* ***intrinsic*** *physical or behavioral* ***traits***". (Wikipedia)
  - **Main** properties: **Universality, Uniqueness, Permanence** and **Collectability;**
  - **Complementary** properties: **Performance, Acceptability** and **Circumvention;**

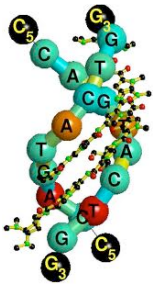
Automated Data Analysis

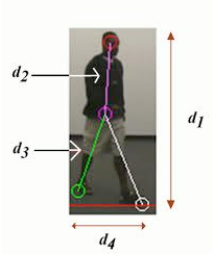
12.612	0.1721	215.19
0.1219	1.1913	0.1283
146.19	1677.1	210.14
2161.9	0.1812	0.8283

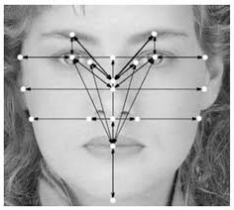
Biometric Signature

Biometrics II

- "The process by which a **person's unique physical** and other traits are detected and **recorded** by an electronic device or system as a **means of confirming identity**". (*Dictionary.com*)







Biometrics: Aditional Definitions

- "The study of automated methods for uniquely **recognizing humans** based upon one or more **intrinsic physical** or **behavioral** traits" (*Wikipedia*)
- A method of verifying an individual's identity based on **measurements** of the individual's **physical features** or **repeatable actions** where those features and/or actions are both **unique** to that individual and **measurable** (*Glossary, Data Core Technology*)
- Biometrics is the science and technology of **measuring** and **analyzing** biological data. In information technology, biometrics refers to technologies that measure and analyze **human body characteristics,** such as fingerprints, eye retinas and irises, voice patterns, facial patterns and hand measurements, for **authentication** purposes (*WhatIs Dictionary*)
- Biometrics is the science of measuring **physical properties** of living beings using **suitable body characteristics** (*Bromba's Biometric FAQ*)
- Any automatically **measurable, robust** and **distinctive** physical **characteristic** or personal **trait** that can be used to identify an individual or verify the claimed identity of an individual (*Woodard et al.*)

Biometric Recognition Systems

- Biometric recognition systems are currently used for different purposes with highly satisfactory results.
  - Refugee control, Access to Physical Resources, Airports…

(http://fingerprint-security.net/)

(http://www.lockheedmartin.com)

(http://www.engadget.com/)

"Considering the recent mandates of several governments for the nationwide use of biometrics (…) Pattern recognition systems have never been tried at such large scales nor have they dealt with such a wide use of sensitive personal Information. As pattern recognition researchers, it is a great opportunity and challenge for us to make a difference in our society (…)"

A.K. Jain; "*Biometrics: A Grand Challenge*", ICPR 2004.

Biometric Recognition Systems

- Historically, personal identification numbers (PINs), passwords, names, social security numbers, and tokens (cards, keys, passports, and other physical objects) have been used to recognize an individual;
  - For example, access to an automatic teller machine (ATM) is generally controlled by requiring presentation of an ATM card and its corresponding PIN.
- Sometimes, however, recognition can lead to the denial of a benefit. This could happen if an individual tries to make a duplicate claim for a benefit or if an individual on a watch list tries to enter a controlled environment.
  - **Refugee control** is one of the successful applications of biometrics

Biometric Recognition Systems

- A primary motivation for using biometrics is to easily and **repeatedly recognize** an individual so as to enable an automated action based on that recognition.
- The reasons for wanting to automatically recognize individuals can vary:
  - They include **reducing error rates** and improving **accuracy**, reducing **fraud** and opportunities for circumvention, reducing **costs**, improving **scalability**, increasing physical **safety**, and improving **convenience**.
  - Often some combination of these will apply.

Biometrics Bottom Level I

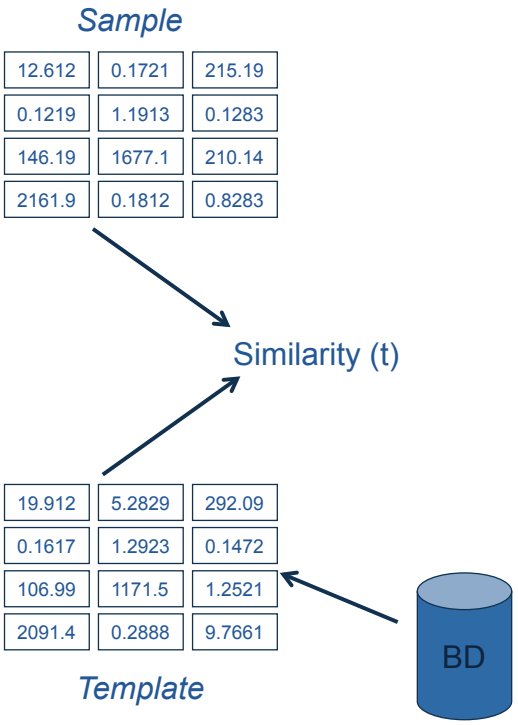
Statistical Analysis →

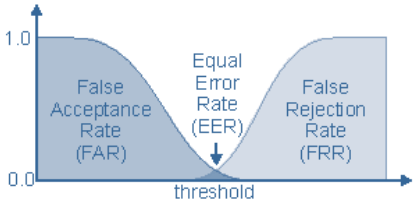
12.612	0.1721	215.19
0.1219	1.1913	0.1283
146.19	1677.1	210.14
2161.9	0.1812	0.8283

Biometric Signature:

•Feature set that is (supposed to be) unique to each individual and allows its identification (verification).

Biometrics Bottom Level II

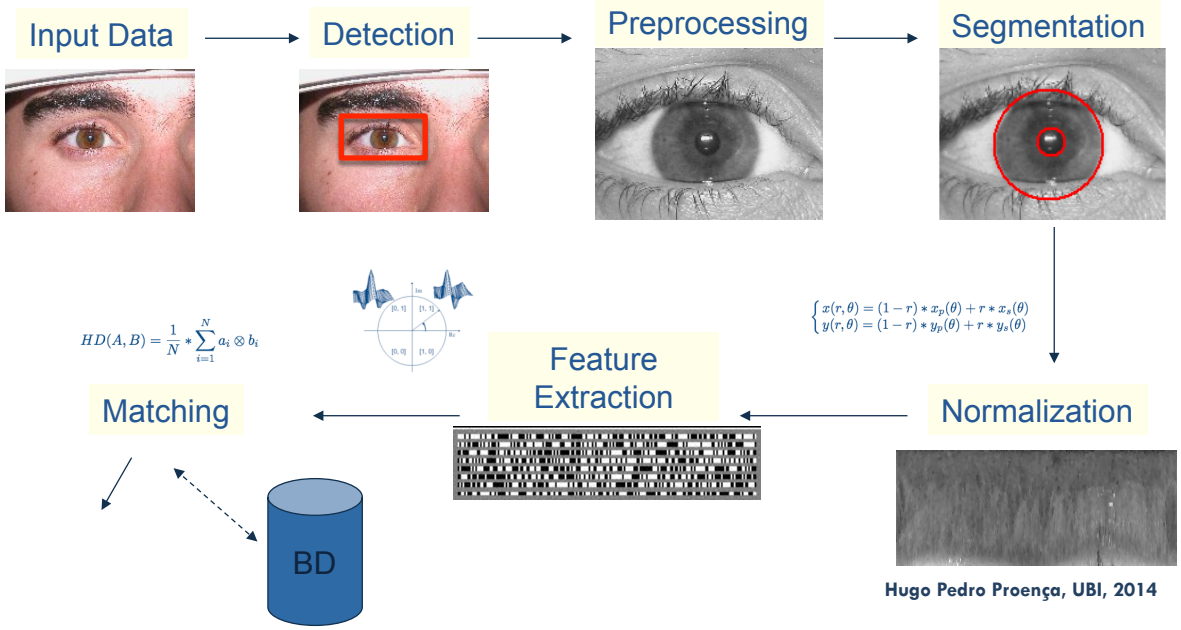




Hugo Pedro Proença, UBI, 2014

Big Picture

$$max_{r,x_0,y_0}\left|G_\sigma(r) * \frac{\delta}{\delta r}\oint_{r,x_0,y_0}\frac{I(x,y)}{2\Pi r}ds\right|$$



Hugo Pedro Proença, UBI, 2014

Biometrics: History

- Let's start discussing a myth:
  - *"The interest on Biometrics is due to the fact of being a novelty, because it's a too **shallow** domain"…*
  - **Question:** Which book do you think it contains the first reference to biometrics?

Biometrics: History

- **Answer:** The **Bible** itself!
  - Revelation 13:16-17, *"He causes all, both small and great, rich and poor, free and slave, to receive a mark on their right hand or on their foreheads, and that no one may buy or sell except one who has the mark or[f] the name of the beast, or the number of his name."*

Biometrics: History

- There are evidences of biometric uses on human history as early as pre-historical age.
  - Estimated 31000 years old caves are adorned with pre-historical pictures apparently signed by fingerprints stamps of authors.
  - Another evidence is the use of fingerprints by Babylonian at 500 B.C.
    - "Babylonian business transactions are recorded in clay tablets that include fingerprints." (Dermatoglyphics, "Hand Analysis", International Institute of Hand Analysis, January 2005)

Biometrics: History

- Even in a non automated way, biometrics has been used - at least - for centuries. In the 14th century, the Portuguese writer João de Barros reported its first known application.
  - He wrote that Chinese merchants stamped children's palm print and footprints on paper with identification purposes.
  - Also, it is believed that the ancient civilizations of Egypt and China performed some type of biometric recognition

Biometrics: History

- In the western world, until the late 1800s the automatic recognition of individuals was largely done using "photographic memory".
  - In 1883, the French police and anthropologist **Alphonse Bertillon** developed an anthropometric system, known as *Bertillonage*, to fix the problem of identification of convicted criminals.
  - This was a quite complex procedure that could take up to twenty minutes and is considered the first scientific system widely used to identify criminals

Biometrics: History

- Bertillonage's basis was the measurement of certain lengths and widths of the head and body and the recording of individual markings, such as tattoos and scars.

Source: http://en.wikipedia.org/wiki/Alphonse_Bertillon

Biometrics: History

- The Bertillon's system faded when it was discovered that some people share the same measures and several people could be treated as one.
- The failure of Bertillon motivated the use of fingerprinting, which is presently almost standardized worldwide.
- In 1880, the British scientific journal Nature published an article by Henry Faulds and William James describing the uniqueness and permanence of finger prints.
  - This motivated the design of the first elementary fingerprint recognition system by Sir Francis Galton and improved by Sir Edward R. Henry.
  - Having quickly disseminated, the first fingerprint system in the United States was inaugurated by the New York State Prison Department in 1903 and the first known convicted due to fingerprint evidences was reported in 1911

Biometrics: Timeline

Year	Description
1858	First systematic capture of hand images for identification purposes is recorded
1870	Bertillon develops anthropometrics to identify individuals
1892	Galton develops a classification system for fingerprints
1903	New York State Prisons begins using fingerprints
1936	Concept of using the iris pattern for identification is proposed
1963	Hughes research paper on fingerprint automation published
1969	FBI pushes to make fingerprint recognition an automated process
1976	First prototype system for speaker recognition is developed
1993	Development of an iris prototype unit begins

Source: http://www.biometrics.gov/documents/biohistory.pdf

Biometrics: Timeline

Year	Description
1994	First iris recognition algorithm is patented
1995	Iris prototype becomes available as a commercial product
1996	Hand geometry is implemented at the Olympic Games
1998	FBI launches CODIS (DNA forensic database)
2001	Face recognition is used at the Super Bowl in Tampa, Florida
2004	US-VISIT program becomes operational
2004	U.S. Presidential directive calls for mandatory government-wide personal identification card for all federal employees and contractors
2004	First statewide automated palm print database is deployed in the US

Source: http://www.biometrics.gov/documents/biohistory.pdf

Biometrics History: Further References

- National Science and Technology Council, Committee on technology, Committee on Homeland and national Security, Subcommittee on Biometrics, "Biometrics History":
  http://www.biometrics.gov/documents/biohistory.pdf

- Dany Huot, Zaman University, "History of Biometrics",
  http://timerime.com/en/timeline/943290/History+of+biometrics/

Biometrics State-of-the-Art: **Iris-On-The-Move**

- A major example of the state-of-the-art engineering behind a biometric recognition system that **almost** does not requires the active participation of subjects in the data acquisition process.

Source: http://gulfnews.com/pictures/news/best-of-staff-october-15-october-21-2013-1.1245778

Hugo Pedro Proença, UBI, 2014

Hugo Pedro Proença, UBI, 2014

[2]

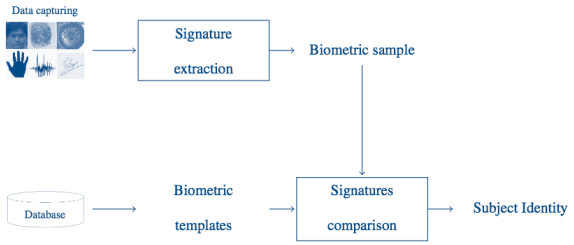
BIOMETRIC RECOGNITION SYSTEMS (MEI)

Syllabus

Theoretical	**Biometrics modes of functioning: verification vs. identification systems; Frequent observers; Classification of biometrics systems: overt vs. covert, habituated vs. non-habituated, attended vs. non-attended, standard vs. non-standard environments, public vs. private, open vs. close; Typical biometric traits: fingerprint, face, iris, palm vein, retina**
Labs	Least-squares basics; Image registration exercises: Using SIFT descriptors and homography matrices

Biometrics

- Regardless the trait used, the biometric applications follow the procedure illustrated below.



- The process begins by the data capturing, where the biometric sample is acquired. Next, through the feature extraction a biometric signature is created, that is further compared with a specific or several biometric signatures registered in the database.
  - These are commonly designated as biometric templates and were collected during the enrollment process and correspond to a verified subject identity.
- If the comparison between biometric signatures has enough similarity, it is assumed that both of these were extracted from the same person, otherwise, they must have been extracted from different persons

Biometrics

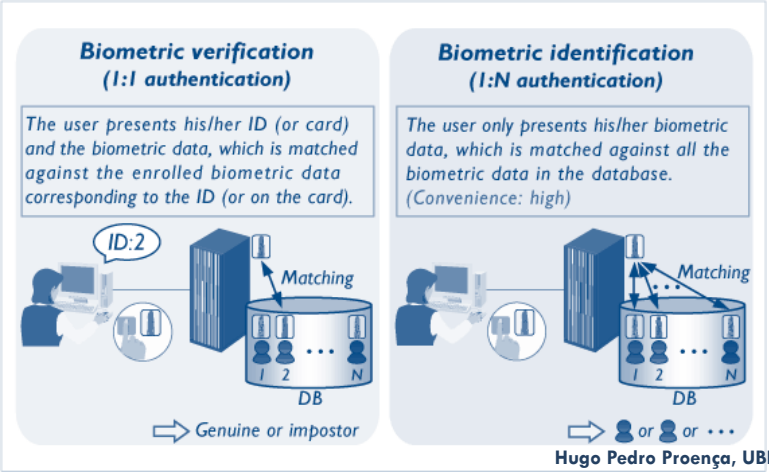
- The number of comparisons between the biometric sample and templates determines a basic distinction between the two modes of performing biometric recognition: **verification** and **identification**.
- In the **verification** mode, the system verifies the authenticity of a claimed enrolled identity, trying to answer the question:
  - **"is this person who he/she claims to be?"**
- This requires that, together with the biometric sample, the subject's id must be given to the recognition algorithm.
- Further, the comparison between the biometric template correspondent to that identity and the sample is performed. If the similarity is high enough, the claimed identity is accepted, meaning that both biometric signatures were extracted from the same person. Otherwise, the identity is denied, meaning that the enrolled and the sample signatures were extracted from different subjects.

Biometrics

- The **identification** mode, often named as negative recognition, tries to answer the question:
  - **"who is this person?"**, or sometimes:
  - **"is this person in the database?"**
- After acquiring the required data and extracting the biometric sample, a comparison is made with the "N" enrolled identities in the database, in order to find the identity from which the sample was collected.
- In this mode, it is usual to output a list of the "k" most probable identities of a biometric sample.

Identification vs. Verification

- Verification can be seen as an iteration of an identification process, where we just compare the biometric sample against one of the enrolled identities.



Source: http://www.hitachi.com/rd/portal/story/idless_biometrics/01.html

Biometrics Classification

- Biometric systems can be classified according to six perspectives, as a function of the characteristics of the recognition procedure itself:
- **Overt / Covert**
  - If the user is aware about the acquisition of his biometric data, the application is defined as overt;
  - otherwise, is defined as covert.
- This is clearly one of the most concerning characteristics of a biometric system, regarding the privacy issue.

Overt Biometric Recognition

- Corresponds to the large majority of the biometric recognition systems deployed.
  - There is the notion of "**user**" of this kind of systems
  - They know that the system is running and **are willing** to be recognized, in order to access to some resource / good.

Source: http://www.macrumors.com/2014/01/21/apple-iris-scanning/

Covert Biometric Recognition

- It is currently restricted for most civilian applications, due to privacy concerns.
- There is not the notion of "user" of the system, as the subjects do not even know they are being recognized by automata
- National Security Agency (U.S.) is said to have research programs about covert recognition

Source: http://www.surveillance-company.com/

Biometrics Classification

- **Habituated / Non-habituated**
  - When the majority of the people that interacts with the biometric system are every-day users, the recognition is performed in the habituated mode.
  - If the average frequency of use from each user is low, the recognition is performed in the non-habituated mode
  - This is relevant to the degree of cooperation and training demanded from the users.

Biometrics Classification

- **Fixed / Indefinite Duration**
  - This characteristic refers to for how long the captured biometric data will be used.
  - It is closely related to the privacy of the biometric data.
    - When the biometric captured data are used for a long period, it has a more privacy risk that the short period of use.
  - It's one of the key questions that anyone interested in deploying a recognition system should answer to authorities.

Biometrics Classification

Optional / Mandatory

- A biometric system that imposes that users provide their biometric data is considered Mandatory.
- In such cases, if one subject refuses to give his biometric data, it may suffer some punishment, i.e., will not access to the desired resource / good.
- In contrary, if a biometric system is optional, the user may decide whether to use or not it, being its identity verified in other ways, such as by presenting a passport or driver's license.

Biometrics Classification

Attended / Non-attended

- If the user is observed and guided by supervisors during the process, the biometric recognition is performed attended;
- Otherwise, the biometric system is considered non-attended.
- Obviously, the easy-of-use of the recognition system is much more relevant in the non-attended mode.
- There is a third family of hybrid systems, whish are attended to non-habituated users and non-attended to the remaining subjects.

Biometrics Classification

- **Standard / Non-standard** environments
  - When all the conditions can be controlled and the recognition takes place indoors within constrained conditions, it is considered that the system is running in a standard environment;
  - Otherwise, the use is called in non-standard environment.

Biometrics Classification

- **Public / Private**
  - If the users are not employees of the organization that owns the recognition system, the application is public;
  - On the contrary, if the users are employees, the application is called private.

Biometrics Classification

- **Template Storage / Identifiable Data Storage**
  - The biometric data may be stored as a mathematical model (template) or as identifiable by human sample (images, audio, etc).
  - Biometric templates are generally only of value when processed through a vendor algorithm, and cannot be linked with a specific biometric characteristic without dedicated processing.
  - On contrary, Biometric images are generally identifiable, and can be associated with a specific individual based on visual or aural inspection.
  - As example, in a fingerprint system can be stored the fingerprint image or only the minutiae information.

Biometrics Classification

- **Open / Closed**
  - This refers to whether the biometrics collected within the system will be shared (open) or remain internal (closed) to the organization carrying out recognition.
  - For example, a fingerprint recognition system used to control the access of the employees to a building and logon to their computer network, may be considered closed if the data aren't shared with other external systems.
  - Other examples are state driver's licenses and entitlement programs.
    - One country may want to communicate with others to eliminate frauds. This would be an open system, in which standard formats of data and compression would be required to exchange and compare information.

Biometrics Classification

- **User Ownership / Institutional Ownership of the Biometric Data**
  - In a user ownership system, the user maintains ownership over his or her biometric information, whereas in other case the public or private institution owns the user biometric data.
    - E.g., by using a smart card
  - This characteristic has to do much with privacy of the biometric data: user control over the collection, usage, and disposal of biometric information is not possible in every deployment scenario, especially in entitlements programs or other public sector uses.
  - Also, it his closely related with the "confidence" a user has on a biometric system
    - *How do I know "for sure" that they will not use my biometric data?*

Biometrics Classification

- **Let's analyze a practical case:**
  - "The Immigration and Naturalization Service's Passenger Accelerated Service System (INSPASS) is a biometric system used since 1990 at many airports of USA and Canada, with the aim to reduce the inspection processing time for authorized travelers. The INSPASS system uses the hand geometry biometric for the verification of the travelers. In order to used this system the traveler must obtain a INSPASS card in the enrolment centre before traveling, there the hand geometry template is captured and stored in the card. When the travelers next arrive to the airport, they go directly to the INSPASS terminal which is able to read the INSPASS card and allow the person to key in their flight details. The terminal is connected to a remote computer which checks the validity of the card and the traveler must then place their hand on the hand geometry reader to give a new sample for comparison with the template printed on their INSPASS card."

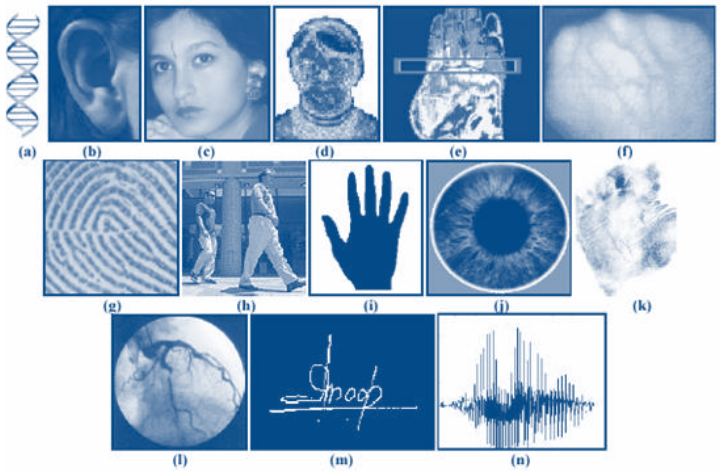
Example taken from: http://www.griaulebiometrics.com/en-us/book/understanding-biometrics/introduction/applications

Biometrics Classification

- **The INSPASS is:**
  - **Overt** because the travelers are aware that they are required to give a biometric measure as a condition of enrollment into this system.
  - **Optional** because the passengers can choose the normal inspection control if they do not want to use the INSPASS system.
  - Of **fixed duration** because the validity of the enrolled data is normally one year.
  - **Public** because enrollment is open to any frequent traveler into the United States.
  - **Closed** because INSPASS does not exchange biometric information with any other system.
  - Runs in **Non-habituated** mode because most international travelers use the system less than once per month.
  - Has **institutional ownership** of biometric data, because the hand geometry templates are also stored on a central computer of INS, each time when a person is verified, the template is updated.
  - It is **template storage** because only the geometric information of the hand is stored, the template has 14 bytes of size.
  - It is **non-supervised** and in a **standard** environment because collection of the biometric will occur near the passport inspection counter inside the airports, but not under the direct observation of an INS employee.

Biometric Traits

- There is an increasingly large number of biometric traits
  - Some are classical (e.g., face, iris, and fingerprint)
  - Many others are quite recent, and are constantly being proposed new possibilities
    - What do you think about the "sound of gait?" ☺



Source, Further Reading: A (possible biased) comparison. http://epic.org/privacy/surveillance/spotlight/1005/irid_guide.pdf

Biometric Traits: DNA

- The **deoxyribonucleic acid (DNA)** is represented through a one-dimensional code, unique for each person. The only exception are identical twins, which can represent a serious problem, regarding security and forensic applications.
  - DNA identification is based on techniques that use the non-coding tandemly repetitive DNA regions. Humans have 23 pairs of chromosomes containing their DNA blueprint. One member of each chromosomal pair comes from the mother, the other comes from the father.
  - Every cell in a human body contains a copy of this DNA. The large majority of DNA does not differ from person to person, but 0.10 percent of a person's entire genome would be unique to each individual, which represents 3 million base pairs of DNA.
- This method is considered to have **some drawbacks,** as the **easy contamination** and **sensitivity**, the **impossibility to perform real-time recognition** and severe **privacy issues,** due to the fact of the DNA analysis can reveal susceptibility to some diseases.
  - Due to many distinguishable characteristics between the DNA and the remaining biometric traits, the discussion about its inclusion as a biometric trait subsists.

Biometric Traits: Ear

- Using ears for the recognition of individuals has been interesting for the research community for, at least, **100 years**.
- During crime scene investigation, ear marks are often used for identification in the absence of valid fingerprints.
- Although in 1989 Iannarelli analyzed over 10000 ears and concluded about the existence of enough dissimilarities to be used for biometric purposes, **the ear uniqueness is questioned;**
- Commonly, there are at least three methods for ear recognition: taking a photo of an ear, taking earmarks by pushing ear against a flat surface and taking thermogram pictures of the ear.
- However, just as a face maybe covered with a scarf, the ears **maybe partially or completely covered by hair or ear muffles**. This implies the requirement of users cooperation, in order to acquire acceptable ear images. Apart from this, changes in perspective are another common problem.
- Despite these problems, there are some advantages that propitiate its use: the requirement of images with smaller resolution, its uniform color distribution and minor variability in shape with respect to expressions

Biometric Traits: Face

- The importance of facial features for human recognition cannot be overstated. Facial images are the most common biometric characteristic used by humans to perform personal recognition.
- This is a **non intrusive** and suitable trait to perform covert recognition. Three types of feature extraction methods can be distinguished: generic methods based on edges, lines, and curves; feature-template-based methods that are used to detect facial features such as eyes and structural matching methods that take into consideration geometrical constraints on the features.
- Although performance of commercially available systems is reasonable, there is still significant room for improvement, since false rejection rate is about 10% and the false accept rate is about 1%. These systems face strong **difficulties** when the faces are captured **under different angles** and uncontrolled ambient illumination. Moreover, it is still questioned if a face itself is sufficient basis for reliably recognition of a subject, as, for instance, **twins have very similar faces.**
- Another problem could be with **counterfeit,** as users can dramatically change the appearance of their face, through decorative objects or even through plastic surgeries.
- As main advantages, it must be enhanced **the high acceptability and universality** of face recognition. Users perceive the recognition system as an automated mechanism that exclusively performs a trivial task.

Biometric Traits: Body Thermogram

- It is possible to capture the pattern of **heat radiated by the human body** with an infrared camera. That pattern is considered to be unique for each person, enabling its potential use for biometric purposes.
- It was observed that the capturing of body images through an infra-red camera produces a unique signature when heat passes through the tissue and is emitted from the skin.
- These signatures are often called thermograms, being claimed that a thermogram is unique to each individual and is less vulnerable to disguises. Face thermograms may depend heavily on a number of factors such as the emotional state of the subjects, or body temperature, and like face recognition, face thermogram recognition is view-dependent [51].
- It is a noninvasive method, although image acquisition is rather difficult when other heat emanating surfaces are near the body.
- Another advantage could be the potential use to perform covert recognition, as well its difficult counterfeit. Even plastic surgery, which does not reroute the flow of blood through the veins, is believed to have no effect on the formation of the face thermogram.

Biometric Traits: Hand Geometry

- Systems that measure hand and finger geometry use a digital camera and light. The aligning of the fingers against several pegs is required to ensure an accurate reading. Then, a camera takes one or more pictures and uses this information to typically determine the length, width, thickness and curvature of the fingers.
- As it is considered a simple technique with high user acceptability and minimal intrusion degree, commercial hand geometry-based verification systems have been installed in hundreds of locations around the world.
- Its relatively low level of uniqueness leads that businesses and schools, rather than high- security facilities, typically use hand and finger geometry readers to authenticate users, not to identify them. Disney theme parks, for example, use finger geometry readers to grant ticket holders admittance to different parts of the park.
- Finger-based biometric systems have a few strengths and **weaknesses**. Since fingers are **less distinctive** than fingerprints or irises, some people are less likely to feel that the system invades their privacy. However, the **geometry** of **fingers** changes over time due to **injury, changes in weight or arthritis** and can be affected by several other factors (e.g., **jewelry**)

Biometric Traits: Fingerprint

- A fingerprint is a pattern of **ridges** and **furrows** located on the tip of each finger. Fingerprints were used for personal identification for many centuries and the matching accuracy is acceptable.
- The uniqueness of a fingerprint can be determined by the **pattern of ridges** and furrows as well as by the **minutiae points**. These are local ridge characteristics that occur at either a **ridge bifurcation or a ridge ending**. The feature values typically correspond to the position and orientation of certain critical points, known as minutiae points.
- The matching process involves comparing the two-dimensional minutiae sample and template patterns.
- Among the main advantages for the use of fingerprints are the higher levels of **acceptability** and their easy of use, as well the fact that it is a **mature** technology with many years of proven effectiveness. Also, the fact that its technology is legally accepted and that **millions** of enrolled **fingerprints exist**, are important.
- As disadvantages, it is considered vulnerable to **noise** and distortion brought on by dirt and twists. Also, since physical contact between the finger and the scanning device is required, the surface can become oily and cloudy after repeated use and reduce the sensitivity. **Hygienic** considerations must be considered too.

Biometric Traits: Gait

- Although it was originally performed through the use of physical devices attached to the subjects' legs, the vision-based gait biometrics has recently received a lot of attention, and the first known effort towards recognition was made by Niyogi and Adelson in the early 1990s.
- The human gait is a **periodic** activity with each gait cycle covering two strides: the left foot forward and right foot forward strides. Each stride spans the double-support stance to the legs-together stance as the legs swing past each other and back to the double-support stance.
- Discriminating information in the basis of gait biometrics can derive from two aspects: shape and dynamics. Based on these two types of information, gait recognition is commonly performed through three distinct approaches: **temporal-aligned based** (analysis of the time-series features as, for instance, of the whole subject silhouette), **static-parameter based** (analysis of parameters that characterize gait dynamics, such as stride length, cadence and speed) and **silhouette shape-based** (emphasizes the silhouette shape similarity and disregards any temporal information).
- However, gait **vulnerability** to changes in the **walking surface**, walking **speed** or in carrying conditions were reported. Due to these, gait-based biometric systems tend to present high false rejection rates. Also, since video-sequence is used to capture the required data, it is considered as one of the most computationally expensive methods.
- The fact that it can be easily and **conscientiously modified by the users**, contributes to a higher probability of circumvention

Biometric Traits: Iris

- The iris begins to form in the third month of gestation and the structures creating its pattern are largely complete by the eighth month. It is the annular region of the eye bounded by the pupil and the sclera (white part of the eye) on either side. Its complex pattern can contain many distinctive features such as **arching ligaments, furrows, ridges, crypts, rings, corona, freckles and a zigzag collarette**.
- Each iris is unique and even irises of identical twins are different. Furthermore, the iris is more easily imaged than retina; it is extremely difficult to surgically tamper iris texture information and it is possible to detect artificial irises.
- To obtain a good image of the iris, recognition systems typically illuminate the iris with near-infrared light, which can be observed by most cameras yet is not detectable by humans.
- Among potential disadvantages for its use remains the **weak public acceptance** of the iris imaging for biometric purposes, due to old-fashioned thoughts about iridology.
- The available results of both accuracy and speed of iris-based identification are highly encouraging and point to the feasibility of **large-scale recognition** using iris information.
- Due to this and to the above described characteristics, it is common to consider iris as one of the best biometric traits, although this evaluation is dependent of the specific purpose

Biometric Traits: Keystroke

- It is believed that each person types on a keyboard in a distinguishable way, such that the analysis of the different rhythms that a subject types in the keyboard can be used for its recognition.
- This technology examines **either dynamics as speed and pressure**, the **length** of **time** each key is held down, the **time elapsed** between hitting certain keys and the **tendencies** to switch between a **numeric keypad** and **keyboard numbers**.
- The extracted features are statistical in nature and specifically designed to characterize the keystroke dynamics over writing samples of 200 or more characters
- The main advantage of the use of keystroke-based biometrics is its potential for **continuous monitoring**. In opposition to other traits, the keystroke information can be continuously analyzed by the recognition system, decreasing the probability of active counterfeit measures.
- Moreover, since users are accustomed to authenticating themselves through usernames and password, most keystroke biometric methods are **completely transparent** and are **well accepted** by users.
- Among potential disadvantages, **privacy concerns** must be considered, as the way a subject strokes can be used to infer information about its **work effectiveness**, for instance.

Biometric Traits: Odor

- Olfaction has an extremely high importance in the human being. Since it is one of the five main senses, many philosophers and scientists have tried to comprehend the sense of smell.
- It is a difficult task, because people often have problems with finding words even to describe their smell sensations. The main problem, associated with odor perception is that there is no physical continuum as sound frequency in hearing or Newton's circle in color vision.
- The odor biometrics is based on the fact that, virtually, each human smell is unique. Common sensorial systems are composed by three main stages: calibration, recording and restoration. The first is continuously performed in order to establish a baseline response, denoted as "response to fresh air". Later, when the signal suffers significant changes, the recording of the subject's data is performed and the final stage corresponds to the time required to restore the sensor to the initial stage.
- Apart from being potentially **affected** by **deodorants** or **perfumes**, it is claimed that the human odor is strongly affected by **seasonal habits**, diets or medication treatments.
- Moreover, the use of body odor sensors brings up the **privacy issue** as the body odor carries a significant amount of sensitive personal information and it is possible to diagnose diseases or activities of the last hours.

Biometric Traits: Retina

- Retinal scan measures the blood vessel patterns in the back of the eye. The pattern formed by veins beneath the retinal surface in an eye is stable and unique and is, therefore, feasible for recognition.
- Digital images of retinal patterns can be acquired by projecting a low intensity beam of visual or infrared light into the eye and capturing an image of the retina using optics similar to a retinoscope. The fact that the retina is small, internal, and difficult to measure makes the capturing of its **image more difficult** than for most of the other traits;
- Retina information procedures usually apply edge enhancing techniques and vessel crossings localization. Other techniques rely on the identification of vessels using adaptive thresholding techniques, followed by graph-matching techniques that find the best match between the vessels localization.
- Since it is protected in the eye itself, and is **not easy to change or replicate the retinal vasculature**, this is considered as one of the most **secure biometric traits**. Retina based systems are used for **high security applications**, as the access to prisons.
- Oppositely, because users perceive the technology as **intrusive**, **unfriendly** and with high cooperative demands, this type of biometric trait has not gained high popularity. Also, it is accepted that retinal vasculature can reveal some medical conditions (e.g., hypertension or diabetes), which is another factor deterring the public acceptance of retinal scan-based biometrics. The **high cost** should be referred too.

Biometric Traits: Signature

- Signature can be regarded as unique and results from both behavioral and hand geometry variations associated to each subject. The way a person signs his or her name is known to be characteristic of that individual since centuries, although the analysis of the signature dynamics is recent.
- There are two major strategies to perform signature recognition: image-based and dynamics analysis
- The first approach is the most classical and is based on the visual appearance of the signature. The latter analyzes both speed, direction and pressure of writing, stroke order and its major weak point results of the specific hardware dependence.
- There are, at least, three advantages over other biometric techniques: it is a **socially accepted recognition** method already in use in bank and credit card transactions, most of the new generation of portable computers and personal digital assistants use handwriting as main input channel and, oppositely to finger, iris or retina patterns, a signature may be changed by the user, similarly to a password.
- However, the use of signature-based biometrics has several weaknesses. Individuals with **muscular illnesses** and people who sometimes sign with only their initials might result in **high false rejection** rates. Often, signatures dramatically **change** over a period of time and are influenced by physical and emotional conditions of the subjects. Additionally, since many users are unaccustomed to signing on a tablet, some subjects' signatures may differ from their signatures on ink and paper, increasing the potential for false rejection.

Biometric Traits: Voice

- Oppositely to the majority of the biometric traits, that are image-based, voice possesses the singularity of **dealing with acoustic information**. The most relevant features of a subject's vocal pattern are determined by physical characteristics as the vocal tracts, mouth, nasal cavities and lips shape. These are **low varying features** over adult lifetime, although the individual behavior and social environments can highly influence the subject's voice.
- Feature extraction techniques typically measure formants or sound characteristics unique to each person's vocal tract and the pattern matching algorithms are similar to those used in the face recognition.
- Speech-based authentication is currently **restricted to low-security applications** because of the high variability in an individual's voice and poor accuracy performance of typical speech-based authentication systems.
- As advantages, the fact that most **existing voice-based systems are designed for use with standard telephone networks**, makes it possible to support a **broad range of deployment environments** for voice based biometric applications.

Biometric Traits: Factors of Influence

- The origin of discriminating information of a biometric trait is typically divided into into three different families:
- **Genotypic**: are traits that are defined by the genetic individual constitution,
- **Randotypic**: are those formed early in the development of the embryo, and claimed that the shape of these are distributed randomly through the entire population. Usually the latter are considered the most valuable features for biometric purposes due to the necessity of absolute uniqueness feature sets per subject.
- **Behavioral**: are those aspects that a subject develops through training or repeated learning. The analysis of these variants enabled us to establish a parallelism between the phenotypic and randotypic-behavioral trait types. The latter can be seen as two sub classes of the former and specify the type of interaction between subject and environment. If it was formed with subject conscientiousness, it is behavioral, otherwise is randotypic

Biometric Traits: Factors of Influence

- If we analyze the classification of the biometric traits afore described and establish our own classification for the origin of the biometric traits, following the schema proposed by Fried and Bomba, we come out with the following Table, where the first identifies the biometric trait and the others contain the influence of the respective type on the trait. Values are percentile and "100%" and "0%" denote, respectively, maximum and minimum influence, according to the above described information sources

Trait	Genotypic	Randotypic	Behavioral
DNA	48%	48%	4%
Ear	40%	40%	20%
Face	40%	40%	20%
Facial Thermogram	20%	50%	30%
Finger Geometry	25%	62.5%	12.5%
Fingerprint	25%	62.5%	12.5%
Gait	30%	10%	60%
Hand Geometry	25%	62.5%	12.5%
Hand Vein	45%	45%	10%
Iris	1%	98%	1%
Keystroke	1%	1%	98%
Odor	37.5%	25%	37.5%
Palmprint	25%	62.5%	12.5%
Retina	1%	98%	1%
Signature	1%	1%	98%
Voice	42%	29%	29%

Biometric Traits: Comparison

- Biometric traits can be compared according to different criteria, among which 7 are highlighted:
- **Uniqueness**: Is probably the most relevant characteristic of a biometric trait. It measures the degree of dissimilarity of the trait between individuals and its capacity of being separable. This feature strongly determines the probability of false acceptances by the system;
- **Universality:** measures the scope of the trait, the number of people where it occurs. Obviously, the optimal biometric trait should occur in as many people as possible;
- **Permanence**: is the quality of being immutable over time, measuring the probability of the biometric trait suffering significant changes over lifetime. This parameter has strong impact in the false rejection rates of the system;
- **Collectability** or measurability: is the characteristic that expresses the technical and humane easiness in the capture of the relevant trait information. This value plays a role in the users' comfort, which can easily dictate between the biometric system adoption or rejection;
- **Performance**: as to do with the time required to perform the biometric recognition, once the data is captured. It is relevant in order to distinguish between the computational requirements of the recognition process;
- **Acceptability** or intrusiveness: is a measure related with socio-cultural users' concerns, as well as with privacy concerns associated with the data capturing;
- **Circumvention**: measures how easy it is to counterfeit the system, which has high relevance in the security compass.

Biometric Traits: Comparison

- This Table gives an average classification of the most typical biometric traits, according to the previously described criteria, being the scores averaged from 10 different report classifications available on the web.

Trait	Uniqueness	Universality	Permanence	Collectability	Performance	Acceptability	Circumvention
DNA	87%	95%	94%	19%	19%	15%	55%
Ear	46%	58%	85%	50%	53%	100%	50%
Face	44%	92%	50%	84%	25%	99%	37%
Facial Thermogram	95%	100%	25%	100%	69%	85%	100%
Finger Geometry	23%	58%	70%	75%	47%	70%	50%
Fingerprint	78%	47%	91%	62%	98%	49%	71%
Gait	25%	50%	25%	100%	21%	100%	50%
Hand Geometry	54%	57%	54%	78%	50%	67%	59%
Hand Vein	57%	52%	53%	52%	50%	56%	97%
Iris	96%	93%	97%	62%	98%	50%	95%
Keystroke	17%	23%	28%	56%	25%	67%	50%
Odor	70%	89%	85%	25%	21%	50%	37%
Palmprint	96%	50%	100%	50%	97%	50%	50%
Retina	94%	86%	66%	29%	98%	23%	100%
Signature	35%	39%	34%	83%	23%	97%	33%
Voice	39%	49%	31%	59%	23%	99%	33%

Hugo Pedro Proença, UBI, 2014

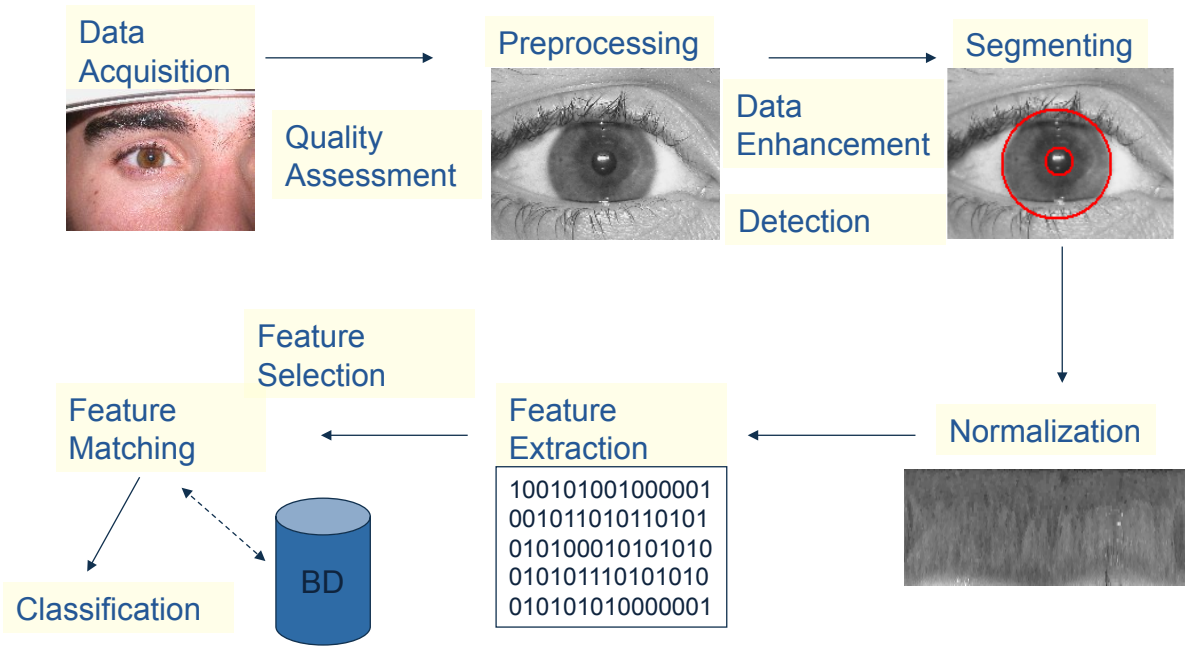
[3]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

Syllabus

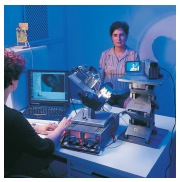
Theoretical	**Biometrics typical processing chain: data acquisition, detection, segmentation, normalisation, feature extraction, classification and fusion; Examples of relevant algorithms for each phase of the processing chain: Adaboost-based detector, Hough transform, Gabor filters, similarity metrics and Bayesian decision rules.}**
Labs	Image filtering basics; Image smoothing; Edge detection: Sobel, Laplacian and Canny detectors

Biometric Processing Chain



Data Acquisition

- The first phase of the recognition is – obviously – the data acquisition phase.
- A imaging device is used, according to a **acquisition setup** and **protocol**, producing a **high-dimensional representation** of the biologic signals (e.g., fingerprint or face images, signature dynamics).
- The acquired data cannot be used directly in the classification module, due to two reasons:
  - Biologic signals suffer of **high variability**, between acquisition sessions. The signal obtained for any person is never exactly the same.
  - To avoid the **curse of dimensionality**, a more compact representation of the signal (the signature) is extracted from the raw signal and will be further used for the comparison.



Data Acquisition Devices

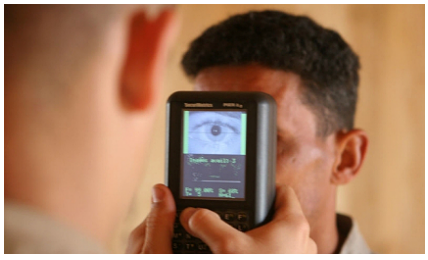
Interflex

HRIndustries

Panasonic

Suprema Inc.

Biolink Solutions



Animetrix

Quality Assessment

- The quality of biometric data refers to the usefulness of a recently acquired sample for recognition
  - The usefulness depends of two factors:
    - Amount of discriminating information (e.g., inherent features acquired)
    - Permanence of information in sample, with respect to templates (e.g., signal-to-noise ratio)
- In short, it consists of an algorithm that analyses a sample and assigns it a quantitative quality score
- A particularly important topic is the detection of fakes, i.e., avoid fraudulent processes;

Example of good / poor quality biometric samples

Quality Assessment

- The main purpose of quality assessment is to avoid that samples are considering for the further processing phases.
  - Possibility 1: They will lead to unnecessary processing, as it will result unavoidably in false rejections
  - **Possibility 2: They might lead to false acceptances**
- There are some traits (e.g., the iris) that guarantee that even for extremely degraded data false matches will not occur
- For such cases, if the processing chain of quality assessment is computationally expensive, the best thing to do is no ignore quality assessment.

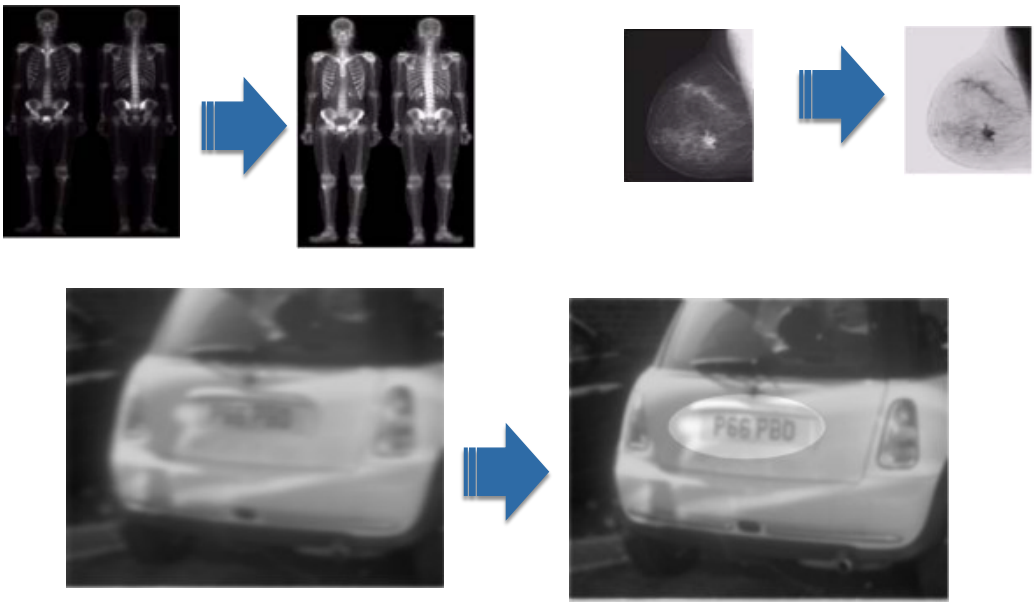
Quality Assessment

- There are, however, 3 main purposes to perform quality assessment
  - Provide quality feedback to users
    - "Please step back"
  - Improve the matching confidence, i.e., by associating the resulting matching score with the quality measured
    - It's not the same thing if two signatures are practically equal, provided that both images from where they were extracted are of good / bad quality
  - Improve the performance, in case of multi-biometric systems
    - Quality score might be used to derive the weights given to each biometric modality
  - The ISO/IEC 29794 is a standard that defines uniform interpretability and inter-operability of biometric quality scores

Data Enhancement

- Data enhancement, is the process of improving the quality of the acquired signal, provided that the quality assessment module reported appropriate scores for the different quality parameters.
  - Knowing the source of degradation, the source is to increase the "signal-to-noise" ratio (SNR)
- Typically, it uses prior knowledge of the acquired signal, to facilitate the subsequent processing phases (e.g., mostly segmentation and feature extraction), or even to provide better visualization
- Different types of normalization (e.g., histogram equalization or histogram stretch) or filtering approaches (e.g., Gabor wavelets) can be employed to separate noise from biometric signals.
- Methods can be divided into two broad families
  - Spatial Domain
  - Frequency Domain
- Most enhancing operations can be done either in the spatial or frequency domain

Image Enhancement: Examples

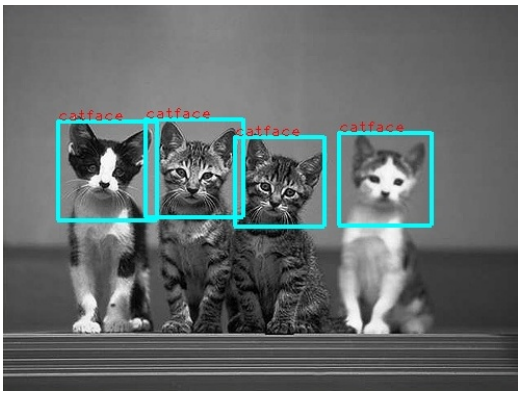


Detection

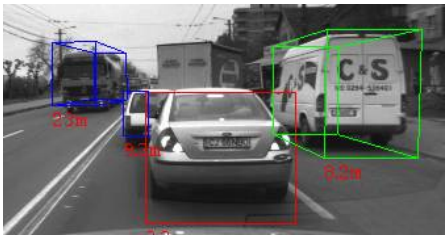
- Detection might be regarded as the first phase of the core recognition process.
- The goal is to roughly localize in the data the instances of the interest object
- It is sometimes considered (erroneously) that its semantic is a bit overlapped with segmentation
  - Detection: "Select the image regions where my interest object is", i.e., define the **Region-of-Interest (ROI)"**
  - Segmentation: "**Determine the boundaries** of my object in that ROI"
- Typically, object detection algorithms extract features and use learning algorithms to recognize instances of an object category.
  - Detection often includes most of the typical phases of a Pattern Recognition process
- Several techniques are typically used in this phase
  - Background subtraction
  - BLOB analysis
  - Keypoint correspondences

Detection: Examples

2D



3D

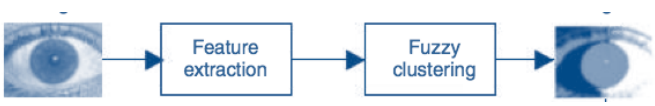


Surveillance

Multiple objects detection

Segmentation

- Segmentation is the process of partitioning a digital image (or preferably a ROI), into segments, assigning a different label to every segment
- A good segmentation algorithm assigns to all pixels that regard the same object, the same label
- The simplest technique to perform segmentation is thresholding, which only works if the intensity / color values between every component of the object are linearly separable.
- Clustering techniques are typically used to reduce the number of colors / intensities in an image, reducing its heterogeneity



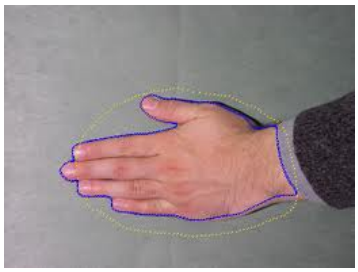
Segmentation

- Compression-based methods reduce the heterogeneity of data, and indirectly group elements into components, by minimizing the coding length
- Histogram-based method are also used to locate clusters in the image, being each cluster regarded as a component
  - They are known as highly efficient, requiring a single pass over the image
- Edge-detection is another typical phase of segmentation
  - By detecting the regions in the image with high variability in small patches, we attempt to highlight the boundaries in the image, that should correspond to object boundaries

Segmentation

- Region growing methods assume that neighbor pixels that regard the same component should have similar color / intensity
  - Watershed is a classical choice for this family of algorithms
- Partial Differential equation-based methods
  - This family of algorithms includes some of the state-of-the-art segmentation approaches (e.g., active contours)
  - The idea is to evolve an initial curve towards the lowest potential of a cost function.
  - However the minimization of the cost function is usually non-trivial and certain smoothness constraints are usually imposed, i.e., geometrical constraints on the evolving curve are added.

Example of active contours:
Yellow: Initial contour
Blue: Optimized contour



Source: https://www.ee.iitb.ac.in/uma/~krishnan/research.html

Feature Extraction

- In this phase, data is processed to extract a set of features, upon two fundamental requirements:
  - The extracted features should have different values for different subjects
  - The extracted features should be as permanent as possible, i.e., for two different sessions of the same subject, the resulting values should be similar
- Features can either have a direct physical counter-part (e.g., minutiae points for fingerprints), or indirectly related to any physical trait (e.g., filter responses for iris images).
- The resulting set is commonly referred to as the biometric signature and will be used as an input for matching.

Feature Selection

- The Feature Extraction phase often produces a large set of features
  - This is particularly evident for strategies with parameterizations that lie in high dimensional spaces
    - Gabor filters
    - Multi-Lobe Differential Filters
- Having a set of features that is too large to be appropriately processed (data storage / processing constraints), the goal is to select only a subset of these features
  - Such subset of features will not lead to substantial decreases in performance
    - Actually, it might even increase performance, due to the curse-of-dimensionality effect
- There is another family of methods that instead of selecting a subset of features, they transform the features into a set of lower cardinality
  - Instead of having x_1 and x_2, they produce $x_3 = K_1x_1 + K_2x_2$
  - Principal Component Analysis (PCA)

Matching

- The matching algorithm compares the signatures extracted from the query sample to the templates stored in the database
  - This produces a score that might represent similarity or dissimilarity
  - Ideally, the matching algorithm should cope with variations in the feature sets, due to noise, occlusions or changes in rotation, pose or perspective
- It is often carried out an alignment of signatures, in order to increase the robustness of the resulting scores against the above factors.
- Integration (fusion) of various feature representations in a matching algorithm or combining different matching algorithms seems to be the most promising way to significantly improve the matching accuracy.

Classification

- In the final stage, having the matching score, it should be provided the decision
  - Both in verification or identification modes, the decision is always the same.
    - "Are the sample and enrolled signatures from the same subject?" YES / NO
- Sometimes, in identification queries, we are not interested in knowing the probable list of enrolled identities for a sample
  - Instead, we are interested in the "most likely" signatures, independently of the absolute value of the score
- The biometric matching algorithms range from simple nearest neighbor algorithms, to sophisticated machine-learning-based methods, such as support vector machines and neural networks.
- Thresholding techniques are used to decide if the distance of the claimed identity (in verification) or first rank (in identification) is sufficient for authentication.

[4]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

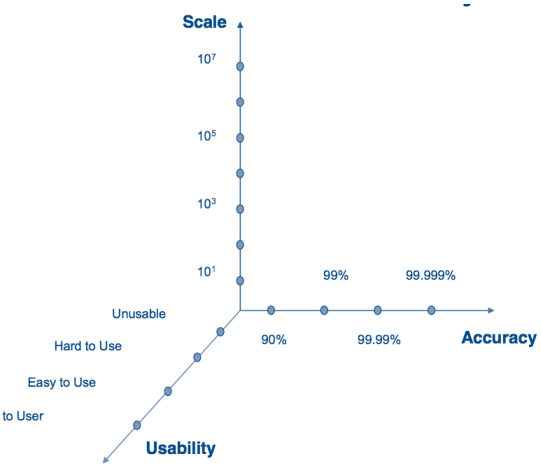
University of Beira Interior, Department of Informatics
Hugo Pedro Proença, 2014/2015

Syllabus

Theoretical	**Notion of system covariate; recognition robustness; robustness against data variation factors: acquisition-noise, changes in translation, scale, rotation, resolution, focus, perspective, occlusions and environmental (lighting) conditions; Challenges behind the development of robust recognition systems. Biometric Menagerie.**
Labs	Color spaces: examples; Global / local image equalization algorithms: histogram equalization and stretch; adaptive histogram equalization; contrast enhancement; Homomorphic filtering.

Biometric Systems

- Any biometric system may be characterized in terms of three main properties:
  - **Accuracy**
    - How often the system makes an error?
  - **Scale**
    - For how many people the system works?
  - **Usability**
    - How usable the system is, i.e., what are the tasks demanded to subjects, in order to enable recognition

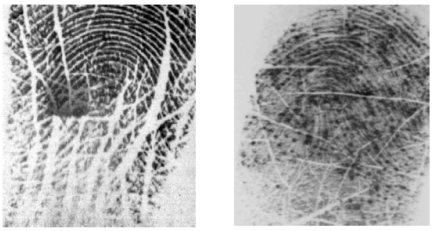


Biometric Systems Limitations

- As expected, **there is no** such thing as the perfect biometric system
  - No matter how effective the system is, there will always be circumstances that lead to failures
  - Luckily, such failures mostly regard false rejections, i.e., legitimate users which the system (erroneously) did not recognized
  - There are three primary reasons for such imperfect performance
  - ***Information limitation***: The discriminating information in the samples are limited due to the intrinsic signal, sensor or acquisition protocol.
    - As we unconstraint the acquisition process, the more variability we will have in data and the higher errors will be.
    - As an example, is it supposed that any face recognizer works in the same way, for such distinct poses as the illustrated below?

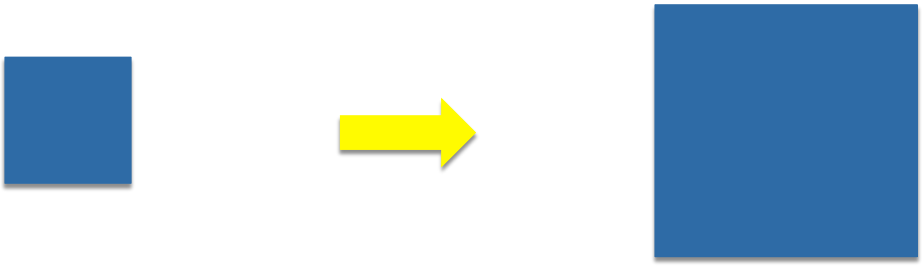
Biometric Systems Limitations

- ***Representation limitation:***
  - Undoubtly, the optimal feature representation should keep all the discriminability between subjects, while reducing the variability due to changes in data, for a given subject.
  - In practice, however, the existing feature extraction systems fail in two ways
    - They do not capture the all the degrees-of-freedom in the biometric signal
    - They are not invariant to many factors, such as scale, translation, or rotation
    - This means that the resulting features will be highly dependent of the signal quality
    - As an example, below we illustrate two fingerprint images from the same subject, which due to their limited quality, are not able to be considered for further processing



Typical Data Variations

- ***Changes in Scale***
  - This type of changes results from different acquisition distances, or from changes in the optical properties of the sensor (e.g., depth of field)
    - Are all the changes inside the object linear with respect to the rate of scale change?
    - If this not happens (deformable object), the problem gets more serious

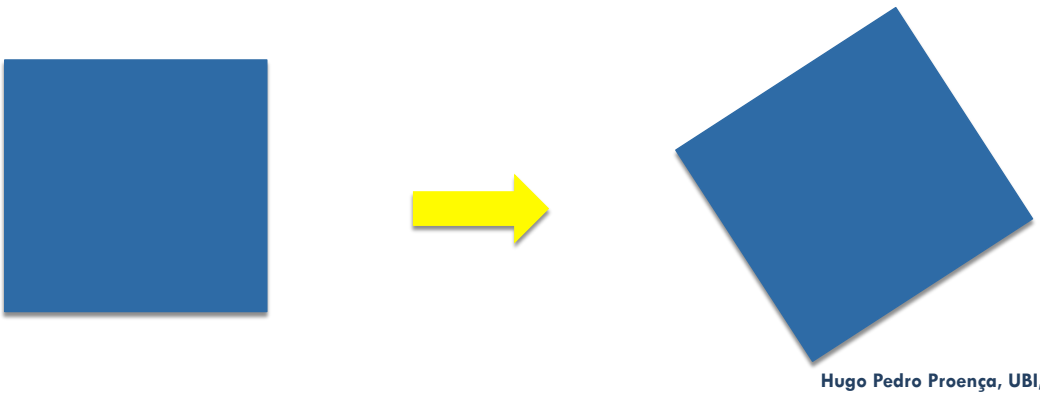


Typical Data Variations

- ***Changes in Translation***
  - This type of changes occurs in result of movements of the object, or of the camera.
  - It is one of the simplest to be compensated, usually simply by a detection module

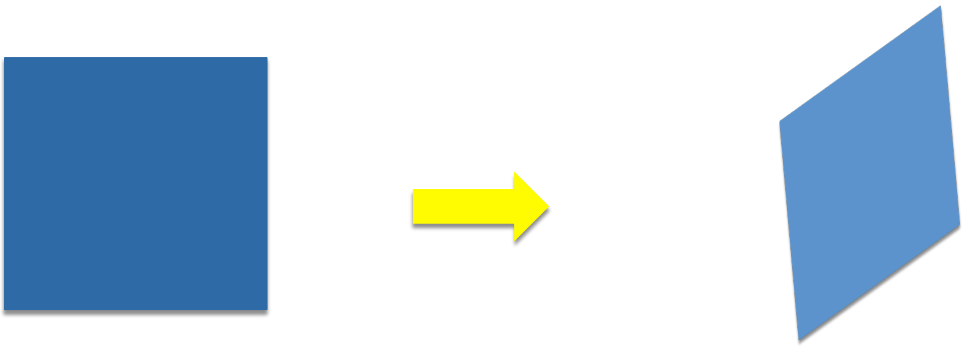
Typical Data Variations

- ***Changes in Rotation***
  - This type of changes occurs in result of (orthogonal) rotation of the object, or of the camera.
  - It is one of the simplest to be compensated, usually simply by a detection module



Typical Data Variations

- ***Changes in Perspective***
  - This type of changes occurs in result of 3D rotations of the object, or of the camera.
  - It is relatively simple to compensate for, if the object is planar
    - Homography



Typical Data Variations

- ***Occlusions***
  - Occlusions might result from other objects that are placed between the interest object and the camera.
  - In practice, it is impossible to compensate for this factor, as the information is not available anymore

System Covariates

- In biometrics research, this term is typically used to refer a variable that possible predicts the outcome of a process
  - In practice, if I have an extremely blurred image, the recognition process will likely result in a NON-MATCH
- Any biometric recognition system is affected by external factors and intrinsic subject features
- Identify these factors and understand their behavior on performance can serve two important goals:
  - to predict the performance of the algorithms at novel deployment sites and
  - to design appropriate acquisition environments at prospective sites to optimize performance

System Covariates

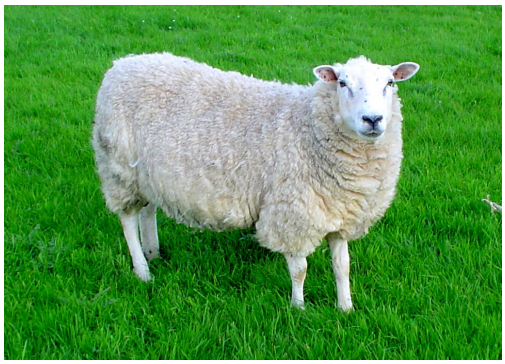
- It is important to understand the influence of the various factors (covariates), such as image quality metrics, population demographic factors and environmental conditions on the performance of biometric recognition systems
- This will determine how biometric systems are designed and implemented in real-world operational scenarios
- System covariates are typically strongly dependent of the biometric trait considered
  - Face: is known to be particularly sensitive to changes in pose, and facial expressions
  - Iris: is known to be particularly sensitive to the levels of iris pigmentation

System Covariates

- Users of a biometric recognition system are not all "**equally hard**" to be recognized.
  - As an example, in the case of iris recognition at visible wavelengths, it is particularly hard to recognize subjects with high levels of pigmentation (dark eyes).
- The concept of **biometric menagerie** consists of eight classifications that categorize subjects that are ideal candidates for biometric recognition
  - Those that provide many true accepts and few false accepts/rejects.
  - Apart these, it will proposes 7 other families of subjects, according to the type of errors that such user propitiates

Biometric Menagerie: Sheeps

- **Sheeps** are the large majority of the biometric system users
  - They provide average levels of false rejections and also average levels of false acceptances
  - Note that "average" refers to the error rates typically obtained by the recognition system



Biometric Menagerie: Lambs

- **Lambs** are legitimate users that are register in the database, but against which there is a large number of other users samples that report a false match
  - This is due to the fact that signatures from many other subjects become identical to the enrolled signature of the "lamb", as acquisition conditions or protocol slightly change

Biometric Menagerie: Wolves

- **Wolves** are users that can be falsely matched against a relatively large number of the enrolled entities
- This is due to the high variability in their biometric signatures, when extracted under slightly different protocols, or acquisition environments

Biometric Menagerie: Goats

- **Goats** are users hat are particularly hard to match against themselves
  - It is not likely that their signature is falsely matched against another's
  - The problem is that the trait being analyzed (in their case) suffers of large variability
  - This means that they will produce always biometric signatures with high variability, which decreases the probability of getting a true-match

Biometric Menagerie: Phantoms

- **Phantoms** are users that hardly match with signatures extracted from their selves or others
  - As a result, they lead almost invariable to a NON-MATCH response

Biometric Menagerie: Chameleons

- **Chameleons** are a relatively rare family of users of a biometric recognition system:
  - They provide always high similarity scores, independently of being a comparison between signatures from the same subject or not
  - They can be confused with every other user of the biometric system

Biometric Menagerie: Worms

- **Worms** are the worst type of users of a biometric recognition system:
  - Provide a high number of false rejections, i.e., given two images of a subject classified as a "dove", there is a relevant probability that the system outputs a NON-MATCH
    - Probably due to anatomical features of the subject trait
    - Or to inconvenience of the data acquisition protocol
  - Provide a high umber of false acceptances, i.e., it is likely that their signatures are confused with signatures from other subjects

Biometric Menagerie: Doves

- **Doves** are the optimal users of a biometric recognition system:
  - Provide a residual number of false rejections, i.e., given two images of a subject classified as a "dove", it is almost sure that the system will produce a MATCH
  - Provide a residual probability of false acceptances, i.e., their signature is never confused with another signature from a different subject

Biometric Menagerie: Handling

- Several attempts have been made to attenuate the effect of biometric menagerie families:
  - Specific users acceptance thresholds
  - Specific users use specific traits
  - Score normalization techniques, according to users history
  - Multimodal biometrics score-level fusion, using users specific weights
  - Feature weighting / selection per subject
  - Specific feature matching algorithms, using weighted components per user

[5]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

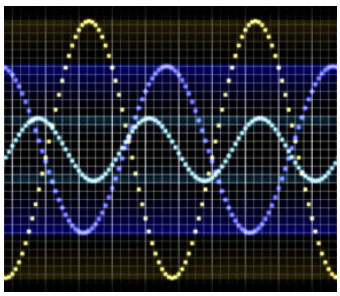
University of Beira Interior, Department of Informatics
Hugo Pedro Proença, 2014/2015

Syllabus

Theoretical	**Signal Processing Basics: linear systems, signal decomposition / superposition. The frequency domain, Discrete Fourier Transform and Fast Fourier Transform. Convolution and filters.**
Labs	Signal processing on MATLAB . Examples: image smoothing and edge detection

Signals and Systems

- What is a "**signal**"?
  - A signal can be regarded as a description how a parameter varies (dependent variable) with respect to another one (independent variable).
- In 1D signals, the voltage of an electric charge varies with respect to time
- In 2D signals, the intensity of a pixel varies with respect its location in image
- Tipically, signals are denoted by upper case letters.
  - Discrete signals are denoted by [].
    - E.g., X[n], Y[k]
  - Continuous signals are denotes by ()
    - E.g., X(i), Y(j)



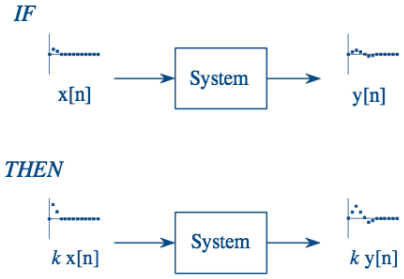
Source: http://www.embedded.com/collections/4396125/Digital-signal-processing-

Linear Systems

- A system is said to be **linear** if it complies two mathematical properties:
  - **Homogeneity.**
  - **Additivity.**
- **Shift invariance** is not a strict requirement for linearity, but it is mandatory for most of the pratical Digital Signal Processing techniques.
- Characterizing the complete input / output properties of a system by **exhaustive measurement** is **impossible.**
- When a system is linear it is possible to use the **responses to a small set of inputs** to **predict** the response to **any input.**
- This saves an amount of work, and turns it possible to **characterize the system completely.**

Linear Systems

- **Homogeneity:**
  - f: R→R, such that f(x)=y. If z=**k**x then f(z)=**k** f(x).
  - In practical terms "*a system is homogenous if an amplitude change in the input corresponds to an identical amplitude change in its output*".



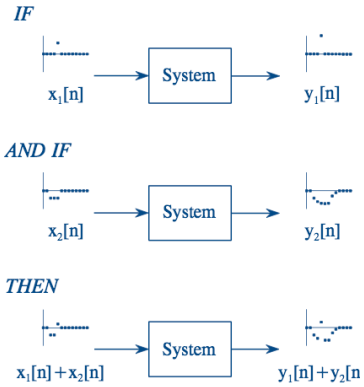
Source: http://www.dspguide.com/

Linear Systems

- Exercise:
  - Consider the following system f:R^2→R, such that:
    - f(x,y)=2x-4y+2
    - Determine the **homogeneity** of f.
  - Now, consider the following system g:R→R, such that:
    - g(x)= exp(x)
    - Determine the homogeneity of g.

Linear Systems

- **Additivity**:
  - f,g: R→R, such that $f(x_1)=y$ and $f(x_2)=z$. If $x_3=x_1+x_2$ then $f(x_3)=f(x_1)+f(x_2)=y+z$
  - In practical terms "*a system is additive if added signals pass by the system without interacting*".



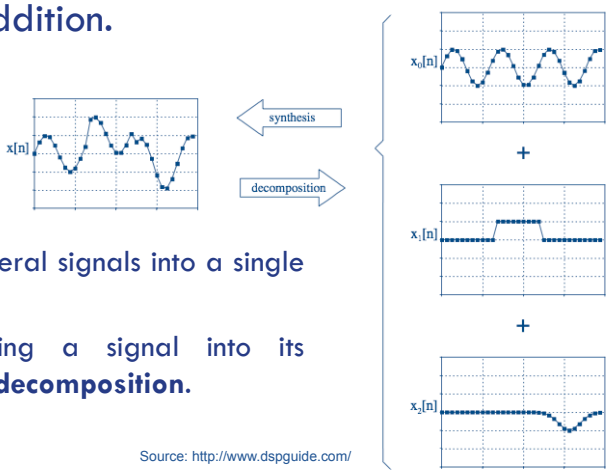
Source: http://www.dspguide.com/

Linear Systems

- Consider the following systems. Evaluate their additivity:
  - f: R→R, such that
    - f(x) = x
  - g:R→R, such that
    - g(x)=0
  - h:R^2→R, such that
    - h(x,y)=xy
  - z:R^2→R
    - z(x,y)=x+3y

Superposition of Signals

- When we are dealing with linear systems, the only way signals can be combined is by scaling (multiplication of the signals by constants) followed by addition.



  - The process of combining several signals into a single one is called **synthesis**
  - The inverse process, broking a signal into its fundamental parts, its called **decomposition**.

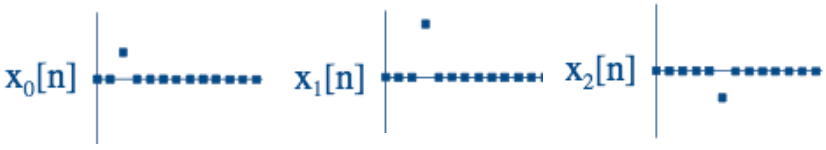
Source: http://www.dspguide.com/

Superposition

- It's the **heart** of signal processing system.
  - It gives the overall strategy to understand how systems and signals are analyzed:
- Having one input signal:

Source: http://www.dspguide.com/

- We decompose it into simpler signals:

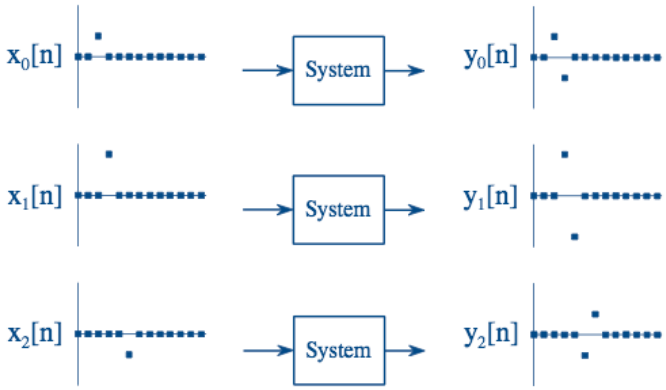


Source: http://www.dspguide.com/

- Remember that our goal is **to understand the system**

Superposition

- Next, each input signal **component passes individually** through the system:



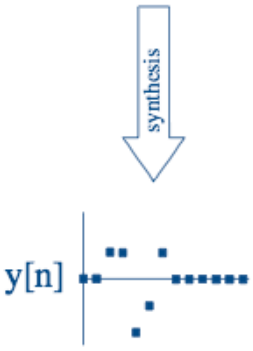
Source: http://www.dspguide.com/

- Yielding the corresponding **output signal components.**

Superposition

- Instead of trying to understand how complicated signals pass through the system, all we need to know is how their components are affected by the system.

- Finally, the output signal components are summed and we get the signal output, exactly equal as if the original signal was passed through the system.



Source: http://www.dspguide.com/

Signal Decomposition

Impulse Decomposition

- Decomposes the original signal "x" (length N) into N signals, where each component contains only one non-zero value:

  $x_k(k)=x(k)$

  $x_k(j)=0,\ j<>k$

- Impulse decomposition is important because it allows signals to be examined one sample at a time.
- By knowing how a system responds to an impulse, the system's output can be calculated for any given input. This approach is called **convolution** and will be the subject of further discussions.

Exercise: Consider the following signal, represented in time-domain:

[2,3,-4,1,0,5,2,4]

- Use impulse decomposition to the above signal and extract the resulting impulses.

Signal Decomposition

Fourier Decomposition

It resulted from a very important finding, by J. Fourier

- "*Any periodic signal can be decomposed by a (potentially infinite) sum of simpler periodic signals*".

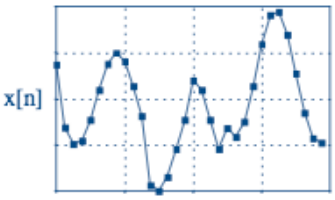
In practice, it decomposes any N length signal into N+2 signals, half of them sin waves and the remaining ones cosine waves.

- The first cosine component has fundamental frequency 0. The second has fundamental frequency 1, …
- Similar observations for the sin waves.

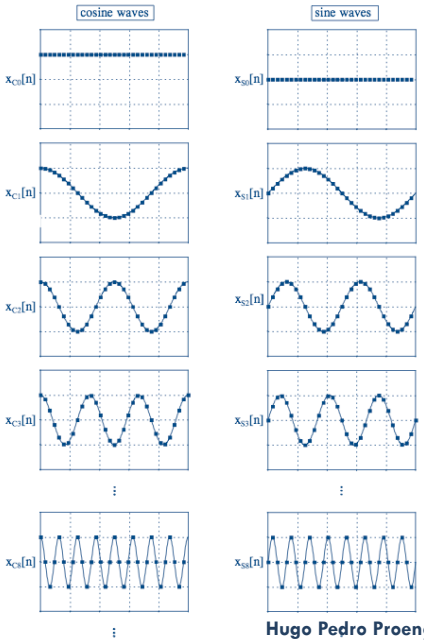
Since the frequency of each component is fixed, the only thing that changes for different signals being decomposed is the amplitude of each of the sine and cosine waves.

Signal Decomposition

- **Fourier Decomposition**

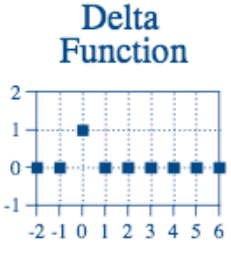


Source: http://www.dspguide.com/



Signal Decomposition

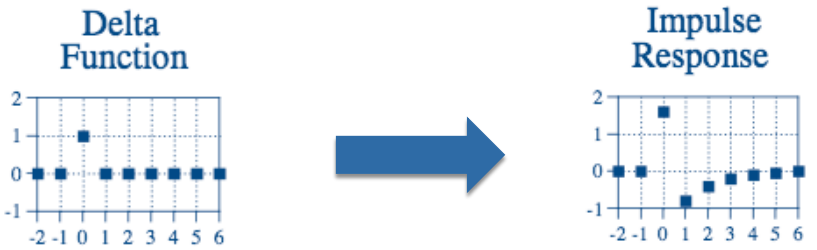
- **Impulse Decomposition**
  - The notion of "**Delta function**" (d) is extremely important, when using impulse decomposition. A delta function has the central component equal to 1and the remaining ones equal to 0.
  - Let $f_k(x)$ be a signal resultant of input decomposition of f(x).
    - $f_k(x) = k\ d(x+t)$. Every input is s scaled and shifted version of the delta function



Source: http://www.dspguide.com/

Signal Decomposition

- **Impulse Decomposition**
  - According to the above discussion, the output signal can be found by adding the output of these scaled and shifted impulse responses.
  - In practical terms, if we know the response of a linear system to an impulse, we know everything about that system.



Source: http://www.dspguide.com/

Source: http://www.dspguide.com/

Signal Decomposition

- **Exercise:**
  - Consider the following impulse response of a 1D signal in a system "f" (centered at index "0").
    - 0, 0, -1, 0, 1, 0, 0
- **Determine:**
  - f(1,2,4,0,-1,1,0,2,3,1,0)

- In general signal processing domain, the impulse response of a system is called "filter kernel", or "convolution kernel".
- In image processing, it is called **point spread function (PSF).**

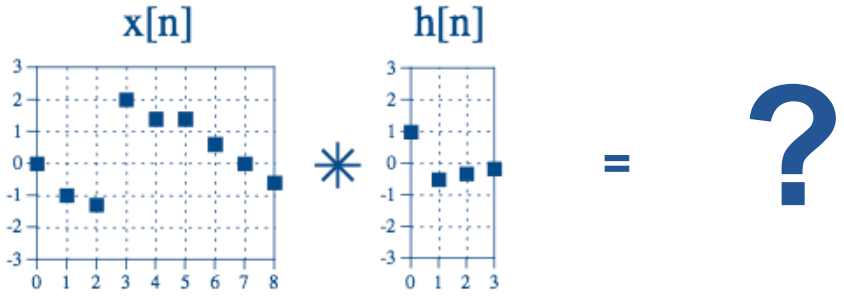
Convolution

- It is a mathematical operation that describes the relationship between three signals:
  - Input signal
  - Impulse response
  - Output signal
- As it combines addition (+) with multiplication (x), it is usually denoted by "*".
  - Y[k]=h[k]*x[k]

$$y[i] = \sum_{j=0}^{M-1} h[j]\,x[i-j]$$

Convolution

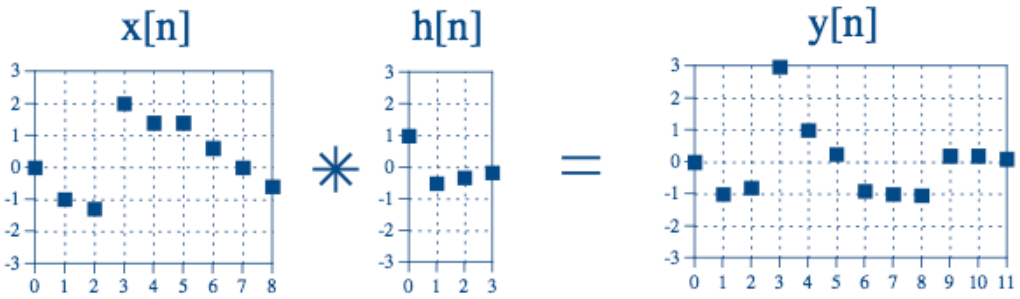
- Exercise: Obtain the result of the convolution of the following signals:



Source: http://www.dspguide.com/

Convolution

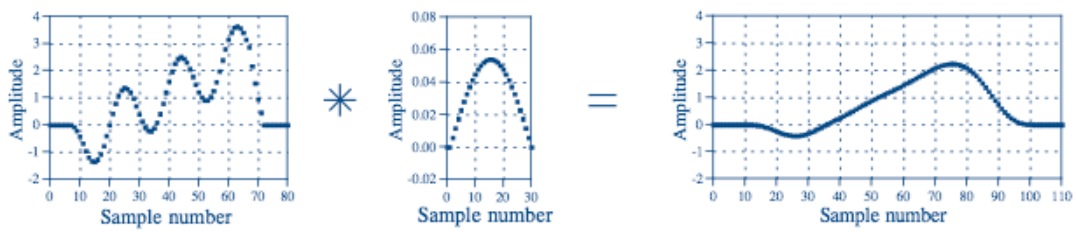
- Exercise: Obtain the result of the convolution of the following signals:



Source: http://www.dspguide.com/

Convolution: Examples

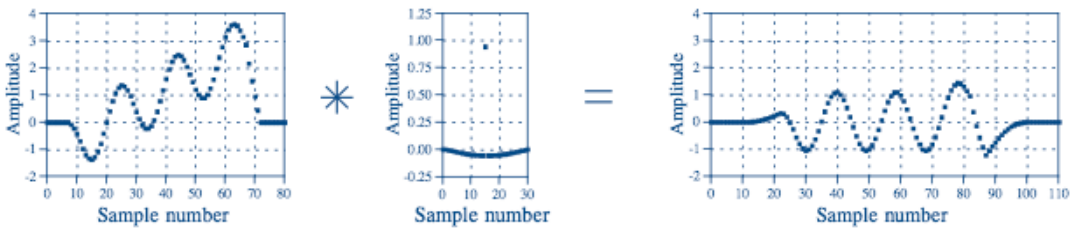
- Example of a low-pass filter, i.e., a filter that keeps only the lowest frequency components.
  - Disregards the details of the signal, i.e., the high-frequency information



Source: http://www.dspguide.com/

Convolution: Examples

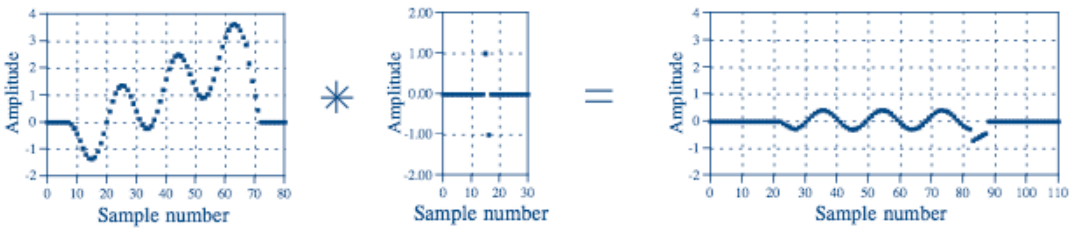
- Example of a high-pass filter that does exactly the opposite of the previous filter, i.e., it keeps only the detail of the signal.



Source: http://www.dspguide.com/

Convolution: Examples

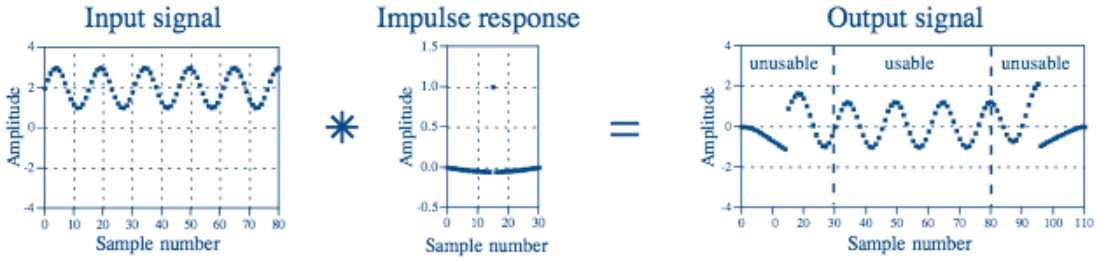
- Example of a discrete derivative filter, i.e., a filter that obtains the local differential between filter components.



Source: http://www.dspguide.com/

Convolution: Caution

- When an input signal is convolved with an impulse response of length "M", then the first and last "M-1" components are not usable.
  - Why is this?



Source: http://www.dspguide.com/

Frequency Domain

- Any signal can be represented by a linear combination of basis-functions.
- In case of 2D images (f):

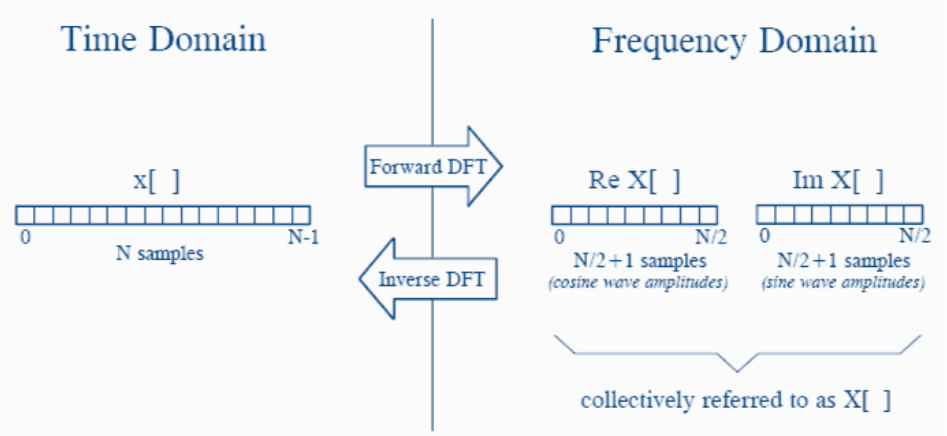
$$f(x,y) = \sum_k a_k \Psi_k(x,y)$$

- Here, a_k are the contributions of each basis-function to the original image.
- Basis functions are exponentials, complex and expressed in terms of harmonic functions (sin and cos):

$$\Psi_k(x,y) = \exp(i(\mu_k x + \nu_k y)) \qquad \exp(i\theta) = \cos(\theta) + i\sin(\theta)$$

Discrete Fourier Transform

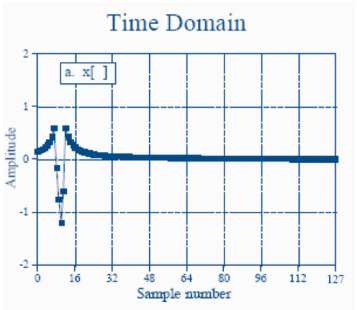
- We can build the following correspondence between any signal represented in the time (space) and frequency domains:



Source: http://www.dspguide.com/

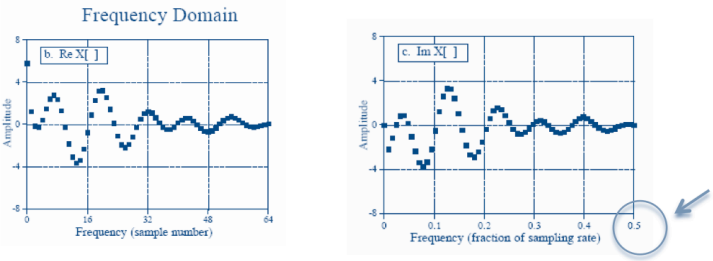
Discrete Fourier Transform

- Suppose we have the following signal, represented in the time-domain:



Source: http://www.dspguide.com/

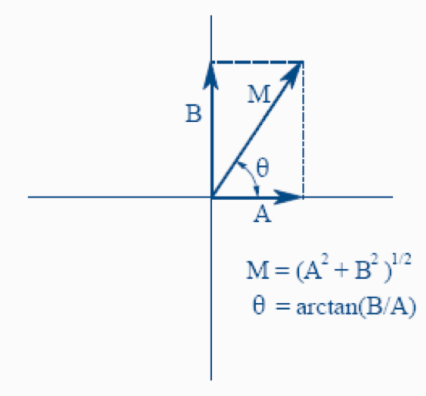
- By using the DFT algorithm, we are able to express it in the following way:



Source: http://www.dspguide.com/

Discrete Fourier Transform

- Often, it is more understandable to express the output of the DFT in Polar coordinates, rather than in the original "real" and "imaginary" components:
  - Mag(X[k])=sqrt(Re(X[k])^2+Im(X[k])^2)
  - Phase(X[k])=arc tan(Im(X[k])/ Re(X[k]))



Source: http://www.dspguide.com/

Convolution Theorem

- The convolution of two signals in a given domain (either spatial or frequency) corresponds to the point-by-point multiplication in the complementary domain.

Convolution

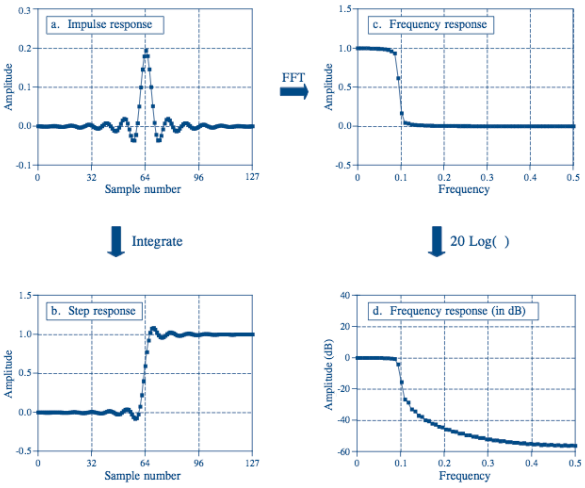
$$H(x)=f(x) * g(x) \leftrightarrow H(x)=F(x) \times G(x)$$

Point-by-point multiplication

- This is extremely important in modern DSP and in practical terms, enabled the existence of most state-of-the-art technologies and devices
  - TV, radio, computer

Filters

- According to the convolution theorem, the convolution in time | frequency domain corresponds to multiplication in frequency | time domain.
  - Each filter has an **impulse response**, a **step response** and a **frequency response**:



Source: http://www.dspguide.com/

Filters

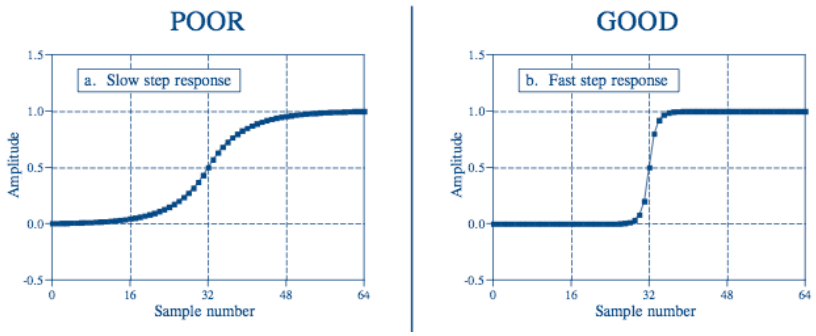
- Impulse Response
  - Output of a system to an impulse
- Step Response
  - Output of the system when the input is a step.
  - It can be obtained without passing any signal thorugh the system.
    - By integrating (running sum in discrete mathematics) the impulse response.
- Frequency Response
  - It can plotted in liner or logarithmic scales (decibels).
  - Corresponds to the Fourier Transform of the Inpulse Response

Filters

- The step response is often osed to measure how well a filter performs in the time domain, mostly in terms of:
  - **Transition speed**. In order to discriminate components of a signal, the duration of the step should be shorter that the spacing of events. Thus, the transition speed should be as fast as possible.
  - Usually expressed by the proportion of samples between a low a high amplitude levels (10 and 90%).

Filters

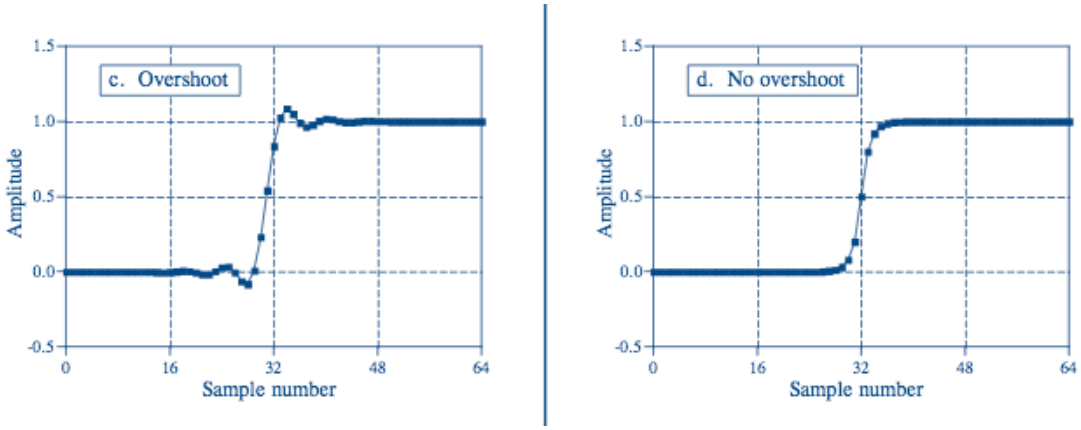
- Example:



Source: http://www.dspguide.com/

Filters

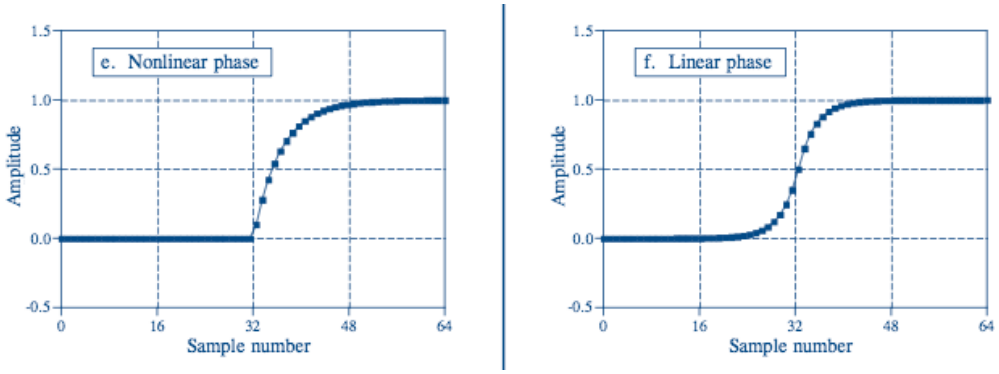
- **Overshoot.** It corresponds to inverse variations to the major variation of step response.
- It changes the signal amplitude non-homogeneously.



Source: http://www.dspguide.com/

Filters

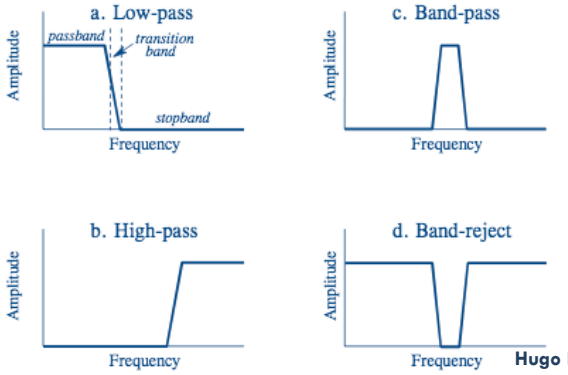
- Linear Phase. Usually it is desired that the upper half of the step response is symmetrical to the lower half. This property is called **linear phase**.



Source: http://www.dspguide.com/

Filters

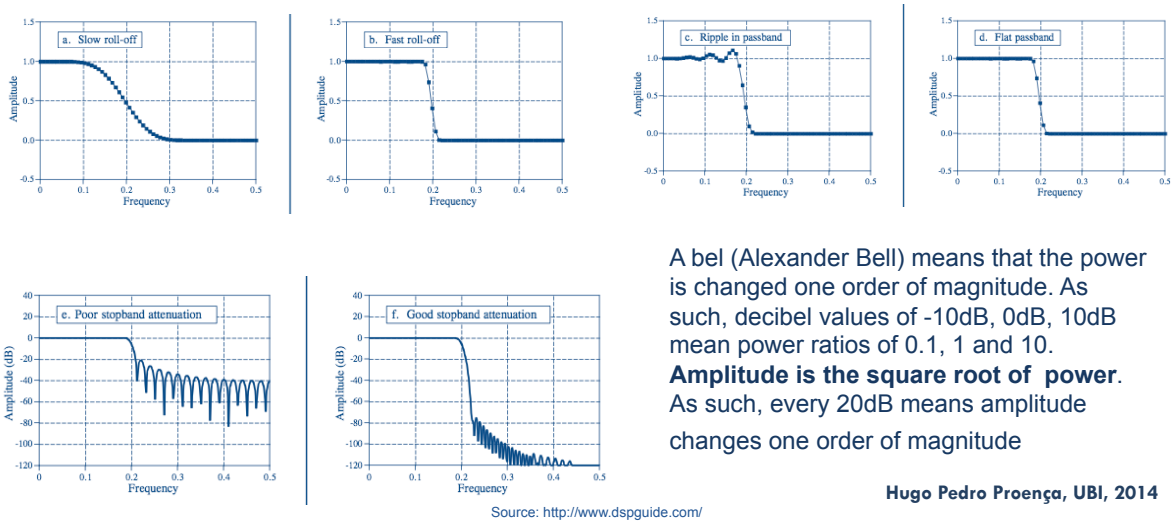
- When analyzing a system in terms of its frequency response, the most important factor is to observe the amount of frequencies that are blocked or passing through the system.
  - The **passband** refers to the range of frequencies that pass trough the system
  - The **stopband** gives the frequencies that are blocked
  - The transition band is on the boundary



Source: http://www.dspguide.com/

Filters

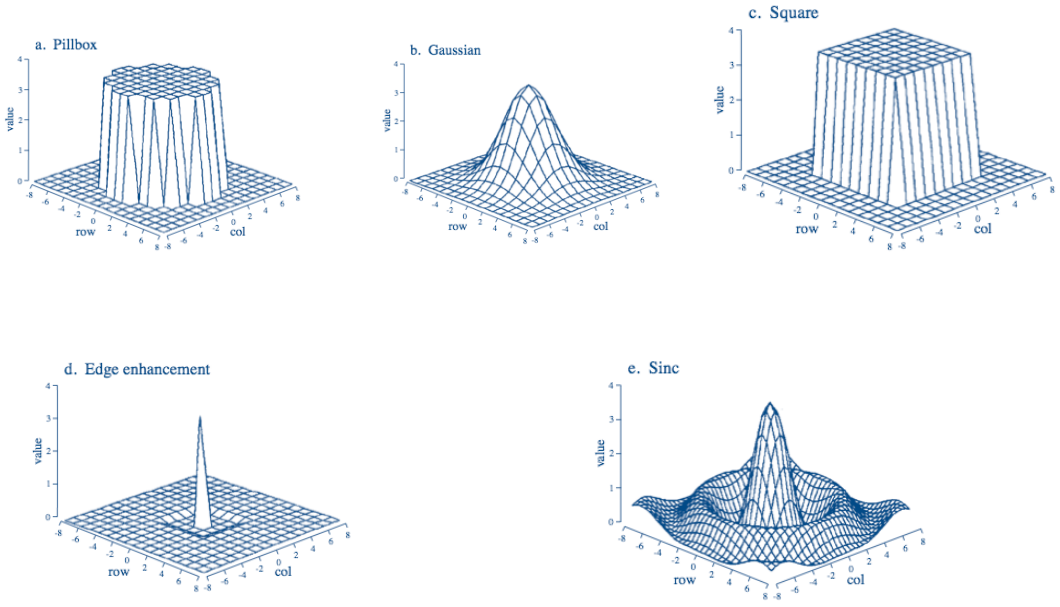
- We are usually interested in filters that have a fast **roll-off** (short transiction band) and without **ripples**. Finally, in order to actually block frequencies, we want to keep good **stopband attenuation** (expressed in logarithmic scale).



A bel (Alexander Bell) means that the power is changed one order of magnitude. As such, decibel values of -10dB, 0dB, 10dB mean power ratios of 0.1, 1 and 10. **Amplitude is the square root of power**. As such, every 20dB means amplitude changes one order of magnitude

Source: http://www.dspguide.com/

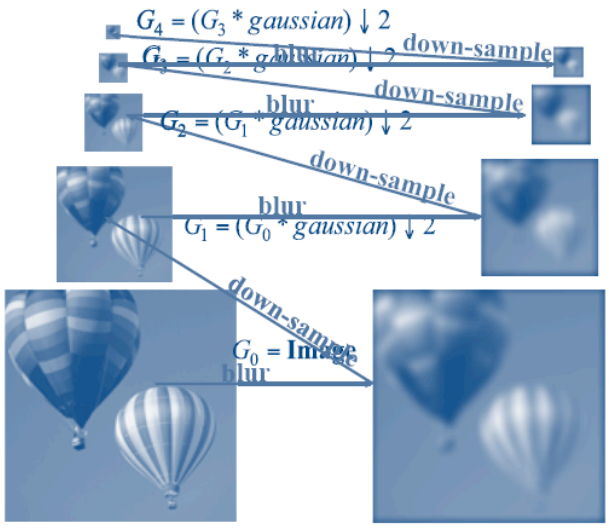
Filters: Examples



Source: http://www.dspguide.com/

Multiscale Analysis

- Usually, features occur in signals at different locations, scales (rotations, …). A widely used strategy is to build a **data pyramid**.
  - Different versions of the same data, each one represented at different scale
  - According to the Nyquist Theorem, only two samples are required to reconstruct a signal with 1 cycle.
  - By blurring each scale with different Gaussian kernels, there is useless information at each scale.
  - That useless information is simply removed from the signal.



Source: http://opencv-python-tutroals.readthedocs.org/

Image Scale-Space

- It is a theory for representing signals at multiple scales.
  - One parameter (t) family of smoothed images. "t" is defined as the scale parameter
  - All structures smaller that sqrt(t) are smoothed away from the "t"-level of the scale-space.
- Let f(x,y) be an image, represented in the spatial domain:
  - f: N x N → N
- Let g(x,y,t) be a family of Gaussian kernels, such that:
  - g:N x N x N → N

$$g(x, y; t) = \frac{1}{2\pi t} e^{-(x^2+y^2)/2t}$$

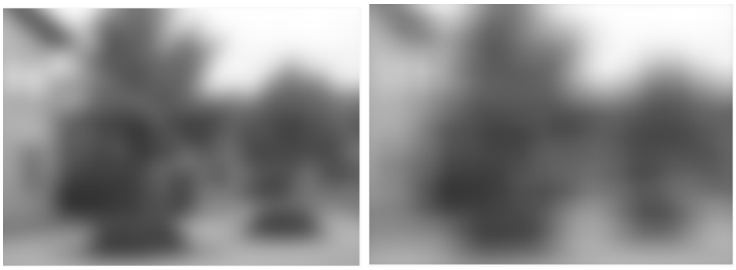
Image Scale-Space

- The scale-space representation at level "t" is given by:
  - L(x,y,t) = f(x,y) * g(x,y,t)
- At the limit, for t=0, g(x,y) becomes an impulse function.
  - As such, L(x,y,0)=f(x,y)
- Example:
  - **L(x,y,0)**

Source: http://opencv-python-tutroals.readthedocs.org/

Image Scale-Space: Example

L(x,y,1) L(x,y,4) L(x,y,16)



L(x,y,64) L(x,y,256)

Source: http://opencv-python-tutroals.readthedocs.org/

Image Scale-Space: Interest Points

- Let $L_x = \vartheta_x / L(x,y,t)$ and $L_y = \vartheta_y / L(x,y,t)$
- An edge at scale "t" is a local maximum of the gradient magnitude:
  - $L_v = \text{sqrt}(L_x + L_y)$
- Similarly, a BLOB (Binary Large Object) at a given scale "t" corresponds to the extreme values of the determinant of the **Hessian Matrix**:
- $\det(H(L(x,y,t)) = L_{xx}L_{yy} - L_{xy}^2$

Image Scale-Space: Interest Points

- Formally, the Hessian matrix is the square matrix of the second order partial derivatives of a function.
- Let $f(x_1, x_2, \ldots x_n)$ be a real-valued function. Its Hessian matrix is given by:

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Hessian Matrix

- Exercise: Consider the following image.
  - a) Obtain its Hessian matrix.
  - b) Locate its edges
  - c) Locate its BLOBs

9	14	23	43	23	26	0
12	16	18	98	187	26	1
12	87	19	47	21	18	17
128	128	200	173	77	76	14
171	38	1	23	2	187	14
72	23	29	93	99	28	200

Edges

- **Edges** are often used to feed the segmentation algorithms.
  - Regions with notorious changes in brightness (color).
  - Areas of high contrast values.
- These areas correspond to the objects boundary, which is used to parameterize its perimeter.
- Edge detectors can be classified into two major classes:
  - First order derivatives
    - Roberts, Prewitt, Sobel, Canny
  - Second order derivatives
    - Laplacian, Marr

Edges: First Order Derivatives

- Sobel
  - This operator calculates the opposite of the gradient at each point, giving the direction of the largest change (light → dark) and the magnitude of that change.
  - The image gradient is defined by a 2D vector, corresponding to horizontal and vertical derivatives.
    - At each image point, the gradient vector points in the direction of largest possible **intensity increase**, and the length of the gradient vector corresponds to the rate of change in that direction.

$$\nabla f = \frac{\partial f}{\partial x_1}\mathbf{e}_1 + \cdots + \frac{\partial f}{\partial x_n}\mathbf{e}_n$$

Edges: First Order Derivatives

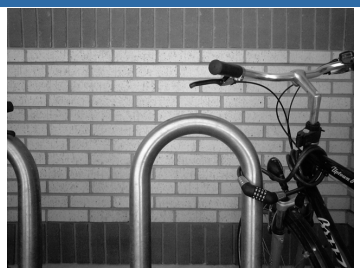
- Sobel Impulse Responses:

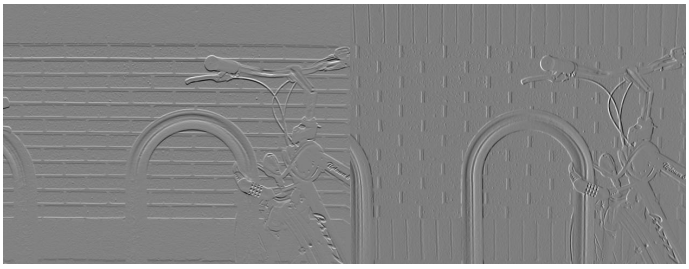
$$\mathbf{G}_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \qquad \mathbf{G}_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}$$

- The gradient magnitude is given by:
  - $G=\text{sqrt}(G_x^2+G_y^2)$
- The opposite of the gradient direction is given by:
  - $\Theta=\text{atan2}(G_y,G_x)$

Edges: First Order Derivatives

- Example:





Y-gradient

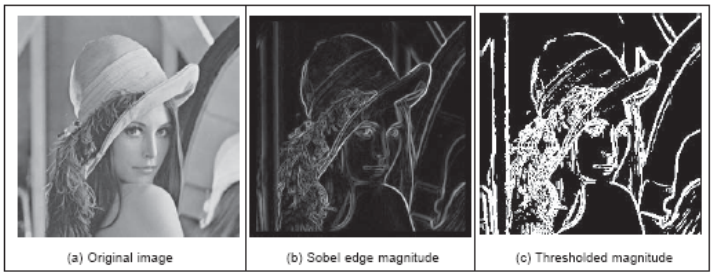
X-gradient

Gradient magnitude

Source: http://johncostella.com/edgedetect/

Edges: First Order Derivatives

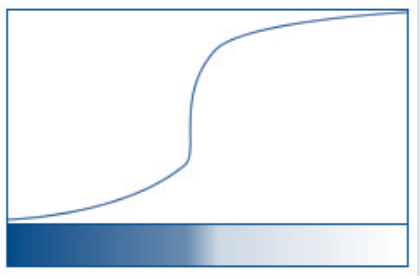
- Finally, the gradient magnitude is thresholded, so that magnitudes higher than a value are considered edges.

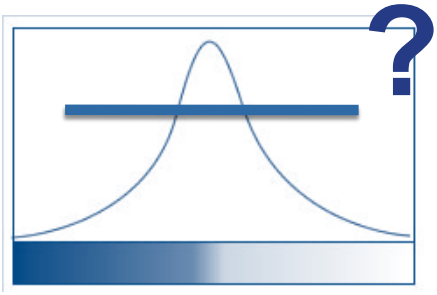


Source: http://johncostella.com/edgedetect/

Edges: Zero-crossing

- Let's consider the following illustration about pixels intensity. Suppose we sample the pixels intensity acorss a row, column of the image.
- As previously seen, first order derivatives consider "edges" all regions where the gradient magnitude is high enough.
  - What's "enough"?

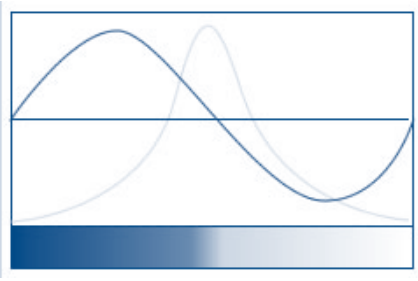




Edges: First Order Derivatives

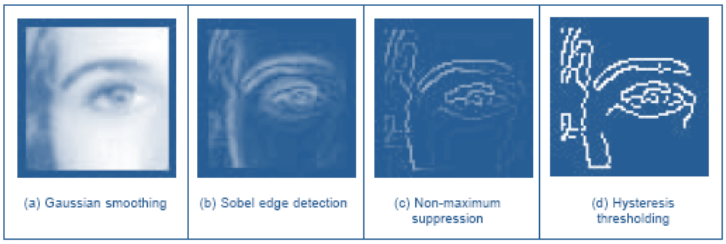
- There has to be a threshold above which a peak is classified as an edge, else it must be considered part of noise.
- The next key idea: On the left (where the curve is rising), the slope is positive. On the right, the slope is negative. So there must exist a point where there is a zero crossing (point with second derivative equal to 0). That point is the edge's location.
- This is the key idea of Laplacian edge detectors:

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$



Edge Detection

- The Canny edge detector is known to provide the most reliable estimate of edges in images.
  - It starts by approximating the image gradient (Sobel).
  - It performs non-maximal suppression
    - Points with gradient values lower than any element in its neighborhood, are discarded.
    - Finally, an hysteresis process, connects edge pixels.



Edge Detection: Exercise

- Consider the following image.
  - Obtain the Sobel approximations to the vertical and horizontal gradients.
  - Find the set of edge points (for a particular threshold value).
  - Find the set of edge points, according to the Laplacian operator.

102	139	197	136	178	129	27	64	126	242
216	128	209	29	240	231	61	190	37	60
195	20	174	190	133	196	172	168	13	128
122	242	9	153	39	79	189	119	207	118
202	39	255	193	67	126	230	71	217	11
75	44	6	241	130	64	171	118	88	111
36	166	21	170	8	250	67	219	33	171
103	70	199	253	150	14	131	229	117	159
78	135	32	221	100	207	125	220	194	228
33	113	88	32	79	34	7	87	67	165

Hugo Pedro Proença, UBI, 2014

Hugo Pedro Proença, UBI, 2014

[6]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

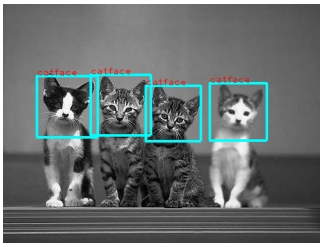
University of Beira Interior, Department of Informatics
Hugo Pedro Proença, 2014/2015

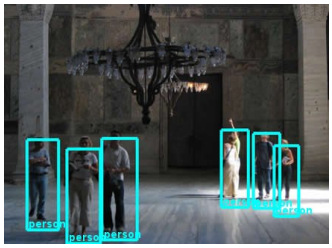
Syllabus

Theoretical	**The Viola and Jones object detector: weak features, integral image, Adaboost-based model learning, object detection**
Labs	Object detection according to MATLAB Computer Vision toolbox; Learning detectors.

Object Detection

- It is usually one of the earliest phases of biometric recognition systems.
- The goal is to **roughly detect some regions-of-interest** (**ROI**) that might contain **instances** of the object to be handled by the system.





Object Detection

- There are multiple varying factors in the acquired data:
  - Lighting
  - Shape (scale, translation, rotation)
  - Perspective
  - Background
  - Intrinsic object deformations
  - Occlusions
- The object detector module **should be able to handle appropriately these variations.**
  - A flexible and robust detector is desired.
  - Ability to handle overlapping instances (post-processing).

Object Detection

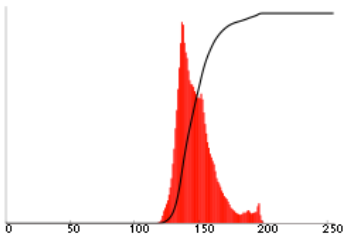
- The **Object Detection** phase might itself contain some of the typical remaining phases of Pattern Recognition Systems. E.g.:
  - Image Preprocessing
    - Data Normalization
    - Local rectification
  - Overcomplete feature set representation
    - **Complete basis** do not have linear dependence between **basis elements** and have the same number of elements as the original space.
    - **This is not the usual case** in feature representation for image detection purposes, as it is hard to find a complete basis.
  - Machine-Learning based techniques to build appropriate models
  - Postprocessing to fuse multiple detections.

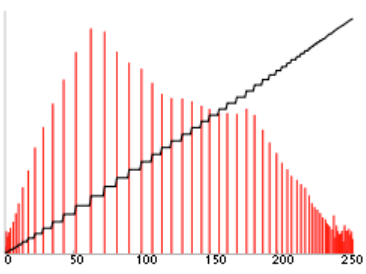
Preprocessing

- It is often though as complementary, but it has notorious impact o performance.
- The goal is to compensate as much as possible for environmental variations in the acquired data.
- **Examples** of techniques used in this phase:
  - Histogram stretch, equalization
  - Homomorphic filtering
  - Center-surround filter

Histogram Equalization

- **Example:**





Homomorphic Filtering

- **Example:**

Source: http://www.mathworks.com/matlabcentral/fileexchange/21357-homomorphic-filtering

Adaboost Object Detector

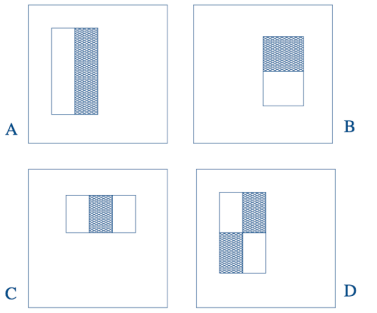
- Originally proposed by Viola and Jones, it is the most popular object detection algorithm:
  - *Paul A. Viola, Michael J. Jones: Robust Real-Time Face Detection. International Journal of Computer Vision 57(2): 137-154 (2004).*
  - It requires a set of training data
    - Labeled positive and negative samples.

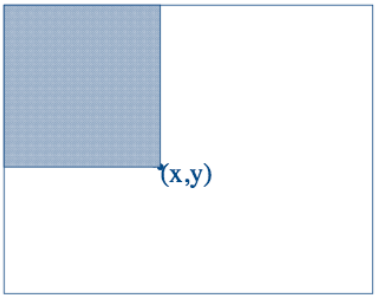
  - Builds a strong detection from a set of very simple (weak) detectors.
  - It exploits the correlation between weak detectors.
  - Able to operate in real-time environments.

Adaboost Object Detector

- Features are reminiscent of Haar-basis functions:
  - The value of each feature (A,B,C,D) is given by the difference between sum of intensities in rectangular regions.
  - In order to obtain each value in an efficient way, the concept of integral image "ii(x,y)"

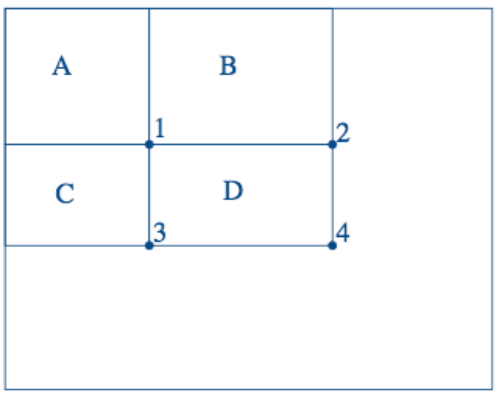
$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x', y'),$$





Adaboost Object Detector

- **Integral image:**
  - According to its definition, it is possible to obtain the sum of any rectangular image region, just by summing (subtracting) four different values:
  - **Region D=4+1+(2-3)**



Adaboost Object Detector

- **Integral Image Exercise.**
  - Consider the following image. Obtain the corresponding integral image
  - Create a MATLAB function that receives one grayscale image and returns the corresponding integral image.

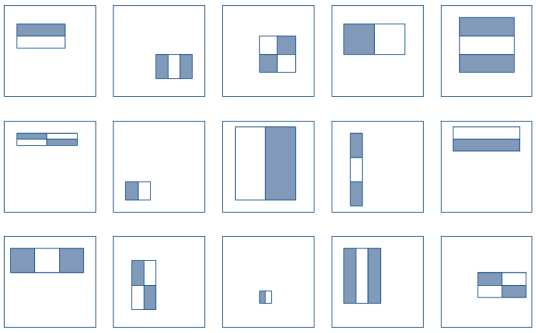
9	14	23	43	23	26	0
12	16	18	98	187	26	1
12	87	19	47	21	18	17
128	128	200	173	77	76	14
171	38	1	23	2	187	14
72	23	29	93	99	28	200

Adaboost Object Detector

- **Weak Classifiers:**
  - Each weak classifier h(x,y,p,t) consists on the analysis of one of the initial types of features, centered at a given position (x,y), where the most discriminating power between negative and positive samples occur at threshold "t" and comparison sign "p"
    - $H(x,y,p,t) = p\ f(x,y) < p\ t$
- For each possible featurea weak classifier is built.
- A set of weights are initialized, according to the total of negative and positive samples
- At each iteration choose the best weak classifier (the one with lowest error)
- Update the weights of the remaining classifiers, so that those that do not fail in the cases where selected weak classifiers fail are previleged
- The final strong classifier is a function o each weak classifier and the corresponding weight.

Adaboost Object Detector

- **Feature Sets.** According to the original formulation of Viola and Jones, for a 24 x 24 region there will be around **180,000** possible **weak features.**
  - Impossible to evaluate the entire feature set at test time (real-time)



Adaboost Object Detector

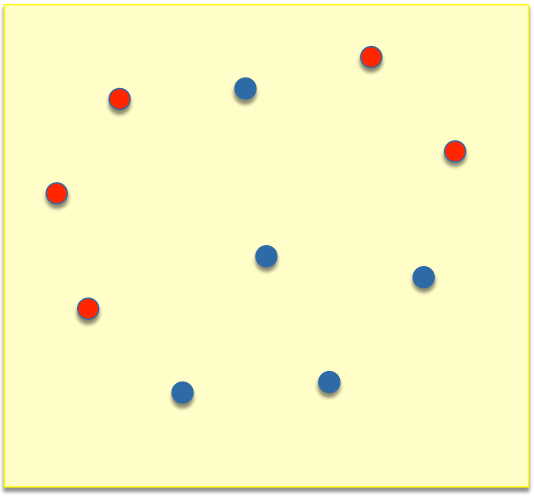
- **Dimensionality Reduction.**
  - The main challenge is to create a good detector, **using only a subset of these features**
- **Problem:** Which features to select?
- Select the best features in an **orthogonal** (independent) way is known to **produce poor results**
  - Due to feature correlation issues
- **Boosting** is the key to this part of the problem
  - Y. Freund and R. Schapire; A short introduction to boosting, Journal of Japanese Society for Artificial Intelligence, 14(5): pag. 771-780, September, 1999.

Adaboost

- Consider that each feature is a "**weak classifier**"
  - A weak classifier is only slightly better than random classification.
- Select the best (actually the least bad) weak classifier (**1st classifier**)
- Classify the data
- Look at where it makes errors
- Reweight the data so that the inputs where errors were observed get increased weights
- Select the best weak feature on the weighted data (**2nd classifier**)
- Combine the 1st and 2nd classifier and weight the inputs according to where they make errors
- Learn the **3rd classifier** on the weighted data
- …
- Learn the **Nth classifier** on the weighted data
- Final classifier is the combination of the set of N classifiers

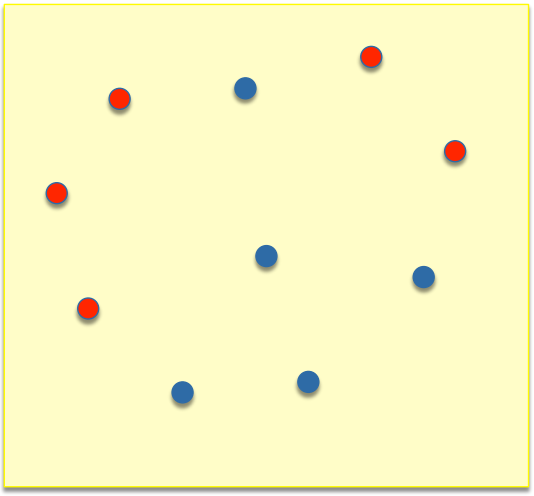
Adaboost: Key Insight

- Suppose that we have a 2D feature set {A,B}, belonging to 1 of two classes, where instances can be graphically represented in an intuitive way:



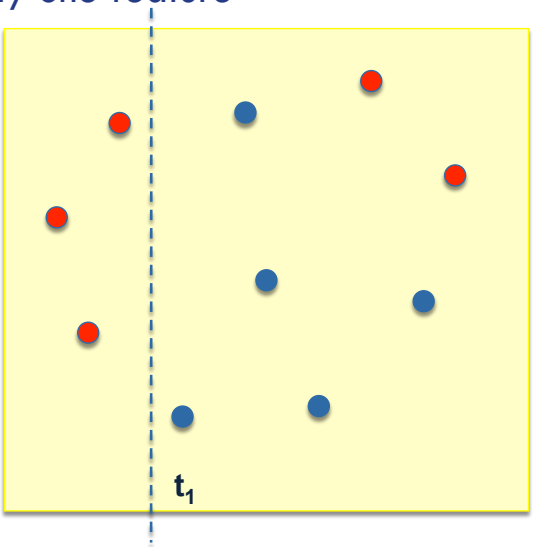
Adaboost: Key Insight

- Let the weight of each instance be represented by the radius of the corresponding circle
  - At the beginning, all eelements have the same weight



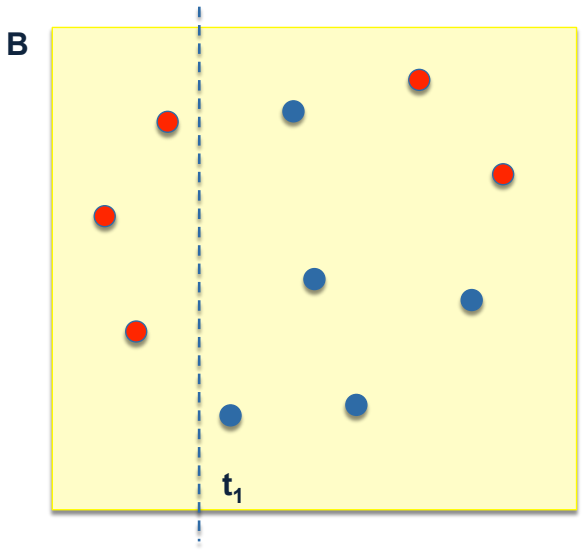
Adaboost: Key Insight

- The best weak classifier is illustrated below.
  - Note that this extremely simple classifier regards exclusively one feature



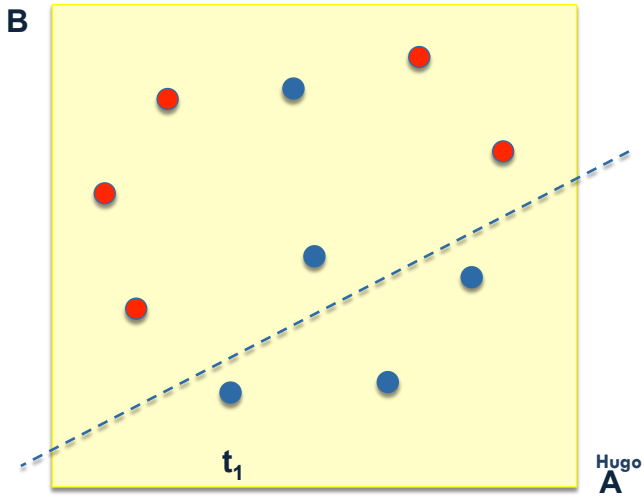
Adaboost: Key Insight

- The first rule states that "IF A < t_1 THEN [RED]"



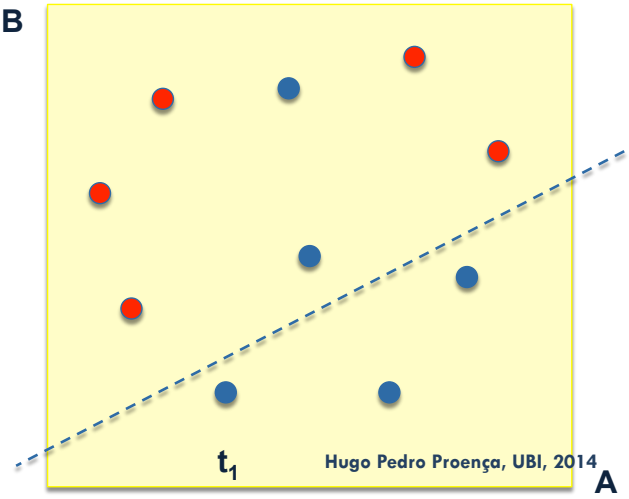
Adaboost: Key Insight

- **Question:** Would the weak classifier illustrated below be possible? Why?



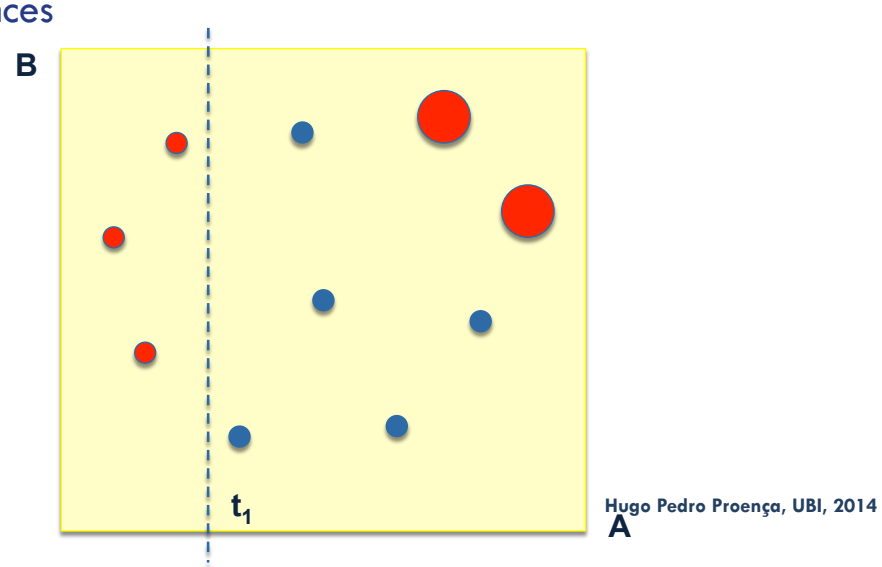
Adaboost: Key Insight

- **Answer:**
  - **No:** It is not a single feature classifier (regards "A" and "B")
  - **Yes:** For the state-of-the-art implementations every weak classifier actually analyzes more than one feature



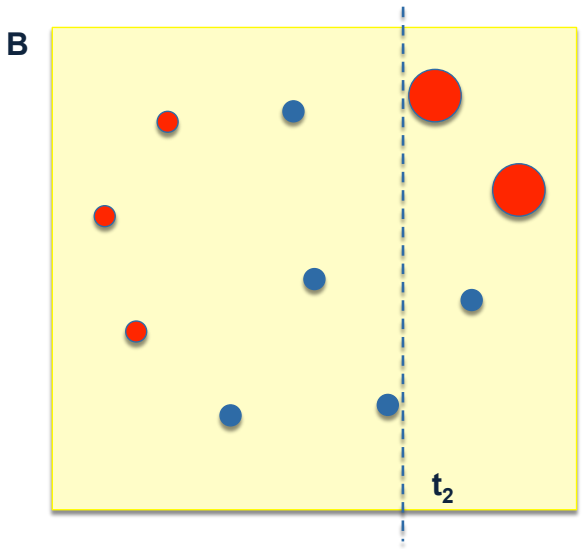
Adaboost: Key Insight

- The first rule states that "IF A < t_1 THEN [RED]"
  - Increase the weight (relevance) of the bad classified instances



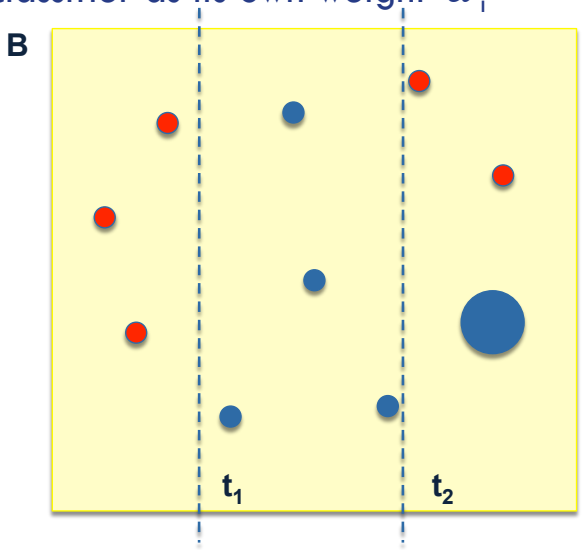
Adaboost: Key Insight

- The second weak classifier will be "IF A > t_2 THEN [RED]"



Adaboost: Key Insight

- Update weights according to the combination of the first two classifiers
  - Each classifier as its own weight α_i



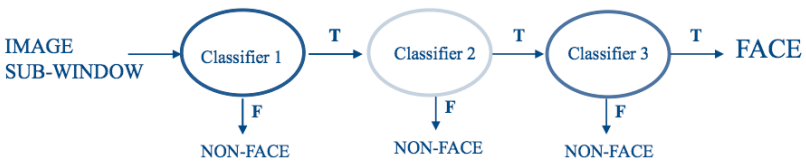
Adaboost: Key Insight

- The final strong classifier **"f"** (ensemble) is a weighted (α) linear combination of the responses given by the T weak classifiers **h**:

$$f^T(\cdot) = \text{sign}\left[\sum_{t=1}^{T} \alpha_t h_t(\cdot)\right]$$

Adaboost: Cascade Classifiers

- Another key contribution of the detector due to Viola and Jones is the **"attentional cascade"** of weak classifiers, which is actually the key for real-time performance
  - For each image patch, the **first classifiers** in the cascade guarantee that the patch **is not** a face
    - This is the **large majority** of the cases, even in images with many faces



Adaboost: Cascade Classifiers

- In practical terms, and regardless the type of objects we want to detect, the idea is to gather a large set of "**positive instances**"
  - Viola and Jones used 5,000 24 x 24 faces
  - Frontal images
  - Normalized for scale, translation

Adaboost: Cascade Classifiers

- Finally, collect a large set of "negative instances", i.e., images containing any object except te interest object.
  - Usually this set should have (at least) twice elements as the set of positive instances
  - Sometimes, large images of background textures are used, together with an automated way to crop patches of such data, which are used as negative instances.
  - It is important that the negative data contains elements that correspond o **the typical false-match cases** of the strong classifier.
    - **Implies iterative learning.**

[7]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

University of Beira Interior, Department of Informatics
Hugo Pedro Proença, 2014/2015

Syllabus

Theoretical	**Parametric vs. non-parametric segmentation methods; form fitting algorithms; introduction to active contours and deformable models: snakes; geodesic active contours.**
Labs	Notion of parametric space: Hough transform; Line, circular and elliptical Hough transform; Snakes implementation and testing; "*Active contours without edges*"

Summary: Segmentation

- As previously seen, the role of **segmentation** is to find a boundary of the object of interest, as close as possible to that object.
  - If the boundary is not too close to the object, we are considering non-object regions for further processing phases.
  - On the opposite side, if the boundary disregards parts of the object, we might be wasting valuable information.

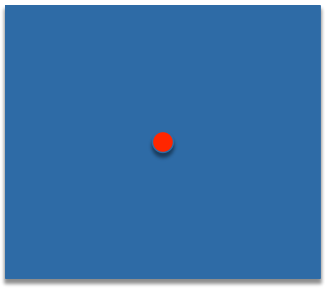
Source: http://note.sonots.com/SciSoftware/NcutImageSegmentation.html

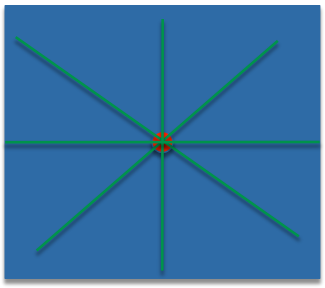
Segmentation

- **Parametric vs. Non-Parametric Approaches**
  - Several fundamental statistical concepts are helpful prerequisite knowledge for fully understanding the terms "parametric" and "nonparametric."
  - These include random variables, probability distributions, parameters, population, sample, sampling distributions and the Central Limit Theorem.
- In the specific case of image segmentation, parametric statistical procedures rely on assumptions about the type of the distribution in the underlying population
  - In practice, they assume a shape for the object, even if it is an approximation.
  - Example: The boundaries of the iris are circular. Hence they can be parameterized by 3 values: {x,y,radius}
- Nonparametric statistical procedures rely on no or few assumptions about the shape or parameters of the population distribution from which the sample was drawn.
  - In practice, the shape of the object to be segmented is not previously defined
    - It will depend of the data.

Form Fitting: Hough Transform

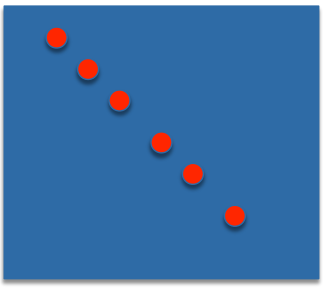
- The goal of the **Hough transform** is to find a previously knwon shape in the image
  - Hence, it is a **parametric approach**
- The key idea is to establish a **voting procedure,** in which every element (pixel) that potentially belongs to a shape votes in the **family of compatible shapes**

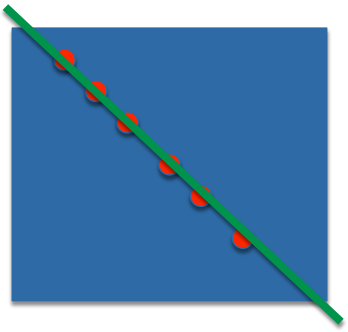




Form Fitting: Hough Transform

- If the image contains the shape we are searching for, then a **large number of elements** will **vote** on the **same parameterization of the shape**





Form Fitting: Hough Transform

- Algorithm: Condensed Perspective
  - Create a discrete **Parameter Space**
  - For each element in the image, count a vote in every **compatible bin** of the parameter space
  - Find the bin(s) with the most votes

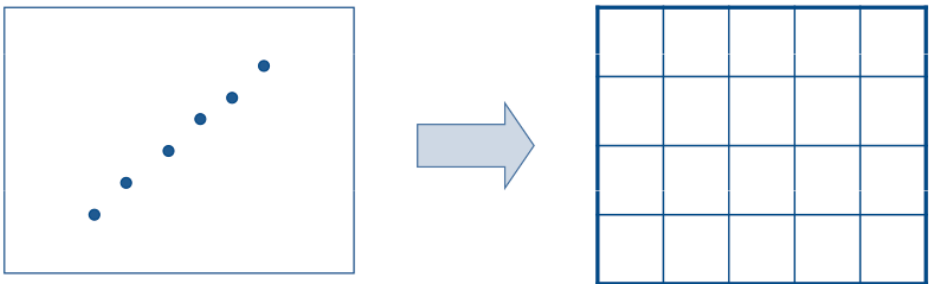


Image Space

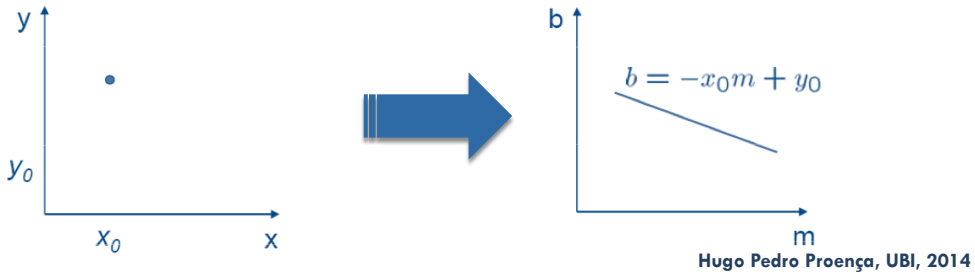
Parameter Space

Form Fitting: Hough Transform

- **Strong Points of Hough Transform**
  - Conceptually simple;
  - Easy to implement;
  - Robust to noisy data.
  - Can be adapted to many types of forms
- **Weaknesses of Hough Transform**
  - Computationally complex for objects with many parameters.
  - It has specific implementations (parameter spaces) for each type of objects Looks
  - Objects that share parameterizations cannot be discriminated
    - E.g., collinear line segments cannot be separated

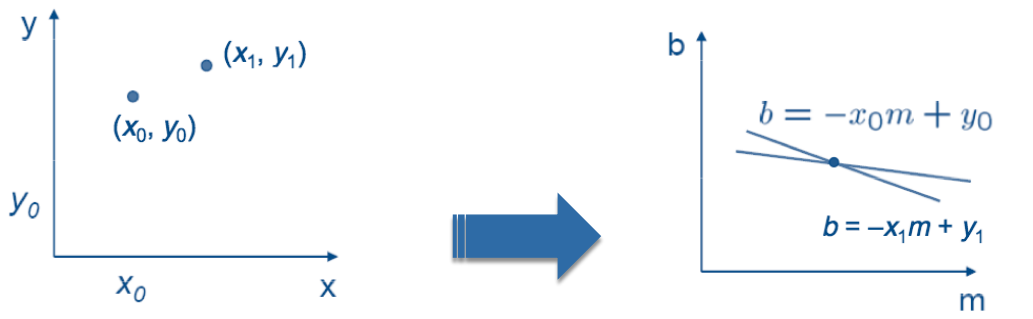
Hough Transform: Example

- Suppose we are searching for lines in image.
  - Decided to use the linear Hough transform
  - Every pixel that potentially belongs to a line has a position (x_0, y_0)
  - The family of lines that are compatible with that pixel is given by:
    - $b = -x_0 m + y_0$
  - **The key is that a point in the image space corresponds to a line in the parameter space;**
    - I.e., the set of lines to which the point at (x_0, y_0) might potentially belong



Hough Transform: Example

- Two points (x_0,y_0) and (x_1,y_1) in the image space will intercept in the parameter space
  - Corresponding to the parameterizations (b,m) of the straight line that passes through (x_0,y_0) and (x_1,y_1) in the image space



Hough Transform: Example

- Parameter Space Issues
  - Should be bounded
  - All bins should have apriori the same probability of getting votes
- Polar parameter spaces are often used

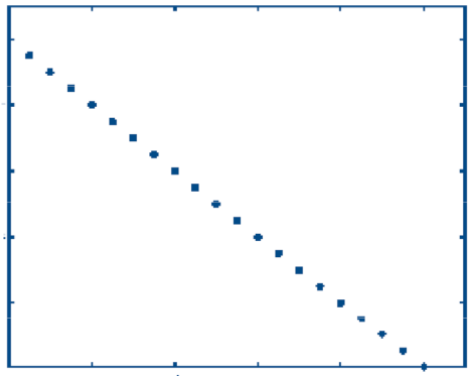
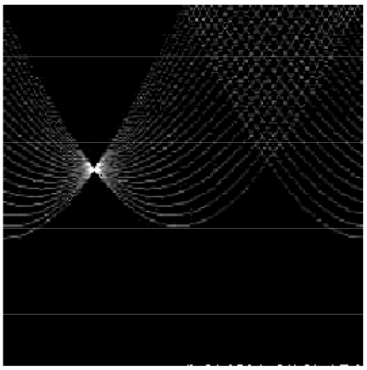


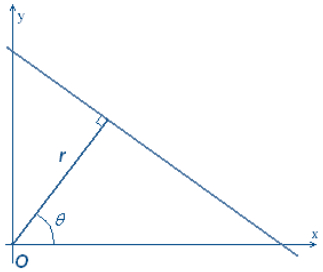
Image Space



Parameter Space

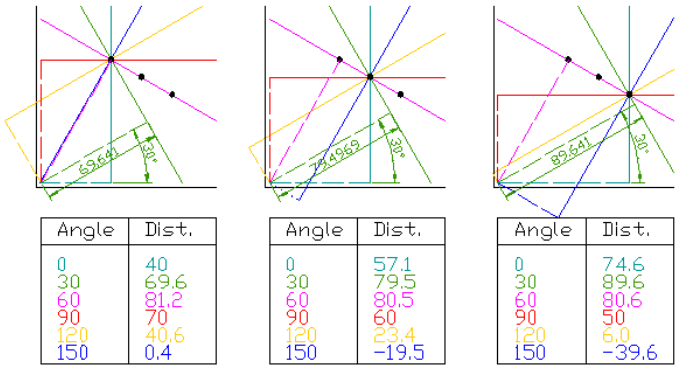
Hough Transform: Example

- Example: **Line Parameterization.**
  - Each "xy" position can be represented in polar coordinates by two parameters: "r" is the distance to origin, and "Θ" is the angle of the vector from origin to xy.
  - $r = x\cos(\Theta) + y\sin(\Theta)$



Hough Transform: Example

- Example: **Line Parameterization.**
  - Let (x_0,y_0) be an edge point in the image. An infinite number of straight lines potentially pass by that point, but all these lines share the following relationship:
  - $r = x_0 \cos(\Theta) + y_0 \sin(\Theta)$
- Illustration:



Angle	Dist.
0	40
30	69.6
60	81.2
90	70
120	40.6
150	0.4

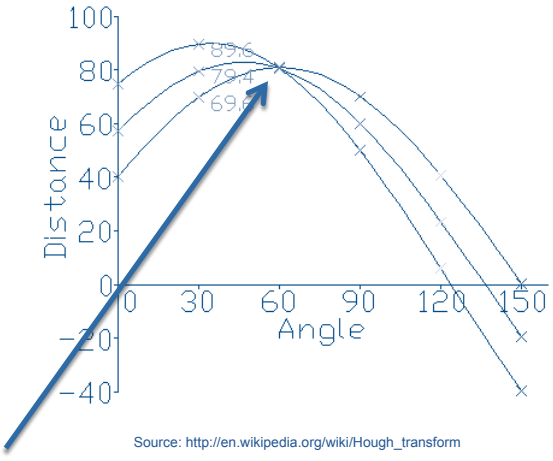
Angle	Dist.
0	57.1
30	79.5
60	80.5
90	60
120	23.4
150	-19.5

Angle	Dist.
0	74.6
30	89.6
60	80.6
90	50
120	6.0
150	-39.6

Source: http://en.wikipedia.org/wiki/Hough_transform

Hough Transform: Example

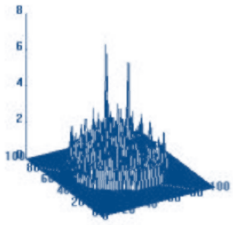
- Example: **Line Parameterization.**
  - The Hough transform algorithm uses an array, called an accumulator to store the votes of each point for each particular shape.
  - The dimension of the accumulator is equal to the number of unknown parameters of the Hough transform problem.
  - For example, the linear Hough transform problem has two unknown parameters: the pair (r, θ).
  - The maximum value at a specific position corresponds to the most probable parameterization for the desired shape.



Source: http://en.wikipedia.org/wiki/Hough_transform

Hough Transform: Remarks

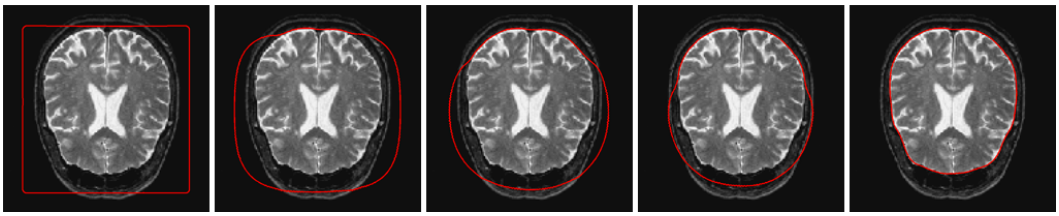
- It is specially important to find a parameter space where all dimensions vary linearly.
- Also, it should be considered small variations in data, so that these do no disturb the algorithm functioning.
  - *"A straight line is never a straight line"* in practical CV tasks.
- Usually, each pixel votes not only in a specific set of positions, but also on its neighbors.
- It is specially expensive in terms of computation, for more complex shapes.
- A circle is parameterized by:
  - $x=x_0+r\cos(\Theta)$
  - $Y=y_0+r\sin(\Theta)$



- Even assuming that we know the radius of the circle, this would give raise to a heavy search on the parameter space. Heavier if we don't know radius (3D search)

Active Contours

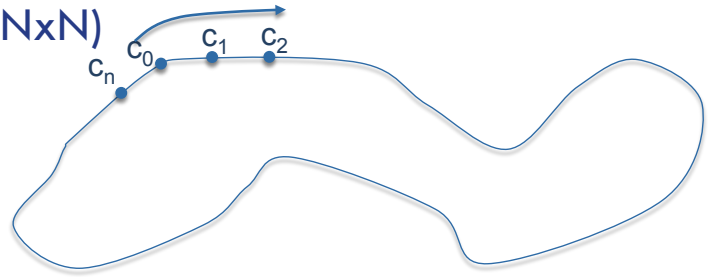
- Are a specially successful algorithm for image segmentation. Although boundary-driven (as the previously seen Hough transform), they are model-free and bring knowledge-based information to the problem.
- Key idea: Suppose we have an elastic material around the object of interest and let it press freely.
  - It is supposed that such elastic material stop when he encounters a "*strong enough wall*".



Source: https://www.ceremade.dauphine.fr/~peyre/teaching/manifold/tp1.html

Active Contours: Snakes

- Having a parameterized quantized contour (curve C:R→NxN)



  - $||c_i - c_{i-1}|| = k$
  - Curve length = n k
- A co

$$E[(C)(p)] = \alpha \int_0^1 E_{int}(C(p))dp + \beta \int_0^1 E_{img}(C(p))dp + \gamma \int_0^1 E_{con}(C(p))dp$$

Active Contours: Snakes

- Cost function terms:

$$E[(C)(p)] = \alpha \int_0^1 E_{int}(C(p))dp + \beta \int_0^1 E_{img}(C(p))dp + \gamma \int_0^1 E_{con}(C(p))dp$$

being "p" a point along the contour.

  - The **internal term** gives information about desired regularity/smoothness along the curve and has two components (resisting to stretching and bending).
  - The **image term** attracts the active contour towards the object (strong gradients).
  - The **external term** (optional) can be used to account for user-defined constraints, or prior knowledge on the structure to be recovered
  - The lowest potential of such a cost function refers to an equilibrium of these terms.
    - We seek to minimize the cost function

Active Contours: Snakes

- The internal term is function of the first and second derivatives.

$$E_{int}(C(p)) = w_{tension}(C(p)) \left| \frac{\partial C}{\partial p}(p) \right|^2 + w_{stiffness}(C(p)) \left| \frac{\partial^2 C}{\partial p^2}(p) \right|^2$$

  - The first-order term makes the snake act like a membrane and second-order term makes it act like a thin plate.
  - Large values of $w_{tension}$ increase the internal energy of the snake as it stretches more and more, whereas small values will make the energy function insensitive to the amount of stretch.
  - Large values of $w_{stiffness}$ will increase the internal energy of the snake as it develops more curves, whereas small values of will make the energy function insensitive to curves in the snake.

Active Contours: Snakes

- The image term has three components: lines, edges and extreme points:
  - The image enery value is a linear combination of these:

$$E_{img}(C(p)) = w_{line}E_{line}(C(p)) + w_{edge}E_{edge}(C(p)) + w_{term}E_{term}(C(p))$$

  - E_{line} corresponds to image intensity: I(x,y). The sign of w_{line} will dictate if the contour should evolve towards bright or dark regions.
  - E_{edge} is the negative image gradient: $\nabla | I(x,y) |^2$ and will make the snake attract towards contours with large image gradients.

Active Contours: Snakes

- The extreme points term analyze the curvature of level lines in a slightly smoothed image, detecting corners and terminations in an image (potential interest points).

- $E_{term} = \frac{C_{yy}C_x^2 - 2C_{xy}C_xC_y + C_{xx}C_y^2}{(C_x^2 + C_y^2)^{3/2}}$

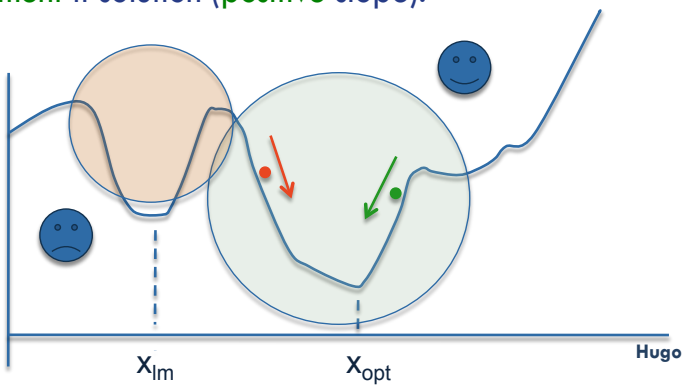
being C_x and C_y the image gradients

Active Contours: Evolution

- We are interested in finding the parameters of the curve (specified by the position of its control points), that minimize the value of the cost function.
  - Suppose we have 10 control points, each one defined by (x_i,y_i) coordinates, in an image 800x600 px.
  - Each control point will have 480 000 potential positions.
  - The number of possibilities for the final contour will be $480\,000^{10}$!
    - Henceforth, brute force is out-of-question

Active Contours: Optimization

- **Gradient-descent optimization** can be used to minimize the cost function of the snake.
- Example.
  - Consider a function of one variable. If we have a starting guess at the value of the solution, we can look at the slope at that point and decide to increment the value of our solution (negative slope) or decrement it solution (positive slope).



Active Contours: Optimization

- We will start from an initial guess for the position of the snake. Then, at each iteration, we can adjust the position of each control point, in the following way:

$$x_{t+1} = x_t + \gamma \frac{df}{dx}(x_t) \text{ and } y_{t+1} = y_t + \gamma \frac{df}{dy}(y_t)$$

being "f" the cost function.

$(x,y)_{t+1} = (x,y)_t + \nabla E(C(p))$

- $E(C(p)) = \alpha \nabla E_{int}(C(p)) + \beta \nabla E_{img}(C(p))$
- E_{img} is often referred as the "external" (ext) force.

$$\nabla E_{internal}(s) = \nabla \left[(\alpha(s) \|\mathbf{v}_s(s)\|^2 + \beta(s) \|\mathbf{v}_{ss}(s)\|^2)/2 \right]$$

$$\nabla E_{internal}(s) = \left[\left(\alpha(s) \nabla \left\| \frac{d\bar{v}}{ds}(s) \right\|^2 + \beta(s) \nabla \left\| \frac{d^2\bar{v}}{ds^2}(s) \right\|^2 \right)/2 \right] = \alpha \frac{\partial^2 \bar{v}}{\partial s^2} + \beta \frac{\partial^4 \bar{v}}{\partial s^4}$$

Active Contours Optimization

- The derivatives can be approximated using finite differences
  - The first derivative using 2 adjacent points on the contour
  - The second derivative using 3 adjacent points on the contour
  - The "n-th" derivative using n+1 points on the contour.
- At each iteration, the values of each (x_i,y_i) are given by:

$$\bar{x}_i =\leftarrow x_i - \gamma\left\{ w_{internal}\left[\alpha\frac{\partial^2 x}{\partial s^2}(\bar{v}_i) + \beta\frac{\partial^4 x}{\partial s^4}(\bar{v}_i)\right] + \frac{\partial}{\partial x}E_{ext}(\bar{v}_i)\right\}$$
$$\bar{y}_i =\leftarrow y_i - \gamma\left\{ w_{internal}\left[\alpha\frac{\partial^2 y}{\partial s^2}(\bar{v}_i) + \beta\frac{\partial^4 y}{\partial s^4}(\bar{v}_i)\right] + \frac{\partial}{\partial y}E_{ext}(\bar{v}_i)\right\}$$

- The process is repeated until no significant changes in the contour occur.

Geodesic Active Contours

- An extremely popular Active Contours approach is due to Chan and Vese ("Active Contours Without Edges", IEEE Transactions on Image Processing, vol. 10, issue 2, pag. 266-277, 2001).
- The goal is to find a partition that divides the image into "n" regions, such that homogeneity is maximized within each partition and minimized between different partitions.
- Notation: Image: u_0, Regions: Ω_i, Contour: C

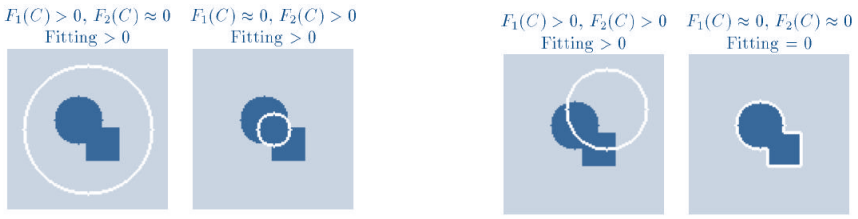
$$inf\left\{ F^{MS}(u,C) = \int_{\Omega}(u-u_0)^2 dxdy + \mu\int_{\Omega\setminus C}|\nabla u|^2 dxdy + v|C|\right\}$$

- A reduced case of this model (u=constant inside each component) was handled by Chan and Vese.

Active Contours Without Edges

- Given a curve C= $\delta\ \omega$, two unknown constants c_1 and c_2=mean (u_0) in Ω_i. The first region comprises the object $\Omega_1=\omega$ and the second region the background $\Omega_2=\Omega-\omega$.
- The following cost function (energy function) should be minimized with respect to c_1, c_2 and C.

$$F(c_1,c_2,C)=\underbrace{\int_{\Omega_1=\omega}(u_0(x,y)-c_1)^2dxdy}_{F_1}+\underbrace{\int_{\Omega_2=\Omega-\omega}(u_0(x,y)-c_2)^2dxdy}_{F_2}+v|C|$$



Source: http://scialert.net/fulltext/?doi=itj.2008.490.496

Active Contours Without Edges

- Formulating the problem as a level-set: C={(x,y) | Φ(x,y)=0}. The energy function becomes:

$$F(c_1,c_2,\phi)=\int_\Omega(u_0(x,y)-c_1)^2H(\phi)dxdy+\int_\Omega(u_0(x,y)-c_2)^2(1-H(\phi))dxdy+v\int_\Omega|\nabla H(\phi)|$$

where H(.) is the Heaviside function. (H(x) = 0 iff x<0; H(x)=1 otherwise).

- In order to find the minimum value of "F", it is required to find its derivatives and set them to zero.
- Using Euler-Lagrange equations, c1, c2 and Φ will be updated recursively.

Active Contours Without Edges

- Updating:

$$\begin{cases} c_1(\phi) = \dfrac{\int_\Omega u_0(x,y)H(\phi(t,x,y))dxdy}{\int_\Omega H(\phi(t,x,y))dxdy} \\ c_2(\phi) = \dfrac{\int_\Omega u_0(x,y)(1-H(\phi(t,x,y)))dxdy}{\int_\Omega 1-H(\phi(t,x,y))dxdy} \end{cases}$$

$$\frac{\partial \phi}{\partial t} = \delta(\phi)[v\ div\left(\frac{\nabla \phi}{|\nabla \phi|}\right) - (u_0 - c_1)^2 - (u_0 - c_2)^2]$$

Being δ(.) the Dirac function (δ(x)=0 iff x <> 0, δ(x)=1 iff x=0).

Active Contours Without Edges

- For implementation purposes, it is important that both the Heaviside and Dirac functions are continuous, being the following approximations used. ($\varepsilon = 10^{-5}$)

$$H_\varepsilon(z) = \begin{cases} 1, if\ z > \varepsilon \\ 0, if\ z < -\varepsilon \\ \frac{1}{2}\left[1 + \frac{z}{\varepsilon} + \frac{1}{\pi}\sin\left(\frac{\pi z}{\varepsilon}\right)\right], if\ |z| \le \varepsilon \end{cases}$$

$$\delta_\varepsilon(x) = H'^{(x)}_\varepsilon = \begin{cases} 0, if\ |z| > \varepsilon \\ \frac{1}{2\varepsilon}\left[1 + \cos\left(\frac{\pi z}{\varepsilon}\right)\right], if\ |z| < \varepsilon \end{cases}$$

Active Contours Without Edges

- Approximation of curvature:

$$div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) = \frac{\phi_{xx}{\phi_y}^2 - 2\phi_{xy}\phi_x\phi_y + \phi_{yy}{\phi_x}^2}{(\phi_x + \phi_y)^{3/2}}$$

- Using finite differences:

$$\begin{cases} \phi_x = \phi_{x+1,y} - \phi_{x,y} \\ \phi_y = \phi_{x,y+1} - \phi_{x,y} \\ \phi_{xy} = \phi_{x+1,y+1} - \phi_{x,y+1} - (\phi_{x+1,y} - \phi_{x,y}) \end{cases}$$

Hugo Pedro Proença, UBI, 2014

Hugo Pedro Proença, UBI, 2014

[8]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

University of Beira Interior, Department of Informatics
Hugo Pedro Proença, 2014/2015

Syllabus

Theoretical	**Texture / Shape descriptors: histograms of oriented gradients (HOGs), local binary patterns (LBPs) and basic of Fourier descriptors; Basics of image filtering: convolution and correlation. Examples of kernels: Gaussian, Gabor, and multi-lobe differential filters.**
Labs	The Fourier transform; Fast Fourier Transform (FFT); Signal processing fundamental theorem; Gabor and Multi-lobe differential image filtering.

Image Features

- Having a segmented object "O", the next processing phase aims at it description, mostly for recognition purposes.
- Hence, the goal of the feature extraction phase is to convert the object of interest into a **feature vector**, which can be regarded as a point in a multi-dimensional space (where recognition occurs).
  - $F=(f_1,f_2,\ldots f_n)$
- Ideally, F should have some important properties:
  - **Discriminability** (it should contain enough information to distinguish between the set of objects to be handled);
  - **Permanence** (each f_i should keep similar values, under typical data transformations: scale, translation, rotation, perspective, blur, occlusions);
  - **Collectability** (the properties to be measured should be possible to infer from the acquired data).

Image Features

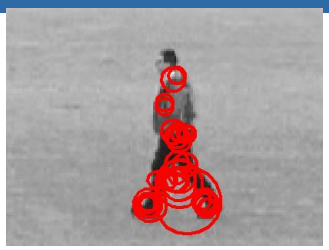
- There are many different possibilities to extract image features, which can be divided into major categories:
- The choice of any type of image features for a specific application is completely dependent of the application and can hardly be completely formalized.
  - **Shape Features**: when the shape of the object boundary is important to subsequent processing phases.
  - **Texture Features**: features that analyze the distribution of intensities inside the object.
  - **Local Features**: when the extracted f_i value regard properties from a sub-region of the object.
  - **Global features**: when f_i was extracted using the whole object.

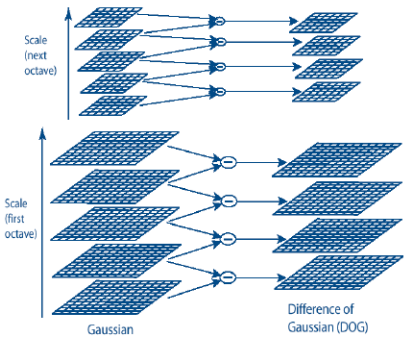
Image Features: Interest Points

- An earliest difficulty is to choose from where features can be extracted, inside a segmented object.
- If local-based features are used (the most common strategy), its extraction inside the whole object might be computationally expensive.
- Hence, the identification of **interest points** in images has been gaining significant popularity over the last few years.
  - An **interest point** marks a region of the segmented object that potentially deserves more attention than its surroundings.
  - Most of the state-of-the-art object encoding/matching (recognition) strategies heavily rely on the detection of interest points in data.

Interest Points: SIFT

- The interest points from the SIFT approach are called “keypoints”.
  - Let L(x,y,Š) be one image I(x,y) smoothed by a Gaussian kernel (G) with sigma Š.
  - The Difference of Gaussian (DOG) image is given by: : L(x,y,Š) - L(x,y,KŠ).
  - A pyramid of images is built, with increasing sigma values and grouped into “octaves”.
    - One octave corresponds to the product of sigma by “2”

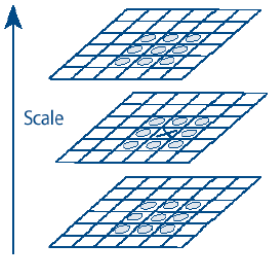




Lowe, David G. (1999). "Object recognition from local scale-invariant features". *Proceedings of the International Conference on Computer Vision*. **2**. pp. 1150–1157. doi:10.1109/ICCV.1999.790410.

Interest Points: SIFT

- Next, the detection of potential interest points is given by local extrema:
  - These are estimated by comparing the DoG(x,y) to all its neighbors, in a n x n x n radii.
  - In the right illustration, the pixel marked with “X” will be compared against the 26 values marked by green dots.
  - If the value is either a local maximum or minimum, it is considered a candidate.

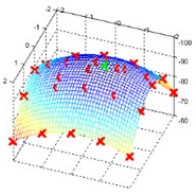


Source: https://courses.cs.washington.edu/courses/cse576/08sp/projects/project1/artifacts/xingli08/artifact/Project1_Xing%20Li.htm

Interest Points: SIFT

- The detection of local extrema produces many unstable candidates. Next, all these keypoints are filtered, in order to discard points with low contrast and edge responses.
- Initially, the position of the keypoint is accurately determined by interpolation of nearby data.
  - The actual maxima/minima almost never lie exactly on a pixel. They lie between pixels. So, we must locate the sub-pixel location.
- Interpolation is done using the quadratic Taylor expansion of the DoG scale-space function, D(x,y,σ) with the candidate keypoint as the origin:

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}$$



- The actual extremum is determined by taking the derivative of this function with respect to "x" And setting it to zero.

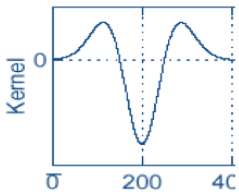
Source: https://courses.cs.washington.edu/courses/cse576/08sp/projects/project1/artifacts/xingli08/artifact/Project1_Xing%20Li.htm

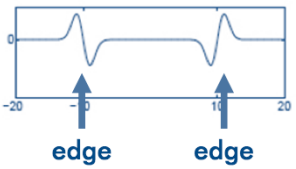
Interest Points: SIFT

- Next, the keypoints candidates with low contrast are discarded.
- The idea is that, if the magnitude of the intensity at the current pixel in the DoG image is less than a certain value, it is rejected.
- Because we have subpixel keypoints (we used the Taylor expansion to refine keypoints), we again need to use the Taylor expansion to get the intensity value at subpixel locations.

Interest Points: Scale Detection

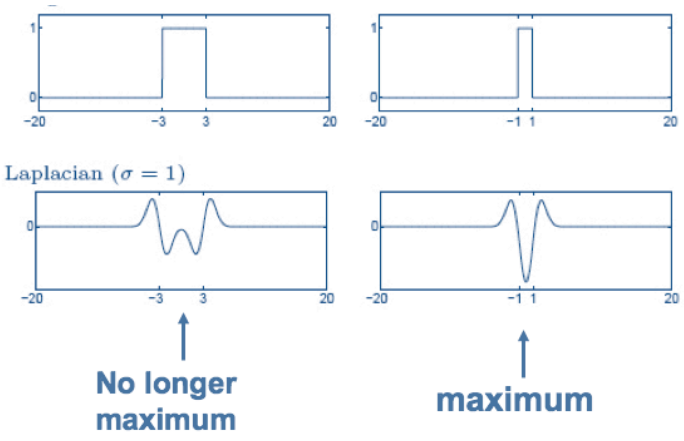
- Both the previously seen SIFT and SURF approaches provide a description of the scale at which an interest point is detected.
- This is important, because it will provide an estimate of the size of the neighborhood where the feature encoding process should occur.
- Every time we do not have such estimate, it is desirable to perform au automated scale selection process, which can be done by zero-crossing analysis:





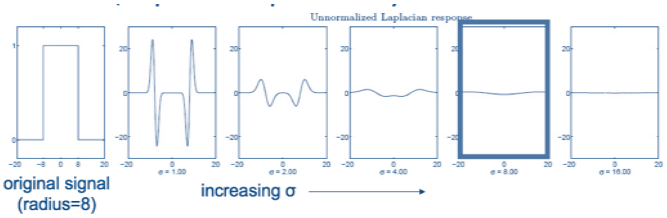
Interest Points: Scale Selection

- The basic idea is that the magnitude of the Laplacian response will achieve a maximum at the center of the object, provided the scale of the Laplacian is matched against the scale of the object.

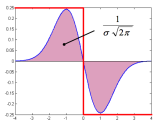


Interest Points: Scale Selection

- So, we will have the scale of the object by convolving it with Laplacians at different scales, looking for the maximum response.
  - It is important to note that some preprocessing should have occurred so to transform the object in a BLOB.

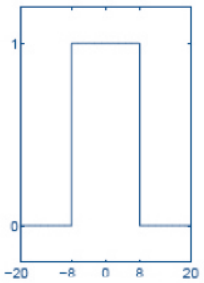


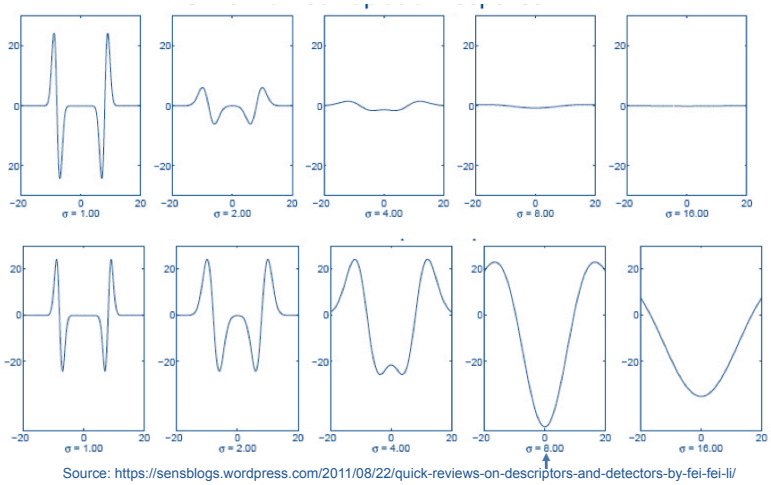
  - The problem is that the response of a Laplacian kernel decays as scale increases. This is due to the fact that the response of a derivative of a Gaussian filter decreases as σ increases.



Interest Points: Scale Selection

- To keep the magnitude of the response scale invariant, we must multiply it by σ^2 (Laplacian is the second derivative).
- Effect of scale normalization:

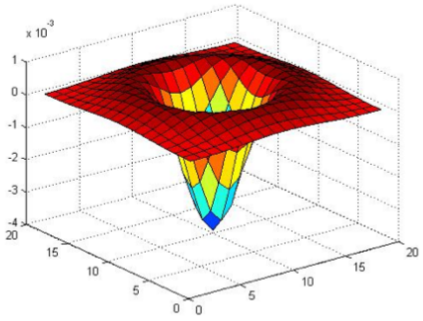




Source: https://sensblogs.wordpress.com/2011/08/22/quick-reviews-on-descriptors-and-detectors-by-fei-fei-li/

Interest Point: Scale Selection

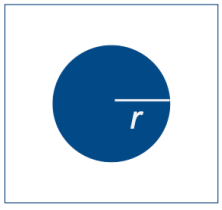
- Hence, the isotropic (circularly symmetric) Laplacian of Gaussian operator (weighted by σ^2) can be used for automated scale detection:

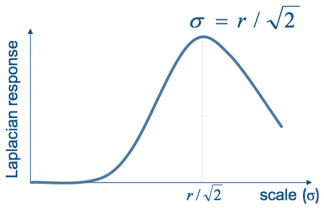


$$\nabla^2_{\text{norm}} g = \sigma^2 \left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \right)$$

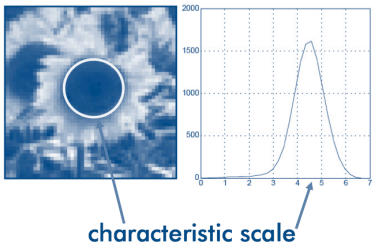
Interest Point: Scale Selection

- For a binary circle of radius r, the Laplacian achieves a maximum at:





- We define the "characteristic scale" as the scale that produces the maximum among the Laplacian responses:



Source: https://sensblogs.wordpress.com/2011/08/22/quick-reviews-on-descriptors-and-detectors-by-fei-fei-li/

Image Decriptors: Classification

- There are different types of feature extraction processes, either based on:
  - **Shape**
  - **Region (texture)**
- Descriptors can either work on the entire image or or image patches, which determine a major division
  - **Global**
  - **Local**

Region Descriptors

- The simplest region descriptor is given by the area of the object: $A(S) = \int_x \int_y I(x,y)dy\,dx$
  - In the discrete domain, it is obtained by simply summing all pixels belonging to the object, provided that it is binary.
- The object perimeter is obtained from the parametric coordinates of the object (x,y) $P(S) = \int_t \sqrt{x^2(t) + y^2(t)}dt$
  - In the discrete domain, it is approximated by the accumulated difference between adjacent boundary pixels

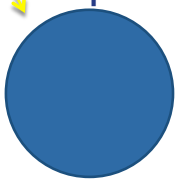
$$P(S) = \sum_i \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$$

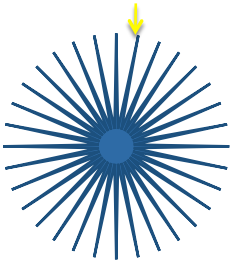
Region Descriptors

- Having both the area and perimeter of an object, its compactness can be obtained:

$$C(S) = \frac{A(s)}{P^2(s)/4\pi}$$

- This descriptor gives values closed in [0,1] interval. Near 1 for compact objects, near 0 for non-compact objects.



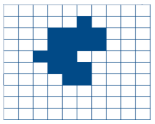


Region Descriptors

- Dispersion is another measure of the object regularity.
  - Measures the proportion between the highest length inside the object and its area:

$$I(S) = \frac{\pi \max((x_i - \bar{x})^2 + (y_i - \bar{y})^2)}{A(S)}$$

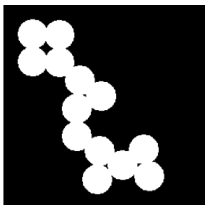
- Exercise:
  - 1. Obtain the area, perimeter, compactness and dispersion of the following object:



  - 2. Consider that you have a MATLAB function named "getBoundary(I)", that returns the boundary pixels of a binary object: [r_1, c_1; …; r_n, c_n], being r_i,c_i the row/column coordinates.
  - Implement MATLAB functions for the above referred region descriptors.

Region Descriptors

- The number of possibilities for region descriptors (as others) is quite large.
- It's up to the person that is developing the system to decide wich might be useful for the given problem, considering the fundamental premises of a good feature: discriminability, permanence and collectability.
- The "regionprops" MATLAB function implements most of the well knwon region descriptors.
- Examples:
  - Poincaré features are given by the total number of holes inside an object.
  - The Euler number is the difference between the total connected regions and the number of holes.



Poincaré: 4; Euler: -3

Region Descriptors: Image Moments

- Image moments are global descriptors of objects. They are extremely popular due to their invariance properties.
- They are associated to the statistical variant of image processing, as they assume that there are not occludded parts of the object.
- In 2D data, the "p-q" order moments are given by:
- $$m_{pq} = \sum_x \sum_y x^p y^q I(x,y) \Delta A$$
- ΔA is the offset (typically 1). The invariance yields from normalized and centered moments:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x,y) \Delta A \qquad \bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

Image Moments

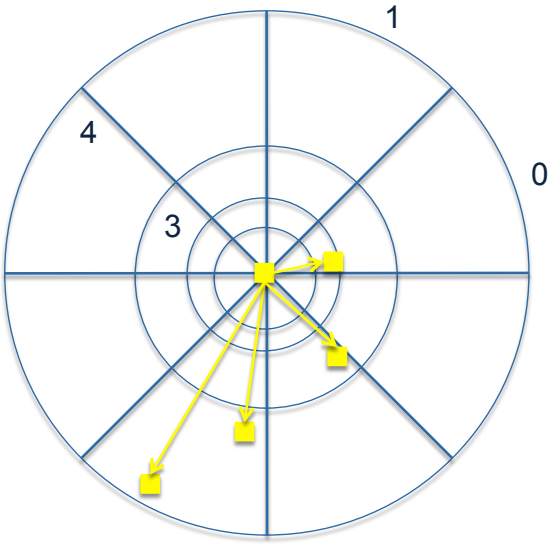
- Among the different possibilities to obtain image moments, **Hu moments** are well known, due to the observed invariance to some traditional data variation factors:
- Invariance to scale is obtained in the following way:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \qquad \gamma = \frac{p+q}{2} + 1$$

- Typical moments:

$$M1 = \eta_{20} + \eta_{02}$$
$$M2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$
$$M3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$
$$M4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$
$$M5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2)$$
$$M6 = (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$
$$M7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})(3(\eta_{12} + \eta_{30})^2 - (\eta_{21} + \eta_{03})^2)$$

Shape Context: Example

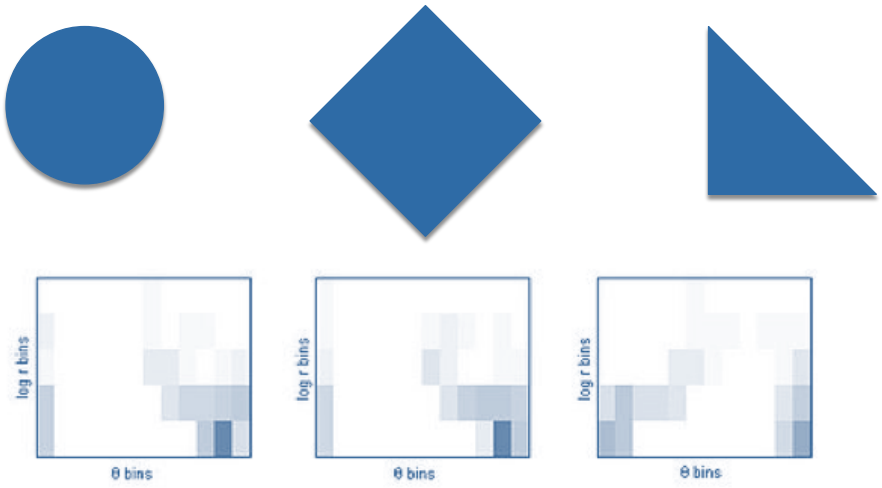


The histogram will be updated:
(1,0): +1
(4,5): +2
(3,6): +1

Exercise: Implement a MATLAB function that returns the "Shape Context" of a given object

Shape Context: Example

- Carefully analyze the "Shape Context" representations of the following objects:



Shape Context: Matching

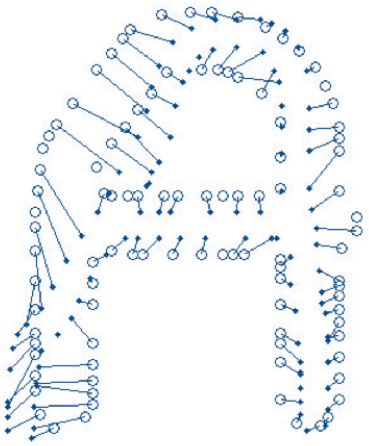
- Having one histogram describing the object boundary, the natural choice to match "Shape Context" representations is the Chi-square distance.
- Having "g" and ""h" histogram representations, the distance is given by:

$$C_S = \frac{1}{2}\sum_{k=1}^{K}\frac{[g(k)-h(k)]^2}{g(k)+h(k)}$$

- Apart from a global distance measure, it is possible to obtain a representation of the "local costs" for every point on the boundary of "g" to adapt to "h" boundary.
- Specially useful in matching deformable objects.
- This is done by a Linear Assignment Problem, that aims at matching each point on "g" to one in "h", so to minimize the total cost of match

Shape Context: Deformable Matching

- In pratice, a matrix "n x n" is built, having all distances from elements of "g" to elements of "h".
- For instance, using the "Hungarian method", " a one to one " correspondence is found.
- Then, it should be decided whether the required movements are plausible.
  - If they are, both images might regard the same class of object.



Source: http://commons.wikimedia.org/wiki/File:Matching_example_using_the_Shape_Context_descriptor.jpg

Hugo Pedro Proença, UBI, 2014

Hugo Pedro Proença, UBI, 2014

[9]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

University of Beira Interior, Department of Informatics
Hugo Pedro Proença, 2014/2015

Syllabus

Theoretical	**Feature normalisation: min/max and mean/var rules; Typical probability distributions: uniform, normal and binomial distributions; Notion of eigenvector; analysis of principal components; PCA and LDA; non-linear classification algorithms: neural networks and support vector machines; Fusion at different levels: data, features, scores and responses; Dempster-Shaffer fusion theory.**
Labs	Analysis of eigenvectors and principal components: toy examples; Image features PCA; Linear classification through LDA.

Statistics

- **Standard Deviation**: Measures the average distance between a set of data points and their mean. It's a data dispersion measure:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}}$$

- **Kurtosis**: degree of peakedness of a distribution, defined as a normalized form of the fourth central moment of a distribution.
  - The i[th] central moment is taken around the mean: $\mu_k = \mathrm{E}\left[(X - \mathrm{E}[X])^k\right]$

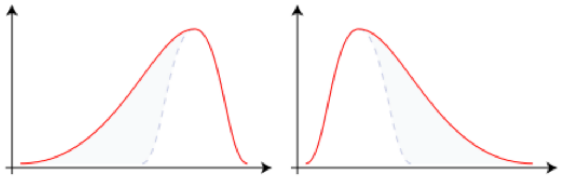
$$\beta_2 \equiv \frac{\mu_4}{\mu_2^2},$$

Statistics

- **Skewness**: It's a measure of the assymetry of the probability distribution of a real-valued random variable:

$$g_1 = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^{3/2}},$$

  - Positive skew: the right tail is longer
  - Negative skew: the left tail is longer
  - Zero: symetric



Statistics

- All these measures are one-dimensional.
  - They denote a specific property of a component in a – most likely higher dimensional – data set.
- Most times we are dealing with high dimensional data.
  - Apart from interested in the individual statistics of each component, we might be interested in the relationships between the different dimensions.
  - Examples:
    - The height of a given subject has some influence in its weight?
    - The color of a subjects' eyes has some influence in its Mathematics skills?
- **Covariance** it's a measure that describes the variability of each dimension of a data set, with respect to the remaining dimensions.

Statistics

- Covariance is obtained for any two dimensions of a high-dimensional feature space. In case of more dimensions, multiple bivariate values should be obtained. (i.e., for (x,y,z), it's possible to measure the covariance between "x, y", "x, z" and "y, z").

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

- In pratical terms, the distance between sample points and their mean is multiplied. Then, the result is divided by (n-1)
- The results is meaningful, essentially by analysing it's sign:
  - Positive: Both dimensions vary directly.
  - Negative: Both dimensions vary inversely.
  - Zero: Dimensions are independent.

Statistics

- The **Covariance Matrix** contains all covariance values between every possible dimension of a feature space :

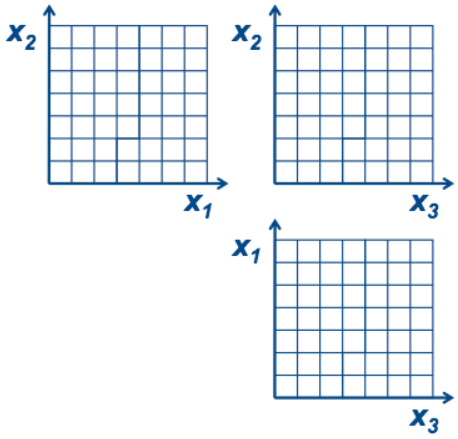
$$C^{n\times n} = (c_{i,j},\ c_{i,j} = cov(Dim_i, Dim_j))$$

$$C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix}$$

- Values along the main diagonal describe the variance of the corresponding dimension.
- Based on its definition, it is obvious that cov(x,y)=cov(y,x), i.e., the covariance matrix is symetric with respect to its main diagonal.

Statistics

- Exercise. Obtain the covariance matrix for the given data set:



Obs.	X1	X2	X3
1	2	2	4
2	3	4	6
3	5	4	2
4	6	6	4

	x_1	x_2	x_3	$x_1-\mu_1$	$x_2-\mu_2$	$x_3-\mu_3$	$(x_1-\mu_1)^2$	$(x_2-\mu_2)^2$	$(x_3-\mu_3)^2$	$(x_1-\mu_1)(x_2-\mu_2)$	$(x_1-\mu_1)(x_3-\mu_3)$	$(x_2-\mu_2)(x_3-\mu_3)$
1												
2												
3												
4												

Eigenvectors and eigenvalues

- Consider the multiplication of a matrix by a vector:

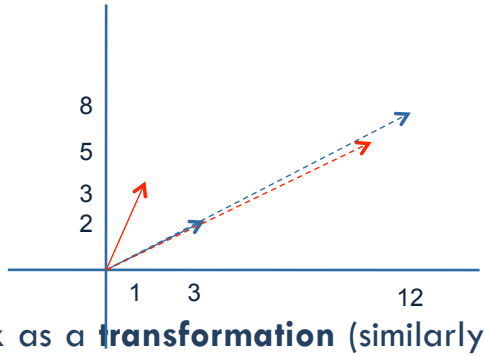
$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

- In the first case, the resulting vector is not a multiple of the original vector.
- Oppositelly, in the second case, the resultant vector (12,8) is a multiple of the multiplier. As such, the latter is an **eigenvector**.
  - The correspondong **eigenvalue** is "4"

Eigenvectors and eigenvalues

- By analysing the direction of the original and resultant vectors:



- Regarding the matrix as a **transformation** (similarly to the previosuly seen transformation matrices), it can be concluded that in the second case, the direction was not changed. **This is the key insight the notion of eigenvector.**
  - The given matrix does not change the direction of its eigenvectors.

Eigenvectors and eigenvalues

- The notion of **eigenvalue** is strongly related to the **eigenvector**.
- It is the value that should by multiplied by the eigenvector to obtain the original vector.
- In the above example, 4 was the eigenvalue that corresponds to the given eigenvector.
- As such, eigenvalues and eigenvectors come in pairs and are always inter-related.

Eigenvectors and eigenvalues

- As a summary, **the eigenvectors of a matrix correspond to the directions that are not changed by the transformation matrix.**
- Not all matrices have eigenvectors.
- Matrices have to be square.
- A (n x n) matrix has – at most – "n" eigenvectors.
- The set of eigenvectors of a matrix (image) is widely used to describe the spatial content of that image (feature).
- In MATLAB, this eigenanalysis is made by the "eig()" function:
  - [V,D] = eig(A)
  - Returns the eigenvectors (D) and corresponding eigenvalues (V) of matrix A.

Eigenvectors and eigenvalues

- There is an important property to be stressed: the eigenvectors of a matrix are orthogonal. This is to say that they form an orthogonal basis of the matrix.
  - We are able to express every point of a data set by linear combinations of its basis-vectors.
  - **This is specially usefull for the analysis of principal components (PCA).**
  - It is usual to determine the eigenvectors/eigenvalues in their normalized version, i.e., with length normalized to 1.
  - As previously seen, the length of a vector does not affect its property of being (or not) an eigenvector.
  - Hence, having an eigenvector $(x_1, ..., x_n)$ it is usual to divide each component by the norm of this vector, in order to obtain length "1":
    - $||(x_1,..., x_n)|| = \text{sqrt}(x_1^2 + ... + x_n^2)$

Eigenvectors and eigenvalues

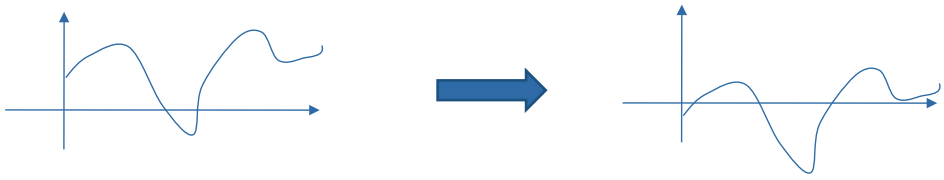
- **Exercise**
  - Determine, from the following vectors, which are eigenvector of the matrix given below and, if positive, determine the corresponding eigenvalue.
    - Matrix: $\begin{pmatrix} 3 & 0 & 1 \\ -4 & 1 & 2 \\ -6 & 0 & -2 \end{pmatrix}$
    - Vectors: $\begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \\ 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$

Principal Component Analysis

- **The Principal Component Analysis (PCA)** it's a well known way to detect patterns on data, by expressing it on a way that enhances similarities or differences.
- Detecting patterns on high dimensional data is a hard task, either for humans or machines.
  - Requires huge amounts of data. An empirical rule says that at the minimum, d^2 instances are required to analyze a d-dimensional data set.
- PCA is also used to compress data (reduce dimensionality), without loosing substantial information.

Principal Component Analysis

- **Step 1.** The analysis of principal components requires a data set (with dimension n) and cardinality (k).
- **Step 2.** Removal of energy. For each dimension, the corresponding mean is subtrated to each component. As such, all dimensions of the data set have zero energy.



Principal Component Analysis

- **Step 3.** Calculus of the covariance matrix. Here, the relationships between independent components are detected, together with an assessment of the data dispersion in each dimension (by analysing the main diagonal components).
- **Step 4.** As the covariance matrix is square, it is possible to obtain the set of eigenvectors and corresponding eigenvalues.
- **Step 4.1.** Eigenvectors normalization. All eigenvectors are normalized to have norm equal to 1.

Principal Component Analysis

- **Step 5.** Selection of components. The set of eigenvectors is sorted by decreasing order, considering the corresponding eigenvalues. From this set, the "k_1" principal components are selected.
  - This is the step that performs the reduction of dimensionality.
- **Step 6.** A transformation matrix is built, by concatenating the eigenvectors selected in the previous step.
  - This matrix will be used to represent all points in the reduced dimensionality feature space. MAT=[vect1, vect2, ... Vectk$_1$]

Principal Component Analysis

- **Step 7. Data Transformation. As the transformation matrix has "d" lines (corresponding to the dimension of the original feature space and** k_1 columns (corresponding to the dimension of the new feature space), when multipling each original data point by the transformation matrix, we obtain a vector of k_1 components. These are the new representation of the data points, in the principal components space.

$$[1 \times d] \times [d \times k_1] = [1 \times k_1]$$

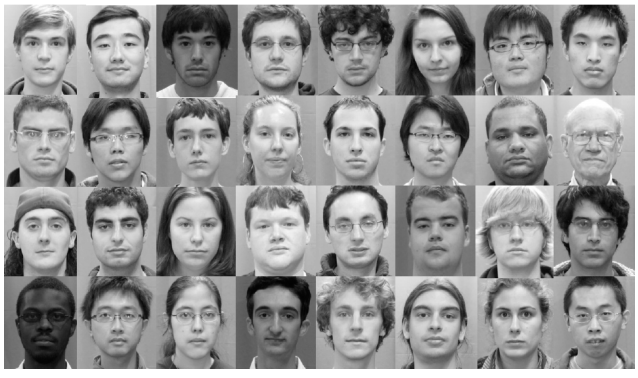
Principal Component Analysis

- **How to choose the value of k_1?**
  - The previously described process does not give any information about a strategy to select the dimensionality of the principal components feature space.
  - There is no formal rule. However, some heuristics about what is generally better exist.
- Usually, the variation in magnitude of consecutive eigenvalues (after sorting) is measured. When changes in magnitude are higher than a threshold, the selection process is stopped.
- Most frequently, the proportion of the data variability that is kept by the selected components is measured.
  - We are interested in keeping around 90-95% of the original data variability.
  - The analysis can be done by measuring the proportion of the sum of eigenvalues λ_i :
  - **Variability:**

$$= \frac{\sum_{i=1}^{M} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

PCA: Example

- Having a set of 128 face images (with dimensions 64 x 64).
- Each face can be regarded as a point represented in a feature space of 4096 dimensions (64 x 64).



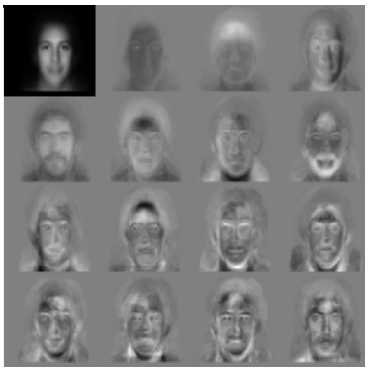
Source: http://commons.wikimedia.org/wiki/File:Matching_example_using_the_Shape_Context_descriptor.jpg

PCA: Example

- The PCA algorithm is used to select the principal components.
  - In pratice, the eigenvectors (with dimension 4096) with largest corresponding eigenvalues will be selected.
- As an example, the facial recognition process can be done in the new feature space of (much more) reduced dimension.
- Or, the PCA can be also used to represent a face, with much less information.

PCA: Example

- Example of the 16 principal components (eigenvectors with the largest eigenvalues) from the above data set:



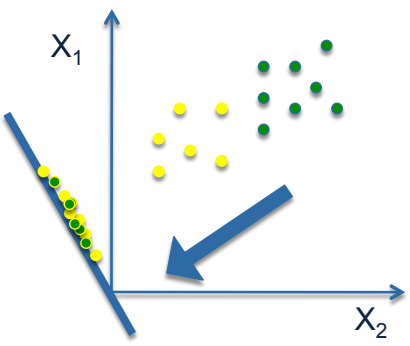
Source: http://commons.wikimedia.org/wiki/File:Matching_example_using_the_Shape_Context_descriptor.jpg

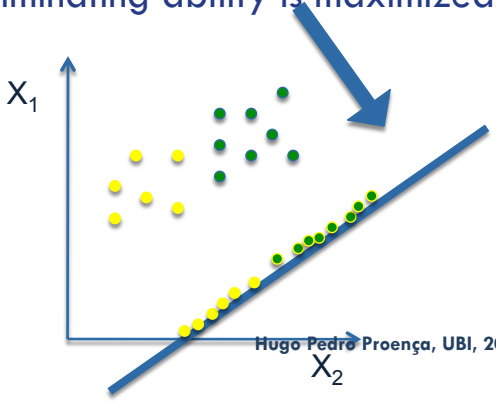
LDA: Linear Discriminant Analysis

- The main goal of LDA is to reduce the dimensionality of data sets.
  - The same goal of PCA?
- The main difference between PCA and LDA is that here we want to maximize the discriminating ability of the selected data.
  - In opposition to PCA, where the class of each instance is not considered.
  - Recall that the main idea of PCA to re-express the available dataset to extract the relevant information by reducing the redundancy and minimizing noise.

LDA: Linear Discriminant Analysis

- Having a pattern classification problem. We have T_1 and T_2 total instances, respectivelly belonging to classes C_1 and C_2.
- Each element of C_i is represented in a F^n feature space (**X**), i.e., by a set of "n" features.
- The main idea is to projectall instances into a feature space F^m (**Y**), such that m < n and the discriminating ability is maximized.





LDA: Linear Discriminant Analysis

- In order to find the best projection vector, it is required to define a "measure of goodness" of each projection.
  - So we have a way to select some projection, and discard the remaining.
- We start by obtaining the mean vector of each class in the original (X) and transformed (Y) feature spaces.

$$\mu_i = \frac{1}{N_i}\sum_{x\in\omega_i} x \quad \text{and} \quad \tilde{\mu}_i = \frac{1}{N_i}\sum_{y\in\omega_i} y = \frac{1}{N_i}\sum_{x\in\omega_i} w^T x$$

$$= w^T \frac{1}{N_i}\sum_{x\in\omega_i} x = w^T \mu_i$$

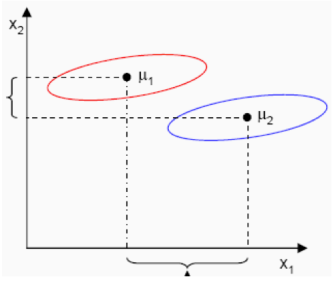
LDA: Linear Discriminant Analysis

- The objective function must consider the distance between the projected means for classes C_1 and C_2.

$$J(w) = \left|\tilde{\mu}_1 - \tilde{\mu}_2\right| = \left|w^T\mu_1 - w^T\mu_2\right| = \left|w^T\left(\mu_1 - \mu_2\right)\right|$$

- This measure is not good enaough, because it doesn't take into account the standard deviation within both classes:

Better separability
Smaller distance



Worst separability
Largest distance

LDA: Linear Discriminant Analysis

- The solution proposed by Fisher is to maximize a function that represents the difference between the means, normalized by a measure of the within-class variability, or the so-called *scatter*.
- For each class, the scatter is given by variance:

$$\tilde{s}_i^2 = \sum_{y \in \omega_i} (y - \tilde{\mu}_i)^2$$

- Measures the variability per class, atter projecting all instances to the reduced feature space (Y).
- Hence, the variability for both classes is given by the sum of corresponding values:

$$\tilde{s}_1^2 + \tilde{s}_2^2$$

LDA: Linear Discriminant Analysis

- Step 3: Obtain within-class scatter matrix:

$$S_w = S_1 + S_2 = \begin{pmatrix} 1 & -0.25 \\ -0.25 & 2.2 \end{pmatrix} + \begin{pmatrix} 2.3 & -0.05 \\ -0.05 & 3.3 \end{pmatrix} = \begin{pmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{pmatrix}$$

- Step 4: Obtain between-class scatter matrix:

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T = \left[\begin{pmatrix} 3 \\ 3.8 \end{pmatrix} - \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix} \right] \left[\begin{pmatrix} 3 \\ 3.8 \end{pmatrix} - \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix} \right]^T = \begin{pmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{pmatrix}$$

LDA: Linear Discriminant Analysis

- Step 5: The LDA projection is obtained by the solutions of the generalized eigenvalues problem:

$$S_W^{-1} S_B w = \lambda w$$

$$\Rightarrow \left| S_W^{-1} S_B - \lambda I \right| = 0$$

$$\Rightarrow \left| \begin{pmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{pmatrix}^{-1} \begin{pmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = 0$$

$$= (9.2213 - \lambda)(2.9794 - \lambda) - 6.489 \times 4.2339 = 0$$

$$\Rightarrow \lambda^2 - 12.2007\lambda = 0 \Rightarrow \lambda(\lambda - 12.2007) = 0$$

$$\Rightarrow \lambda_1 = 0, \lambda_2 = 12.2007$$

LDA: Linear Discriminant Analysis

- Step 6: According to the properties of eigenvectors and eigenvalues:

$$\begin{pmatrix} 9.2213 & 6.489 \\ 4.2339 & 2.9794 \end{pmatrix} w_1 = \underbrace{0}_{\lambda_1} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

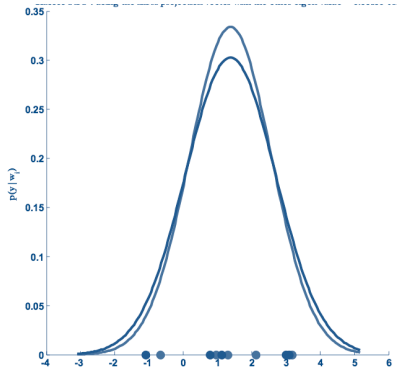
$$\begin{pmatrix} 9.2213 & 6.489 \\ 4.2339 & 2.9794 \end{pmatrix} w_2 = \boxed{\underbrace{12.2007}_{\lambda_2} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}}$$

$$w_1 = \begin{pmatrix} -0.5755 \\ 0.8178 \end{pmatrix} \quad and \quad \boxed{w_2 = \begin{pmatrix} 0.9088 \\ 0.4173 \end{pmatrix} = w^*}$$

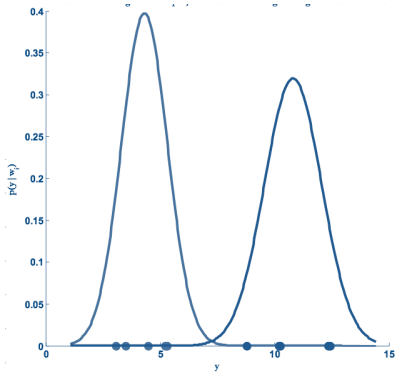
LDA: Linear Discriminant Analysis

- Separabilities obtained when using the smallest and the largest eigenvalues:

(Smallest)



(Largest)



Hugo Pedro Proença, UBI, 2014

Hugo Pedro Proença, UBI, 2014

[10]

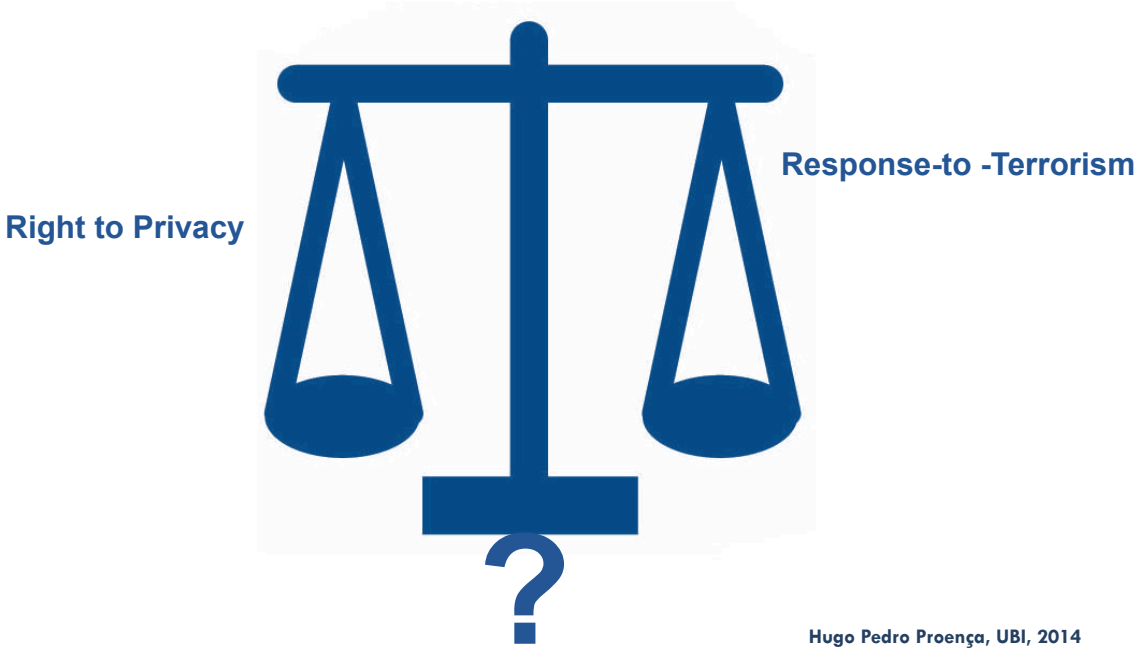
BIOMETRIC RECOGNITION SYSTEMS (MEI)

University of Beira Interior, Department of Informatics
Hugo Pedro Proença, 2014/2015

Syllabus

Theoretical	**Ethics in Biometrics: *bios* vs. *zoe* discrimination; Right to privacy vs. right to anonymity; Social / cultural impact of biometrics; (anti)Spoofing techniques; Security vs. privacy trade- offs; Overview of official ethical reports: *RAND*, *BIOVISION*, *OECD* and *NSTC*.**
Labs	Exercices MATLAB Neural Network and Classification toolboxes; Bagging, Boosting and Random Forest examples.

Biometrics: Ethics and Privacy

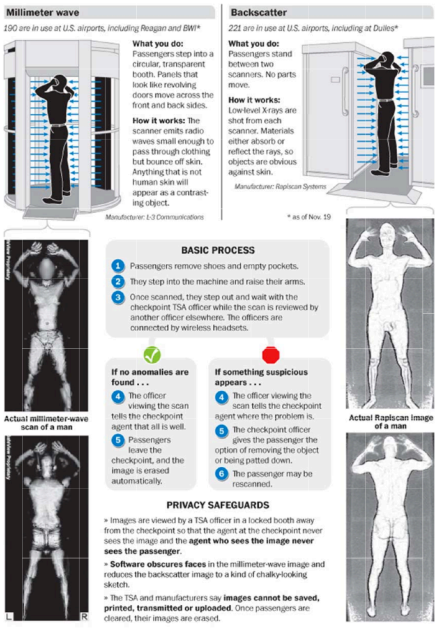


Biometrics: Ethics and Privacy

- Biometrics is said to analyze automatically our **bios**, for recognition purposes
- However, the correct designation should perhaps be "zoemetrics".
- **Zoe** is the Greek term to refer human life according to the main properties that we share with all living beings.
  - In opposition, bios refers to a cultural perspective
- Biometrics deals primarily with human zoe, not with human bios.
  - But can we clearly separate both phenomena?
- Aren't biological differences, our genotype and phenotype, intimately related with cultural ones?
  - Are we able to fully analyze our bios?
  - Are we able to fully analyze our zoe?

Biometrics: Ethics and Privacy

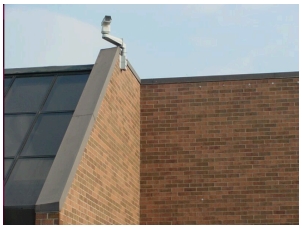
- Is this a reasonable procedure?
  - What if prevents that a terrorist enters and explodes a plane?
  - On the other way, is it mandatory to expose myself completely just to enter a plane?



Source: http://en.wikipedia.org/wiki/Full_body_scanner

Biometrics: Ethics and Privacy

- As in many other domains, the difficulty is to establish a precise limit, for the "*acceptable*" use of a biometric technology.
  - This is particularly notorious for covert biometric recognition systems
  - For what cases is it reasonable to get my biometric data collected?



Nearby a public building?

At the supermarket?

At the gas station?

Source: http://commons.wikimedia.org/wiki/File:Matching_example_using_the_Shape_Context_descriptor.jpg

Biometrics: Ethics and Privacy

- The existing regulations unavoidable contain subjective terms that difficult the objective perception of the cases where biometric recognition systems are acceptable
- In short, **there is no** clear **constitutional barrier** to government deploying a **national biometrics identification and tracking system**

Biometrics: Ethics and Privacy

- Civil libertarian entities fear is that the increasingly high desire to be safe will unavoidable push the balance toward the side of "security", instead of "privacy"
- However, they cite Benjamin Franklin
  - ***"They that can give up essential liberty to obtain a little temporary safety deserve neither liberty or safety"***

Biometrics: Ethics and Privacy

- However, the context in B. Franklin's life was completely different. As stated by Neil Young (musician):
  - ***Benjamin Franklin said that anyone who gives up essential liberties to preserve freedom is a fool, but maybe he didn't conceive of nuclear war and dirty bombs***

Biometrics: Ethics and Privacy

- A set of reports were published, by official entities, on the ethical and social implications of biometrics:
- **RAND Report**: Army biometric applications: : identifying and addressing sociocultural concerns
- Working Paper of the data protection working party of the European Commission: biometrics
- **BIOVISION Report**: BIOVISION –Roadmap to successful deployments from the user and system integrator perspective
- Organisation for Economic Co-operation and Development (**OECD**) Report: Biometric-based technologies
- Report of the Institute for Prospective Technological Studies – European Commission Joint Research Centre: Biometrics at the frontiers: assessing the impact on society
- National Biometrics Challenge Report (and other Reports) of the National Science and Technology Council (**NSTC**) of the United States.

RAND Report

- The RAND Institute was asked by the US Army to examine the legal, ethical and sociological issues raised by biometrics.
- The report concluded about three areas of ethical and social concern raised by biometric technology:
  - **1) Information privacy**
    - Creep, tracking and data misuse
  - **2) Physical privacy**
    - the stigma associated with some biometrics;
    - the possibility of actual harm to the participants by the technology itself;
    - the concern that the devices used to obtain or "read" the biometric may be unhygienic;
  - **3) Religious objections**
    - Some Christian groups consider biometrics to be the brand of the Evil on the basis of a interpretation of the Revelation book

Working Paper of Data Protection Working Party of the European Commission

- The EU Commission Advisory Body on Data Protection and Privacy addressed the issue of privacy.
  - The document lists reasons for concern and sets some basic principles.
  - First of all the working paper emphasizes that biometrics identification technology must respect the so called "purpose principle" according to which no personal data can be collected without explicit and legitimate purposes.
  - A second principle is the respect for proportionality. Biometric data may only be used if adequate, relevant and not excessive. "A respect for the principle of proportionality – concludes the working party – imposes a clear preference towards biometric applications that do not process data obtained from the physical traces unknowingly left by individuals or that are not kept in a centralized system".
  - Finally, they consider the risk that biometric data may contain more information than that which is necessary for identification or authentication/verification functions.

BIOVISION Report

- The project identified some human and social elements relevant to biometrics
  - "*Initially, it appears that people have a spontaneously positive attitude towards biometrics*" –
  - "At a second glance there is a tendency to be skeptical, especially with regard to the privacy issues when using biometrics.
- The report mostly focuses medical implications:
  - "One of the issues that causes concern in the application of biometrics is that there may be a direct or indirect medical risk.
  - As biometrics gains prominence, we anticipate that curiosity or speculation could make potential users question the direct or indirect effects of biometric techniques on their health." While concerns for direct medical are largely irrational, concerns for indirect medical implications deserve to be discussed more in depth.

OECD: Biometric-based technologies

- The Organization for Economic Cooperation and Development (OECD) issued a report on biometric-based technologies, discussing the privacy concerns in biometrics.
- The report highlights three areas of privacy concern
  - the potential of "function creep" in biometric systems;
  - the risk that these systems may become an infrastructure of surveillance;
  - that consent and transparency may be optional in certain biometric implementations.
- The OECD document argues that it is (almost) never acceptable to store biometric template in a centralized databank;
- The document regards the very nature of biometric data, considering it extremely sensitive

Institute for Prospective Technological Studies

- The issues report considers five areas of possible public concern:
  - **Privacy:** "One could argue that the use of a part of oneself (the biometric feature that is being digitized, stored and compared) as one's identity is eliminating the space that we traditionally place between our physical selves and our identity."
  - **Social aspects.** They concern the need to prevent function creep and to address factors such as age, ethnicity, gender, diseases or disabilities (including natural ageing), which could impair usability in certain categories of people.

Institute for Prospective Technological Studies

- **Economic aspects.** They address the issue of "optimal identity". "The strongest identity protection is not necessarily the optimal one identity errors and abuse may become less frequent, but when they happen, they could potentially be more dangerous";
- **Legal aspects.** Consider data protection rules, transparency and privacy. The report warn against risks carried by wider implementation and "about the failure to protect individuals from their inclination to trade their own privacy with what seems to be very low cost convenience";
- **Medical aspects.** They distinguish between direct and indirect medical implications., as in the BIOVISION report

National Science and Technology Council Report

- This report describes the major challenges that must be addressed by the biometrics community. They consider biometrics one of the most promising identity management tools.
- They underline the need of a synergic work to overcome the challenges: such a work will lead the community to meet evolving operational requirements while being supported by a robust biometrics industry

Security / Privacy

- The trade-off is clear:
  - How much security we are allowed to loose, to get a bit more privacy (and *vice-versa*)?
- There are standards that tend to enforce ethics
- **AITP Standards of Conduct**
  - *Protect the privacy and confidentiality of all information entrusted to me.*
  - *Inform the public in all areas of my expertise.*
  - *Insure that products of my work are used in a socially responsible way.*

Security / Privacy

- **ACM Code of Ethics**
  - *Respect the privacy of others.*
  - *Ensure that those affected have their needs clearly articulated.*
  - *System must be validated to meet requirements.*
- **Article 29 Working Party**
  - It regularly issues opinions, publishes working documents and resolutions on different topics related to the protection of privacy and personal data, aiming towards a harmonised application of these directives in the 27 EU Member States.

[11]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

University of Beira Interior, Department of Informatics
Hugo Pedro Proença, 2014/2015

Syllabus

Theoretical	**The Universal declaration of human rights; National regulations: the Comissão Nacional para a Proteção de Dados; International regulations and Standards: 8 CFR 103.16 - Collection, use and storage of biometric information; European data protection directive 95/46/EC, European ART 29 - Working Party on the Protection of Individuals with regard to the Processing of Personal Data.**
Labs	Hands-on biometric system I: choosing the trait and acquisition environment; Data acquisition setup and protocol; Expected data variations; Data collection.

Biometrics Regulations

- Due to its sensitiveness, biometrics is regulated by many different entities, standards and rules.
- One of the problem is that most of these regulations are not mandatory, due to national laws constraints and specificities.
- Above all, one the of the regulations that prevails is the Universal declaration of Human Rights, which has two articles that touch the biometrics domain

The *Universal Declaration of Human Rights* (UDHR)

- It is a declaration adopted by the United Nations General Assembly on 10 December 1948, after the WWII.
- The Article 12 states that:
  - "No one shall be subjected to arbitrary interference with his **privacy**, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the **right to the protection of the law against such interference or attacks**."
- The Article 3 states that
  - "Everyone has the right to life, liberty and **security** of person."

The *Universal Declaration of Human Rights* (UDHR)

- Both the articles 3 and 12 assure the right no privacy and security.
- But, what if at some case, both rights are impossible to assure in a jointly way?
  - "If you want to catch that plane safely, you must perform some kind of biological analysis that you may find as offending your privacy".
- At what specific circumstances, privacy rules over security?
- Is the right to privacy sometimes confused with the right to right to anonymity?
  - Are both concepts similar?

Comissão Nacional de Proteção de Dados

- It is the Portuguese legal entity that regulates the functioning of biometric recognition systems.
- Presently, anyone that wants to deploy a biometric system has to fill an electronic form, describing the conditions under which the system will run
  - Questions:
    - Who will process the information?
    - What's the main purpose of the system?
    - What's the biometric trait used?
    - Will you store the biometric data or the signature?
    - Do you have a centralized or local database?
    - Are biometric data supposed to be interchanged with third-party entities?
    - What are the physical and logical security plan, regarding the access to the database?

Directive 95/46/EC of the European Parliament

- Issued seven directives for the protection of personal data:
  - **Notice**: data subjects should be given notice when their data is being collected;
  - **Purpose**: data should only be used for the purpose stated and not for any other purposes;
  - **Consent**: data should not be disclosed without the data subject's consent;
  - **Security**: collected data should be kept secure from any potential abuses;
  - **Disclosure**: data subjects should be informed as to who is collecting their data;
  - **Access**: data subjects should be allowed to access their data and make corrections to any inaccurate data; and
  - **Accountability**: data subjects should have a method available to them to hold data collectors accountable for not following the above principles.

8 CFR 103.16 - Collection, use and storage of biometric information.

- Issued by the U.S. government, it regulates the way biometric data may be used and stored, both for American and foreign citizens (immigrants)
- A document of similar purposes was issued by the European Commission
  - The **Article 29 Data Protection Working Party** was set up under the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the **protection of individuals** with regard to the processing of personal data and on the **free movement** of such data.

Biometrics: Examples of Successes

- The National Research Council (NRC) released a report entitled Biometric Recognition: Challenges and Opportunities.
  - The main conclusion that they stated that any biometric system is inherently fallible, mostly because an automated match between two biometric records is based on similarity scores that represent "probabilistic" results.
  - The term probabilistic was misunderstood, to illustrate that there is a large probability of failure in any biometric system
  - Moreover, probability does not depend of the user, i.e., whether he behaves / not correctly during the data acquisition phase, i.e., it is intrinsic to the system
- IBIA: International Biometrics & Identification Association answered that report, by stating that:
  - Experience over the past decade has shown that biometric technology significantly enhances the effectiveness of many identity-based systems and constitutes an important tool in protecting borders, reducing entitlement fraud, enforcing laws, securing networks and facilities and protecting personal information from unauthorized access.

Biometrics: Successful Examples

- Moreover, IBIA illustrated 3 examples of success of biometric recognition technology:
  - **Tens of millions** of notebook computers shipped in the U.S. now **include biometric sensors** as an embedded feature to protect the owner's sensitive files and personal information from unauthorized access. Similarly, **biometrically-enabled smart phones** and other mobile devices that provide biometric data protection are now being introduced in the U.S. after achieving **widespread acceptance** in international markets like **Asia**.
  - According to a release from the Department of Homeland Security (DHS), the Secure Communities initiative, which uses biometric information and services to identify and remove criminal aliens in state prisons and local jails, **has resulted in the arrest of more than 59,000 convicted criminal** aliens this year alone, including **more than 21,000 convicted of major violent offenses like murder,** rape, and the sexual abuse of children. We can thank biometrics for the fact that these criminal aliens are no longer walking around freely in our society.
  - The Department of Defense (DoD) directive 8521.01E issued in 2008 states that "Biometrics is an important enabler that shall be fully integrated into the conduct of DoD activities to support the full range of military operations". **Biometric technology is being used more and more by the military** to protect U.S. bases from unauthorized entry by evil-doers carrying fake credentials.
    - In addition, U.S. combat troops **rely on biometric technology to** help identify **insurgents** and combatants in places like Afghanistan.

Biometrics: Successful Examples

- Computerworld magazine published a report about a specific case of deploying a biometric recognition system at a hospital.
- At a glance, they give three tips for the successful deployment of a biometric recognition system:
- **Test, test, test**
  - "Concerned about false positives and other problems with the readers, McQuaid knew that testing extensively was the only way to ensure the system would be working well from the start. Before the readers were rolled out to staff, they had been tested for weeks to make sure they were easy to use and ready to be implemented. "
  - "I imagined every way to break it," says McQuaid. "You only get one chance with physicians and nurses. They way they see it the first time is kind of what it is."
- **Train extensively**
  - The second step McQuaid recommends is to have a comprehensive training program for a biometric-based system. At Parkview, even per-diem workers are given detailed instructions on the system during training.
  - "Every time there is an orientation, we go in and work with them to make sure its done right," he says. "If you do it right the first time, it will save you a ton of calls on the back end."
- **Plan for backup**
  - The last thing McQuaid suggests before implementing biometrics is a strong backup plan. Every employee has several fingers scanned so that several can be used in case one finger is cut or not being read properly. As a last resort, the system is password enabled, so a staffer can enter the password if the bio reader fails to respond.
  - While the system isn't being used throughout the entire facility yet, the plan is to roll it out in other areas, like labs and the operating room, in the very near future.

Hugo Pedro Proença, UBI, 2014

[12]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

University of Beira Interior, Department of Informatics
Hugo Pedro Proença, 2014/2015

Syllabus

Theoretical	**Biometrics systems engineering: development; High-level design; Prototyping biometric recognition systems; Rapid application development frameworks; System interfaces; Performance optimization frameworks; Debugging, testing and validation; Proof-of-concept system.**
Labs	Hands-on biometric system III: Feature extraction, normalisation and classification.

Biometrics Systems Engineering

- A software development methodology framework emerged during the 1960s, due to increasing complexity of the computational systems.
  - Before that, it was up to the programmer (developer) to implement the software in a single-shot, i.e., the system was being modeled at the same time the software was being implemented;
  - The main goal is "*to pursue the development of information systems in a very deliberate, structured and methodical way, requiring each stage of the life cycle from inception of the idea to delivery of the final system, to be carried out in rigidly and sequentially*" (Geoffrey Elliott (2004) *Global Business Information Technology: an integrated systems approach*. Pearson Education. p.87.)

Source: http://commons.wikime

Biometrics Systems Engineering

- **Software systems engineering** is an engineering branch associated with the development of software, using well-defined scientific principles, methods and procedures.
  - The expected outcome is an efficient and reliable software product.

- IEEE defines software engineering as:
  - The application of a **systematic, disciplined, quantifiable approach** to the development, operation and maintenance of software;

Biometrics Systems Engineering

- There are several approaches to software systems engineering, such as:
  - **Waterfall:** a linear framework, in which development is seen as flowing steadily downwards (like a waterfall) through the phases of requirements analysis, design, implementation, testing (validation), integration, and maintenance;
  - **Prototyping**: an iterative framework characterized by the development of intermediate prototypes, i.e., incomplete versions of the software program being developed;
  - **Incremental:** a combined linear-iterative framework;
  - **Spiral:** combines elements of both design and prototyping-in-stages, in order to get the combine advantages of top-down and bottom-up concepts;
  - **Rapid application development (RAD)**: an iterative framework that involves iterative development and the construction of prototypes. The key is the fast development and delivery at a relatively low investment cost;

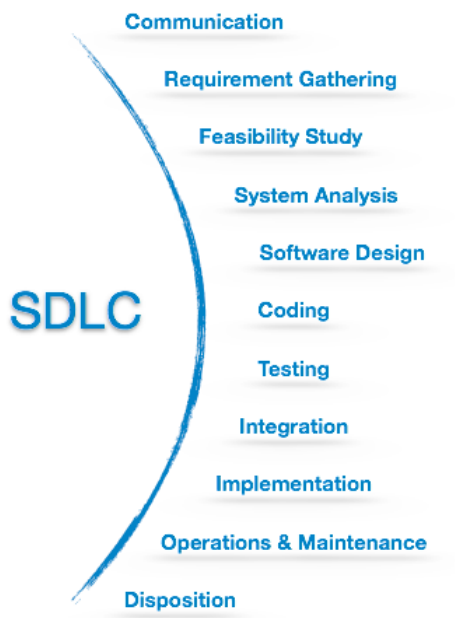
Biometrics Systems Engineering

- Biometrics is a particularly sensitive domain where the concepts about systems engineering should be faithfully addressed.
  - **Privacy** / **Security** issues
  - Access to **fundamental** (vital) features

Source: http://commons.wikimedia.org/wiki/

SDLC

- SLDC refers to the Software Development Life Cycle.
- It is a well-defined, structured **sequence of stages** in **software engineering** to develop the intended software product;
- Particularly fits the requirements / constraints of a biometric recognition system



Source: http://www.tutorialspoint.com/software_engineering/software_development_life_cycle.htm

SLDC: Communication

- Communication is the initial step, where someone requests a desired software product.
  - There should be some kind of contact to the provider, and the initial terms are negotiated;
  - In Biometrics (and not only), it is particularly important to perceive / make clear the assumptions:
    - Are the acquisition devices already available?
    - Are there appropriate SDKs for such devices?
    - What are the "degrees-of-freedom" of the system?

SLDC: Requirement Gathering

- Interviews and discussions with the various stakeholders in the problem domain should occur.
  - Try to bring out as much information as possible on their requirements. It typically involves:
    - Studying the existing or obsolete system and software,
    - Conducting interviews of users and developers,
    - Check the company database, if allowed
    - Collecting answers from the questionnaires.
- In biometrics engineering, it is important to perceive the desired system in terms of the traditional systems classification:
  - overt vs. covert, habituated vs. non-habituated, attended vs. non-attended, standard vs. non-standard environments, public vs. private, open vs. close;

SLDC: Feasibility Study

- This step is often neglected, which might lead to financially disastrous decisions.
- The team designs a draft of the software development process and of the resulting system.
  - Will the resulting system fulfill all requirements of stakeholders?
  - Is there any possibility to go beyond the requirements?
  - Is the project financially, practically and technologically feasible, given the available resources?
  - Specifically for biometrics, **is it legal to implement / deploy the proposed system?**
  - **What are the ethical / privacy considerations in the project?**

System Analysis

- Having positively concluded about the feasibility of the project, the team…
  - designs a roadmap of the plan and
  - try to select the most adequate software model suitable for the project.
- This step includes understanding the limitations of the resulting system,
- Also, in the internal perspective, the impact of project on organization and personnel should be evaluated here.
- At the end of this step, there should be:
  - A list of the functional requirements of the biometric system. E.g.,
    - "*[F1]: The system shall allow new users to enroll, by creating an account on the web site*".
    - "[F2] The system should not allow that users modify their username fields".

Software Design

- This phase joins all pieces of information collected during the previous phases:
  - The major system requirements;
  - The types of users;
  - The users input;
  - A list of actions, for any input;
  - The output of this phase is divided into:
    - **Logical plan.**
      - Description of meta-data and data dictionaries, logical diagrams, data-flow diagrams and in some cases pseudo-code.
    - **Physical plan.**
      - Description / specification of the data acquisition sensors, data processing hardware, network architecture and data storage plan

Software Design

- What are the actors in the system?
- What are the actions allowed for each actor?
- **Use Cases:**
  - Describe **objectively** what the system should do in **response** to any **action** from one **actor**, and how will this **meet** a particular **functional requirement.**
  - E.g., [UC1], ID: "Display Account Balance", Send information about the user account to the screen, provided that he completed successfully the authentication process, i.e., the matching similarity score was above a threshold T_1

Coding

- This is the fundamental **programming** phase.
- In this phase, the implementation of the software actually starts in terms of writing program code, according to the programming language and integrated development environment previously found the most suitable.
- In Biometrics development, it is particularly important to know how will the resulting modules be evaluated, for error catching.
  - **Do we have access to real data?**
  - **If so, has the data the same features of the one acquired during system running?**

Testing

- The initial phase of testing is more devoted to catching programming errors, and to guarantee that the resulting software runs as planned during the previous phases.
- In biometrics, there is a set of standards test procedures that should be carried out on any resulting system
- **Biometric Data Interchange**
  - Fingerprint Minutiae, Standard: ANSI INCITS 378-2004; Test method: ANSI INCITS 423.2-2008
  - Fingerprint Minutiae, Standard: ISO/IEC 19794-2:2005; Test Method: ISO/IEC 29109-2:2010
  - Finger Image, Standard: ANSI INCITS 381-2004; Test Method: ANSI INCITS 423.4-2009
  - Finger Image, Standard: ISO/IEC 19794-4:2005; Test Method: ISO/IEC 29109-4:2010
  - Face image, Standard: ISO/IEC 19794-5
  - Iris Image, Standard Iris Exchange (IREX)
- **Biometric Application Program Interface**
  - Standard: ANSI INCITS 358-2002; Test Method: ANSI INCITS 429-2008
  - Standard ISO/IEC 19784-1:2006; Test Method: ISO/IEC 24709-2:2007

Integration

- The resulting system should need to be integrated with the libraries, databases and other program.
- It is particularly important to decide the way the **software and database** interaction is carried out
  - Will the DBMS compromise system security?
  - Will the communication be done over secure channels?

Operations Maintenance

- Having the system running (at least experimentally), it should be tested for portability and adaptability and integration related issues.
  - "What if <something> happens?"
  - Train a groups of elements that will operate the system in a daily basis
  - They will be the front-end of the system, and essential to keep it functional
  - Design a maintenance plan
    - Schedule
    - Frequency
    - Tasks
    - Alert and their severity

Biometrics System Maintenance

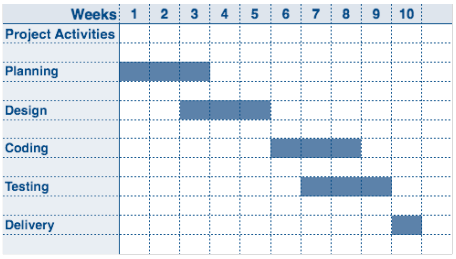
- Keep in mind that the a biometrics system is as good as the weakest link in the system.
  - Maintenance of the database, networks, doors, locks, readers, cameras and related systems and applications that support, or are controlled by, the biometrics system is vital.
- At first, the accurate enrollment is most times essential, i.e., assure that good quality data of fingerprint or iris scans are entered into a database.
- Administration of the database is always an issue.
  - What if someone leaves the company?
- All the physical components of the system need to be regularly assessed for mechanical or other failure.
  - If the biometrics acquisition device gets grimy or is damaged through use, exposure to harsh environments or the elements (rain, dust, heat) it will collapse the entire system.
- Regular maintenance every three months, or as often as the use demands, could eliminate most of the risks and extend the life of the system
  - Keep a track record of the qualified staff.

Biometrics Project Management

- The risk and uncertainty rises exponentially with respect to the size of the project, even when the project is developed according to set methodologies.
  - In projects of moderate / big dimension, there are certain tools that become **fundamental**, i.e., in practice terms, the project **cannot** be successfully carried out without using them.
    - **Gantt** charts;
    - **PERT** charts;
    - **Resource histograms**;
    - **Critical path analysis**;

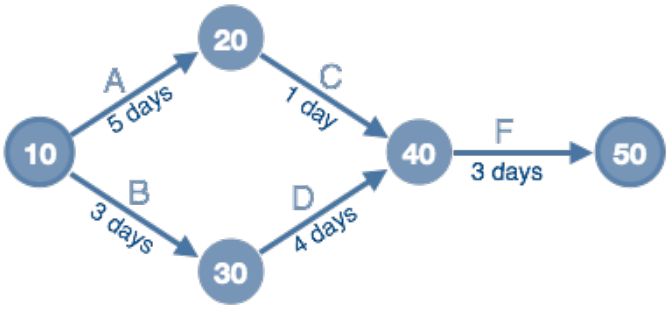
Biometrics Project Management

- **Gant charts.**
  - Were devised by Henry Gantt (1917). They represent the project schedule with respect to time periods. Consist of a horizontal chart with bars representing activities and time scheduled for the project activities. Activities (tasks) appear in the vertical axis;



Biometrics Project Management

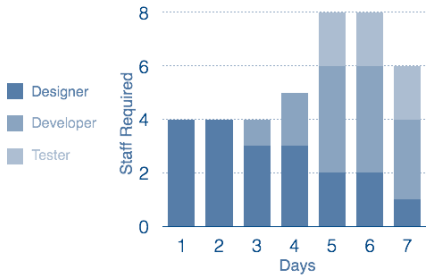
- **PERT charts.**
  - It is an acronym for "Program Evaluation & Review Technique" and it depicts the project as graph, representing the main events in both parallel and consecutive way.



Resource Histograms

- Are graphical tools that represent the quantity of resources required over time for a project phase (task).
- Are effective tools for staff **planning, coordination** and for the **project budget.**

Staff	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Designer	4	4	3	3	2	2	1
Developer	0	0	1	2	4	4	3
Tester	0	0	0	0	2	2	2
Total	4	4	4	5	8	8	6



Hugo Pedro Proença, UBI, 2014

Hugo Pedro Proença, UBI, 2014

[13]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

Syllabus

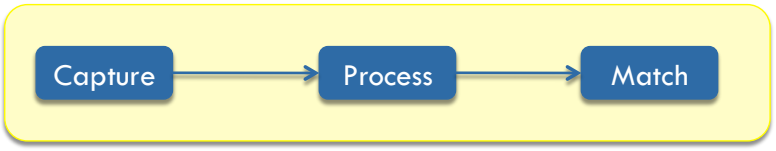
Theoretical	**Biometrics systems engineering: integration and maintenance; Data storage and interface management; Security configuration; Job scheduling; Preventive maintenance plans; System breakdown plans. Biometric Interfaces**
Labs	Hands-on biometric system IV: system test and maintenance plan.

Biometrics Interfaces

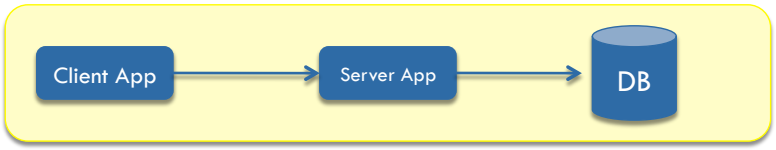
- Biometric interfaces refers to the methods by which one component of the biometric system communicates with another.
  - These components may be devices, software, or entire external systems.
  - A particular sensitive point is the definition of policies for exchange of information.
  - Interfaces are key elements of biometric system architecture and design and provide the basis for interoperability.
- Biometric systems are composed of subsystems and components
  - As in any the general case of computational systems, having strong components does not guarantee a strong system
  - Components must interact with one another, using the defined interfaces

Biometrics Types of Interfaces

- Interfaces belong to one of three main families:
  - Inter component



  - Intra application



  - Inter system

Data Interchange

- Biometric data exists in several forms. Even though they are all sensitive, some are particularly important to concern about
- Raw data captured by a sensor
- Partially processed data (e.g., ROI detected, contrast enhanced,…)
- Biometric signature, suitable for matching
- Any piece of data may be formatted and encoded in different ways (e.g., compressed or uncompressed)
- It can also exist as a single sample or be packaged together with other like or unlike samples.
- It may be in proprietary format or (more dangerous) in a standard format, with or without associated metadata.

Data Interchange

- The format information about biometric data must be (obviously) met by the sender / receiver
  - This is obviously easier if both ends are inside the same entity
- At the end, in a completely closed system the data format can be simply "whatever works better".
  - This leads to high customized formats
- The problem arises in open systems
  - Formats need to be standardized, in order to be understood by a large number of external systems
- Data interchange formats exist nowadays for most modalities, including structures, meta-data, headers and the definition of security information
  - Example: Common Biometric Exchange Formats Frame- work (CBEFF) and ANSI/NIST ITL1-2007

Device Interfaces

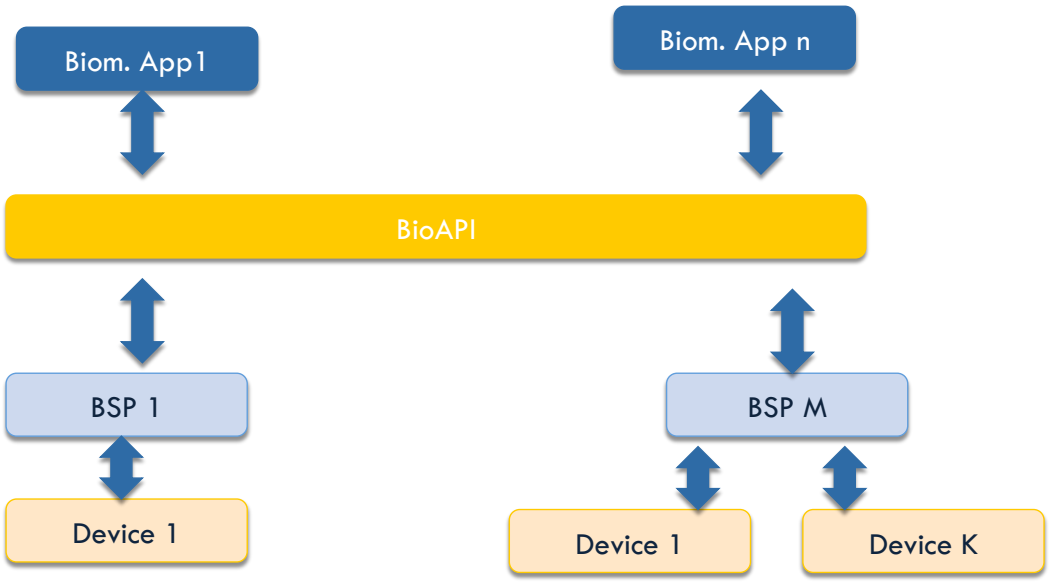
- Biometric sensors capture data and often provide additional processing / storage capabilities
- In order to integrate a biometric device in a computational application, an interface to that device must be designed. This includes:
  - the physical interface;
  - the communications protocol;
  - the data/ message exchange.
- Industry standards define both the physical interface and communications protocols. Because biometric data samples can be very large, an interface that provides adequate speed and bandwidth is desirable.
  - Today, the Universal Serial Bus (USB) or IEEE 1394 ("Firewire") are the most commonly used.
- In addition to the device drivers, biometric device manufacturers usually provide software developer kits (SDKs) to manage the interaction between the device and third-party applications.
- Applications interface to SDKs via a defined application programming interface (API) as described in the following.

Software Interfaces

- These are the entities that provide a set of biometric functions or capabilities via a software interface.
  - It might include biometric processing and matching algorithms, or the control of a biometric device.
    - Reusable software packages are called SDKs.
    - Biometric SDKs with standardized interfaces are called biometric service providers (BSPs).
- APIs can be either be provided at high or low level
  - A high level API provides a set of abstract, generalized functions (e.g., "Enroll" or "Recognize")
  - A low level API provides more specific, atomic functions (e.g., "Capture Data" or "Set Contrast level").
- The most well known biometric API is the **BioAPI**, originally developed by a group of over 100 organizations from industry, government, and academia and published in 2000 as an open systems industry specification.
  - Subsequently, version 1.1 was published as an American National Standard (INCITS 358) and version 2.0 as an international standard (ISO/IEC 19784

Software Interfaces: BioAPI

- The idea in BIOAPI is to provide a common architecture for working with heterogeneous biometric applications and service providers:



Communications Interfaces

- In order to transmit information between systems or module of the same system, a communication interface must be used
  - The best known are those used by the justice community.
    - The FBI's Electronic Fingerprint Transmission Specification (EFTS);
    - The Interpol Implementation (INT-I)
  - Both utilize the ANSI/NIST ITL1-2000 standard to define transactions (request and response messages) with their respective systems
- ANSI/NIST ITL1-2000/2007 defines the content, format, and units of measurement for electronically encoding / transmitting biometric data, and the associated biographic information.
- This standard defines a series of "record types", each one for a type and format of data.
  - For example, a Type-4 record contains a high resolution grayscale fingerprint image;
  - A Type-9 record contains minutiae data

Service Interfaces

- A recent paradigm in systems engineering is the "service oriented architecture" (SOA)
- Under this paradigm, the information systems are decoupled from systems providing the biometric recognition
  - This augments the modularity of the resulting systems. For instance, it is much simpler to change the type of trait used, while keeping the majority of system features working.
- The most often used protocols for such services are XML over Hypertext Transmission Protocol (HTTP) or Simple Object Access Protocol (SOAP) over HTTP.
  - SOAP services are defined in terms of Web Services Definition Language (WSDL) and frequently utilize a set of existing web services standards.
  - Service providers may post their WSDL to a directory which can be read by potential users or, in closed systems, may be provided directly to known requesters.
- Although today most services interfaces are system specific, a project known as Biometric Identity Assurance Services (BIAS) is in progress to standardize a set of generic biometric Web services.

Biometric Data Management Standards

- This is concerned with the broader issues of management and procedural measures for protecting biometric data.
  - Includes awareness training and accounting and auditing procedures as well as a reference to technical measures.
- This kind of standard originated from the ANSI X9 subcommittee in the US X9.84 Standard – Biometric Information Management – Security (2003)
  - Later on, it evolved to the development of ISO 19092-2008 – Financial services – Biometrics – Security Framework
  - ISO 19092 is a biometric-specific extension of ISO 17799, the Code of Practice for Information Security Management

Biometric Data Security Standards

- The standard for biometric data storage and exchange is ISO/IEC 19785 CBEFF (*Common Biometric Exchange Format Framework*).
- CBEFF is a multi-part standard where Part 4 "*Security block format specifications*" provides the information regarding the protection of biometric data integrity and confidentiality.
  - It is based on a basic block of biometric data: **Biometric Information Record** (BIR)
  - BIR is subdivided into a standard **Block Header** (SBH), a **Biometric Data Block** (BDB) and the **Security Block** (SB)
    - The SB security block contains relevant security information such as cryptographic checksums, digital certificates, and data encryption algorithm specifications etc. that are used to guarantee the integrity and confidentiality of the data.

Biometrics Security Evaluation Standards

- ISO/IEC 19792: Information Technology – Security Techniques – Security Evaluation of Biometrics
  - This standard was developed in the scope of the Standards Committee 27.
- It is not targeted at a specific evaluation scheme;
- Instead, it aims to provide guidance to developers and evaluators on security concerns for biometric systems and to specify a generic methodology for their evaluation.
- Gives detailed information on potential threats, countermeasures, and evaluation requirements.
  - It assumes that evaluators are familiar with the broader IT security evaluation issues and does not address these.
  - 19792 covers biometric-specific security issues of the system as a whole as well as threats and potential vulnerabilities of the component parts.
  - Describes technical and nontechnical threats and how these may be reduced or eliminated by appropriate counter- measures.
  - It guides the evaluators on testing and the assessment of potential vulnerabilities and countermeasures, and it defines the responsibilities of vendors and evaluators in the evaluation process.

Biometric System Engineering

- Biometric System Architecture
  - It is primarily defined by the storage location of the templates and the physical place of the recognition module
  - Centralized / Distributed server
  - Workstation at the client side
  - Portable devices at the user side (e.g., smartcard)
- How the recognition module will work?
  - Match on card?
  - Match on device?
  - Match on server?
- The preliminary recognition phases usually take place near the sensor, in order to reduce computational burden of data storage and transmission.

Hardware / Software Design

- Having an architecture designed, the subsequent issues are:
- **Sample Acquisition**
  - Design the data acquisition protocol
  - Define the policy to deal with poor quality samples
- **User Interface**
  - Ergonomy
  - Demographic issues of the users population
  - Cultural constraints
- **Biometric Processing Components**
  - Used SDKs
- **Communication Channels**
  - Secure communication links?
  - Challenge-response mechanisms

Hardware / Software Design

- **Database Design**
  - Normalization form
  - Distributed architecture
  - Backup plan
  - Mirroring
- **Interoperability**
  - Complied standards
  - Operating system compatibility
  - Version issues
  - Software incompatibilities

Maintenance Policies

- **Integrity of Enrolment**
  - Guarantee that fake users are not able to enroll
- **Quality of Enrolled Data**
  - Users training
  - Minimum data quality required
  - Exception procedures (no biometric system is universal)
- **System Configuration**
  - Adjustment of Acceptance Thresholds
  - Raised alarms
- **Privacy**
  - Prevent inside and outside illegitimate access to data
  - Log control
  - Long integrity

Hugo Pedro Proença, UBI, 2014

Hugo Pedro Proença, UBI, 2014

[14]

BIOMETRIC RECOGNITION SYSTEMS (MEI)

University of Beira Interior, Department of Informatics
Hugo Pedro Proença, 2014/2015

Syllabus

Theoretical	**Types of errors in biometric recognition: false acceptances, false rejections; Confusion matrices; Acceptance thresholds; Equal error rate (EER); Receiver operating characteristic (ROC) curves, Detection-error tradeoff (DET) curves; Rank-N curves; Area under curve (AUC) and decidability index.**
Labs	Bootstrapping performance evaluation; Students cross-evaluation of hands-on biometric systems: spoofing and anti-spoofing strategies.

Biometrics Performance Evaluation

- Typically, performance metrics take the form of rates;
- Regardless the used metric, it should be noted that the measured error rate during laboratorial evaluation is distinct from the expected rate that occurs in the deployed biometric systems;
- The most typical performance evaluation measures depend of the **mode** the system is functioning
  - Verification (1:1)
    - It's up to the system to confirm that a claimed identity is the correct one
  - Identification (1:N)
    - The biometric system should find the closest signature in the database
- **Verification + Identification = Recognition**

Biometrics Performance Evaluation

- For the **Verification** mode, the most typical evaluation metrics are:
  - **False Match Rate (FMR)**: it's the probability that a system will incorrectly assume as "genuine" a comparison between biometric signatures extracted from different subjects.
    - It is the most concerning measure, in terms of security
    - It might give assess to unauthorized personal, in some sensitive resource
  - **False Non-Match Rate (FNMR):** it's the probability that a system will incorrectly assume as "impostor" a comparison between biometric signatures extracted from the same subject.
    - It is mostly concerning in terms of user convenience
    - It might imply that the system denies the identity of a valid user.

Biometrics Performance Evaluation

- There are performance measures that apply to both the Verification and Identification modes.
- **Failure-to-Acquire (FTA)**: it's the proportion of recognition attempts in which a biometric system fails to detect or acquire a biometric signal of adequate quality
  - Due to failures related to user presentation, sample segmentation, feature extraction, or quality control.
- **Failure-to-Enroll (FTE):** its is a measure that is similar to FTA, but applies to the enrollment process, i.e., when the acquisition conditions and protocol should be optimal.
  - Sample quality, and user-system interaction have influence in FTE

Biometrics Performance Evaluation

- **Generalized FAR:** This is a "functional" version of the FAR. It combines the enrollment, sample acquisition and matching errors. It is given by:

$$GFAR = (1\text{-}FTA) \times FMR_{threshold} \times (1\text{-}FTE)$$

Samples are acquired OK

Subjects are enrolled

- **Generalized FRR:** Combines the enrollment, sample acquisition and matching errors. It is given by:

$$GFRR = FTA + (1\text{-}FTA) \times FTE + (1\text{-}FTA) \times (1\text{-}FTE) \times FNMR_{threshold}$$

Samples are not acquired OK

Samples are acquired OK, subjects are enrolled but a matching error occurs

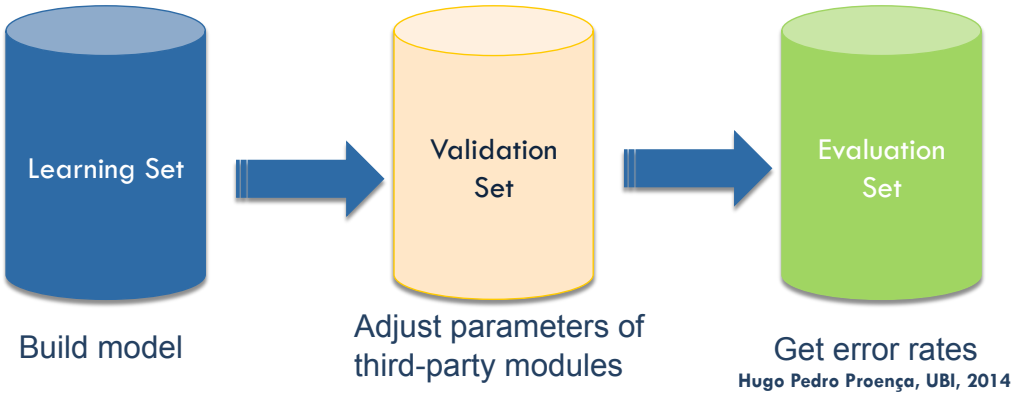
Samples are acquired OK but subjects not enrolled

Biometrics Performance Evaluation

- **True Acceptance Rate** (Sensitivity): it's the probability in which a biometric system successful matches two biometric signatures from the same person
- **True Rejection Rate** (Specificity): it's the probability in which a system successfully detects that two biometric signatures are not from the same person
- **Equal Error Rate:** It is the error value when both the FMR and F NPR have approximately the same value
  - It's often known as **Crossover Error Rate**
- **Half-Total Error Rate (H-TER):** this is an aggregate of FMR / FNMR (used not too often)
  - H-TER=(FMR + FNMR)/2
- **Weighted Error Rate (WER):** it's a generalization of H-TER, in which weights given to each type of error can vary

Biometrics Performance Evaluation

- It is critical that the performance measures are taken from a data set that is completely disjoint of the data used to build the model.
  - This is a classical rule of Artificial Intelligence, but that is still often neglected, producing expected error rates that are far below the error rates actually observed when systems are deployed

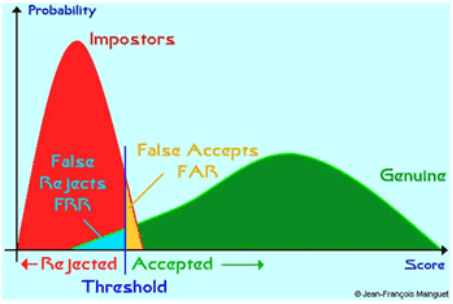


Biometrics Performance Evaluation

- The notion of **Acceptance Threshold** is critical for any biometric system, and determines all the operating points a system could have.
- Suppose that we are matching (Hamming distance) binary biometric signatures with 100 bits. We collect a set of N sample signature (from K subjects), and observed the typical number of agreeing bits between signatures.
- We divide our observations into two families:
  - **Genuine Comparisons:** if both signatures are from the same person
  - **Impostor Comparisons:** if both signatures are from different persons
- It is expected that – on average – matching signatures from the same subjects will produce the most number of agreeing bits.

Biometrics Performance Evaluation

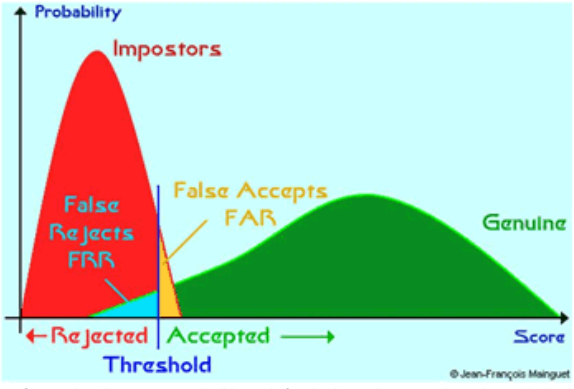
- **Decision Environment:**
  - If we estimate the density of the matching scores obtained for both types of distributions, we would have:



Source: http://www.recoware.hu/biometria/fogalmak_meghatarozasok_szohasznalat.html

Biometrics Performance Evaluation

- **Decision Environment:**
  - How many bits will be required to report a match?
  - If the value is too high, it will be really difficult to get a match, but almost impossible to observed a false match
  - On the other way, if the value



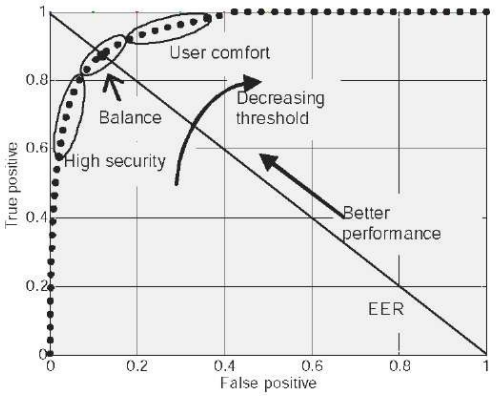
Source: http://www.recoware.hu/biometria/fogalmak_meghatarozasok_szohasznalat.html

Is too low, false matches will be likely, but almost for sure all genuine comparisons will produce matching scores higher than such value

  - The "number of agreeing bits" is actually the acceptance threshold and determines the possibilities for the working modes of the system

Biometrics Performance Evaluation

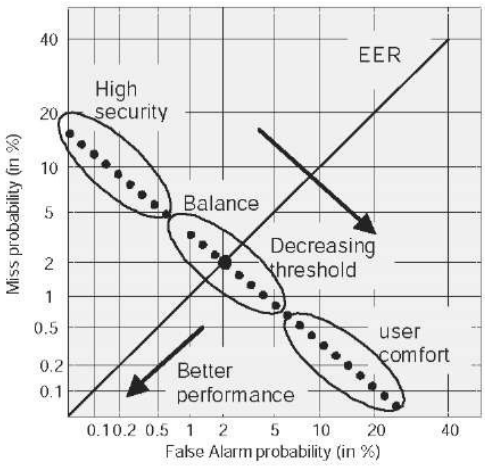
- **Decision Environment:**
  - **Each value for acceptance threshold** corresponds to a pair of FMR / FNMR values, or equivalently TPR / FMR.
  - By joining all points, we obtain the **Receiver-Operating Characteristic (ROC) Curve**
  - It is one of the most important ways to report the system performance, because it reflects all the operating points (and not just one)



Source: http://what-when-how.com/artificial-intelligence/biometric-security-technology-artificial-intelligence/

Biometrics Performance Evaluation

- Similarly, we can obtain another widely used plot, by relating the FNMR and the FMR: **The Detection-Error Trade-off Curve**

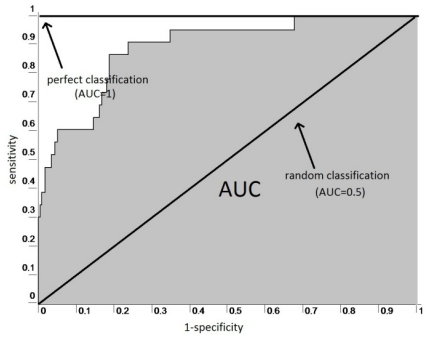


Source: http://what-when-how.com/artificial-intelligence/biometric-security-technology-artificial-intelligence/

Biometrics Performance Evaluation

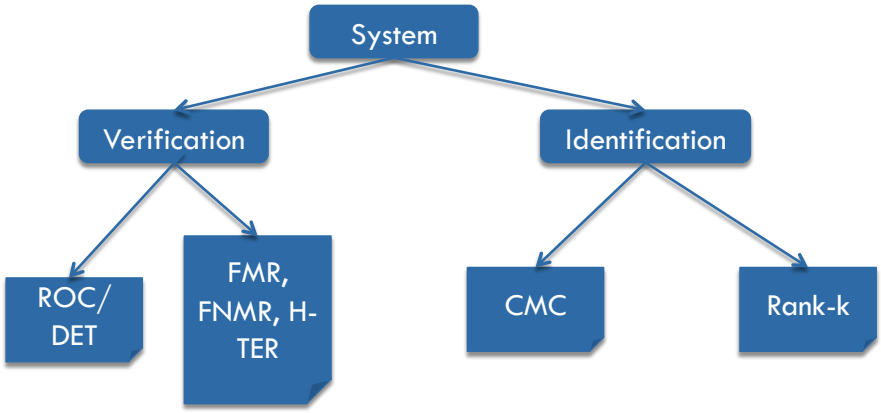
- Typically, based on the ROC curve we get a widely used summarizing measure: the **Area Under Curve (AUC)**
- The advantage of AUC is that it consists of a single value
- It is closed in the [0,1] interval.
  - "1" means that the system is optimal (error-free)
  - "0" denotes the worst possible system

(actually, the lowest we get is 0.5, for binary classification systems



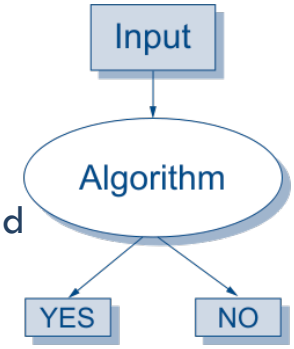
Biometrics Performance Evaluation

- **Most typically, the following error rates are used:**



Biometrics Performance Evaluation

- A Biometric recognition system, can be regarded as a binary classification problem.
  - At every step, it's up to the system to decide when a comparison between 2 biometric signatures regards (or not) the same subject.
- This is called a "**Decision Problem**"
- For this type of problems, there is a particularly relevant measure that measures how separated are the genuine and impostors distributions



Source: http://en.wikipedia.org/wiki/Decision_problem

Biometrics Performance Evaluation

- The **Decidability index (d-prime)** is given by:

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{1}{2}(\sigma_1^2 + \sigma_2^2)}}$$

being μ the means of both distributions and σ the corresponding standard deviations.

- Large values for the d-prime occur when the means are far from each other and the standard deviations small.
  - This corresponds to decision environments well separated.

Biometrics Performance Evaluation

- The International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) identify three primary categories of biometric performance evaluation:
  - **Technology evaluation;**
  - **Scenario evaluation;**
  - **Operational evaluation.**

Biometrics Performance Evaluation

- **Technology evaluation**
  - Technology evaluations are used to compare the offline effectiveness of algorithms that work with the same biometric modality.
  - Dataset selection is an important aspect of technology evaluations;
  - Dataset samples should be collected using sensors that are equally compatible with all algorithms under evaluation, and collected data should not be accessed by any algorithm prior to testing.
  - The use of "fixed" datasets helps ensure that technology evaluation results are repeatable.

Biometrics Performance Evaluation

Scenario evaluation

- Scenario evaluations are used to test the integrated performance of a prototype system, examining how samples collected from real test subjects are processed in real time in a modeled environment.
- As in technology evaluations, scenario evaluations examine the effectiveness of algorithms but also include the opportunity to assess hardware components (i.e. sensors) and user-system interaction.
- The acquisition sensors involved in scenario evaluations are unique to each test; therefore, each tested system will collect slightly different data.
- If multiple systems are undergoing comparison, scenario evaluations must control for environmental and population factors across all data collection.

Biometrics Performance Evaluation

Operational evaluation.

- Operational evaluations assess the performance of a biometric system in a precise application environment, using a specific target population.
- Unlike Technology and Scenario Evaluations, offline testing might not be feasible in operational evaluations.
- Also, it is extremely hard to get repeatability of results, as data is acquired live
- Ground-truth information might also be a problem

Biometrics Performance Evaluation

- **Test Population**
  - How many persons will be used in our performance evaluation test?
    - As many as possible might be too vague
  - Will the subjects be guided, or trained?
  - At what extent do the elements in that test population represent the definitive users of the biometric system?

Test Population

- **Test Subject Control**
  - Performance evaluations employ varying degrees of test subject behavior control, which can take the form of system training and instruction, feedback and guidance during system use, supervision, and permitting time for user habituation;
- **Size of Test Subject Population**
  - Typically, the reliability of results increases as the number of test subjects increases.
  - Tests conducted on small subject populations are likely to understate a system's capabilities or fail to identify the impact that outlier users have on system performance.

Test Population

- **Composition of Test Subject Population**
  - Test populations should reflect an application's intended user base with respect to user gender, ethnicity, age, physiology, and level of acclimation to the suggested technology
  - It is important that the main feature of the test population are similar to the corresponding features of the definitive users
- **Physiology and Behavior of Test Subjects**
  - It is equally important that both physiological and behavioral aspects of the test population resemble as close as possible the corresponding features on the definitive users

Hugo Pedro Proença, UBI, 2014

Further Readings: http://www.usma.edu/ietd/docs/BiometricsMetricsReport.pdf

Hugo Pedro Proença, UBI, 2014

[15]

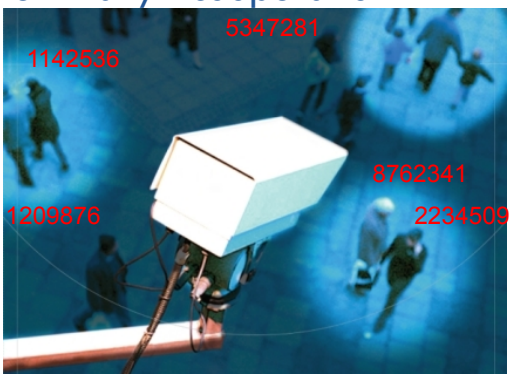
BIOMETRIC RECOGNITION SYSTEMS (MEI)

Syllabus

Theoretical	**Challenges behind the development of robust recognition systems; Hardness of non-cooperative recognition scenarios; Requirements in large-scale deployment; Bridging the gap between biometrics and visual surveillance or "*are we approaching to George Orwel's Big Brother?*"**
Labs	Ideas for biometrics research; overview of the state-of-the-art; cohesive perspective of innovations

Non-Cooperative Recognition

- To develop an automaton able to robustly recognize human beings, without requiring them any cooperation in the acquisition process.
  - Perhaps contrary to usual belief, such automata are still confined to science Fiction.



(Source: http://www.prisonplanet.com/)

  - Book "*Nineteen Eighty-Four*" by George Orwell (1948) gave birth to the idea of "*Big Brother*".

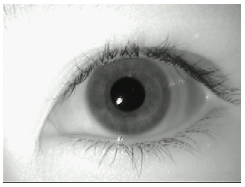
Main Goal

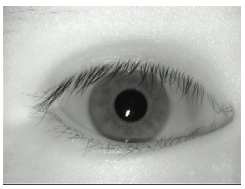
- A possibility: Use the iris / ocular region
  1. Naturally protected internal organ;
  2. Supports contactless data acquisition;
  3. Uses the lowest and middle-low frequency components to perform recognition;
  4. Its regular shape turns easier its detection and segmentation;
  5. It is not deformable, and does not change with expressions;
  6. Its planar shape turns easier to compensate for off-angle acquisition.

Main Goal

- Unconstraining the image acquisition protocol and not asking for the cooperation of subjects leads to serious decreases in the average quality of data:

Constrained/Cooperative Acquisition

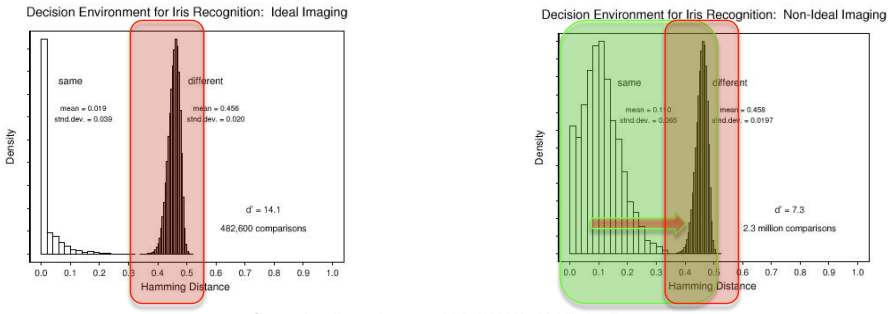




Unconstrained/Non-Cooperative Acquisition

Main Goal

- **Problem: Degradations in Performance**
  - Current biometric systems achieve remarkable performance when analyzing good quality data.
  - Performance significantly decreases when data is degraded, due to problems in the acquisition process.

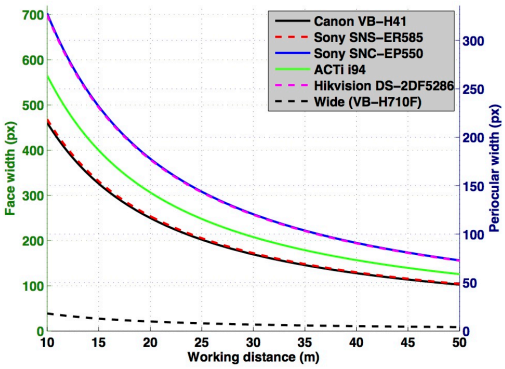


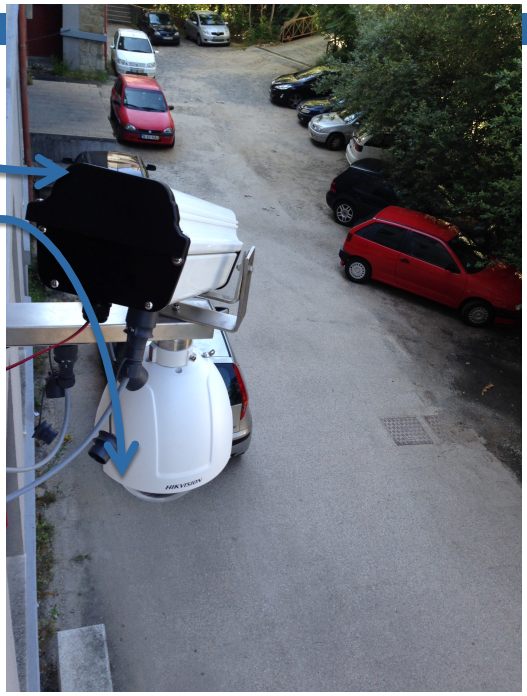
(Source: http://www.cl.cam.ac.uk/~jgd1000/decidability.html)

… and gets much worse on *heavily degraded* data

QUIS-CAMPI

- **Recognition Framework II**
  - **Outdoor environment**
  - **Surveillance Scenario**
  - **Two cameras:**





Example Video

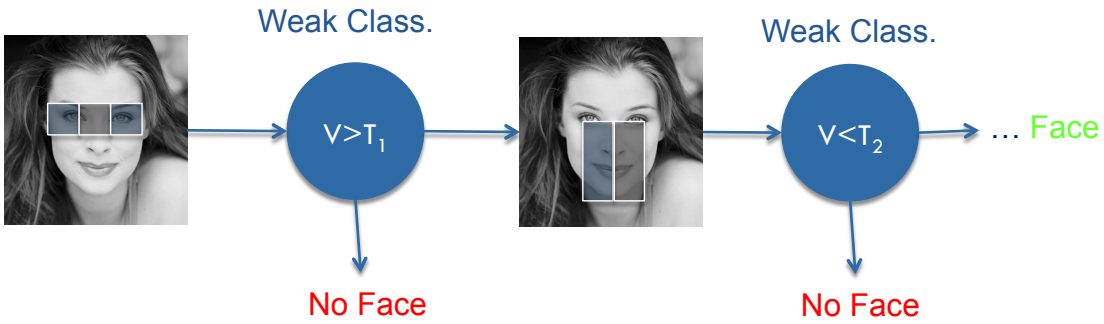
QUIS-CAMPI: Phase 1

- Face Detection
  - The breakthrough in face detection happened with Viola & Jones.
  - Uses a cascade of "weak-classifiers" (Haar features) and a Boosting ensemble scheme.

P. Viola and M. Jones; "Robust Real-Time Object detection", IJCV, 20 (11), 1254-1259, 2001

QUIS-CAMPI: Phase 1

- Face Detection: Cascade of Weak Classifiers.
  - Each classifier is based in a single feature, that measures the difference in intensity between two regions of a face.
  - In order to be classified as a "Face", an instance must pass over "N" weak classifiers.

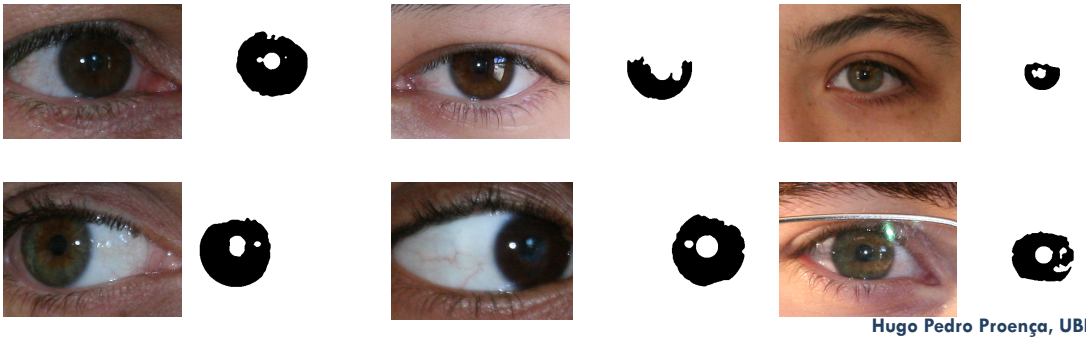


QUIS-CAMPI: Phase 2

- Segmentation of VW Degraded Data
  - State-of-the-art segmentation approaches follow the "**Texture-Classification-First**" strategy.
    - In opposition to "**Boundaries-First**" strategies:

$$max(r, x_0, y_0)\left| G_\sigma(r) * \frac{\partial}{\partial_r} \oint_{r, x_0, y_0} \frac{I(x,y)}{2\pi r}\, ds \right|$$

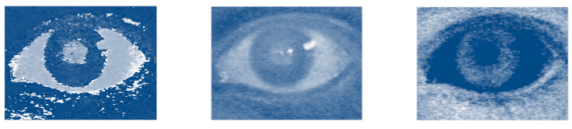
    - H. Proença. "Iris Recognition: On the Segmentation of Degraded Images Acquired in the Visible Wavelength", IEEE-TPAMI, 32:8, 1502-1516, 2010.
    - Z. He et al. "Towards Accurate and Fast Iris Segmentation for Iris Biometrics", IEEE-TPAMI, 31:9, 1670-1684, 2009.



QUIS-CAMPI: Phase 2

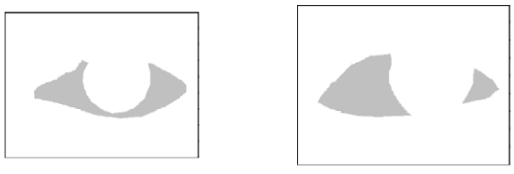
Texture-First Strategies: Example

- H. Proença. "Iris Recognition: On the Segmentation of Degraded Images Acquired in the Visible Wavelength", IEEE-TPAMI, 32:8, 1502-1516, 2010.
- The sclera is particularly easy to be detected in color data.



(a) Hue component. (b) Blue chroma (blue - luminance) component. (c) Red chroma (red - luminance) component.

- Based on the sclera map, an estimate of the iris pixels is possible to obtain (neural networ

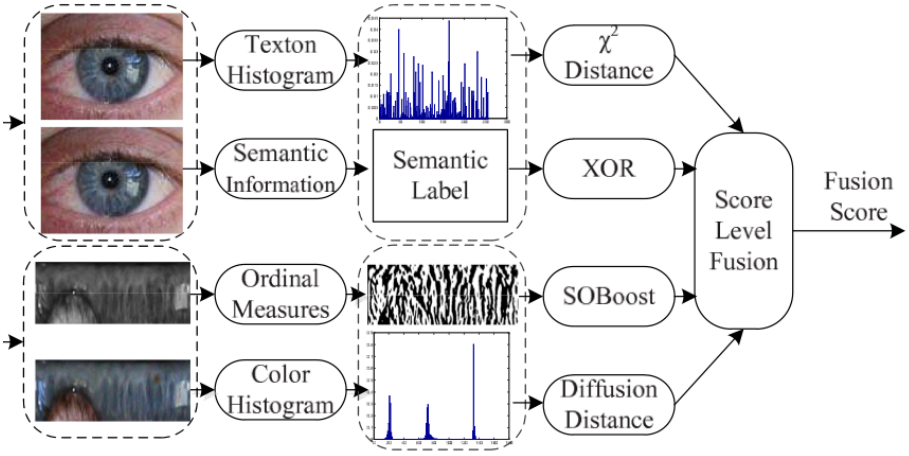


QUIS-CAMPI: Phase 2

- Examples of Segmented VW Degraded Images
  - Given a periocular region, the segmentation of the noise-free iris data is a relatively solved problem

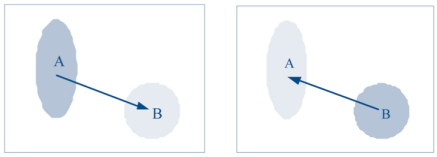
QUIS-CAMPI: Phase 3

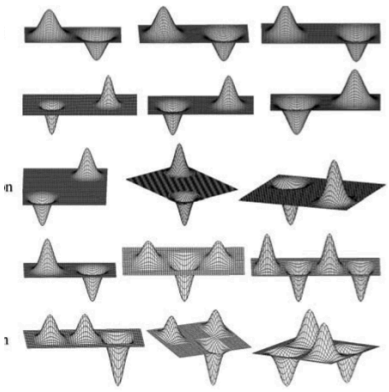
- Feat. Extraction/Matching on VW Degraded Data
  - Tan *et al. "Noisy Iris Image Matching by Using Multiple Cues", Pattern Recognition Letters, 33, 970-977, 2012.*



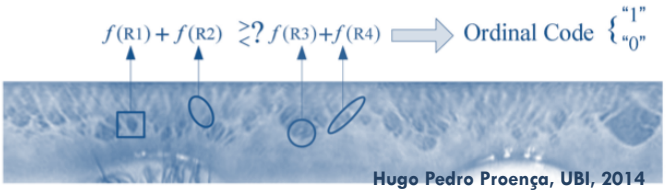
QUIS-CAMPI: Phase 3

- **Ordinal Measures are important for the extraction of robust discriminating features from degraded data.**
  - ***Z. Sun and T. Tan; "Ordinal Iris Measures for Iris Recognition", IEEE-TPAMI, 31 (12), 2211-2226, 2009.***



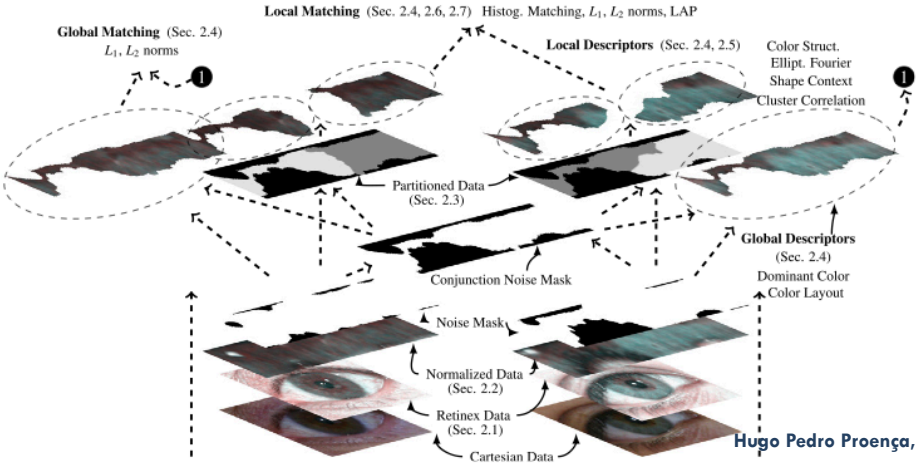


$$MLDF = C_{\mathrm{p}} \sum_{i=1}^{N_{\mathrm{p}}} \frac{1}{\sqrt{2\pi\delta_{\mathrm{p}i}}} \exp\left[\frac{-(X-\mu_{\mathrm{p}i})^2}{2\delta_{\mathrm{p}i}^2}\right] - C_{\mathrm{n}} \sum_{j=1}^{N_{\mathrm{n}}} \frac{1}{\sqrt{2\pi\delta_{\mathrm{n}j}}} \exp\left[\frac{-(X-\mu_{\mathrm{n}j})^2}{2\delta_{\mathrm{n}j}^2}\right],$$



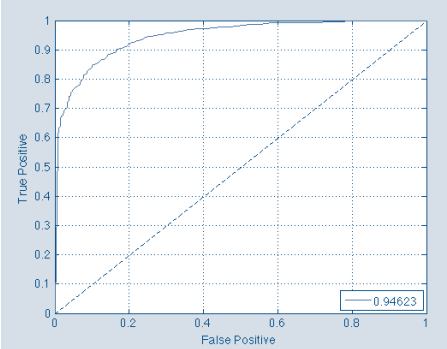
QUIS-CAMPI: Phase 3

- Feat. Extraction/Matching on VW Degraded Data
  - H. Proença, G. Santos; "Fusing Color and Shape Descriptors in the Recognition of Degraded Iris Images Acquired at Visible Wavelength", Computer Vision and Image Understanding, 116, 167-178, 2012.
    - MPEG.7 Color Descriptors: Dominant Color, Color Layout, Color Structure
    - Shape Descriptors: Shape-Context, Elliptic Fourier



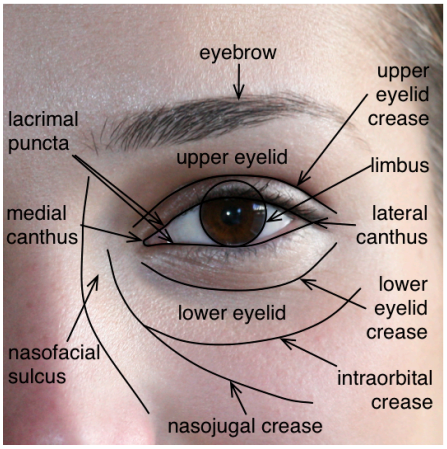
QUIS-CAMPI: Phase 3

- Iris Recognition: State-of-the-Art Performance
  - Decidability index: $d' = \frac{|\mu_E - \mu_I|}{\sqrt{\frac{\sigma_I^2 + \sigma_E^2}{2}}}$
    - Tan et al: d'=2.57 (also considers the eyes' surrounding).
      - d' > 3 is considered a "good" pattern recognition system.
    - Using exclusively the iris (d'=2.12).
  - Receiver Operating Characteristic Curve (AUC=0.94623)



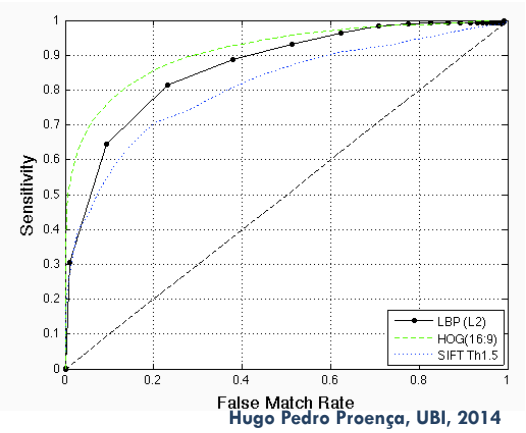
Periocular Recognition

- Devices for collecting iris images acquire not only the region corresponding to the iris, but also all the regions surrounding the eye.
- In a unconstrained setup, there is a lack of discriminating information (particularly for heavily pigmented irises)
- Why not to use information that we already have (eye vicinity) and can augment (in the worst case, only slightly), the amount of discriminating information?



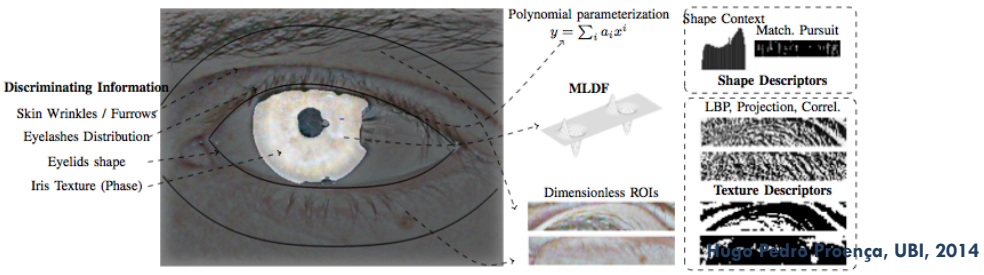
Periocular Recognition

- Park *et* al. Periocular Biometrics in the Visible Spectrum," IEEE-TIFS, 6(1), 96 – 106.
  - Acquisition of the periocular data is expected to require less cooperation and permits a larger stand-off distance than other ocular biometrics.
  - There is a relatively short number of published works
  - Based in:
    - SIFT
    - Local Binary Patterns
    - Histograms Oriented Gradients



Periocular Recognition

- **Main problem:** Currently developed periocular recognition algorithms are holistic, i.e., they use the same feature encoding / matching strategies, without considering the biological components.
  - Results in highly correlated features.
  - Performance too sensitive to the same image covariates.



Periocular Labeling

- **Main Goal:** Develop new periocular recognition algorithms that apply specialized feature encoding / matching algorithms with respect to each biological component in the periocular region.
  - (1) Hair; (2) Eyebrows; (3) Eyelashes; (4) Sclera; (5) Iris; (6) Glasses and (7) Skin

Hugo Pedro Proença, UBI, 2014

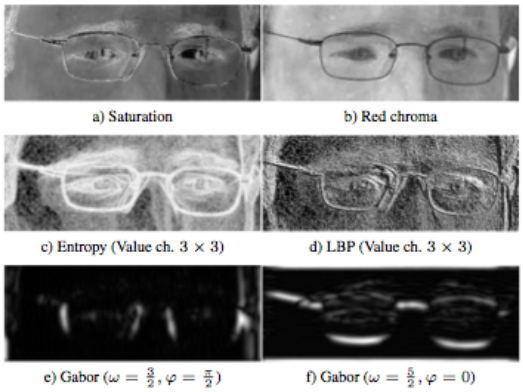
Periocular Labeling

- Example Algorithm

  (Presented at IJCB 14)

$$\gamma(x,y) \in \mathbb{R}^{34}$$

1. Extract a set of (well known) local texture descriptors.



Hugo Pedro Proença, UBI, 2014

Periocular Labeling

- Example Algorithm

  (Presented at IJCB 14)

  2. Learn seven binary classification models, each one specialized in detecting one component. ω_i

  $$\eta_i : \mathbb{R}^{34} \rightarrow [0,1]$$

  Use the response of the classification models to obtain the likelihood of each class.

  $$p\Big(\eta_i\big(\gamma(x,y)\big)|\omega_i\Big)$$

  According to the Bayes rule, the posterior probability functions are given by

  $$P\Big(\omega_i|\eta_i\big(\gamma(x,y)\big)\Big) = \frac{P\Big(\eta_i\big(\gamma(x,y)\big)|\omega_i\Big)}{\sum_{j=1}^{7} P\Big(\eta_j\big(\gamma(x,y)\big)|\omega_j\Big)}$$

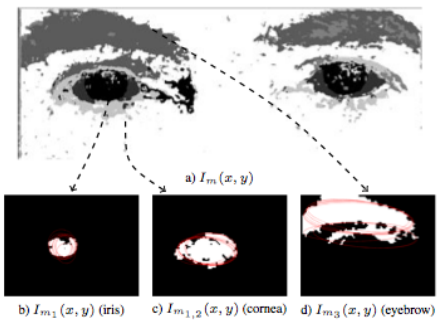
Periocular Labeling

- Example Algorithm

  (Presented at IJCB 14)

  3. Find the indexes of the maximum posterior probabilities at each point in the image:

  $$I_m(x,y) = \arg\max_j p\Big(\omega_j|\eta_j\big(\gamma(x,y)\big)\Big)$$

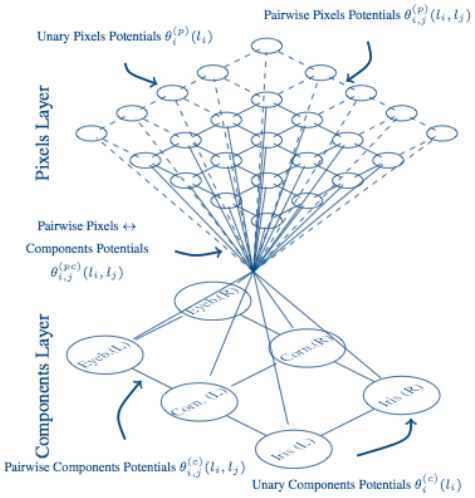


Periocular Labeling

- Example Algorithm

  (Presented at IJCB 14)

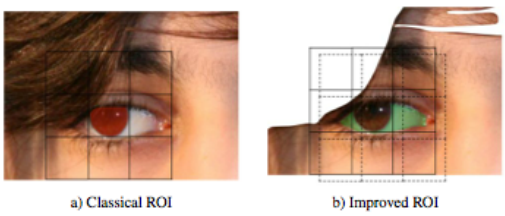
  4. Use a form fitting algorithm to find possibilities for each component (elliptical shape)
  5. Use a Markov Random Field to infer the optimal configuration (minimum cost)



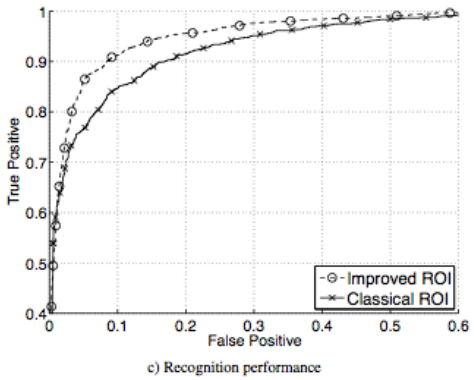
Periocular Labeling

- Example Algorithm (Presented at IJCB 14)

  Substantially increases performance by improving the way the ROI is defined
  - No hair
  - No glasses
  - Center of the cornea as reference point



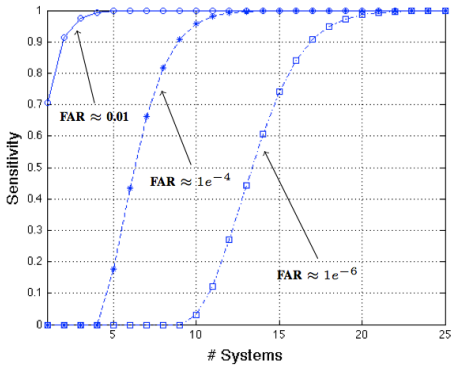
a) Classical ROI b) Improved ROI



c) Recognition performance

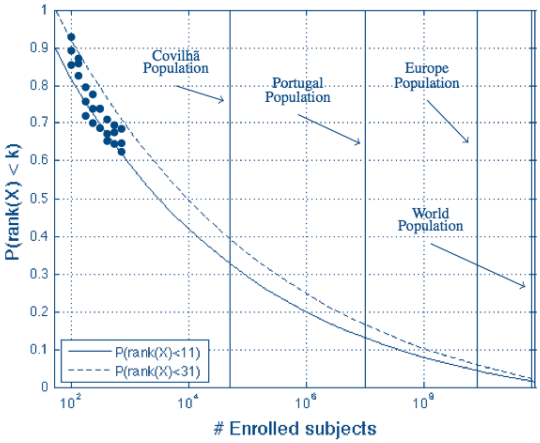
Further Work: Non-Cooperative Recognition

- Suppose that we are able to install several consecutive recognition systems.
  - Working on different conditions, to minimize correlation.
    - Lighting conditions: different types of illuminants
    - Acquisition perspectives and distances
    - Analyzing different regions
    - Using different algorithms
  - Number of independent systems required for a given sensitivity, for different levels of performance.



Further Work: Non-Cooperative Recognition

- Current Performance
  - Probability that a query returns the true identity in the first 10 results, with respect to the number of subjects enrolled in the system.



Further Work: Non-Cooperative Recognition

Frequent Observer

- In security purposes, it is not only of interest to identify whether a person is already in the gallery, but also to see how often it appears in front of the camera. A frequent observer or a traveler in airport may be a subject of interest. With a surveillance camera, a sequence of images of a person at a particular time can be captured and used for identification.

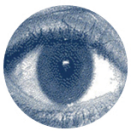
Source: http://www.subservientchicken.com/

Further Work: Non-Cooperative Recognition

Recognition From Moving Aerial Devices

- Also with security purposes, the goal here is to implement autonomous agents that are able to perform the classical phases of biometric recognition, using images acquired from the air, with a moving camera.

A.3 Laboratorial Sheets



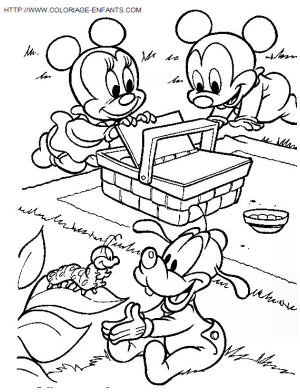
Biometrics

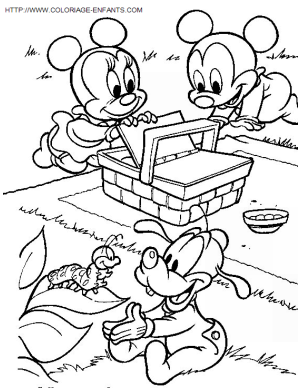
1. Using the MATLAB[R] command line, load the image "begin.jpg" (http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/begin.jpg).

a. Obtain the image dimensions (height x width x depth).
b. Select the first color chanel of that image and plot it, using "imshow"
c. Program one MATLAB[R] function that counts the number of pixels with a certain color, specified as a parameter of the function
d. Convert the image into grayscale and plot it.
e. Invert the grayscale image.
f. Rotate the image in 180º.
g. Rotate the image 45º. What happens to the dimensions of the transformed image?
h. Make an image flip, with espect to the horizontal axis
i. Get the histogram of the grayscale image.

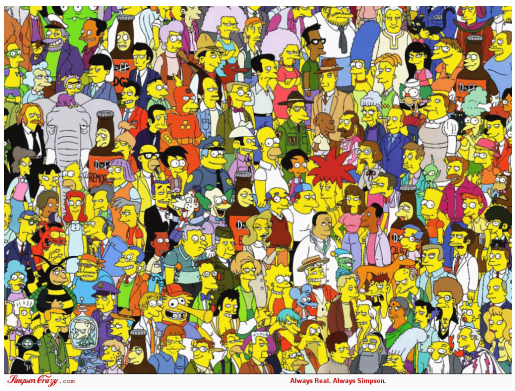
2. Find the differences. The pairs of images below have several differences between them. Implement a function that highlights (with a circle) the regiosn with differences. Note that circles should be as small as possible, in rder to contain the whole region with differences..

a. http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/difs1.tif e
http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/difs1_mudada.tif





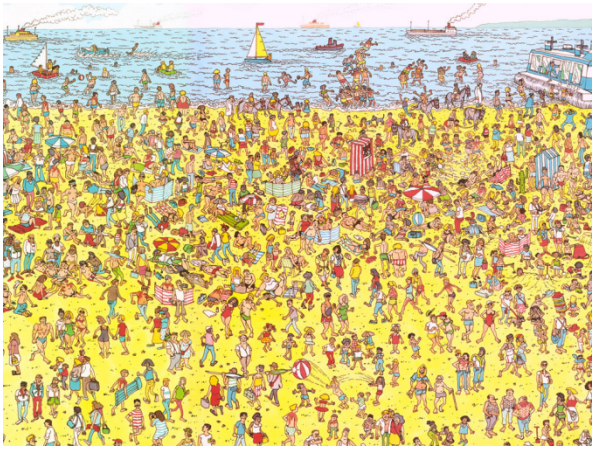
b. http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/difs2.tif e
http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/difs2_mudada.tif



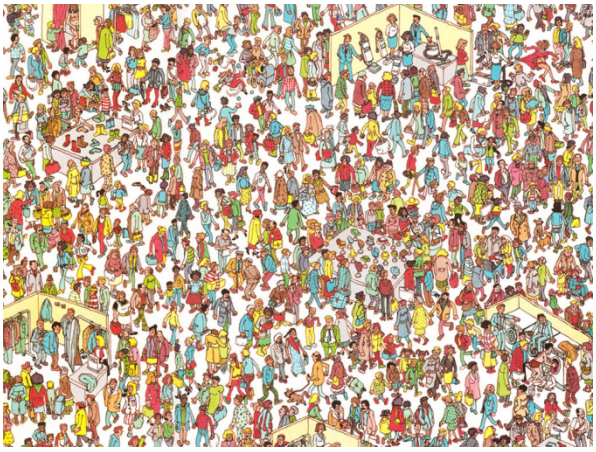
c. http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/difs3.tif e
http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/difs3_mudada.tif

3. Where is Wally? Find the image at the left inside the larger image (and draw a rectangle at the correponsing image patch) .

a. http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/WhereIs.tif e
http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/search_WhereIs.tif

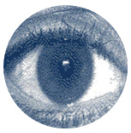


b. http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/WhereIs2.tif e
http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/search_WhereIs2.tif



c. Now, serach for the Waldo character in the following image http://www.di.ubi.pt/~hugomcp/visaoComp/imgs/wally.tif. If you are not able to do it, there is no big problem! Perhaps this could be a Project for the whole semester.

Biometric Systems, 2014/15
Pratical Sheet 2



Biometrics

1. Find a least squares solution of the system Ax=b, being

   A=[4 0; 0 2; 1 1] and b=[2 0 11]

2. Find a least squares solution of the system Ax=b, being

   A=[1 1 0 0; 1 1 0 0; 1 0 1 0; 1 0 1 0; 1 0 0 1; 1 0 0 1], and b=[-3 -1 0 2 5 1];

3. determine the least squares errors obtained in the exercises above.

4. Implement a MATLAB^R function to obtain the least squares solution of inconsistent systems, together with the least squares error corresponding to the solution

5. Chose one of the available implementations of SIFT in MATLAB^R and test it, using pairs of images that share parts of a scene.

6. Get two (frontal) images (A,B) of the same person and one image (C) of a notoriously different subject. Get the SIFT correspondences between A, B; A,C and B,C. What can you conclude?

7. repeat the exercise in 6, using different acceptance thresholds. Draw a plot, containing the ration between the number of correspondences between A,B and A,C. use that plot to infer the optimal acceptance threshold for a recognition system.

8. Download the dataset available at: (http://www.di.ubi.pt/~hugomcp/visaoComp/ficha2.zip).

   a) Implement MATLAB^R functions to apply the image transforms below, both for grayscale and color images. Use the concept of homogeneous coordinates.

      i. Translation

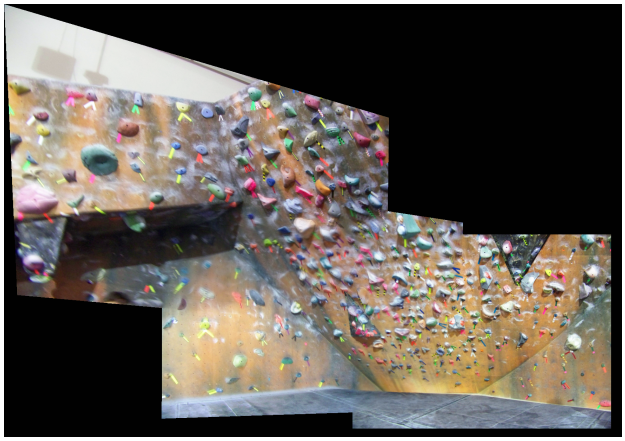
      ii. Rotation

      iii. Scale

      iv. Projective Transform

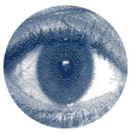
   b) [Image Alignment]. Implement MATLAB^R function that receives two images and aligns them, according to a set of keypoint correspondences.

2. [Mosaic] Implament a MATLAB function that…



i. receive, by mouse input, a set of keypoint correspondences.

ii. Using the concept of projective transform, aligns both images



Biometrics

1. Consider the following matrix

1	1	1	5
1	1	5	5
1	5	5	8
5	5	8	10

Convolve this image with the following filter [0 0 1; 0 0 0; 0 0 0]. What is the actual effect of the filter?

2. Consider the filter with impulse response [1 0 0; 0 1 0; 0 0 1]. Write this impulse response as function of a sum of shifted delta functions.

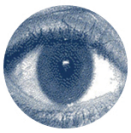
3. Write a MATLAB^{R} function that receives a filter and returns a cell vector with the set of shifted delta functions corresponding.

4. Look at the following image, that can be downloaded from http://www.di.ubi.pt/~hugomcp/biometrics/multimedia/humanBody.jpg. Enhance the image, in order to be easily distinguishable the regions that contains bones and the remaining parts of the image.

5. Use the image above (and others), to compare the output of the Sobel, Canny and Laplacian filters.

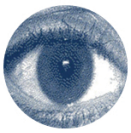
6. Binarize the image above, using different thresholds. Conclude about the optimal threshold that will discriminate bones.

7. Create a 2D blur filter of size 10x10 which sums to 1. Convolve it with the image and display the new image. Create a 1D filter of size 10 which sums to 1. Convolve it with the lines of the image. Transpose the filter and convolve it with rows of the filtered image. Can



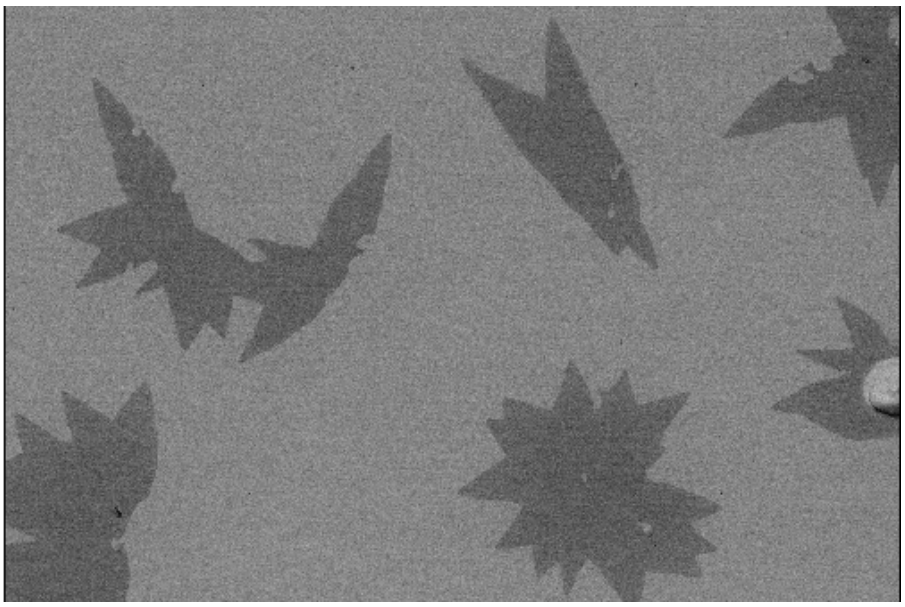
you understand and explain what is happening?

Biometric Systems, 2014/15
Pratical Sheet 3



Biometrics

1. Look at the following image, that can be downloaded from http://www.di.ubi.pt/~hugomcp/biometrics/multimedia/gray.jpg.



a) Obtain the histogram of the image.

b) Stretch the histogram to scale [0, 255]

2. Implement a MATLAB^R function that receives an image and a pair of values [a,b], returning the corresponding image with histogram stretched to [a,b] interval.

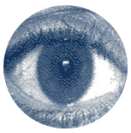
3. Equalize the histogram of the image. What can you conclude about the practical effect of histogram stretching / equalization?

4. Local histogram equalization. Write a MATLAB^R function that receives one grayscale image and a positive integer "r". Then, it returns the corresponding image with the histograms in patches r x r equalized.

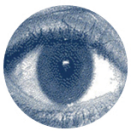
5. Get a homomorphic filter for MATLAB. There are many available in the web. Use the image above to conclude about the purposes / effectiveness of homomorphic filters.

6. Repeat the exercise in 5. using an image with locally heterogenous levels of contrast.

Biometric Systems, 2014/15
Pratical Sheet 3



Biometric Systems, 2014/15
Pratical Sheet 5



Biometrics

1. Learn how to properly use the following MATLAB^R functions
    a. fft
    b. fft2
    c. ifft
    d. ifft2
    e. fftshift f. ifftshift

2. Using MATLAB^R, verify in practice terms the theorem fundamental of convolution, i.e., f[n] * g[n] = F[n] x G[n].

3. Confirm practically that the convolution operation is commutative, i.e., é: f*g = g*f.

4. Confirm in a practical way that: f*(ag + bh) = a(f*g)+ b(f*h), being a and b scalar.

5. Use a grayscale image and visualize the real and imaginary parts of its Fourier transform.

6. Visualize the magnitude of the Fourier transform resulting image.

7. Visualize the phase of the Fourier transform resulting image.

8. Create a Gaussian filter. Convolve the filter with a grayscale image. Draw the possible conclusions about the Gaussian kernel you used.

9. Visualize the magnitude of a Gaussian filter represented in the frequencies domain. What can you conclude about the shape of filters in the frequency domain with respect to their pass-band?

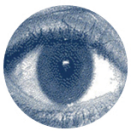
10. What's the result of: d/dx (f(x) * g(x))? Tip: use the convolution theorem.

11. Create an image with a solid black background. The, create a filter with two white points at the image center, regularly spaced What should be the result of the convolution of an image with this kernel, i.e., f= i*g?

    a. How will I, G, F, look like in the frequencies domain?

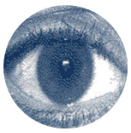
12. Create a pyramid of Gaussian filters with length 2, 5, 10, 20, 50,100 and 200.

    a. Choose a grayscale image with resolution 800 x 600
    b. Convolve the image with the pyramid of Gaussian kernsls. Conclude about the limit above which convolving using the FFT becomes advantageous, with respect to spatial convolution.



13. Let f(x,y) be one image with dimensions M x N. Let F(u,v) be its Fourier transfom. Let g(x,y) be an image with dimensions (2M) x (2N) , with corresponding Fourier denoted by G(u,v). G(u,v) é generated from F(u,v), by inserting one line with 0 among consecutive lines of F, together with a column of 0 between two consecutive columns of F. What's the relation between f and g?

Biometric Systems, 2014/15
Pratical Sheet 3



Biometrics

1. Take the tutorial from Mathworks[R] about object detection. Get it at: http://www.mathworks.com/discovery/object-detection.html

2. Try the prebuilt detectors for frontal faces. Test it using:

a) frontal faces
b) profile faces
c) faces with arbitrary pose
d) faces partially occluded
e) blurred images
f) poor-resolution images

What can you conclude about the robustness of the detector? What are the most concerning covariates?

3. Use the learning option of the MATLAB[R] Computer Vision toolbox to create a detector for an arbitrary type of object.

a) Select a set of positive instances with "N" elements

b) Select a set of negative instances with 2N elements

c) Train a detector with 5 levels, FMR=0.01, TMR=0.995

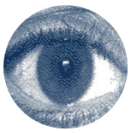
d) Test the detector built in c) and conclude about its effectiveness

e) Augment the number of faces detected, by increasing the TMR parameter

f) reduce the number of non-faces detected, by decreasing the FMR parameter

g) Improve the detector effectiveness, by increasing the number of levels.

Biometric Systems, 2014/15
Pratical Sheet 3



Biometrics

1. Write a MATLAB^R function that receives a grayscale image and returns a version of the image with the "N" pixels with highest magnitude in terms of local gradient having value "1". All the remaining pixels should 0.

2. Write a MATLAB^R function that receives one binary image (with edges denoted by 1) and uses the linear Hough transform to find the best line in the image. The function should return the (b,m) parameterization of that line.

3. Change the function written in 2., in order to return a cell vector with "K" parameterizations corresponding to the best "K" lines in the image, sorted decreasingly according to the value in the accumulator space.

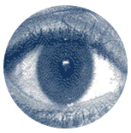
4. Repeat the exercise in 3. But for the circular Hough transform, returning the (x,y,r) parameterizations.

5. Random Elliptic Hough Transform. Write a MATLAB^R function that implements the method for detecting ellipses described at: http://www.bmva.org/bmvc/1988/avc-88-041.pdf

a) Evaluate the typical changes in the results with respect to slight changes in each of the method parameters.

6. Use one of the simplest implementation about snakes, available at: http://www.cb.uu.se/~cris/blog/index.php/archives/217. Run it and test its fitting effectiveness.

Biometric Systems, 2014/15
Pratical Sheet 8



Biometrics

1. Create a random binary vector in MATLAB^{R} , using the “randi” function. The vector should have 2^{n} components.

2. Get the FFT of the above created vector. Visualize the real, imaginary, magnitud and phase components.

3. Using the function “repmat”, get a longer (repeated) version of the vector created above. Apply the FFT to the resulting vector and compare with the results obtained for the original vector. What can you conclude about that?

4. Use the function “ifft” to the real/imaginary components resulting from “fft”. Check that the original data is obtained.

5. Repeat the exercises in 1-4, for 2D data, using the fft2 and ifft2 functions. Tip: You might need the fftshift function.

6. Filtering in the frequencies domain. Get the FFT of an image. Apply the “fftshift” function to the result. Set to 0 all elements that are farther than “K” from the image center. Convert the resulting data into the space domain, by using ifft2. Check the difference between the original image and the reconstructed version.

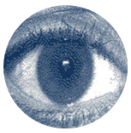
7. Write a function that creates a pair of Gabor filters, according to what’s described in http://en.wikipedia.org/wiki/Gabor_filter

8. Write a function that receives one image and a bank of Gabor filters and returns a cell matrix with the responses of the original image to each of the Gabor filters given as parameter.

9. Create a Multi-Lobe differential filter (MLDF), using exclusively differences between Gaussian blurred images.

10. Write a function that receives one grayscale image and a bank of MLDFs, returning a cell vector with the responses to each MLDF.

Biometric Systems, 2014/15
Pratical Sheet 9



Biometrics

1. Write a function that creates a set of 2D random instances, where the first and second components are correlated at a predefined level (0 means independence).

2. Generate a 2D data set with 1000 elements, using the function written in 1.

3. Apply the PCA MATLAB^{R} built-in function to the generated dataset.

4. Plot the 2D dataset. By superposition, draw a line that represents the direction with highest variability.

5. Use the well know “Iris” data set and apply PCA to it.

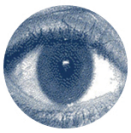
6. Analyze how many features you will need to keep 95% of the data variability

7. Transform the “Iris” data set into the orthogonal PCA subspace, keeping 95% of the variability.

8. Transform the “Iris” data set, using only the top-2 bets features

9. Plot the resulting points from 8. What is the correlation between the components in the horizontal and vertical axes?

Biometric Systems, 2014/15
Pratical Sheet 10



Biometrics

1. Use the "Iris" data set. Load it into MATLAB^{R} and start the Neural network Pattern Classification tool. Use one of the features as the class to be predicted and the remaining features as the predictors. Learn a neural network and get its effectiveness on the test test.

2. Repeat the exercise in 1, each time using a different column as the class. Using the performance scores of the NN in the test set, infer the column that is the most correlated to the remaining ones.

3. MATLAB offers a set of data sets for illustration / testing purposes. These are available at: http://www.mathworks.com/help/nnet/gs/neural-network-toolbox-sample-data-sets.html Use them, to learn neural networks for:

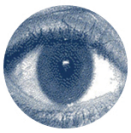
a) classification tasks

b) regression tasks

c) Pay particular attention to the parameterization used, specifically the number of epochs, the training algorithm, the number of neurons in the hidden layer and the type of transfer functions used.

4. Compare the effectiveness you are able to obtain with Neural networks, against some of the stat-of-the-art algorithms included in the statistics toolbox (http://www.mathworks.com/help/stats/support-vector-machines.html) : Support vector machines, Bagging, boosting and random forest classifiers.

Biometric Systems, 2014/15
Pratical Sheet 13



Biometrics

1. Write a MATLAB^R function that generates a random 2D matrix with a set of scores that regard the similarity between genuine and impostor comparisons:

$$[x_1, 0|1; x_2\ 0|1; \ldots x_n\ 0|1]$$

In this matrix, the first column is the similarity score and the second is flag denoting whether the corresponding line regards an impostor (0) or genuine comparison (1).

2. Write a function that receives one matrix as the one generated in 1) and plots the corresponding Receiver-Operating Characteristic curve.

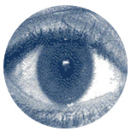
3. Write a function similar to the written in 2, but to obtain the Detection-Error Trade-off curves

4. Add the AUC and EER statistics o the return values of the functions written in 2-3.

5. Write a function that returns the decidability of a recognition system, receiving as input a matrix as the one generated in 1.

6. Bootstrapping-like evaluation. Write a function that repeatedly draw "N" sub-samples of a set of scores (specified by a matrix similar to the above). Then, it obtains the performance evaluation measures as in 2-5 for each sub-set and returns the means and standard deviation values obtained for each performance indicator.

A.3.1 Additional Exercises



Supervised Classification of Iris Data

Consider the set of normalized iris data, available at: http://www.di.ubi.pt/~hugomcp/doc/VISUM. Download the "Part1.zip" file. Every image is represented in "tiff" format, and the corresponding binary mask that discriminates between the noise-free and occluded regions of the iris is stored in "bmp" format. Examples:

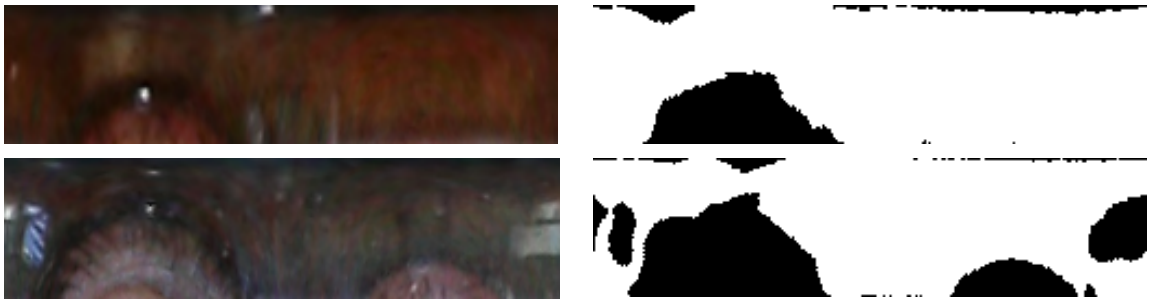


Fig. 1: Examples of two normalized iris images, together with the corresponding binary masks that discriminate between the noise-free and the occluded regions of the iris.

Unzip the "supervised.zip" file. This file contains a folder with visible wavelength data from the UBIRIS.v2 set.

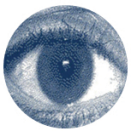
That folder is subdivided into "Gallery" and "Probe". Inside the Gallery folder, there is one image per subject. Every image has its ID associated with the filename.

Use the following function to get information from the ID and image corresponding to an image name:

- [ID, frame]=getInfImg_UBIRISv2(filename)"

Inside the "Probe" folder, there is a set of images with unknown ID. Your task is to establish a unique correspondence between every image in the probe folder and one ID in the Gallery. It is assured that all probes have the corresponding identity in the gallery data.

Biometric Systems, 2014/15
Hands-on II



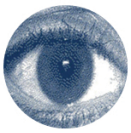
Unsupervised Classification of Periocular Data

As discussed in the theoretical module, the concept of periocular recognition has been gaining relevance in the literature, particularly to work in unconstrained data acquisition scenarios, where data is likely to be degraded. The basic idea is to use not only the iris to perform recognition, but also all the information in the vicinity of the eye, in order to improve the amount of discriminating information. Consider the data set from "Part.2.zip", available at: http://www.di.ubi.pt/~hugomcp/doc/VISUM. It contains a set of periocular data ("tiff" format"), with the corresponding iris masks (same name, but "bmp" format). Examples are given below:

Fig. 1: Examples of periocular images, together with the corresponding binary masks that localize the iris. Images from the upper two rows are from the same subjects, whereas the image given at the bottom row is from a different individual.

The problem is that we don't know the identities of the subjects in the data set, not even how many subjects it contains. Hence, in this second "Hands On" part, the goal is to divide the data set into "N" groups, where each group contains the images that regard the same individual.

A.3.2 Hands-on Assignment



Biometrics Hands-On

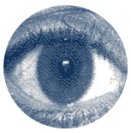
The following text is an excerpt of an interview that the leader of "***Little Jumps***" club given to our company, with regard to a biometric system that we are trying to sell them. The "***Little Jumps***" club was founded about 30 years ago, and is located at the Northwest part of the city. It mostly organizes amateur horse races for children, but sometimes it opens their activities to seniors, during the Summer festival that they organize in the earlier days of August.

"In the earliest days of 1983, around middle January, myself, Tom, Julie and Peter were taking our – then- young kids to some small horse rides in the countryside, which was great for both kids and parents. These were really good times, where we had the opportunity to relax and contact with nature, and the amazing landscape that we have nearby. As time was passing by, another parents joined our team and in less than one year there were over 40 persons that meet every weekend for horse riding. Sometimes we even organized amateur races, in order to keep everyone motivated and in good shape.

Then, by the Christmas of 1983 we felt that maybe a formal club will be a good opportunity for us, mostly because one of the farmers in the surroundings gave us an old barn, that we decided to start using as our headquarter. I was the elected the first president of the club and things went so well that I never left the position. Sometimes guys even call me the dictator, but it should be noted that every two years we conduct the democratic election, where both kids and parents are allowed to vote.
Every week, we start at 9:00 and organize horse races for our club members, having even started to give some small prizes, donated by companies and official entities nearby. We keep track of the results achieved by both jockeys and the horses, and by the end of the season we organize the "Annual Pony Awards", where the outperforming elements of the club are awarded and recognized by every body else.

The number of partners of our club has been increasingly continuously and we are now over 1,000 persons and 800 horses. In result of our activities the "National Association of Horse Races" contacted us to establish a partnership and to set our club as the regional entity that selects the candidates to inclusion in the National league, which is know to provide huge financial support to the best jockeys / horses, opening the opportunity to professionalism.

However, since then, there has been some controversy in the club and our oldest partners complaint that the original spirit was lost and that the club is not offering the original services anymore, at least in the same romantic way. Now, during the most attractive races, there are jockeys and horses that no one know (but present themselves as members) that join the races. Usually they take the best prizes and disappear up to the next event.



There are even certain members that complaint that some other partners are delegating their participation to professional jockeys / horses, at some key races, hoping to improve their ranks, in order to be included in the national league team.

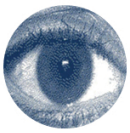
The situation has been turning even more difficult recently, and that's the main reason behind our contact with your company. We want a system that is able to keep track of all the sensitive information about members / jockeys / horses / races in our club, but most of all we want a system that prevents that humans / horses join our races using the identification from other elements, which we hope that will bring peace to our club again."

The task of your team is to design one integrated computational system with a biometric recognition module, in order to handle the needs and security concerns of the club members. The resulting report should give all the details about:

- The way your team will organize to implement the computational system "***Little Jumps***"
- The cohesive perspective of the solution
- The selected biometric trait(s)
- The data acquisition constraints
- The devised acquisition setup and protocol
- The analysis of requirements
- The stakeholder identification
- The complete requirements-list
- The measurable goals
- The development framework
- The validation procedure
- The installation and maintenance plan
- Third party systems: interface plans
- The privacy and legal regulations report
- Any thing else that you may find appropriate

A.4 Exams

Biometric Systems, 2014/15
Exam



Remarks: 1. Read all the questions carefully before providing your answers. **2.** Take into account that this exam should be completed in 90 minutes at most. 3. Any fraud attempt will determine the immediate suspension of your exam.

1. **[Ethics]** Consider the following case, corresponding to a request of a chief-executive-officer of a company, concerned about the effectiveness of the company employees:

   "*Most of our employees are completely devoted to the company. But, as in any other big company (we have over 10,000 employees), there are some exceptions, that I suspect mainly regard employees from the computer programming department. My idea is to setu up a set of automatic surveillance devices in their room, in order to continuously scan their effectiveness and committment to the work. I think that there is anothing wrong with this innitiative, and any person that does not agree with it, will be fired, as it is obvious that is one of the lazy employee*"

   a) **[10 points]** Comment the ethical / privacy issues behind the development of a system that responds to the CEO requests.

   b) **[10 points]** Describe in a brief way, the type of solution you will propose to the CEO.

2. **[General, 20 points]** Contextualize the following statement, in the scope of interfaces to biometrics recognition systems:

   "Any system is as strong as its weakest module / component"

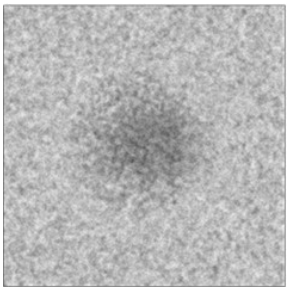
3. **[Traits]** Give examples, and justify them appropriately…

   a. **[10 points]** Of one practical case where the use of the iris as main trait of a biometric would be preferable than the fingerprint.

   b. **[10 points]** The opposite case, i.e., one practical scenario where the iris as main biometric would be worse than fingerprint

4. **[Signals, 20 points]** Consider the following function **f(x,y)=x+y+xy**. Is it possible to regard this function as a linear system? Justify your answer in a formal way.

5. **[Segmentation, 20 points]** Consider the following image. Justify why such image will be hard to segment according to the traditional snakes model. Alternativelly, propose a segmentation strategy that would probably be effective in thist ype of data. Justify your answer.



6. **[Feature Extraction, 30 points]** Implement a MATLAB function to obtain the Hessian matrix of a point in an image. If its not possible to obtain the Hessian, the function should return an empty matrix.

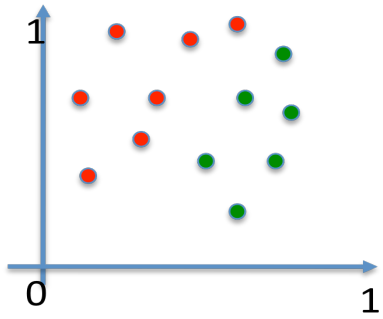
Prototype: hessian getHessian(img, row, col)

7. **[Feature Extraction, 20 points]** Determine, from the following vectors, which are eigenvectors of the matrix given below and, if positive, obtain the corresponding eigenvalues.

1	2	1
6	-1	0
-1	-2	-1

a) [1, 6, -13];
b) [-1, 0, 13];
c) [1, -6, 3];

8. **[Classification]** Observe the following 2D data set. Obtain two equations for straight lines that correspond to:



a) **[10 points]** A *good* projection into 1D space, for binary classification purposes

b) **[10 points]** A *bad* projection into 1D space, for binary classification purposes.

9. **[Feature Selection / Transform 20 points]** Suppose that the following matrix is the output from the MATLAB "princomp" function, meaning that it transforms data in an original space to the condensed PCA version. Based on the following set of original data (1 instance per row), write the PCA consensed version of this data.

PCA MATRIX

1	0	-1
0	0	1
1	1	1
1	-1	0

Original Data: [10 20 -10 0; 10 -20 -10 10; 0 0 -10 10; 20 10 10 -10];

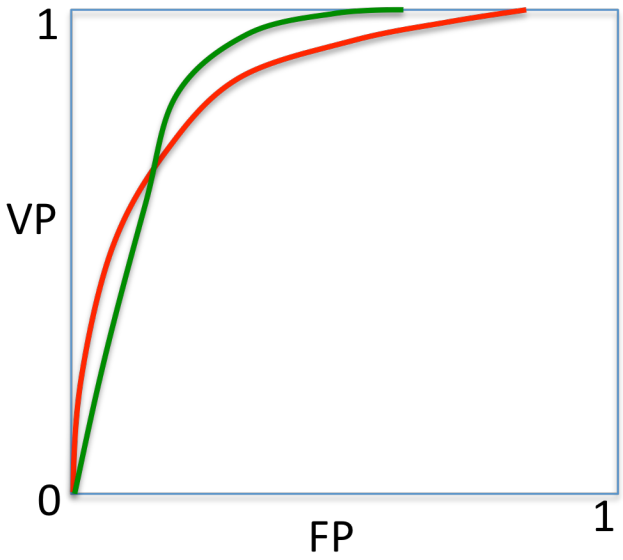
10. **[Performance Evaluation 20 points]** Consider the following Receiver Operating Characteristic plot, that compares the relative effectiveness of an two biometric traits, for a particular scenario. Trait "A" is represented by the red curve, and trait "B" by the green curve. Comment the following statements:

a) The trait "A" is better than trait "B"

b) For high security scenarios, trait "A" should be prefered over "B"

c) Trait "A" will be much more convenient for users than trait "B"

d) Both systems have AUC higher than 0.5



Bibliography

[1] T. Angelo and K. Cross. Classroom assessment techniques: A handbook for college teachers. Jossey-Bass, ISBN: 978-1555425005, San Francisco, CA, 1993.

[2] Anna University. Master in Engineering on Biometrics and Cyber Security. http://www.annauniv.edu/academic_courses/WSA/04.\%20I\%20&\%20C/17.\%20Bio\%20M\%20\&CS.pdf, assessed on September 9, 2014.

[3] Biometric recognition Group AVTS. Universidad Autonoma de Madrid. http://http://atvs.ii.uam.es/, assessed on September 9, 2014.

[4] Biometric Research Lab. Sabanci University. http://biometrics.sabanciuniv.edu/people.html, assessed on September 9, 2014.

[5] Biometric Research Lab.. Indian Institute of Technology Kanpur. http://www.cse.iitk.ac.in/users/biometrics/, assessed on September 9, 2014.

[6] Biometrics Institute. Biometrics Institute Intensive Course. http://www.biometricsinstitute.org/events.php/523/biometrics-institute-intensive-course-2014, assessed on September 9, 2014.

[7] Biometrics Research Group. Michigan State University. http://biometrics.cse.msu.edu/, assessed on September 9, 2014.

[8] Biosecure. Paris Sud University. http://biosecure.it-sudparis.eu/AB/index.php?option=com_content&view=article&id=28, assessed on September 9, 2014.

[9] Brooklyn Polytechnic University. Biometrics. http://archive.poly.edu/epoly/media.doc, assessed on September 9, 2014.

[10] Canberra Institute of Technology. Biometric technologies - certificate IV. http://cit.edu.au/study/courses/biometric_technologies_certificate_iv, assessed on September 9, 2014.

[11] Center for Biometric Research. Chinese Academy of Sciences (China). http://lab.datatang.com/1984DA173065/EnIndex.aspx, assessed on September 9, 2014.

[12] Center for Biometric and Security Research. Chinese Academy of Sciences. http://www.cbsr.ia.ac.cn/english/index.asp, assessed on September 9, 2014.

[13] T. Chan and L. Vese. Active contours without edges. IEEE Transactions on Image Processing, vol. 10, issue 2, pag. 266-277, 2001.

[14] Columbia University. Biometrics. http://www.cs.columbia.edu/~belhumeur/courses/biometrics/2013/biometrics.html, assessed on September 9, 2014.

[15] CyLab. Biometrics Center. Carnegie-Mellon University. https://www.cylab.cmu.edu/research/center-biometrics.html, assessed on September 9, 2014.

[16] da/sec - Biometrics and Internet-Security research group,. Hochschule Darmstadt University of Applied Sciences. https://www.dasec.h-da.de/, assessed on September 9, 2014.

[17] DISI - Department of Computer Science and Engineering. University of Bologna. http://biolab.csr.unibo.it/home.asp, assessed on September 9, 2014.

[18] Faculdade de Engenharia da Universidade de Porto. Processamento e Análise de Imagens em Engenharia. http://sigarra.up.pt/feup/pt/ucurr_geral.ficha_uc_view?pv_ocorrencia_id=335119, assessed on September 9, 2014.

[19] I^{2}SIS, Institute for the Interdisciplinary Studies in Identity Science. University of North-Carolina. http://www.faceaginggroup.com/, assessed on September 9, 2014.

[20] IDIAP Research Institute. Ecole Polytechnique Fédérale de Lausanne. http://www.idiap.ch/scientific-research/research-groups/biometric-person-recognition, assessed on September 9, 2014.

[21] Indian Institute of Technology Delhi. Biometrics. http://www4.comp.polyu.edu.hk/~csajaykr/myhome/teaching/biometrics/eel851.htm, assessed on September 9, 2014.

[22] Indiana-Purdue University. Biometrics. http://cs.iupui.edu/graduate/degrees/certificate/biometrics, assessed on September 9, 2014.

[23] Instituto Superior Técnico da Universidade de Lisboa. Sistemas de Biometria. https://fenix.tecnico.ulisboa.pt/homepage/ist12170/disciplinas-de-doutoramento/sistemas-de-biometria, assessed on September 9, 2014.

[24] Instituto Superior Técnico da Universidade de Lisboa. Modelação e Classificação de Sinais Biomédicos. https://fenix.tecnico.ulisboa.pt/homepage/ist12170/disciplinas-de-doutoramento/modelacao-e-classificacao-de-sinais-biomedicos, assessed on September 9, 2014.

[25] La Trobe University. Biocomputing and Biometric Security. http://www.latrobe.edu.au/handbook/2013/postgraduate/science-tech/coursework/smbbs.htm, assessed on September 9, 2014.

[26] D. Lowe. Object recognition from local scale-invariant features. Proceedings of the International Conference on Computer Vision, vol. 2, pay. 1150–1157, 1999.

[27] Maryland Cybersecurity Center. University of Maryland. http://www.cyber.umd.edu/faculty, assessed on September 9, 2014.

[28] Michigan State University. Biometrics. http://www.cse.msu.edu/Students/Current_Grad/GradHandbook.php, assessed on September 9, 2014.

[29] National Biometric Test Center. San Jose State University. http://www.engr.sjsu.edu/biometrics/, assessed on September 9, 2014.

[30] National Research Council. Biometric Recognition: Challenges and Opportunities. Washington, DC: The National Academies Press, 2010.

[31] New York Polytechnic School of Engineering. Biometrics. http://engineering.nyu.edu/academics/course/CS9093, assessed on September 9, 2014.

[32] Northwestern University. Biometrics. http://eecs.northwestern.edu/course-list/757-eecs-395-495-biometrics, assessed on September 9, 2014.

[33] Notre Dame ReSource: Biometrics Research. Notre-Dame University. http://news.nd.edu/news/8394-notre-dame-resource-biometrics-research/, assessed on September 9, 2014.

[34] Krzysztof Rutkowski. Biometrics - Tip of the Iceberg. Frost & Sullivan Market Insight, May, 2012.

[35] SOCIA: Soft Computing and Image Analysis Lab.. University of Beira Interior. http://socia-lab.di.ubi.pt, assessed on September 9, 2014.

[36] Technical University of Denmark. Biometric Systems. http://www.christoph-busch.de/teaching-biometric-systems.html, assessed on September 9, 2014.

[37] The University of Auckland. Forensics and Biometrics. http://www.ece.auckland.ac.nz/en/about/our-research/research-areas/forensicsandbiometricsresearch.html, assessed on September 9, 2014.

[38] The University of Winnipeg. Biometrics. http://www.acs.uwinnipeg.ca/7201/, assessed on September 9, 2014.

[39] Universidade Aberta. Master in Biostatistics and Biometry. https://www2.uab.pt/guiainformativo/planoestudos2.php?curso=64&m=45, assessed on September 9, 2014.

[40] University at Buffalo. Biometrics and Image Analysis. http://www.cedar.buffalo.edu/~govind/CSE666/, assessed on September 9, 2014.

[41] Universidade da Beira Interior. Mission, Vision and Principles. https://www.ubi.pt/Pagina.aspx?p=missao_e_visao, assessed on September 9, 2014.

[42] Universidade de Coimbra. Video-Vigilance and Biometrics. http://apps.uc.pt/courses/EN/unit/9270/15327/2014-2015?type=ram&id=323, assessed on September 9, 2014.

[43] Universitá of Siena. Biometric Systems. http://homes.di.unimi.it/scotti/ita/md_sistemibiometricidot/index.htm, assessed on September 9, 2014.

[44] Universiteit Twente. Introduction to Biometrics. http://www.sas.ewi.utwente.nl/open/courses/intro_biometrics/, assessed on September 9, 2014.

[45] University of Applied Sciences, Salzburg. Biometric Systems. http://www.aisp-salzburg.ac.at/en/joint-degree-master/curriculum-structure/, assessed on September 9, 2014.

[46] University of Calgary. Biometric Technologies. http://pages.cpsc.ucalgary.ca/~marina/601/, assessed on September 9, 2014.

[47] University of California, San Diego. Biometrics. http://cseweb.ucsd.edu/classes/wi12/cse190-c/, assessed on September 9, 2014.

[48] University of Heltfordshire. Biometrics and Cybersecurity http://www.postgraduatesearch.com/university-of-hertfordshire/52982270/postgraduate-course.htm, assessed on September 9, 2014.

[49] University of Kent. Information Security and Biometrics. http://www.kent.ac.uk/courses/postgraduate/256/information-security-biometrics, assessed on September 9, 2014.

[50] University of Louisville. Introduction to Biometrics. http://www.cvip.uofl.edu/wwwcvip/education/ECE523/index2.html, assessed on September 9, 2014.

[51] University of Southhampton. Biometrics. http://www.ecs.soton.ac.uk/module/ELEC6120#overview, assessed on September 9, 2014.

[52] University of Surrey. Biometrics. http://www.surrey.ac.uk/cvssp/research/biometrics/, assessed on September 9, 2014.

[53] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pag. 511-518, 2001.

[54] West Virginia University. Introduction to Biometrics Systems. http://www.csee.wvu.edu/~natalias/biom426/biom426_fall05.html, assessed on September 9, 2014.

[55] G. Wiggins and J. McTighe. Understanding by Design. McHill Education / ASCD College Textbook Series, Pearson, ISBN: 978-0131950849, 2005.