



**UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH**

**Departament de Teoria del Senyal
i Comunicacions**

FACE RECOGNITION BY MEANS OF ADVANCED CONTRIBUTIONS IN MACHINE LEARNING

PhD Thesis Dissertation

by

VIRGINIA ESPINOSA DURÓ

Submitted to the Universitat Politècnica de Catalunya
in partial fulfillment of the requirements for the PhD degree

Supervised by Dr. Enric Monte Moreno and Dr. Marcos Faúndez-Zanuy

PhD program on Signal Theory and Communications

May 2013



Acta de qualificació de tesi doctoral

Curs acadèmic: 2012-2013

Nom i cognoms

Virginia Espinosa Duró

DNI / NIE / Passaport

33944683H

Programa de doctorat

Teoria del Senyal i Comunicacions

Unitat estructural responsable del programa

Departament de Teoria del Senyal i Comunicacions

Resolució del Tribunal

Reunit el Tribunal designat a l'efecte, el doctorand / la doctoranda exposa el tema de la seva tesi doctoral titulada
FACE RECOGNITION BY MEANS OF ADVANCED CONTRIBUTIONS IN MACHINE LEARNING

Acabada la lectura i després de donar resposta a les qüestions formulades pels membres titulars del tribunal, aquest atorga la qualificació:

☐

APTA/E

☐

NO APTA/E

(Nom, cognoms i signatura)		(Nom, cognoms i signatura)	
President/a		Secretari/ària	
(Nom, cognoms i signatura)	(Nom, cognoms i signatura)	(Nom, cognoms i signatura)	(Nom, cognoms i signatura)
Vocal	Vocal	Vocal	Vocal

_____, _____ d'/de _____ de _____

El resultat de l'escrutini dels vots emesos pels membres titulars del tribunal, efectuat per l'Escola de Doctorat, a instància de la Comissió de Doctorat de la UPC, atorga la MENCIÓ CUM LAUDE:

☐

SI

☐

NO

(Nom, cognoms i signatura)		(Nom, cognoms i signatura)	
Presidenta de la Comissió de Doctorat		Secretària de la Comissió de Doctorat	

Barcelona, _____ d'/de _____ de _____

On ne voit bien qu'avec le cœur. L'essentiel est invisible pour les yeux.
-Només es veu bé amb el cor. L'essencial és invisible a la vista-
Le Petit Prince. Antoine de Saint-Exupéry.

A mi Madre,

Por su lazo umbilical perenne y su indeleble impronta. Perquè la seva grandesa em va convertir per sempre en “la petita de la Isabel”.

A mi Padre,

Por dejarme conectar con la persona que hay detrás de él y transmitirme su amor por la naturaleza, la cultura y el arte.

A los dos, por inculcarnos que la mejor manera de afrontar esta vida, era con una buena dosis de educación y cultura.

A mi abuela Virginia, por legarme su nombre y regalarnos su porte y su savoig faire.

A mi tía Virginia, por ser madre a horas convenidas...muchas.

A las dos, por dejar el listón muy alto.

A mi abuelo Diego, por ser un pequeño gran hombre y quererme irremediabilmente.

A mi bisabuelo Antonio, por creer en un sueño y no doblegarse más que para labrar la tierra, plantar la vid y embotellar la verdad, aunque fuera tras un roble francés. Porque la ciudad de la luz le debe una parte.

A los dos, por mirar atrás y poder respirar profundamente.

Abstract

Face recognition (FR) has been extensively studied, due to both scientific fundamental challenges and current and potential applications where human identification is needed. FR systems have the benefits of their non intrusiveness, low cost of equipments and no user-agreement requirements when doing acquisition, among the most important ones. Nevertheless, despite the progress made in last years and the different solutions proposed, FR performance is not yet satisfactory when more demanding conditions are required (different viewpoints, blocked effects, illumination changes, strong lighting states, etc). Particularly, the effect of such non-controlled lighting conditions on face images leads to one of the strongest distortions in facial appearance.

This dissertation addresses the problem of FR when dealing with less constrained illumination situations. In order to approach the problem, a new multi-session and multi-spectral face database has been acquired in visible, Near-infrared (NIR) and Thermal infrared (TIR) spectra, under different lighting conditions.

A theoretical analysis using information theory to demonstrate the complementarities between different spectral bands have been firstly carried out. The optimal exploitation of the information provided by the set of multispectral images has been subsequently addressed by using multimodal matching score fusion techniques that efficiently synthesize complementary meaningful information among different spectra.

Due to peculiarities in thermal images, a specific face segmentation algorithm has been required and developed. In the final proposed system, the Discrete Cosine Transform as dimensionality reduction tool and a fractional distance for matching were used, so that the cost in processing time and memory was significantly reduced. Prior to this classification task, a selection of the relevant frequency bands is proposed in order to optimize the overall system, based on identifying and maximizing independence relations by means of discriminability criteria.

The system has been extensively evaluated on the multispectral face database specifically performed for our purpose. On this regard, a new visualization procedure has been suggested in order to combine different bands for establishing valid comparisons and giving statistical information about the significance of the results. This experimental framework has more easily enabled the improvement of robustness against training and testing illumination mismatch.

Additionally, focusing problem in thermal spectrum has been also addressed, firstly, for the more general case of the thermal images (or thermograms), and then for the case of facial

thermograms from both theoretical and practical point of view. In order to analyze the quality of such facial thermograms degraded by blurring, an appropriate algorithm has been successfully developed.

Experimental results strongly support the proposed multispectral facial image fusion, achieving very high performance in several conditions. These results represent a new advance in providing a robust matching across changes in illumination, further inspiring highly accurate FR approaches in practical scenarios.

Resum

El reconeixement facial (FR) ha estat àmpliament estudiat, degut tant als reptes fonamentals científics que suposa com a les aplicacions actuals i futures on requereix la identificació de les persones. Els sistemes de reconeixement facial tenen els avantatges de ser no intrusius, presentar un baix cost dels equips d'adquisició i no la necessitat d'autorització per part de l'individu a l'hora de realitzar l'adquisició, entre les més importants. De totes maneres i malgrat els avenços aconseguits en els darrers anys i les diferents solucions proposades, el rendiment del FR encara no resulta satisfactori quan es requereixen condicions més exigents (diferents punts de vista, efectes de bloqueig, canvis en la il·luminació, condicions de llum extremes, etc.). Concretament, l'efecte d'aquestes variacions no controlades en les condicions d'il·luminació sobre les imatges facials condueix a una de les distorsions més accentuades sobre l'aparença facial.

Aquesta tesi aborda el problema del FR en condicions d'il·luminació menys restringides. Per tal d'abordar el problema, hem adquirit una nova base de dades de cara multispectral i multiespectral en l'espectre infraroig visible, infraroig proper (NIR) i tèrmic (TIR), sota diferents condicions d'il·luminació.

En primer lloc s'ha dut a terme una anàlisi teòrica utilitzant la teoria de la informació per demostrar la complementarietat entre les diferents bandes espectrals objecte d'estudi. L'òptim aprofitament de la informació proporcionada pel conjunt d'imatges multiespectrals s'ha abordat posteriorment mitjançant l'ús de tècniques de fusió de puntuació multimodals, capaces de sintetitzar de manera eficient el conjunt d'informació significativa complementària entre els diferents espectres.

A causa de les característiques particulars de les imatges tèrmiques, s'ha requerit del desenvolupament d'un algorisme específic per la segmentació de les mateixes. En el sistema proposat final, s'ha utilitzat com a eina de reducció de la dimensionalitat de les imatges, la Transformada del Cosinus Discreta i una distància fraccional per realitzar les tasques de classificació de manera que el cost en temps de processament i de memòria es va reduir de forma significativa. Prèviament a aquesta tasca de classificació, es proposa una selecció de les bandes de freqüències més rellevants, basat en la identificació i la maximització de les relacions d'independència per mitjà de criteris discriminabilitat, per tal d'optimitzar el conjunt del sistema.

El sistema ha estat àmpliament avaluat sobre la base de dades de cara multiespectral, desenvolupada pel nostre propòsit. En aquest sentit s'ha suggerit l'ús d'un nou procediment de visualització per combinar diferents bandes per poder establir comparacions vàlides i donar informació estadística sobre el significat dels resultats. Aquest marc experimental ha permès més fàcilment la millora de la robustesa quan les condicions d'il·luminació eren diferents entre els processos d'entrenament i test.

De forma complementària, s'ha tractat la problemàtica de l'enfocament de les imatges en l'espectre tèrmic, en primer lloc, pel cas general de les imatges tèrmiques (o termogrames) i posteriorment pel cas concret dels termogrames facials, des dels punt de vista tant teòric com pràctic. En aquest sentit i per tal d'analitzar la qualitat d'aquests termogrames facials degradats per efectes de desenfocament, s'ha desenvolupat un últim algorisme.

Els resultats experimentals recolzen fermament que la fusió d'imatges facials multiespectrals proposada assoleix un rendiment molt alt en diverses condicions d'il·luminació. Aquests resultats representen un nou avenç en l'aportació de solucions robustes quan es contempen canvis en la il·luminació, i esperen poder inspirar a futures implementacions de sistemes de reconeixement facial precisos en escenaris no controlats.

Agraïments

Al gener del 2006, caminant pel barri de Kowloon, a Hong Kong, vaig sentar-me a descansar en un dels pocs espais buits que vaig trobar (el barri en qüestió té el demèrit de ser la zona amb la major densitat de població del planeta) i vaig aprofitar l'avinentsa per encetar una bossa de "fortune cookies". La galeta que vaig agafar, anava acompanyada d'un paperet que portava escrita la següent frase: *"Per tenir bons records, fa falta quelcom més que tenir bona memòria"*.

La veritat és que una vegada arribat aquest moment, miro enrere i una col·lecció de persones em ve a la memòria i, amb les llavis arquejats, corroboro les paraules que un dia em va regalar l'atzar. La vida m'ha regalat a tots vosaltres. Gràcies sinceres.

Al Dr. Enric Monte, home del renaixement, referent holístic indiscutible i co-director d'aquesta tesis, per la seva sempre disponibilitat, el seu temps, els seus savis consells, les seves aportacions i les nostres llargues converses. Pels seus feedbacks sempre positius, i més enllà d'això, per la seva impecable col·locació en el no sempre fàcil paper de co-director.

Al Dr. Marcos Faúndez, co-director d'aquesta tesis, per comptar en mi des del primer moment per formar part del Grup de Recerca de Tractament del Senyal de l'EUPMt-UPC. Per "oblidar-se casualment" a la taula del meu despatx un llibre de l'Umberto Eco titulat *"Como se hace una tesis"*..., engrescar-me en la seva realització i oferir-se reiteradament en la co-direcció de la mateixa. I ja en procés d'elaboració de la tesis, per les múltiples aportacions donades en les etapes de conceptualització i d'execució. Perquè quan les sumes es van convertir en restes, vam saber acabar multiplicant, però sobretot, per donar-me una cop de ma determinant, quan els dies em queien de genolls.

Al Dr. Jordi Solé, pels seus comentaris engrescadors sobre el projecte de tesis i les seves interessants aportacions.

To Dr. Stan Z. Li, from the Center for Biometrics and Security Research (CBSR), from his fine and uplifting workshop on face recognition technologies at the Hong Kong Polytechnic University.

To Dr. Gordon Thomas, former Director of the Police Scientific development branch in the UK Home Office and current Chair of the IEEE-International Carnahan Conference on Security Technology EC, for their assistance, kindness and support in every moment, and for his proposal for appointing me as member of the Executive Committee of the IEEE-ICCST. And last but not least, for his fine English revisions of some parts of the document.

To Dr. Henry Oman, editor in Chief of the IEEE Aerospace and Electronic Systems Magazine, to support and publish three of my research works and also for reporting them in many of his talks and reports.

To Jiří Mekyska, a Czech guy with a extremely difficult name to pronounce but also with an unbelievable capability of making things easy. Thanks for being the perfect colleague and friend and also for taking care of everything in such a perfect way during our stage at Check Republic.

A la Beni, secretària de doctorat, per tenir sempre cura dels aspectes logístics del procés i amenitzar-ho amb una dosis d'empatia.

A en Juan García, per deixar-se atropellar pel passadissos dia sí, dia també, donar-me un cop de ma amb l'anglès i no donar-li importància. I per suposat, per colorejar les reentrées a la feina amb un festival de productes de la terra!

A en Jordi Ayza, company de feina i de despatx, per tenir el comentari adient en el moment oportú i oxigenar-me el camí, però sobretot, per sobreportar els meus moments d'histerisme com sempre: com un senyor.

A en Miquel Roca i en Vicenç Delos (el *Vice-Rector*, que vaig entendre jo...), per fer-me un lloc a l'Escola des del primer dia i enfilat-me a les muntanyes a fer la cabreta quan St. Tomàs d'Aquino disposava.

A en Robert Safont, per colar-se al meu despatx, la majoria de vegades passades les nou, fer-me cinc cèntims del que passava pel món i retornar-me al nivell del mar.

Als companys, amics i doctorands, Xavi, Enric (ja Dr.) i Pablo, perquè quan es fa una travessia pel desert, és millor fer-la en bona companyia. A la comitiva de Síria en ple, perquè va ser divertit, gratuït i inesperat. I perquè no, perquè em vàreu fer sentir reina mora per uns dies ;)

A la Maria, la Montse i la Lina, per fitxar-me a la secció femenina de tallats i alegrar-me el dia.

A en Toni i en Moisès, per saber-ho ser tot: amics, alumnes, projectistes, companys de feina, salvavides de laboratori i sobretot, compis de volei!!!

A en Narcís Rovira, pels seus constants mails nocturns farcits de links amb les respostes a totes les nostres preguntes. Per les seves transfusions d'energia.

A la Sussana Rivero, exprojectista i responsable del servei de material d'audiovisuals, per tenir-ho sempre tot a punt i a temps.

A l'Anna Llacher, per què la pregunta obligada de "*Què, com la portes?*", sempre anava acompanyada d'un vigorós, "*ànims!!!!*".

A la Francina, per estar també des del primer dia.

Als companys del Departament d'Electrònica i Automàtica i també a la resta de companys de l'EUPMt, per formar part del mateix vaixell. En especial, i de nou, a tots aquells que van col·laborar i cedir el seu temps i la seva imatge, per poder elaborar la base de dades multispectral, gènesis d'aquesta tesis. Gràcies a tots "*per donar la cara*";-)

Als meus alumnes passats, presents i futurs, per construir i compartir cada dia un espai comú on poder créixer.

A les dones de la feina, compis d'hores intempestives, i sobretot, a la Isabel, per preguntar sempre per me mare.

I per últim, però no menys important:

Als meus germans, per ser fills únics, tots.

Al meu germà Xevi, per fer que la meva infància, pescant al riu, anant en bicicleta o a sobre d'uns esquís, fos un somni.

A la meva germana Maribel, perquè el món sense ella no seria el mateix.

Al meu germà Toni, perquè quan jo era petita, ell va ser el meu germà gran.

Al meu pare, perquè als primers estius, encara trobava espai al cotxe per embotir "a les dues germanes franceses"...Quina generació...Encara ara em pregunto com s'ho feien...

A la meva mare, per baixar de dos en dos les escales per anar a treballar, anar a comprar entre hores i llaurar el nostre futur. Per totes les tardes d'estiu amb gust a pa amb xocolata.

A la meva cosina Bet, perquè a part de competir, també ens vam estimar. A el meu cosí Miquel, perquè un dia els tres vam ser "*todos para uno y uno para todos*"!

A en Joan Miquel, per aterrar un dia a casa amb aquell llibre de l'any 33 sobre els mètodes d'identificació personal...Compleixo avui la promesa de referenciar-el a la meva tesis. Per tot el que hi ha al darrera d'aquestes dues línies...

Als meus nebots, per ser els fills que no he tingut i omplir-me de l'orgull que suposo es té quan es tenen fills com ells.

A en Nil, per excedir totes les expectatives. Pel whatsapp que em va enviar quan es va assabentar que dipositava la tesi... Si fos un post-it, l'emmarcaria. A l'Adri, per tot el que acompanya al seu somriure. Per viure en primera línia els últims coletazos d'aquesta tesis, i amenitzar el caos de casa i del moment amb el dia a dia. Prometo estar una mica més a l'alçada la propera vegada. A la Maria, perquè la única vegada que li vaig fer realment de tieta, va ser una passada! A en Xevi, perquè també li guardo ganes. A la Clàudia, per la seva meravellosa frescura. Per aquell Nadal que em va fer un lloc a la seva banqueta i vam acabar xapurrejant el piano a quatre mans.

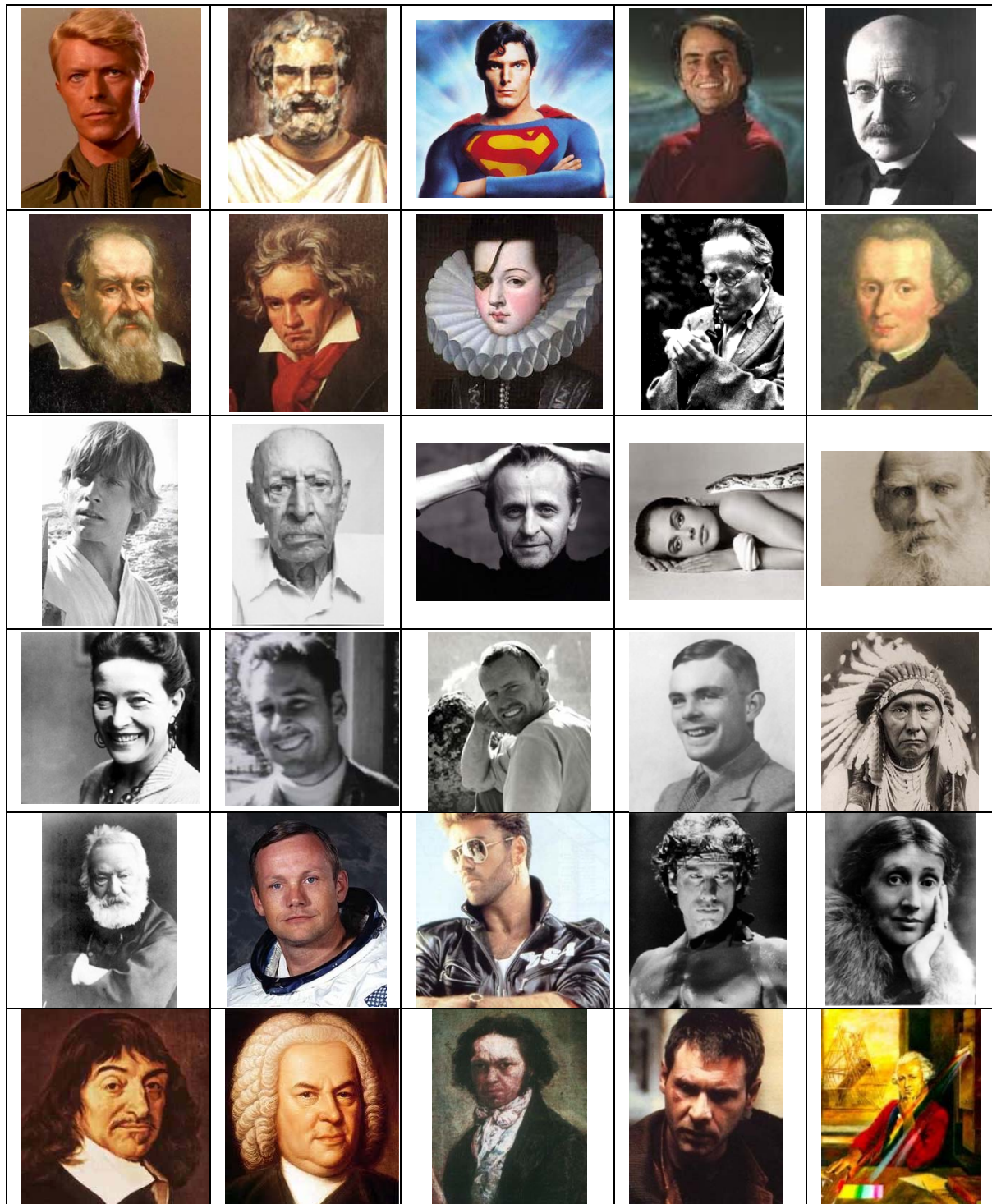
A la Penya, Ana, Txus, Mercè, Mònica, Aurora i Magda, per ser la meva altra família. Us estimo nenes!

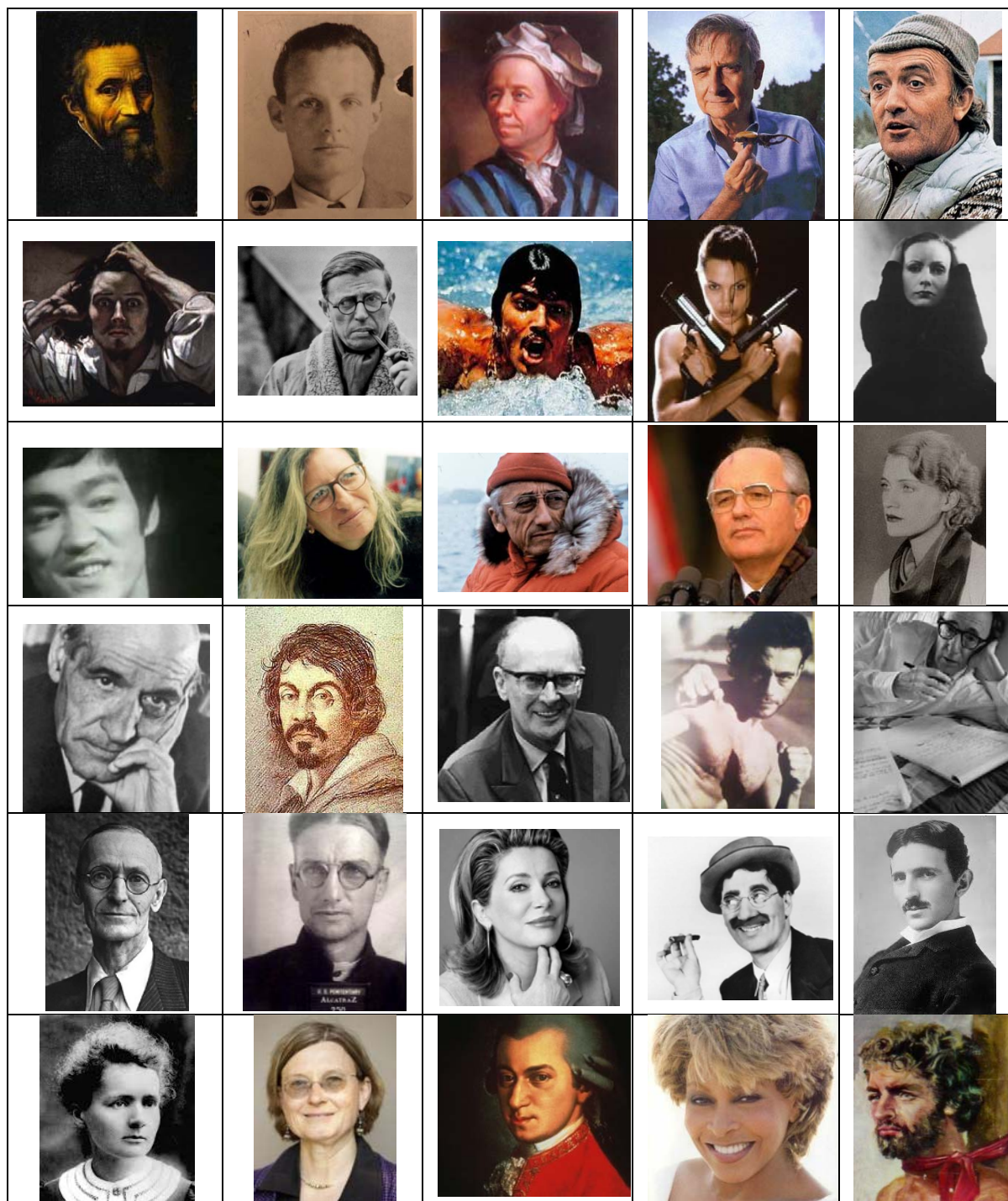
A les meves gates, Sasha i Kloe, perquè quan van estar, ho van omplir tot. A en Whisky, perquè el dia que el vam recollir a la carretera, la loteria ens va tocar a nosaltres. Al seu cosí-germà, l'Sticky, per fer més festes que ningú. Al Tango, per ser el millor gos del món! A les plantes de casa, de la feina i de camí a la feina, però sobretot, a la meva palmera, per escoltar-me. A en Ruscus, perquè només li falta respirar.

A totes les persones i criatures d'aquest planeta que m'han commogut en algun moment de la meva vida.

I especialment i sobretot, a tu, per ser els meus ulls i els meus batecs quan em van faltar.

And... one way or another, also thanks to:







Contents

Contents	I
List of Figures	V
List of Tables	X
List of Terms	XI
1 Introduction	1
1.1 Motivation	1
1.2 Research Objectives	4
1.3 System Overview	4
1.4 Summary of Contributions and Publications.....	5
1.5 Outline of the Dissertation	7
2 Biometric Fundamentals	9
2.1 Introduction	9
2.2 Identification vs Verification.....	11
2.3 Biometric Technologies	14
2.4 Fusion Systems	20
2.5 Technical Challenges	21
2.5.1 Scarcity of Data and High number of Classes	22
2.5.2 Vector Dimension Reduction	22
2.5.3 The Curse of Dimensionality	23
2.6 Ethical Aspects	24
3 Face Recognition	27
3.1 Problem Statement.....	27
3.2 Constraints and Challenges	31

3.3 Face Recognition Technology.	36
3.3.1 Face Recognition by Humans	37
3.3.2 Holistic Approaches: Projections, Transformations and Classifiers based Techniques.....	40
3.4 Infrared Approaches.....	52
3.4.1 Face Recognition in the Near Infrared Spectrum.....	53
3.4.2 Face Recognition in the Thermal Spectrum.	55
3.5 Face Databases	59
4 Visible, Near-Infrared and Thermal Face Imaging.	63
4.1 Introduction.....	63
4.2 Background Fundamentals	64
4.2.1 Electromagnetic Spectrum and Atmospheric Influence.	64
4.2.2 Principles of Infrared Thermometry	66
4.2.3 Face Image Model.....	68
4.2.4 Photometric and Thermal Sensor Models.....	70
4.3 Visible Imaging.....	72
4.3.1 Acquisition Systems in the Visible Spectrum	72
4.4 Near-Infrared Imaging	74
4.4.1 Acquisition Systems in the Near-Infrared Spectrum	74
4.4.2 Near-Infrared Faces.....	77
4.5 Thermal Infrared Imaging	78
4.5.1 Acquisition Systems in the Thermal Spectrum.....	81
4.5.2 Thermal Infrared Faces.	84
5 On the Relevance of Focusing in Thermal Image.	91
5.1 The Focusing Problem	91
5.2 Focusing Approaches	97
5.3 On the Focusing of Thermal Images	98
5.3.1 Focus Measures	99
5.3.2 Materials and Methods.....	102

5.3.3 Experimental Results and Conclusions.	103
5.4 Contribution of the Temperature of the Objects to the Problem of Thermal Imaging Focusing	105
5.4.1 Materials and Methods.....	105
5.4.2 Experimental Results and Conclusion.....	108
6 Multispectral Face Database.	111
6.1 Why a new Multispectral Database is required.....	111
6.2 Previous Design decisions.....	113
6.2.1 Frequency bands of the Multispectral System.....	113
6.2.2 Sensors Arrangement	114
6.2.3 Further Considerations	115
6.3 Acquisition Scenario	117
6.3.1 Visible and Thermal Acquisition System.....	117
6.3.2 Near-Infrared Acquisition System	117
6.3.3 Lighting Conditions.	120
6.3.4 Acquisition Protoco.	122
6.3.5 Database Features	123
7 Information Analysis of Multispectral Images	125
7.1 Introduction.	125
7.2 Background on Information Theory	126
7.3 Our Proposal: Information Theory-Based Fisher Score.....	128
7.4 Experimental Results.	132
7.5 Conclusions.....	136
8 Proposed Face Recognition Approach.	139
8.1 System Overview	139
8.2 The Proposed System	141
8.2.1 Face Segmentation Algorithm.....	141
8.2.2 Feature Extraction Algorithm	142
8.2.3 Feature Selection Algorithm.....	145

8.2.4 Classification.....	147
8.2.5 Fusion Method.....	148
9 Experiments and Results.....	151
9.1 Introduction.....	151
9.2 Experimental Results.....	152
9.2.1 Experimental Results of the Face Segmentation Preprocessing Step	152
9.2.2 Experimental Results with Different Illumination Conditions	156
9.2.3 Experimental Results for a Specific Sensor	158
9.2.4 Experimental Results in Mismatch Conditions	161
9.2.5 Experimental Results Using Multi-Sensor Score Fusion	163
9.3 Conclusions.....	167
10 Conclusions and Future Research	169
10.1 Conclusions.....	169
10.2 Future Research.....	172
References.....	175

List of Figures

1.1	The quartet of Liverpool with different levels of occlusions due to the strong illumination conditions of the scene.....	1
1.2	A quick visual comparative of the evolution of facial thermograms along the last fifteen years.....	3
2.1	A frame of the film <i>“Les Charlots dans le grand bazar”</i> of 1916, showing two individuals with the same resemblance.....	12
2.2	General scheme of a generic biometric recognition system.....	12
2.3	DET plot (dotted line): False Rejection Rate vs False Acceptance Rate.....	14
2.4	(a) Example of a fingerprint performed using the classic method of ink over paper. (b) Parameterization of the two most prominent local ridge characteristics.....	15
2.5	(a) AFIS 20001 Printrak system working during the first visit to the Catalan Police in October 2000. (b,c) A sample of different inkless fingerprints scanners.....	15
2.6	A personal card of an oriental owner.....	16
2.7	The author’s right hand used to access to a sport club by means of a hand acquisition device.....	17
2.8	(a) Iris (colored portion of the eye) pattern and retina pattern (b). (c) topography of the Corneal right eye (c) of the author acquired in 2010.....	17
2.9	Acquisition graphic tablet provided with the customized pen (a) azimuth and altitude angles of the pen with respect to the plane of the tablet (b). (c) Off-line information: image of the written signature. (d) On-line information: pen trajectory, pen pressure and pen azimuth/altitude.....	19
2.10	Marker position acquisition for gait analysis.....	20
2.11	A footprint acquisition of a newborn (a) and a guy with a code bar tattooed in his neck, acquired in the underground of Barcelona last spring (b).....	25
3.1	(a,b) Sample of two couple of different people much alike between them. (c) A pair of tattooed twins (photographed by Ken Probs) as an extreme case of low intersubject distance. (d) Set of different pictures of the actor John Travolta with a very low resemblance, conducing to a very high intrasubject distance.....	28
3.2	(a) Picture called “Multiple personalities” where all people are the same person (published in NY Times magazine, Section 6, pp 48-49. September 1, 1996). Anonymous guy (b1) characterized as a famous international model (b2) (make up by the great make-up artist, Kevyn Aucoin), and the true model (b3), Linda Evangelista. The author (c1) disguised as a curious aged man (c2).....	29
	Differences between face verification developed system using a single snapshot,	

3.3	and using the combination of 5 different ones respectively. The results were evaluated using the FAR & FRR indicators.....	31
3.4	Set of pictures of the author's nephew, since baby to eighteen revealing large variations in shape face over time.....	33
3.5	Occlusions given by different unusual (a) (waterpipe, camera, gun, microphone, and glass of wine) and usual (b) objects.....	33
3.6	(a) Curious frontal and profile face. (b) Full set of different faces of the same subject (Alan Turing) with different pose taking along the same session; Different levels of occlusions are given in rotations over 30° and profile views.....	34
3.7	Faces of the same subject under varying light conditions of the Harvard face database.....	35
3.8	A simultaneously image with a real 3D subject and its 2D projected face.....	35
3.9	Facial image rearranged as a vector. Each pixel responds to a coordinate in the high-dimensional image space.....	40
3.10	Face image expressed as a linear combination of EigenFaces.....	41
3.11	Classical graphical interpretation [Bis95] where the projection on the fisher direction, which is vertical, clearly shows the clustered structure of the data, whereas the projection on the PC (horizontal) contains most of the energy of the signal (the maximum variance of the data is in this direction), fails to classify this structure.....	42
3.12	The first-level WT decomposition of a face (and the four generated subbands). Note that H0h*H1v area records vertical features such as the outline of the face, whereas H1h*H0v area fix changes of the image along horizontal direction, such as mouth, eyes and eyebrows.....	47
4.1	(a) EM spectrum as function of the wavelength. (b) Atmospheric Transmittance in the region of the IR spectrum. Note the atmosphere strongly absorbs between 5 and 8μ due to water molecules in the atmosphere. (c) IR Channels.....	65
4.2	Illustration of the Plank's Law.....	67
4.3	A cup of hot coffee and the concerned three forms of heat transfer. Yellow zones marks heat transfer by conduction while the difference of temperatures among points 1 and 2 is due to the convection and radiation heat transmissions (coffee was previously removed by a spoon). [Image acquired with a FLIR SC620 thermal imager; resolution of 640x480 and NETD <40mK].....	68
4.4	3D face acquisition. (a) The author just standing in front of the screen to acquire and see 3D acquired face. (b) Generating the 3D face image. (c) Rendered image..	70
4.5	Basic operative principle of the active and passive sensors. Notice that the passive sensor senses the energy that is directly radiated by objects.....	71
4.6	Spectral sensitivity response of a general purpose CCD. Note the cut effect of the UV filter over 400nm and the sensitivity showed beyond both sides of the visible spectrum.....	72
4.7	Direction and quality of the light: 1st row: frontal (a), lateral (b), zenithal (c) and background light (d). 2nd row: hard (e), diffuse (f) and soft (g) light.....	73
4.8	NIR focus shift by the focal plane difference among VIS and NIR spectrums.....	75

4.9	A comparative of two night vision surveillance cameras provided with a standard lens and an IR corrected lens. (a) Scene acquired in a low-light situation with the standard lens. (b) Same scene and lighting conditions, acquired with the IR corrected lens.....	75
4.10	Some special applications: Behaviour against scattering: Same scene taken in VIS (a) and SWIR (b) [© FLIR]. (c),(d) Rescue tasks at night near Formentera island, the last winter taking advantage of the nightglow phenomenon.....	76
4.11	Same scene taken in VIS (a) and NIR (b) spectrums with the concerned films (in the second case, an especially sensitized to the region between 700 and 1200nm film, has been used). Strong differences in human skin reproduction can be found, as well as the fruit color reproduction. Extracted from [Gra79]. In addition, veins map may be also distinguished in (b) due to NIR deepness ability.	77
4.12	Eyes of a human and a dolphin illuminated with NIR light. (a) Sclera appears often as dark as the iris. (b) Strong reflection perceived in a dolphin's eye due to the presence of the referred biologic reflector tappetum lucidum system; Frame extracted from The Cove documentary [Psi09].....	78
4.13	Different results obtained when acquiring hair in different spectra.....	78
4.14	A sample of a simultaneously scene taken in visible and thermal IR spectrum.....	79
4.15	A sample of some applications in thermography.....	80
4.16	Same scene simultaneously taken in visible and thermal IR spectrum by a thermal imager, from three different distances: 1st row: TIR images taken at a distance of: 40cm (a), 30cm (b) and 20cm (c). 2nd row: VIS images taken at a distance of: 40cm (d), 20cm (e) and 10cm (f).....	83
4.17	Optical geometric description of parallax error for the worst case discussed in Figure 4.16.....	83
4.18	The actor Robert de Niro in different moments of the film Bull Ranging (He gained weight 27Kg to play with great effect the boxer Jack LaMotta). Minimum differences can be appreciated in forehead-temples and frontal zone of the nose among both images.....	86
4.19	Temperature profile of a human eye. (a) Thermogram of an eye. (b) Temperature profile graph. A Maximum of 36,1°C and a minimum of 34,7°C are detected, while the average temperature is 35,4°C.....	87
4.20	A set of different images of an author's colleague, wearing glasses in VIS (a), NIR (b) and TIR (c) spectrums. Facial thermogram of the author's sister wearing contact lenses (d).....	87
4.21	Same subject taken with two different thermal imagers with a spatial resolution of 160x120 pixels and NETD of 100mK (a) and 60mK (b). In the second sample, a better degree of detail is perceived (see both fringes, e.g).....	88
4.22	Thermogram of a human being, appreciating the jugular in the neck.....	89
5.1	Optical representation of chromatic aberration.....	93
5.2	Germanium transmission curves. (© Edmund Optics).....	94
5.3	Wave after diffraction through a gap.....	94
5.4	DOF as function of f/ number.....	96
5.5	Example of different focus for the two subsets described.....	103

5.6	Focusing measures obtained with the Heater database.....	103
5.7	Focusing measures obtained with the Face database.....	104
5.8	Used bulb in both Visible (a) and Thermal spectra (b).....	106
5.9	(a) Scenario. (b) TESTO 882 thermal imager with the stepping ring manual adapter.....	106
5.10	Best image of each set, at the eight evaluated temperatures.....	107
5.11	Focusing measures of each of the 96 images per subset.....	119
5.12	A sequence of four thermal faces with different levels of blurriness and their related histograms.....	110
6.1	A sample of a subject acquired in the three spectral bands. Co-registered VIS/LWIR thermal imagery. (Images are depicted once segmented).....	114
6.2	Hardware components of the multispectral imaging system mounted on a tripod.	115
6.3	A sample of the autorange process automatically carried out by the thermal camera.....	116
6.4	Implications of wearing glasses in the visible spectrum.....	116
6.5	Spectral sensitivity of the two visible opaque IR filters, specifically matched to our application.....	118
6.6	Webcam cameras and PCB boards for infrared illumination.....	119
6.7	A subject acquired by the system.....	119
6.8	GUI developed to control the IR illumination and camera settings.....	120
6.9	Spectral power distribution of the different lighting systems employed.....	121
6.10	Overall multispectral face acquisition scenario.....	122
6.11	Full Acquisition plan.....	122
6.12	Database structure. For each user there are four sessions and each session contains three kinds of sensors and three different illuminations per sensor.....	124
7.1	Relation between different measurements.....	128
7.2	TH, NIR, VIS and cross-correlation for person number 1.....	133
7.3	Usefulness of each combination.....	135
7.4	Comparison of R1 and R3.....	136
8.1	Overall purposed processing system. Multi-sensor fusion at score level.....	141
8.2	Thermal Face Segmentation algorithm steps for a sample image.....	143
8.3	Feature extraction after transforming by the DCT, and selecting a subset of the low frequency components.....	145
8.4	15x15 first coefficients M1 ratio for visible images of session 1. It is evident that the highest discriminability power is around the low frequency portion (upper left corner).....	147
9.1	1st row: Examples of badly detected face in VIS (a) and TH spectrum (b) using Viola and Jones. 2nd row: Examples of correctly detected faces in VIS and TH spectrums.....	153
9.2	1st row: Examples of badly detected face in VIS (a) and TH spectrum (b) using Viola and Jones. 2nd row: Examples of correctly detected faces in VIS and TH spectrums.....	155

9.3	Identification rate as function of the square size (N) of selected coefficients for visible (VIS), near infrared (NIR) and thermal (TH) sensors for natural (NA) illumination.....	157
9.4	Identification rate as function of the square size of selected coefficients for VIS, NIR and TH sensors for artificial (AR) illumination.....	157
9.5	Identification rate as function of the square size of selected coefficients for VIS, NIR and TH sensors for near infrared (IR) illumination.....	158
9.6	Identification rate as function of the square size of selected coefficients for Visible sensor and natural (NA), artificial (AR) and near infrared (NIR) illumination.....	159
9.7	Identification rate as function of square size of selected coefficients for Infrared sensor and NA, AR and NIR illumination.....	159
9.8	Identification rate as function of the square size of selected coefficients for Thermal sensor and NA, AR and NIR illumination.....	160
9.9	Trained rule combining VIS and NIR classifiers for NA illumination for training and testing, session 4.....	164
9.10	Example of trained rule identification rates combining three classifiers.	165
9.11	Contour plots when combining VIS, NIR and TH sensors under the following training and testing illumination conditions: NA-NA, NA-IR, NA-AR, IR-NA, IR-IR, IR-AR and AR-NA, AR-IR, AR-AR for session 4 and unnormalized feature vectors.....	166
9.12	Contour plots when combining VIS, NIR and TH sensors under the following training and testing illumination conditions: NA-NA, NA-IR, NA-AR, IR-NA, IR-IR, IR-AR and AR-NA, AR-IR, AR-AR for session 4 and normalized feature vectors.....	166
10.1	Overview of the Face Recognition system performed.....	170
10.2	Multifocusing thermal images proposed goal..... A very poor quality frame from a camouflaged surveillance video, registering two people of different races (the Author and an eastern woman) in the SF	173
10.3	Chinatown neighborhood.....	174

List of Tables

3.1	Computational burden of KLT, DCT and WHT for images of size $N \times N$ and concerned execution time using a Pentium 4 processor at 3GHz.....	46
3.2	Reviewed FR techniques and related recognition performance.....	51
3.3	Reviewed FR techniques in the IR spectrum and related recognition performance.....	59
3.4	Overview of the public domain Face DDBBs in the VIS spectrum.....	61
3.5	Overview of the public domain Face DDBBs in different spectral ranges. Two additional rows (facial details –glasses– and spectrum range) have been included as important properties for analyzing IR database approaches.....	62
4.1	Properties of the FPA detectors.....	82
5.1	Resulting maximum pixel sizes given in μm^2 . Special cases of LWIR and f/1 and f/2 have been highlighted.....	95
5.2	Summary of kinds of focus.....	98
5.3	Computational time (ms) for each method.....	104
6.1	Meaning of the file code name.....	123
7.1	Desirable values for mutual information (and correlation) in all the possible combinations and their implication.....	131
7.2	Implications of different ratio values for combining information from sensors.....	132
7.3	Experimental entropies for a single image (averaged for 10 people, 5 different images per person) for Visible, Near-Infrared and Thermal images. As shown, NIR images have the largest amount of information, followed by VIS and TH.....	134
7.4	Experimental results (averaged for 10 people, 5 different images per experimental result) for Visible, Near-Infrared and thermal images.....	134
7.5	Experimental conclusions using the criterion defined in Table 1.....	134
7.6	Experimental ratios and conclusions.....	136
9.1	Successful detection rates.....	153
9.2	Detection time of 220 images under artificial illumination.....	154
9.3	Detection time of 220 images under IR illumination.....	154
9.4	Detection time of 220 images under natural illumination.....	154
9.5	Optimal results for VIS, NIR and TH sensor under NA, IR and AR illumination conditions	161
9.6	Identification rates (%) under different illuminations, sensors and normalization conditions for testing sessions 3 and 4, labeled on the table as 3 and 4. Best results are marked on bold face.....	162
9.7	Identification rate for the combination of two and three sensors under different illumination conditions (NA=Natural, IR=Infrared, AR=Artificial).....	163

List of Terms

BS	Biometric System
BSS	Blind Signal Separation
CCD	Couple Charge Device
CFA	Color Filter Array
CRE	Cross-Race Effect
DCT	Discrete Cosine Transform
DET	Detection Error Trade-off
DFC	Distance From Centroid
DFT	Discrete Fourier Transform
DLDA	Direct LDA
DOF	Depth of Field
DWT	Discrete Wavelet Transform
ECOC	Error Correcting Output Code
ED	Extra Low Dispersion
EER	Equal Error Rate
EOG	Energy of the Image Gradient
EOL	Energy of Laplacian
FAR	False Acceptance Rate
FD	Focal Depth (also known as Depth of Focus)
FET	Failure to Enroll (also known as FTE)
FF	FeedForward
FL	Focal Length
FLD	Fisher's Linear Discriminant (better known as LDA)
FLDA	Fractional-step LDA
FLIR	Forward Looking Infrared
FOV	Field of View (also known as Field of Vision)
FPA	Focal Plane Array
FR	Face Recognition
FRR	False Reject Rate
FRT	Face Recognition Techniques
ICA	Independent Component Analysis
IdR	Identification Rate
IFOV	Instantaneous Field of View
IR	Infrared
IRCF	IR Cut-off Filter (also known as IRC)
IRED	Infra-Red Emitting Diode
KLDA	Kernel LDA
KLT	Karhunen Loéve Transform

KPCA	Kernel PCA
LDA	Linear Discriminant Analysis (also known as FLD)
LFA	Local Feature Analysis
LSDA	Locality Sensitive Discriminant Analysis
LWIR	Long Wave Infrared
MCT	Mercury Cadmium Telluride (also known as HgCdTe)
MDF	Most Discriminating Features
MEF	Most Expressive Features
MI	Mutual Information
MIR	Medium Infrared
ML	Machine Learning
MLP	Multi Layer Perceptron
MSE	Mean Square Error
MWIR	Mid-Wave Infrared
NETD	Noise Equivalent Temperature Difference (also known as NET)
NIR	Near-Infrared
NN	Nearest Neighbor
NNET	Neural Network
OSH	Optimal Separating Hyperplane
PCA	Principal Component Analysis
PPR	Projection Pursuit Regression
PR	Pattern Recognition
RBF	Radial Basis Function
ROC	Receiver Operator Characteristic
RSM	Random Subspace Method
SBF	Skin Blood Flow
SDR	Successful Detection Rate
SML	Sum-Modified Laplacian
SSS	Small Sample Size
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SWIR	Short Wave Infrared
TFR	Thermal Face Recognition
TIR	Thermal Infrared (also coded as TH)
UFPA	Uncooled Focal Plane Array
VOX	Vanadium Oxide
WHT	Walsh Hadamard Transform

Chapter 1

Introduction

*Cuando creíamos que teníamos todas las respuestas, de pronto,
cambiaron todas las preguntas.*

Mario Benedetti

1.1 Motivation

Face recognition (FR) has been broadly studied by several authors over the last thirty years. As a consequence great progress has been achieved toward developing computer vision algorithms that can recognize individuals based on their facial images in a similar way that human beings do, and leading this technology to reliable personal identification systems. This has been possible due to the increase of computational power of state-of-the-art computers. Nevertheless, in real and non-controlled environments, FR systems still remain an open challenge and major problems remain to be solved. The influence of varying lighting conditions is one of these challenging problems [Zou05]. Figure 1.1 shows a sample of a not inconsiderable effect over the faces of The Beatles, produced by the direction of light.



Figure 1.1: The quartet of Liverpool with different levels of occlusions due to the strong illumination conditions of the scene.

In addition, faces do not fit into the traditional approaches of model based recognition in vision, as they are complex three-dimensional objects whose appearance is also affected by a large number of other factors including identity, aging, facial pose and facial expressions, blocked effects and facial look. Thus, FR becomes one of the most fundamental problems in pattern recognition [Li04, Zha00].

Aside from the above statements, advantages of FR automated systems cannot be dismissed. This technology is turning more popular when compared with other biometric modalities. Thus, unlike iris, retinal, hands-geometry or fingerprint recognition systems, FR does not require high accuracy and expensive image acquisition equipments. Moreover, FR is together with gait recognition and speech, a biometric system performed by means of non-contact measurement, which offers advantages such as non-cooperative and/or camouflaged recognition modalities.

On the other hand, and thanks to progresses in microelectronics, spurred by its falling costs, *focal plane arrays* (FPA) provided with *microbolometers* or quantum detectors are now available with high thermal sensitivities, often smaller than 60mK (while 100mK was an achievement only a few years ago), high spatial resolutions (up to 512x512 pixels) and greater uniformity (up to 99,5%). Therefore, and thanks to this new generation of more sophisticated infrared (IR) cameras, new applications are emerging continuously, especially when developing FPA operating in the thermal spectrum sub bands. Today, *thermography* as is known such powerful technology is a widely used imaging system, ranging high end scientific research and development, medical and veterinary support, materials science, quality control in industry processes, energy conservation, building inspection and defense, among the most important ones. Similarly, thermal IR cameras have a powerful set of properties concerning biometric applications, some of which are detailed below:

- Are not affected by illumination because they acquire heat emission, not illumination reflection. Additionally, they are significantly less sensitive to solar reflections, so they are particularly well suited to outdoor applications, which is still a challenge in visible spectrum.
- Due to the thermal imagers are provided with additional visible cameras, this technology easily enables data fusion (visible and thermal) biometric solutions without requiring additional devices.
- Thermal images are more robust to disguises, make up and plastic surgery (which does not reroute the facial map of veins).

Thus, due to the above reasons among others, FR by means of thermal IR imaging has become a subject of research interest in the last few years, emerging as both an alternative and a complementary source of information for FR in the visible spectrum [Zha06]. However, not much research work can still be found in literature, probably due in part by the lack of widely available data sets, as well as high cost of the thermal IR equipment and its low sensitivity, resolution and noise performance. Figure 1.2 shows a quick visual tour of these performance along the last years.

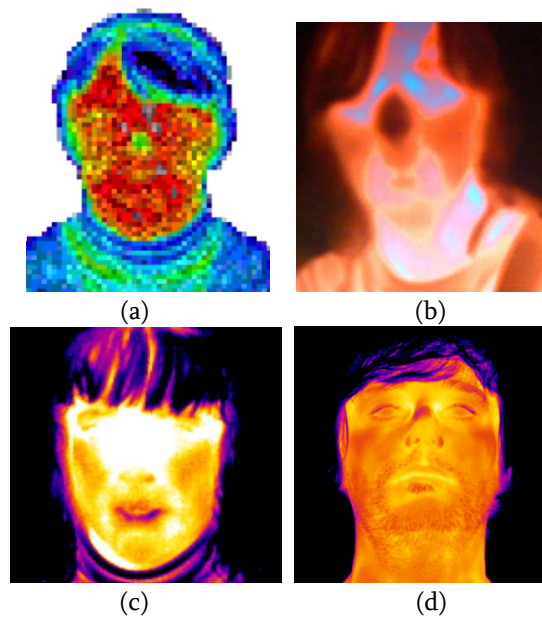


Figure 1.2: A quick visual comparative of the evolution of facial thermograms along the last fifteen years: First row: (a) Image taken in the mid nineties [Extracted from [Hon98]; original source: website www.betac.com available in 1997]. Checkerboard effect is perceived due to sensor coarse spatial resolution. (b) The author's thermogram taken by a general purpose thermal imager at Pompidou center in 2006. (Although technical properties are not available, a low sensitivity is clearly assumed due to the strong appreciated quantization effect). Second row: two additional images acquired in 2009 (c) (160x120; NETD 120mK) and in 2011 (d) (640x480; NETD 30mK) during the thermal IR camera assessment period.

However, although the merits of this new infrared technology allow higher performance, there are also very substantial challenges for such a technology that are still unsolved, both concerning general design aspects and FR related approaches:

- Blur effects due to diffraction and chromatic aberration phenomena.
- Low resolution sensors.
- Limited Depth of Field.
- Parallax error when dealing with both, thermal and visible (also available in the same thermal camera) images.
- Recognition performance is not comparable with that of the broadband images acquired with conventional cameras in the visible and the Near infrared (NIR) spectrum.
- Bulky cameras that are still expensive compared with visible cameras.
- Facial thermograms, shows a strong variability depending on temperature of the environment, specially the nose area (due to its thermoregulatory activity).

- Eyeglasses are fully opaque in the infrared spectrum beyond $3\mu\text{m}$, producing an important face occlusion in the eyes area.

On the other hand, while it is true that human beings can perform biometric recognition based on visible face signals, is not clear enough the more useful information outside the visible and NIR spectrum existing, in form of thermal radiation. In this dissertation, we will appropriately address this open question and we will also consider the use of data fusion architectures in order to deal with this hypothetical amount of relevant information to perform more improved FR systems.

1.2 Research Objectives

The objective behind this research is manifold:

- To investigate technological capabilities of new affordable IR cameras for FR biometric purposes.
- To explore the strength of introducing *heat information of the face* in the face recognition system.
- To mathematically quantify the proportion of redundant and complementary information between visible and thermal facial images, and to extend such comparative to the Near-Infrared human face images.
- To address the problem of multispectral FR systems by means of data fusion and to exploit the overall contributions of the different spectra, visible, near-infrared and thermal images to improve final FR system performance.
- To address the problem of thermal images focusing and the concerned implications when dealing with facial thermograms.

1.3 System Overview

The FR scheme we propose consists on six main stages:

1. Multispectral and multisession facial acquisition stage: This stage has been developed by means of a thermal IR camera, for co-registering the images in both, the thermal and the visible spectrum, and a customized webcam that provides the face images in the NIR spectrum.
2. Face segmentation stage: This stage has been carried out by means of a new algorithm specifically designed for thermographic facial images and fully described in [Mek10].
3. Feature extraction stage: It has been addressed by using the Discrete Cosine Transform.
4. Feature vector coefficients selection stage: This stage has been developed by means of discriminatory criteria.

5. Classification stage: In the involved design, a template matching method has been applied based on distance calculation by using a fractional distance.
6. Data Fusion final step: This fusion has been formulated at matching score level.

1.4 Summary of Contributions and Publications

This Thesis provides the following main contributions:

- The design and acquisition of a novel Multispectral Face Database in the visible, near infrared and thermal spectra [Esp12]. The referred database consists of 41 people acquired in four different acquisition sessions, five images per session and three different lighting conditions. The scope of the utility of this new face database, nicknamed as CARL (*CAtalan Ray Light*) involves mainly the performance assessment in the design of automatic FR systems for civil applications, allowing the development and evaluation of new biometric recognition approaches.
- The redundancy between several spectral bands has been explored and analyzed from the perspective of the information theory [Esp11]. The results reveal that there is complementary information between different wavelengths. This issue emphasizes the growing interest in thermographic imaging applications.
- A new criterion based on the Fisher score for the case of mutual information has been proposed, which allows evaluating the usefulness of different sensor combinations for data fusion and for crossed-sensor recognition (matching of images acquired in different spectral bands) [Esp10].
- The design of a face Segmentation algorithm specially devoted when dealing with facial thermograms [Mek10].
- The development of a multispectral face recognition score fusion system tested on the acquired face database [Esp12]. Experimental results show a significant improvement by combining the three spectra.

Most part of these contributions has been published in indexed journals, international conferences and book chapters, as reported in the following list:

a1 INDEXED JOURNALS

- V. Espinosa-Duró, M. Faúndez-Zanuy, J. Mekyska and E. Monte-Moreno, *A Criterion for Analysis of Different Sensor Combinations with an Application to Face Biometrics*. Cognitive Computation. Vol. 2, Issue 3, pp 135-141. September 2010.
- V. Espinosa-Duró, M. Faúndez-Zanuy and J. Mekyska, *Beyond Cognitive Signals*. Cognitive Computation. Ed. Springer. Vol. 3 pp 374-381. June 2011.

- V. Espinosa-Duró, M. Faúndez-Zanuy and J. Mekyska, *A New Face Database Simultaneously Acquired in Visible, Near Infrared and Thermal Spectrums*. Cognitive Computation. Ed. Springer. July 2012. DOI: 10.1007/s12559-012-9163-2.
- M. Faúndez-Zanuy, J. Mekyska and V. Espinosa-Duró, *On the Focusing of Thermal Images*. Pattern Recognition Letters. Ed. Elsevier. Vol. 32. pp 1548-1557. August 2011.
- M. Faúndez-Zanuy, J. Roure, V. Espinosa-Duró and J. A. Ortega, *An Efficient Face Recognition Method in a Transformed Domain*. Pattern Recognition Letters. Vol. 28, Issue 7, pp 854-858. May 2007. ISSN: 0167-8655.

b1 INTERNATIONAL CONFERENCES

- V. Espinosa-Duró, M. Faúndez-Zanuy and J. Mekyska, *Contribution of the Temperature of the Objects to the Problem of Thermal Imaging Focusing*. Proceedings of the 46th IEEE-ICCST International Carnahan Conference on Security Technology. Pp 363-366. Boston, USA. October 2012. ISBN:978-1-4673-4807-2.
- V. Espinosa-Duró, *Thermal Imaging*. 1st SPLab Workshop. Brno, Check Republic. 27-28 October 2011.
- V. Espinosa-Duró, M. Faúndez-Zanuy and Jiri Mekyska, *Different Sensor Face Images Study from an Information Theory Point of View*. 3rd COST 2102-EUCOGII International school on toward autonomous, adaptive and context-aware multimodal interfaces: theoretical and practical issues. Caserta, Italia. March 2010.
- J. Mekyska, V. Espinosa-Duró and M. Faúndez-Zanuy, *Face Segmentation: A Comparison Between Visible and Thermal Images*. 44th IEEE-ICCST International Carnahan Conference on Security Technology. San José, USA. October 2010. ISBN 978-1-4244-7401-1.
- V. Espinosa-Duró, *Face Recognition using VIS and Near-IR Images: A Comparison*. 8st SCI. MultiConference on Systemics, Cybernetics and Informatics. pp 294-297. Orlando, USA. July 2004. ISBN 980-6560-13-2.
- V. Espinosa-Duró, M. Faúndez-Zanuy and J. A. Ortega, *Face Detection from a Video Camera Sequence*. 38st IEEE-ICCST International Carnahan Conference on Security Technology. pp 318-320. Albuquerque, USA. October 2004. ISBN 0-7803-8506-3.

c1 BOOK CHAPTER

- M. Faúndez-Zanuy, V. Espinosa-Duró, and J. A. Ortega, *Low Complexity Algorithms for Biometric Recognition*. Chapter in Verbal and Nonverbal Communication Behaviours. Lecture Notes in Computer Science-LNCS 4775. Ed. Springer pp 275-285. 2007. ISBN-13 978-3-540-76441-0.

Equally, other interesting publications of the author indirectly related with the dissertation, are also reported. The major ones are listed below:

a2 ADDITIONAL INDEXED JOURNALS

- M. Faúndez-Zanuy, V. Espinosa-Duró and J. A. Ortega, *A Low-Cost Webcam&Personal Computer Open Door*. IEEE AES Aerospace and Electronics Systems Magazine. Vol. 20, Issue 11, pp.23-26. November 2005. ISSN: 0885-8985.
- V. Espinosa-Duró, *Fingerprints Thinning Algorithm*. IEEE AES-Aerospace and Electronics Systems Magazine. Vol.18, Issue 9. pp 28-30. September 2003. ISSN 0885-8985.
- V. Espinosa-Duró, *Minutiae Detection Algorithm for Fingerprint Recognition*. IEEE AES-Aerospace and Electronics Systems Magazine. Vol. 17, Issue 3, pp 7-10. March 2002. ISSN 0885-8985.

b2 ADDITIONAL INTERNATIONAL CONFERENCES

- V. Espinosa-Duró and E. Monte-Moreno, *Face Recognition Approach Based on Wavelet Transform*. 42st International Carnahan Conference on Security Technology. IEEE-ICCST pp 187-190. ISBN 978-1-4244-1816-9. Prague. Czech Republic. October 2008.
- M. Faúndez-Zanuy, V. Espinosa-Duró and J. A. Ortega, *An Efficient Face Recognition Method in a Transformed Domain*. 41st International Carnahan Conference on Security Technology. IEEE-ICCST pp 281-284. ISBN 1-4244-1129-7. Ottawa. Canada. October 2007.
- V. Espinosa-Duró, *Mathematical Morphology Approaches for Fingerprints Thinning*. 36st International Carnahan Conference on Security Technology. IEEE-ICCST. pp 43-45. ISBN 0-7803-7436-3. Atlantic City. USA. October 2002.
- V. Espinosa-Duró, *Biometric Identification using a Radial Basis Network*. 34st International Carnahan Conference on Security Technology. IEEE-ICCST. pp 47-51. ISBN 0-7803-5965-8. Ottawa. Canada. October 2000.

1.5 Outline of the Dissertation

After this brief introductory chapter, the remaining ones are organized as follows:

- The first part of Chapter two covers general concepts about biometric recognition systems, while the second one is focused on the current biometric modalities. Fusion methods, main challenges and ethical aspects concerning the use of biometrics are also discussed.
- Chapter three reviews the field of automated face recognition systems. The first part formulates the FR problem and analyzes the strengths and weaknesses of human faces as biometric authenticators whereas the second one outlines the existing FR approaches and details the state of the art paying special attention to infrared FR technology, where it shows future potential. The chapter ends with a brief overview of the most prominent public face databases.

- Chapter four covers general aspects of the different facial images in the three different concerned spectra, ranging from image properties to related acquisition technologies.
- The first part of chapter five provides an overview of the theoretical optical aspects, discussing limitations and trade-offs in designing cameras for thermal IR acquisition and putting the emphasis on the focusing problem, whereas the second is a more experimental part that concentrates on the relevance of the temperature of the objects when focusing scenes in the thermal spectrum. The chapter ends with the analysis of the focusing problem when dealing with thermal facial images.
- Chapter six is fully devoted to the novel CARL face database specially developed for this research work and fully used in our experiments.
- In Chapter seven, the faces acquired in several spectra are analyzed from the perspective of information theory. The first part of the chapter reviews mathematical theory and introduces the basic analytic expressions, whereas the second presents experimental results and some conclusions are discussed.
- Chapter eight concentrates on the description of the proposed FR approach specifically designed to characterize the power of simultaneously acquired multispectral facial images.
- Chapter nine is devoted to the main experimental results and conclusions, and demonstrates the system performance on the developed DDBB.
- Finally, Chapter ten concludes our work and proposes a discussion of the advantages and limitations of our approach. Some directions for future research in this field are also suggested.

Chapter 2

Biometric Fundamentals

L'home neix amb un signe d'interrogació al seu cor.
Anònim.

In this chapter we summarize the state-of-the-art in biometrics as the general framework of our research work. Section one defines and contextualizes biometrics since the very beginning of the discipline. Section two is devoted to the biometric underlying verification and identification issues as well as the parameters related with the performance assessment. The section third provides an overview of the most established biometric technologies as well as the new trends. Section fourth overviews fusion methods whereas, section fifth and last, incurs in the ethical concerns that should be taken into account when dealing with biometric technology.

2.1 Introduction

Biometry is an old term coined by Sir. Francis Galton, and the discipline itself was firstly conceived as the application of statistical methods to the study of evolution of quantitative characters, and was inspired by Galton's *Natural Inheritance* [Bul03]. In the editorial of the first volume of the journal *Biometrika*, a journal focused to the statistical study of biological problems and founded in 1.901 by Karl Pearson and W. Weldon and F. Galton as editor, the following text to describe the new paradigm can be found [Wpe51]:

The starting point of Darwin's Theory of Evolution is precisely the existence of differences between individual members of a race or species which morphologists for the most part rightly neglect. The first condition necessary, in order that any process of Natural Selection may begin among a race, or species, is the existence of differences among its members.

In this sense, human fingerprints are unique characteristics in the same way face lines of a tiger, or even eyespots of the butterflies are also unique patterns depending of each individual. A few years earlier, namely in 1892, Sir Francis Galton and Sir Edward Henry, had already established the usefulness of fingerprints for identification. Truth be told, the essence idea of recognize different subjects using their physical characteristics is also very old and even used in traditional tales as *the wolf and the seven little kids* tale. Remember the phrase “*Show me your leg below the door*”...

Currently, *Biometrics* refers to automatic human recognition technologies, based on pattern recognition (PR) techniques, using physiological or behavioral characteristics of the persons, called also *biometric authenticators*. Some human traits currently utilized, include fingerprints, iris, retina, facial patterns, speech, hand-written, signature, keystroke patterns and gait [Jai99, Jai04]. Note that one of the major interest of these biological or behavioral traits are that they cannot be forgotten, misplaced, duplicated or stolen.

Going more in depth into biometrics, the first question we should ask is: *Which is the best authenticator to solve a specific biometric recognition problem?* As common sense says, a good biometric trait must accomplish a set of properties. Mainly they are [Jai04, Cla94]:

- *Universality*: every person should have the chosen trait.
- *Distinctiveness* (also refers to *individuality*, *uniqueness* and *univocal* properties): any two persons should be different enough to distinguish each other based on this trait.
- *Permanence* (also refers to *persistence* or *immutability*): the trait should remain constant enough (with respect to the matching criterion) over a period of time.
- *Collectability*: the trait should be acquirable and quantitatively measurable.

However in a real biometric system (henceforth BS), there are a number of related issues that should also be taken into account, including:

- *Performance*: the identification accuracy and required time for a successful recognition must be reasonably good.
- *Acceptability*: people should be willing to accept the BS, and do not feel that it is invasive, dangerous, cause any discomfort, etc.
- *Circumvention*: the possibility to attack and deceive the BS should be negligible.

Since there are numerous BS, it means, not any is perfect one. Each biometric modality has its own advantages and disadvantages with respect to the above mentioned factors and all have sense in an application or other [Nan02]. Using a correct selection criterion based on the different capabilities and performance of the existing biometric systems, the final identification system designed, will be able to prove with reasonable certainty that we are, or are not, someone previously registered in the user’s database [Jai99, Jai04].

In recent years biometric technology are increasingly gaining popularity in a large spectrum of applications, ranging from governmental programs (National ID card, Visa, public security, fight against terrorism,...), commercial applications such as surveillance, security and access control systems (electronic commerce, e-banking,...), to personal applications such as logical and physical access control (computer logon, internet, keyless ignition for cars,...). Although a number of effective solutions are nowadays available, there are still many challenging problems in improving the accuracy, efficiency, robustness and user-friendliness of current biometric systems, being necessary new ideas, algorithms and techniques to overcome some of these limitations. Additionally new problems are also emerging with new applications, e.g. personal authentication on mobile devices such as smartphones, PDAs and other hand held devices.

2.2 Identification vs Verification

Biometric recognition comprises two standard methods of matching the newly captured biometric feature known as *Identification* and *Verification* [Nan02]. We will refer to “*Recognition*” for the general case, when we do not want to differentiate between them. However, some authors consider recognition and identification synonymous.

- *Identification* is determining *who a person is*. It involves taking the measured characteristic and trying to find a match in a database containing records of people and that characteristic. In a more general response, the system will report a list of the most similar individuals in the database. This method may require a large amount of processing power and some time if the database is very large. The majority of the identification is in law enforcement, forensics and intelligence. The system performance is evaluated using an identification rate.
- *Verification (also referred to as Authentication)* is determining if a person is *who he sais to be*. The algorithm either accepts or rejects the claim. Therefore, the algorithm can return a confidence measure of the validity of the claim related with the *verification threshold*, previously defined by the designer. The general process involves taking the measured characteristic and comparing it to the previously recorded data for that person. Obviously, this method requires less processing power and time than the previous one, because the just require *one-to-one* comparison (whereas identification requires *one to N*, being *N*, the number of users in the database). Common authenticators include passwords, private keys, magnetic cards or PINs in order to provide the previous user’s identity to verify (and an additional level of security). Verification is often used for accessing places (physical access control) or information [Ash00]. Figure 2.1 tries to show how verification is not only a difficult task for computers, but also for ordinary people in some cases: Interestingly, Charles Chaplin (the star that popularized the character of Charlotte) was submitted to a contest of Charlotte imitators in San Francisco Theater around 1915, and get tenth place [Mil96].



Figure 2.1: A frame of the film “*Les Charlots dans le grand bazar*” of 1916, showing two individuals with the same resemblance.

The underlying biometric recognition process is similar independent of the biometric problem to satisfy and the final trait chosen. Thus, whatever the system, all exhibit the same common general configuration, depicted in Figure 2.2:

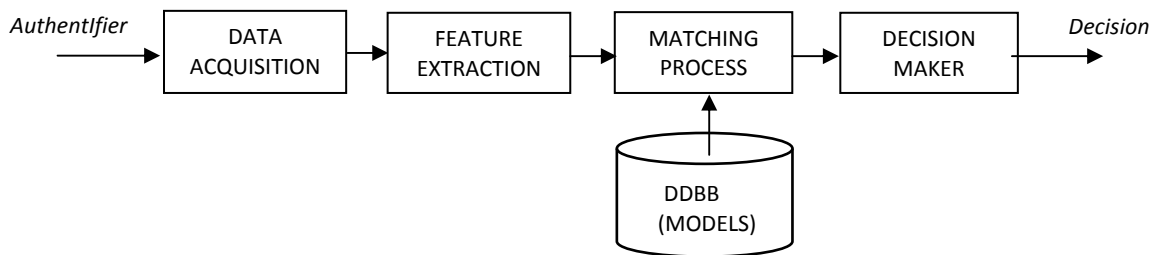


Figure 2.2: General scheme of a generic biometric recognition system.

We summarize the four main steps depicted in the above scheme, as follows:

1. **Data Acquisition:** The physical or behavioral trait sample is acquired by means of a specific purpose acquisition system. This first stage is one of the most sensitive parts, due to the fact that most biometric recognition algorithms strongly depend on the characteristics of the acquired data. Thus, if possible, the quality of the provided signal will be checked. If it is below a previously defined threshold, a new acquisition will be performed. It will also be necessary to capture enough samples in order to achieve a high robustness of the system.
2. **Feature Extraction:** A set of characteristics is extracted from the samples and the user template is extracted by means of several digital signal processing techniques.

3. Matching: Measured parameters of previous step are used to work out a model for the given user. In *enrollment*¹ mode, the whole set of extracted features are stored, and used as model. Once the user has been enrolled during the enrolment phase, a new real-time sample will be taken, and their corresponding model will be matched against all the stored templates of the database in identification mode, or against the user's template in the verification mode. Different distances (Euclidean, Hamming,...), statistics methods (Gaussian Mixture Models), and classifiers (Artificial Neural Networks and Support Vector Machines) have been successfully applied in general approaches to perform this comparison task.
4. Decision: The system decides if the set of features extracted from the new sample is a match or a miss-match.

The final identification system performance can be evaluated using an *Identification Rate* (IdR). The information coming from IdR is straightforward: Proportion of previously enrolled subjects successfully mapped to the correct identity.

For verification, if we have a population of N different people, the system can be assessed using the *False Acceptance Rate* (FAR; those situations where an *impostor* is accepted) and the *False Rejection Rate* (FRR²; those situations where a genuine *user* is incorrectly rejected) indices, also known in Detection Theory as *False Alarm* and *Miss*, respectively. Both errors have a bad effect and need to be weighted carefully to make sure the optimal mix for the required security arrangements is get. This necessary trade-off between them is usually established by adjusting a decision threshold. The concerned performance can be plotted in a *Receiver Operator Characteristic* (ROC) or *Detection Error Trade-off* (DET) plot [Mar97]. The DET curve gives uniform treatment to both types of error, and uses a logarithmic scale for both axes, which spreads out the plot and better distinguishes different well performing systems and usually produces plots that are close to linear (see Figure 2.3). In addition, *Equal Error Rate* (EER) is the value that satisfies the equality $FAR = FRR$ and is often quoted as a summary performance measure.

¹*Enrollment*: In a similar fashion as humans do, the system needs a learning procedure, before being able to recognize (it is obviously hard to recognize a person that has not been seen before). The purpose of enrollment is to have user's characteristics registered for later use.

Additionally, the proportion of individuals for whom the system is unable to generate repeatable templates for a biometric solution is defined as *Failure to Enroll* (FET/FTE) rate [Man01]. FET includes those unable to present the required biometric feature (for example an Iris system can fail to enroll the iris of a blind eye), those unable to produce an image of sufficient quality at enrolment, as well as those unable to reproduce their biometric feature consistently (the system cannot reliably match their template in attempts to confirm that the enrollment is usable).

²In order to work with positive logic the *True Acceptance Rate* (TAR) index also exists, although it is not as much used as FAR one.

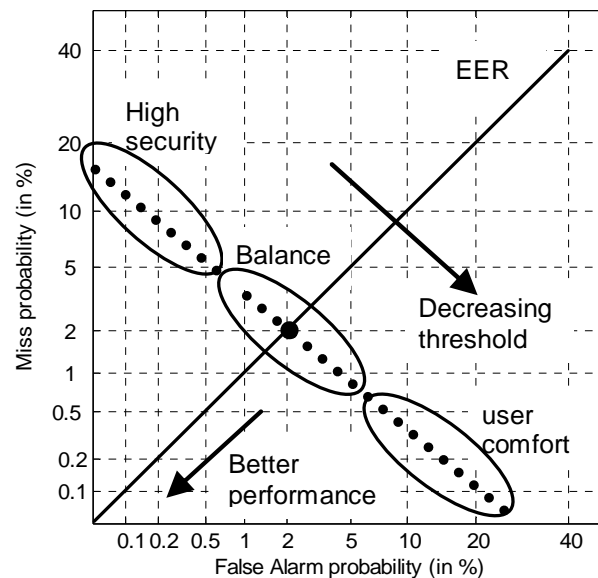


Figure 2.3: DET plot (dotted line): False Rejection Rate vs False Acceptance Rate. If a larger acceptance is allowed so that less people will have a Miss error, the net effect of this would be to have more opportunity for a False Alarm error to occur. This works in reverse as well. This plot uses a logarithmic scale that expands the extreme parts of the curve, which are the parts that give the most information about the system performance.

2.3 Biometric Technologies

A considerable number of Biometric recognition systems based on different kinds of authenticators for covering a large range of biometric applications, exists. Although all these systems satisfy, to a certain extent, the requirements mentioned in Section 2.1, be aware that a biometric system based on physiological characteristics is more reliable than one which adopts behavioral features [Zha00]. This section overviews the most established biometric technologies as well as some of the new trends.

FINGERPRINTS. As has been pointed at the very beginning of this chapter, Sir Francis Galton and Sir Edward Henry concluded in the late-19th century that fingerprints could be used for identification purposes. Curiously, few years earlier, specifically in 1856, William James Herschel, the grandson of the renowned astronomer Sir William Herschel (the discoverer of the infrared spectrum), used for the first time in history, printing traces of heart and right hand index finger, in his private contracts as a proof of identification (in addition to the signature) and published it in the prestigious Nature journal [Her1884, Bea01]. The destiny has wanted that one hundred and fifty years later, biometric systems, have gathered them again.

A fingerprint is formed by a set of ridges and valleys; the ridge flow pattern presents local discontinuities, such as *ridge endings* and *ridge bifurcations* called *minutiae* [Jai97]. As showed in Figure 2.4, the first ones are defined as the point where a ridge ends abruptly whereas the second ones are defined as the points where a ridges forks or diverges into branch ridges [Esp02]; these type location and orientation are features that make each fingerprint unique (*-Two like fingerprints would be found only every 10^{48} years-* [Mal03]).

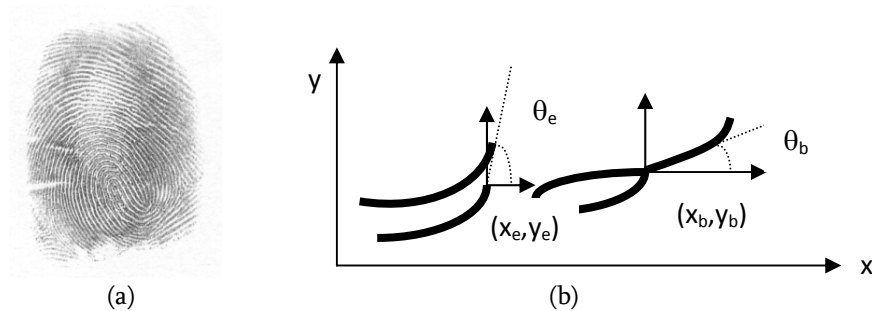


Figure 2.4: (a) Example of a fingerprint performed using the classic method of ink over paper. (b) Parameterization of the two most prominent local ridge characteristics.

Because of the well-known distinctiveness and persistence properties and a last special one as being the only biometric authenticator capable of leave a copy of itself over a set of different surfaces previously touched by the subject, leads fingerprints identification systems the most widely biometric used technique in police field for identifying criminals and victims. These very mature identification systems are known are *AFIS* (Automated Fingerprint Recognition System). Figure 2.5 shows a real performance of the AFIS used by the Catalan Police. In the verification field by its side, fingerprint systems have the additional advantage to require inexpensive standard capturing devices.

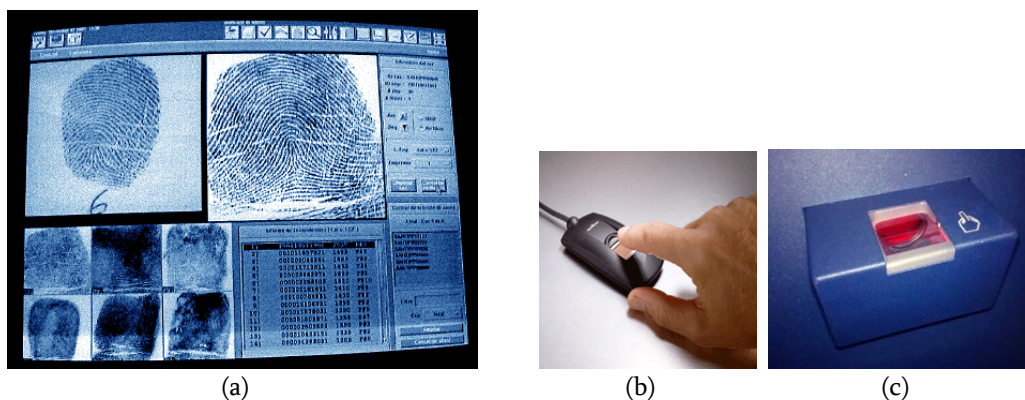


Figure 2.5: (a) AFIS 20001 Printrak system working during the first visit to the Catalan Police in October 2000; Screen shows the results for trial number 6. On the top left, the inkless fingerprint can be seen; on the top right, the rolled-ink fingerprint. Bottom left, there are the most probable fingerprints; and bottom right, the scores for the best candidates in the DDBB. (b,c) A sample of different inkless fingerprints scanners.

FACE. Face recognition is one of the most promising biometric identification methods probably because it is the most natural way to recognize the identity between human beings. Human faces represent one of the most common visual patterns in our environment. Thus, it is usual for people to identify somebody by his face, while would be impossible to do it by means of his fingerprint, iris, retina or even a personal card, due to the large amount of existing languages, with even different characters in worldwide. Figure 2.6 reflects this fact.

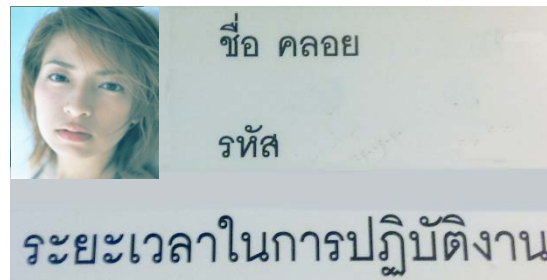


Figure 2.6: A personal card of an oriental owner. Notice that western people may identify the identity of the subject by means of the photograph, but not by the text.

Additionally, it is the second most popular, biometric trait, after fingerprints. One of its advantages is the high social acceptance by users, because it does not have the criminal stigma associated to other biometric traits, such as fingerprints. Although this technology is mainly adapted to applications with cooperative users, an interaction with the people during the overall acquisition process is not always required. This last aspect is especially beneficial for covering surveillance applications in security places such as airports, high security restricted areas...etc. By contrast, the high variability of faces (facial expression, aging, lighting conditions, head rotation, changes of look...) does not make recognition easy. Comparative studies reveal that face recognition cannot perform as accurate as other biometric traits such as fingerprint or iris [Man01]. The research work presented in this Thesis is based on this technology. For this reason, a more extended description of the general framework, involved parameters and recognition methods will be done throughout the document.

HAND-GEOMETRY. In recent years, hand geometry biometric techniques have become a very popular biometric access control, which has captured almost a quarter of the physical access control market [Jai04]. Hand geometry or hand shape technologies have the benefit of being passive, non-intrusive recognition systems, such as both biometric technologies previously covered. With the fast development of 2D and 3D sensors and data processing algorithms, diverse hand shape based biometric systems have been widely deployed in various applications. For the general purpose, several measurements are taken, including finger width and height of the palm, deviation of the fingers from the straight line, angles between them, etc. Figure 2.7 shows a sample of a 2D fully operative device.



Figure 2.7: The author's right hand used to access to a sport club by means of a hand acquisition device. Notice the pegs to properly position the hand.

There are verification systems available that are based on measurements of only a few fingers (typically index and middle) instead of the entire hand. These devices are smaller than those used for hand geometry, but still much larger than those required in some other biometrics (e.g. fingerprints, voice, face).

OCULAR. In spite of the reduced size and sensibility area of the eyes, it provides two reliable ocular biometric methods based on the following physiological characteristics: Iris and Retina (see Figures 2.8 (a) and 2.8(b)). They have the benefit over biometric systems such as those based on fingerprints, face or hand geometry that they are non-reproducible. Last, recent studies carried out by Celia Sánchez-Ramos also reveal the potential of inner corneal surface to become a reliable biometric authenticator in a near future. Figure 2.8 shows a sample of the third referred biological characteristics.

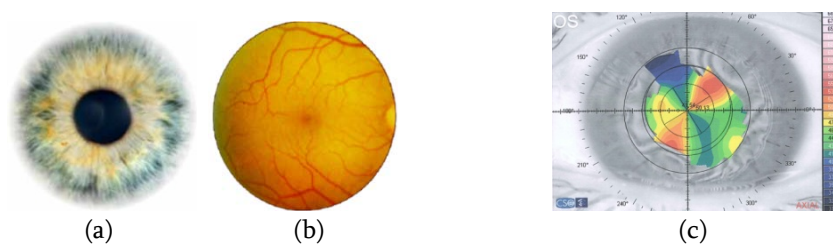


Figure 2.8: On the left, iris (colored portion of the eye) pattern (a) and retina pattern (b). On the right, topography of the Corneal right eye (c) of the author acquired in 2010.

Iris is the biometric characteristic that remains most invariant along time and also presents the highest distinctiveness among biometric features: The combinatorial complexity of the phase information across different people spans about 244 degrees of freedom (fingerprints presents around 100) and generates a discrimination entropy of 3,2bits/mm² [Dau01].

Although iris recognition first purposes was suggested by ophthalmologists and date from 1936, the major iris recognition advancement was first proposed by John Daugman in 1993 [Dau93] with a patent being issued in 1994 [Dau94]. Iris recognition systems not only offers excellent security guarantees reporting extra low FARs but also has the extra benefit of providing an aliveness detection mechanism based on the difference between the iris size of two images sequences (previously to the second acquisition, the system forces to contract pupil -enlarging the iris- of the subject due to a low power light beam applied) in a similar way as the flash of a photography camera works when operating in the red eye reduction mode. Nevertheless, this is not the unique light directly applied to the eye. A near-infrared (NIR) beam is also required in order to illuminate and better reproduce the texture of the eyes. This advantage may become a drawback because human eye does not detect light beam and not respond with the appropriate pupil contraction resulting in possible retina damage for a prolonged period of exposure [Sin97]. Due to this and other reasons, population feels uncomfortable using such kind of systems.

The second current biometric ocular technique use retinal scanners that scan the pattern of the blood vessels of the retina (inside the eye) by illuminating the back of the eye with an IR light beam. Thus, it refers to an intrusive method that requires a high degree of user cooperation. Additionally, this acquisition process may produce secondary effects over the subjects due to the relatively large exposition (about five seconds) to a very low powered laser IR beam becoming the major hurdle for these devices. This technology is very expensive due to sophisticated acquisition devices and not many companies are working on this technology, being EyeDentify who have the world patent for retinal scanners. Military and security applications (research labs, nuclear plants, intelligence agencies, etc) opt for both ocular technologies due to their high performance.

ON LINE SIGNATURE. Also known as *Signature Dynamics* refers to a behavioral biometric property that deals with the way one signs his/her name and has the advantage of being one of most established forms of identification in the financial sector. A pen and a graphic tablet set device or other write-on systems is used as acquisition system providing not only the stylus but also timing and spatial information about the signature as X, Y coordinates, inclination, azimuth, pressure over the tablet and also the track of the pen while is not touching the tablet (these on line features are not as easily accessible as the off line ones). Figure 2.9 shows a sample of a tablet (a) and the treatment of the related angles (b) [Ort03], whereas (c) depicts a sample of an online sign and their related computed parameters (d). A big strength of this technology is that this trait can be changed by the user when desired, for example if the signature is captured by an intruder; this is one singular aspect which is not possible with the rest of biometric authenticators.

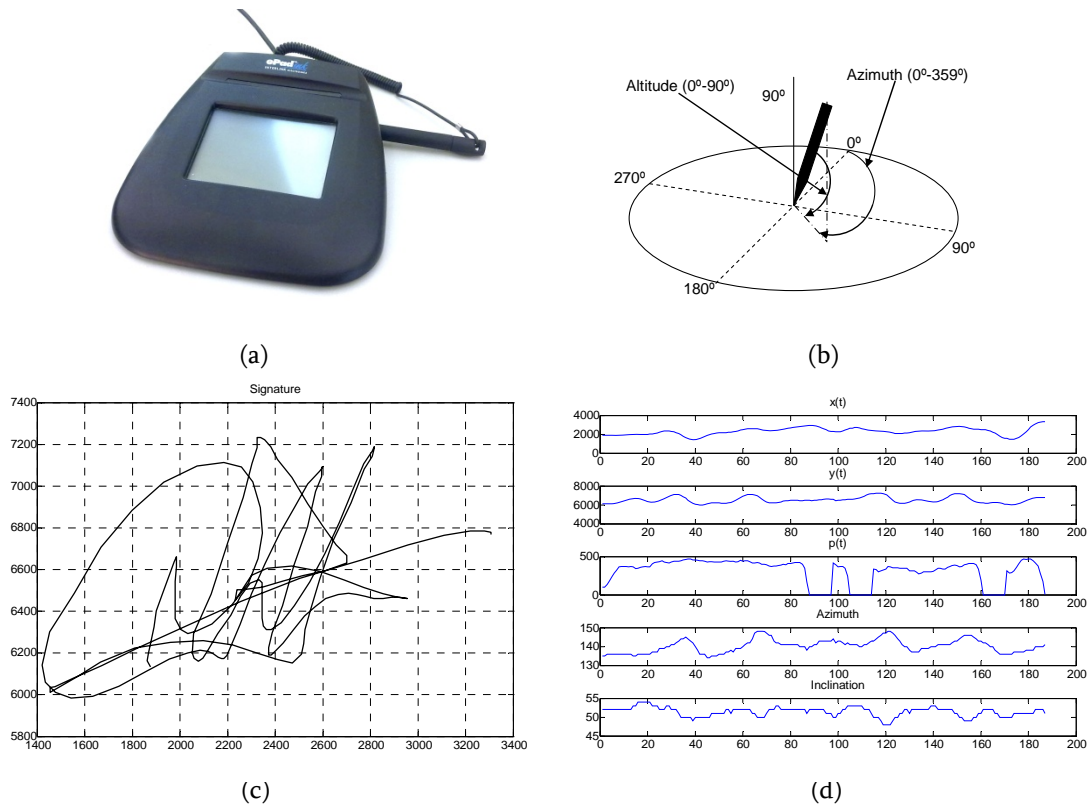


Figure 2.9: Acquisition graphic tablet provided with the customized pen (a) azimuth and altitude angles of the pen with respect to the plane of the tablet (b). (c) Off-line information: image of the written signature. (d) On-line information: pen trajectory, pen pressure and pen azimuth/altitude.

GAIT. Gait Recognition, refers to an emerging biometric modality based on a behavioral authenticator capable of recognizing people by the way of walking. The most important strength of this technology is the potential ability of providing recognition at a distance [Nix06, Zha00]. Furthermore, since gait can be captured with accelerometers (which are already integrated into most smart phones) new nice approaches are currently carried out, as the provided by the authors in [Nic11]. For the other side, the most weaknesses of this authenticator are the large variability in both short and long periods of time. Changes in gait way may be easily due to diseases, weight changes as well as changes in clothing, footwear or surface on which you walk, among others.

In order to compute the required model of the trajectories, the subject is asked to walk at a constant speed (as showed in Figure 2.10) and tracked with three different cameras. Positions of the markers are easily detected by the cameras due to its high level of reflectance. The software computes a fitted skeleton body from the extracted points in order to obtain the required trajectories.



Figure 2.10: Marker position acquisition for gait analysis. (Images analyzed by the Author for a private technical consultancy).

Regarding on the evaluation of these and other less established biometric technologies, independent testing of biometric carried out by independent agencies, are essential. Best known is the work in Sandia National Laboratories, a major US Department of Energy research and development national laboratory, which released the results of its second round of tests on biometric devices in mid-1990 [She92]. Nevertheless, to date, only limited number of additional testing of biometrics has been accomplished in order to accurately establish more real strengths and weaknesses among them.

2.4 Fusion Methods

In many biometric problems, the best performing systems use fusion methods or a combination of matchers [Zha02, Che06]. In this section, we provide an overview of several approaches in order to improve the vulnerability of a biometric system using some kind of fusion.

Usually four levels of fusion are distinguished [Das94, Jai04]:

- a) *Fusion at the Sensor (or Data) level (Low level)*. Data fusion at sensor level implies the use of more than one sensor in order to obtain a more complete knowledge of the person to be acquired. One example is to use two or more image sensors of the same kind but acquiring the 3D scene from different points of view. This example refers to the 3D model construction based on a couple of 2D images. Another example is the use of two or more sensors located in the same (or different) position. In this latter case, the goal is to overcome the limitations of a single sensor. In this dissertation, we are interested on the study of different kinds of sensors in different spectra.
- b) *Fusion at the Feature Extraction level (Intermediate level)*. Each individual feature vector is used to compute the final feature vector by concatenating all of them. Thus, this obtained feature vector has a higher dimensionality and requires some kind of feature reduction and normalization technique. This is one of the main reasons why it is not a common way to apply fusion, but it is a plausible and an easy approximation.
- c) *Fusion at the Matching Score level (Confidence, Opinion or Rank level)*. Each biometric matcher provides a similarity score indicating the proximity of the input feature vector (*template feature vector*) with the model (*pattern feature vector*). These scores can be combined to improve robustness and reliability. Several strategies exists in order to carry out this level of fusion [Fau05b]:

- i. *Fixed rules*: All the classifiers have the same relevance.
- ii. *Trained rules*: Some classifiers should have more relevance on the final result. This is achieved by means of some weighting factors that are computed using a training sequence.
- iii. *Adaptive rules*: The relevance of each classifier depends on the instant time. This is interesting for variable environments. That is: let O_1 and O_2 be the outputs of classifiers number 1 and 2 respectively. Then:

$$O = \omega_1(t)o_1 + (1 - \omega_1(t))o_2 \quad (2.1)$$

The most popular combination scheme is the weighted averaging where the weights can be fixed, trained or adaptive.

- d) *Fusion at the Decision Level (Abstract or High level)*. Decision Fusion fuses decisions coming from different experts. Majority Voting, Ranked List Combination, AND and OR Fusion are the most popular Decision Fusion methods.

On the other hand it is important not to confuse fusion methods applied to BS with a *multibiometric system* where several biometric technologies are used in conjunction with other biometrics in order to take advantage of each system and enhance the final recognition performance [Jai04], and may be use any of the discussed levels of fusion. Resource in multibiometric systems is justified in three cases in particular:

- In high-end security applications when combining several techniques reduces the risk of errors, such as faces and fingerprints. This is the currently adopted solution implemented in U.S airports and border controls (the foreigner is required to provide both face and fingerprints) and also the solution pointed out by the National Institute of Standards and Technology-NIST.
- When using an ergonomic but not sufficiently reliable technique, such as voice recognition, which is then associated with another more reliable method can take over in case of recognition failure.
- Due to enrolment problems, not all biometrics are suited to every people (Asian people for instance have a very weak fingerprints).

2.5 Technical Challenges

The classical PR tools [Dud01, The06] are designed to deal with requirements such as *small input dimensionality*, *small number of classes*, and *reasonable variability of the input*. In the case of biometric recognition these assumptions are clearly not fulfilled, becoming a complex computer vision problem. Therefore, the challenge in biometrics is to develop a machine learning methods related to new structures, and statistical analysis of the performance, from first principles, and from the experimental point of view, capable of dealing with the constraints given by practical applications: high dimensionality of the input, small number of samples per class, and a high number of classes (subjects) [Li04]. This section will appropriately deal with the set of restricted conditions.

2.5.1 Scarcity of Data and High number of Classes

One of the major challenges of Biometrics in general and of face recognition systems in particular, is to be able to identify/verify an identity when only a small amount of training data for each subject (class) is available [Lu03]. On the other hand, biometrics deals with a high number of classes and this characteristic normally degrades monotonically the between-class separation performance, due to the reduction of interclass distances. To this twofold challenge one has to add the requirements that the system is expected to have a minimum performance and to be robust to the variability of the object to be classified.

Notice, in classical PR applications, usually there are a limited number of classes and a great number of training samples. For example in the recognition of handwritten digits from U.S. postal envelopes there are 10 classes (digits) and thousands of samples per class. The situation is clearly reversed in biometrics, where we normally take three to five measures per person during the enrollment (three to five samples per class for training) and the population enrolled in the database is large (much more classes than samples per class). In this situation there are not enough training samples and classical pattern recognition approaches fail to provide a good solution [Fau07a].

2.5.2 Vector Dimension Reduction

Due to the high dimensionality of vectors in use (which also depicts high redundancy between their components), some vector-dimension reduction algorithm must be used. The relevance of feature extraction is twofold:

- The reduction on the number of data that must be processed, model sizes, etc., leads to the consequent reduction on computational burden.
- The transformation of the original data into a new feature space can let an easier discrimination between classes (subjects).

Following broad techniques for dealing with feature vectors of high dimensionality have been used traditionally [Bis95]:

- *Feature Extraction*: aims to reduce the number of data available in a different way by looking for a transformation of the original vector of data that optimizes some appropriately defined criterion of separability among classes. Its approach has the problem of mixing features that might be discriminant.
- *Feature Selection*: Is essentially the selection of a subset of original data that are meaningful to the classification by some criterion. These combinatorial methods can be computationally expensive because of the combinatorial explosion, and some of them are not capable of detecting interactions between features.

Within the context of feature approach it is also possible to deal with a combination of any of the former methods in an attempt to enhance final performance. On the other hand it is not trivial

to decide which the most suitable features are and/ or what the optimal dimension and/or what is the best p value for a given problem is. Even the best set of features for a given classifiers can be suboptimal for a different classifier.

2.5.3 The Curse of Dimensionality

The last challenge is a direct consequence of the two previously stated problems and is considered a major problem associated with pattern recognition. Over fitting and poor generalization in classifiers become more severe when dimensionality of the input is too large relative to the number of training samples. Thus, there is more than one reason for the necessity to reduce the number of features in biometrics to a sufficient minimum. Computational complexity is the obvious one. High correlation between them is another important factor to take into account. However, the major reason is imposed by the required generalisation properties of the classifier, as discussed in [Has01]. The problem at hand can be understood as a problem of function inference in a nearly empty space, i.e. fraction of samples per dimension of the input vector is less than one. In these cases the generalization capabilities of classifiers are difficult to assure, becoming a high complex computer vision problem.

It is known that the geometry of high dimensional spaces behaves quite different to the intuition of our three dimensional world. Since the function has to be defined in the complete feature space, the volume to be described grows exponentially with d , the dimension of the space. This problem is associated with one of the major problems in pattern recognition and refers to the so-called *curse of dimensionality* [Jai00, Dud01] that in principle introduces strong limitations on the performance of the classifiers.

The properties of high dimensional spaces from the point of view of distances are analyzed for instance in [Ham86, Has01, Bis95]. In the case of [Ham86] and, as just pointed out, it is shown that when the dimensionality d is high, with high probability the number of nearly orthogonal directions is 2^d , therefore classifiers based on the dot product as discrimination measure will have poor performance in high dimensional spaces. In [Bis95, Ham86] authors reveal that the fraction of the volume of a sphere of a given radius is concentrated on a thin shell of a thickness that diminishes exponentially with the dimensionality of the space.

Therefore classifiers based on the Euclidean distance will have poor performance in high dimensional spaces, because the data will be on a thin shell in a space aggravated by the fact that all the directions will be nearly perpendicular. Nevertheless these results are pessimistic in the sense that the analysis are done with the hypothesis of a uniform distribution of points in the input space. Although the performance in practice is not bad as expected from these analyses, it indicates that for classifiers based on Euclidean distances or on dot products, an approach for improving performance should be based on reducing the dimensionality of the feature vector.

The following list summarizes the main difficulties of biometric recognition problem as follows:

- Scarcity of training measurements and high number of classes. Thus, this problem will have a ratio number of faces (classes) vs. number of examples per authenticator (examples per class) extremely high, due to the cost of gathering a reasonable number of training samples per person,
- The dimensionality of the input vectors may be extremely high (due to authenticator is most of the times acquired as an image), which forces the use of traditional feature extraction techniques.

Thus, the emphasis will be on providing a robust solution capable of fulfil this strong requirements, which worsens the *curse of dimensionality*.

2.6 Ethical Aspects

In 1.800 the world human population was one billion people. During the 19th century, London became the first "world city" essentially due to the industrial revolution, experiencing a large number of migration movements and becoming the biggest and most inhabited city of the world reaching a total of 4,14 million people in 1.900. The ever increasing population and its mobility caused a parallel increasing identification systems demand. In this context, the, at first sight often reasonable security motives have led biometrics to provide new capabilities contributing over the years to increase the robustness of security systems of the modern globalised society. Similarly, and due to such as increased security requirements, the whole world has been filling with devices and data collection systems capable of tracking and recording our daily activity. Returning again to the capital of the United Kingdom, a place where identity cards are not required, we can speak of another record figure that it currently has: London has the highest density of CCTV surveillance cameras in the world [Luk07] continuously collecting sequences of all over the city. In 2012, a total of 1,85 million surveillance cameras has been reported. (-almost the same number of cameras than half of the population of London 100 years ago...-).

Beyond questions of detail, the fact is that technological advances in biometric systems, have both increased the robustness of security systems, but also provided methods highly susceptible of being able to violate the privacy³ of the people [Sch03]. As a result, we face a tangible threat of declining freedom to lead a private life. Experts in intimacy issues are reasonable worried about this loss of privacy. In particular, they raise doubts about the destination and/or the improper use of the individuals' identity. In more pessimistic scenarios, biometrics may also mean a tool capable of stigmatizing specific groups of people in a similar way as happened to Jean Valjean, the main character of the novel *Les Misérables* of Victor Hugo, an ex-prisoner with the chest marked with hot iron, as it was customary in France at that time [Cal33], who attempted along all his life to rid himself of the stigma such mark supposed. In the worst case, biometrics could come to represent the key of a new Orwellian state enabling institutions and governments to electronically control all their citizens.

³In [Jai11] a full definition of *privacy* is provided: "*Privacy is the ability to lead one's own life free from intrusions, to remain anonymous, and to control access to one's own personal information*".

On this regard, since is true, as the renowned sociologist Z. Baumann points out, that two values without which human life would be unthinkable exists, the *security* and *freedom*, is equally true that found the exact right proportion of both is not a trivial task. It is not a new issue. We are talking about the classic debate between moral right to personal privacy and security of society. In this sense, humankind has been direct witness of both sides of the coin; one the one hand, the increased of mafias, organized crime and terrorism as clear examples of global growing threats, and on the other hand and not minor crimes, the lack of essence of freedom or non-compliance with human rights all over the world ultimately.

Biometrics is located in a thin line between these both paradigms, so, a right trade-off between these security and privacy issues must be required. Sherman establishes a reasonable framework to get it by making a number of statements published in his report of 1992 [She92], summarized as follows:

To be universally accepted, Biometrics must be legally and physically robust, safe to use, not invade the user's privacy, nor be perceived as socially unpalatable. Thus certain powerful methods, such as injected radio tags or tattooed bar codes used to identify livestock and pets, are obviously unacceptable.

Once at this point is bound to ask whether will biometric be used in the future only for the intended purpose or otherwise will be used to other ones such as track people? In this respect, Weiner puts us on notice of what it has meant tip the balance towards security, especially since the S-11 attacks [Wei12]. Although the aim of this thesis is far from answering these questions, maybe it is not so far from doing this kind of reflections, which lead us to finally consider a couple of somehow deeper questions: *How pervasive surveillance should be in free societies?* or better, *How pervasive we should allow it to be?*

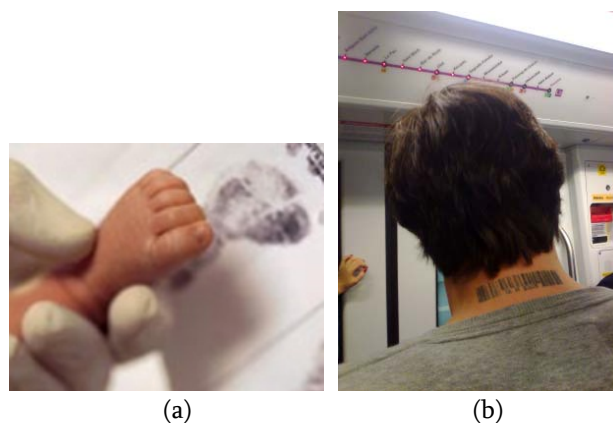


Figure 2.11: A footprint acquisition of a newborn (a) and a guy with a code bar tattooed in his neck, acquired in the underground of Barcelona last spring (b). *A fear future prediction?*

Chapter 3

Face Recognition

Nunca olvido nunca una cara, pero con usted haré una excepción.

Groucho Marx

This chapter gives an overview of face recognition (FR) systems. The first section outlines the FR problem and provides a discussion sketching the reasons for using face recognition despite the large number of drawbacks. Second section addresses both the inherent facial image properties and the processing constraints that difficult the concerned recognition task. In the third section we review the most well-known face recognition technologies and present a literature review, while in the Fourth section we focus on current landscape of Near-infrared FR approaches as well as on new trends in thermal infrared face recognition. The Fifth section and last, briefly describes the most relevant face databases distinguishing among those which only depicts visible faces from those which also depicts faces in other specific spectrums.

3.1 Problem Statement

Whatever the pattern recognition (PR) system, high interclass distances (differences between elements of different classes) and low intraclass or within-class distances (differences between elements of the same classe) are the key issue to obtain a high discrimination capability between classes. In the field of biometric recognition, large intersubject or interclass distances (between-person variability) resulting in good FAR rates whereas small intrasubject or intraclass distances (within-person variability) will make more feasible to obtain a stable model for a given person, also improving FRR.

In the specific field of facial biometric recognition, the performance of the systems has improved significantly in recent years since they were conceived at the beginning of the seventies with the first automatic FR system developed by Kanade [Kan73]. However, high recognition performance is only achieved in fully constrained conditions. When dealing with

more real situations (without no limiting side effect views, lighting conditions, aging, changes of look, blocking effects, cluttered backgrounds...etc), high intersubject scatter and specially low intrasubject scatter becomes a chimera and performance degrades drastically. The pictures collected in Figure 3.1 summarize the mentioned situation. Figures 3.1(a) and 3.1(b) show a sample of two different couples of people with a very similar appearance between them (my parents vs Errol Flynn and Gina Lollobrigida) exhibiting the referred low intersubject scatter situation, that Galton and Darwin knew well [Wpe51], whereas 3.1(c) depicts a couple of twins as the well known extreme case of different characters highly correlated among them, that directly drives the automatic FR to an unapproachable task. On the other hand, Figure 3.1(d) exhibits a set of different photographs of the same subject, showing a very low likeness level between them. This second sample showcases the undesirable high intrasubject scatter discussed and also denotes that this variability is highly nonlinear (NL) and nonconvex (implying non linear separability among individuals) [Li04], limiting the capability to achieve accurate FR systems in real conditions.

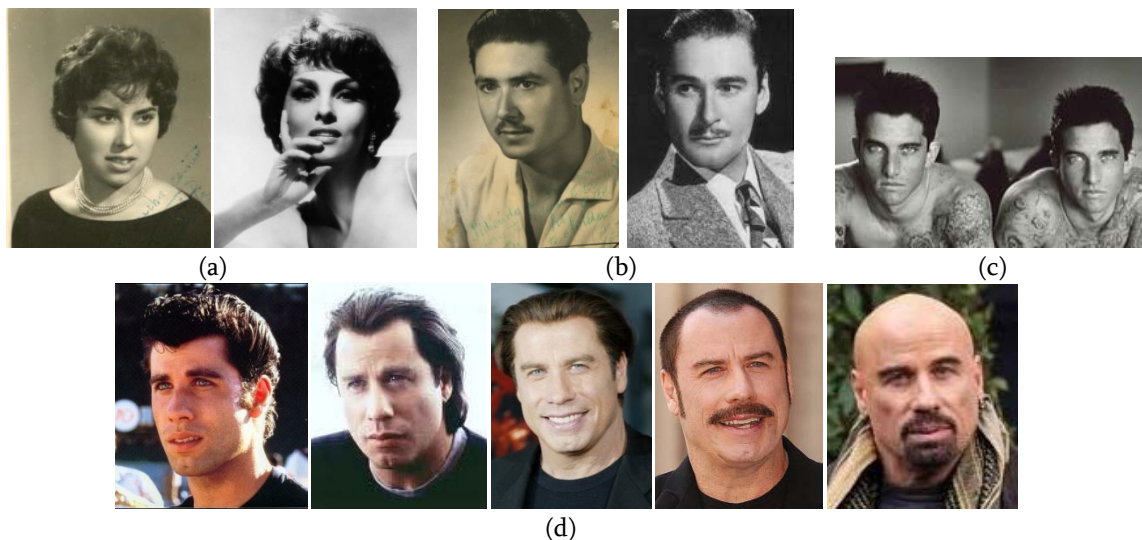


Figure 3.1: (a,b) Sample of two couple of different people much alike between them. (c) A pair of tattooed twins (photographed by Ken Probs) as an extreme case of low intersubject distance. (d) Set of different pictures of the actor John Travolta with a very low resemblance, conducting to a very high intrasubject distance.

In addition to the above, face is not only a biometric feature, but also an important social element subjected to a great number of social canons, such as fashion (that can both, further increase intrasubject variations over time and decrease the intersubject distances among contemporary people), aesthetic canons (in some cases even aided by cosmetic surgery), urban tribes, androgen looks, etc, which also contribute to decrease the differences between people, demanding even more powerful FR algorithms for finally distinguishing among individuals.

Moreover, faces are also subjected to be disguised in order to camouflage the identity or even, for looking like some other person, increasing the recognition system requirements. A sample

of the effects of these factors over faces is shown in Figure 3.2. Although all are extreme cases, is interesting to take into account how accuracy level may be currently achieved in this issue. Figure 3.2(a) exhibits a funny processed picture that shows a scene full of different people that are, in fact, the same person, whereas Figure 3.2(b1, b2, b3) showcases an amazing facial transformation of an ordinary man alike a famous woman just making-up him, revealing how makeup can gives anyone a variety of different faces [Auc00]. Finally, the author has also funny enrolled this practice, disguising as an invented character as showed in Figure 3.2(c1,c2).

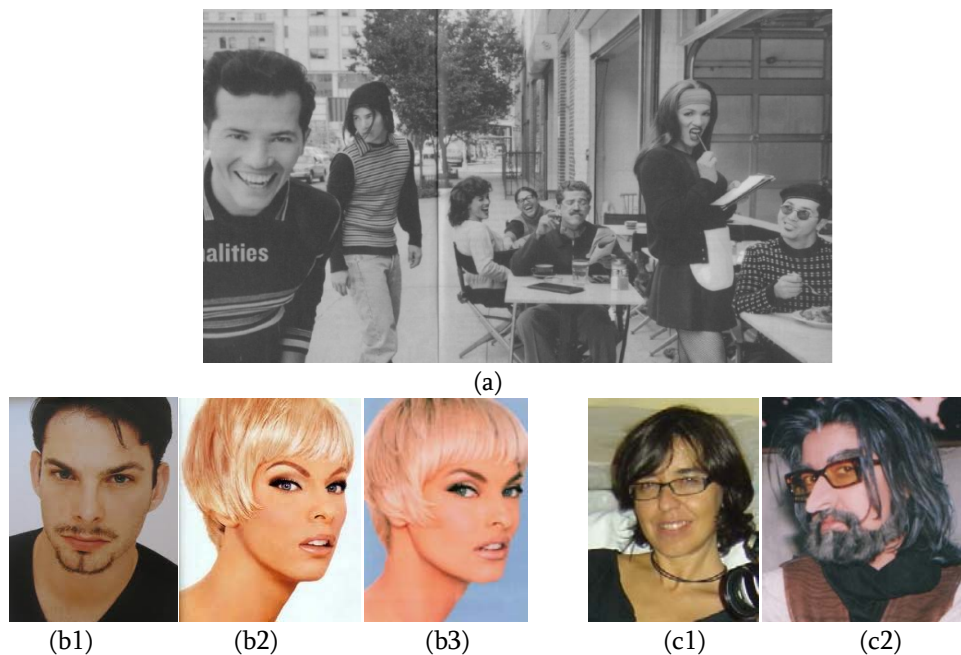


Figure 3.2: (a) Picture called “*Multiple personalities*” where all people are the same person (published in NY Times magazine, Section 6, pp 48-49. September 1, 1996). Anonymous guy (b1) characterized as a famous international model (b2) (make up by the great make-up artist, Kevyn Aucoin), and the true model (b3), Linda Evangelista. The author (c1) disguised as a curious aged man (c2).

Apart from to seek to maximize the between-person variability, and perform templates that reduce within-person variability, there is the fact, as we pointed out in Section 2.5.3, that FR is of this kind of problems where the dimensionality of the input is strongly high, the number of available faces by people are extremely low (10 in some cases, being indeed just one in the worst cases) and the face databases present a large number of different classes (different people). This last set of restrictions leads FR to deeply fall into the curse of dimensionality, driving this kind of BS to a suboptimal ability to generalize. In this overall context, FR remains a complex pattern recognition challenge and major problems are largely unsolved [Li04, Zou05]. Thus, this accurate problem statement, immediately drive us to ask ourselves: *Why use faces as biometric recognition?* The answer is easy: FR becomes interesting because it has the capability to overcome many of the involved biometric problems present in other biometric systems. The major ones may be summarized as follows:

a) About the Face:

- It may be imaged at a distance (allowing human identification at a distance).
- It can be used for aliveness control, since only live people have faces that present variability in consecutive acquisitions (for instance, deformations due to emotions).

b) About the FR BS:

- Presents a widespread acceptance of the population due to basically two factors:
 - FR is the most common way to identify people between them.
 - Is a totally contact-less and non-intrusive biometric recognition system, implying no any direct contact with any acquisition device and no any derived contagion among individuals (note that many people are very protective of their personal space and the idea of this human-device interaction breaking into that space is often quite unacceptable).
- Is easy to set up, particularly for registering.
- Is an affordable system.
- Collaboration with the subject is not required in both biometric stages, acquisition and identification:
 - Enrollment can be done from live images, but also by means of photographs (from passports, ID cards, driver licenses...etc) or even picking out from previous video recordings.
 - In the identification, stage collaboration is either required, since faces can be remotely acquired. In this way, the subject is unaware he is being acquired by the system, making even possible the identification in a camouflaged way (covering the cameras, for instance).

These last two benefits¹ are the most important ones, driving this technology to one of the most appropriate technologies for security and law enforcement applications in conjunction with the use of fingerprints and the most suitable one for surveillance purposes (where the subjects are not-cooperative in nature) and for applications based on human-computer interaction.

While there are exceptions, FR methods are purely two-dimensional (2D) in the sense that subjects are enrolled into the system using 2D images, and the matching process is also carried out using 2D images. In addition, such FR methods can rely on single still images, multiple still images, or video sequence [Kev04]. Although traditionally most efforts have been devoted to the former one, the latest ones are quickly emerging [Zho04], probably due to the reduction of price in image and video acquisition devices. In this context, acquisition by means of video camera, can make possible the development of a multisample biometric by using a sequence of

¹Although gait recognition may also be acquired at distance, as discussed in Section 2.3, this BS always requires a sequence with the respective tracking system (not only a still image to carry out the identification as in FR case). In addition, is not usual to dispose of the required enrollment sequences.

images provided by the system, where the recognition will rely on a set of images, rather than on a single one [Esp04b, Fau05] (see Figure 3.3) or even, address Audio-Visual biometrics, where the audio signal (-speaker recognition-) will be computed jointly with the image signal (-face-) [Pet06, Haz03] and/or the video sequences of the lips area [Jou97] in order to improve the reliability of the final identification system. If a high resolution NIR camera is available as pointed in Section 2.3, a multibiometric system deploying face and iris with collaborative users can also be performed as developed by Wang et al. in [Wan10].

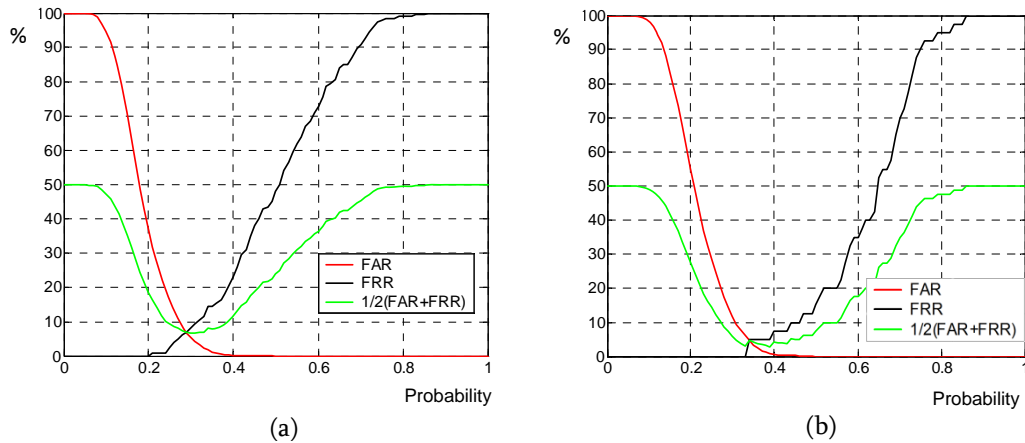


Figure 3.3: Differences between face verification developed system using a single snapshot, and using the combination of 5 different ones respectively. The results were evaluated using the FAR & FRR indicators. Note that the threshold value is less critical for getting a good trade-off among both magnitudes when using five trials (the plots are more separated and specifically FRR curve significantly improves).

3.2 Constraints and Challenges

In order to be able to expand FR towards less restricted conditions, a comprehensive study of the different kind of factors that may degrade facial image as well as the concerned processing constraints that also contribute to difficult the final performance, will be previously done. From this point of view, is stated that faces are subject to a large set of different natural and deliberated sources of variability, arising from lighting conditions, aging, different poses and facial expressions and oclusions as the most important ones [Kum06]. The different sources responsible of variations in the facial appearance that will affect in different levels the FR system performance can be broken down into two types: *intrinsic factors* and *extrinsic ones* [Jaf09, Rom06]:

- a) ***Intrinsic factors*** are due to the human nature and are responsible for both, the differences in the facial appearance of the same person (*intrasubject factors*) and the variation in the facial appearance of different subjects (*intersubject factors*). Some examples of intrasubject factors are growing and being age, degenerative diseases, facial expressions, blocked effects and facial look (glasses, cosmetics, hairstyle and facial hair, tattoos, piercings...etc, which also produce oclusions in major cases), whereas sex, race and identity are clear examples of intersubject factors.

- b) ***Extrinsic factors*** are directly related with the acquisition conditions (lighting conditions, distance and camera viewpoint, resolution and noise introduced by the image sensor, blur effects...etc) and also contribute to increase the differences in the facial appearance of the same person (*intrasubject factors*). Most of these factors will be viewed in the following chapters paying special attention on the defocusing effect.

A brief review of the five most prominent factors pointed in the above classification will be done.

Facial Expressions. Variations produced by this factor are due to the inherent dynamic nature of the faces and arise in daily life during interactions and conversations introducing distortions regarding the neutral expression. Thus, as each basic facial expression is associated to a finite combination of facial muscles there is the possibility of finding invariants and transformations which might help FR systems. In this sense, although FR research techniques have traditionally considered facial expressions as a distortion that negatively influences recognition ratio, we should also take into account that expressions contain reliable information which is used by humans for identifying subjects. In addition, a face will be easier identified at a given local scene, depending on the emotion currently expressed [Mar00]. Emotional facial expressions by its side, refers to a specific subset of the general human expressions. According to Ekman and Friesen [Ekm78] there is a set of only six² basic emotions to which are associated “*six universal facial expressions*” representing *happiness*³, *sadness*, *anger*, *fear*, *surprise* and *disgust*.

Growing and Aging Process. As is well known, facial appearance experiments a gradual non linear variation over time due to child growing and adult aging of the individual as well as other related factors such as health, life style, sun exposure, genetic ones, etc. Figure 3.4 shows a collection of pictures of an individual during the first stage of the human life (since baby to eighteen), revealing the shape change of face profile as the most prominent factor. In addition, it is very difficult to collect face images of the same person over a long time period, and the age related variations are often mixed with other variations due to some other factors [Suo09]. Due to the inexistence of face datasets over a long term of the humans, first current solutions pointed out to generate synthesized face aging datasets. The most well-known one is the FG-Net aging DDBB that produces artificial faces with different levels of wrinkles and facial skin firmness by means of projections of synthesized feature vectors in shape and texture eigenspaces [Wan06]. Nevertheless, although some synthesized face aging datasets exists, there is a lack of quantitative measurements for evaluating the aging treatment in the literature [Suo09].

²Whereas biometric DDBBs present a large number of different classes (1 class by each person), DDBBs of emotional facial expressions just require a limited number of classes (6). In addition, for emotion recognition, a small variation among different snapshots of different people presenting the same emotion will be desired.

³Human beings are the only species in the world that can smile [Rod10b].



Figure 3.4: Set of pictures of the author's nephew, since baby to eighteen revealing large variations in shape face over time.

Blocked Effects. Also known as *occlusions* refers to any usual or unusual objects (outside ones or of the own subject) that partially or fully blocks face image limiting *face-at-a-glance* scenarios. Accessories such as glasses and sunglasses as well as head wearing objects as hats, caps, scarfs or helmets and burkes as the most extreme ones, are the most usual objects that causes this effect.

When the effect is produced by the same subject by means of their own body (long hair and hands basically), then is called *autoblocked effect*. Notice that apart from using hands to eat, drink, smoke or talk by phone, among many other daily activities, they also respond to a very powerful tool for communicating. Hands combined with faces are many times used in communication to express himself, make signs, wave, etc. In addition, deaf language fully fuses face and hands to build the communication system making almost impossible recognize individuals of this collective while they are communicating. Figure 3.5 depicts different types of face degradation due to the blocking and autoblocking effects produced by different unusual (a) and usual (b, c) objects.



Figure 3.5: Occlusions given by different unusual (a) (waterpipe, camera, gun, microphone, and glass of wine) and usual (b) objects.

Viewing Geometry (also known as camera viewpoint). Figure 3.6(a) shows an image of a subject that is simultaneously viewed from frontal and profile points of view, revealing no any practical difference between both viewpoints. Unfortunately, this is a strong exception of the reality. Viewing geometry, *facial pose* or equivalently, *camera viewpoint*, is the concerned factor that introduces different levels of deformation with respect to the frontal view due to the projections, shadows and self blocking. Slight changes in facial pose may lead to large changes in the facial appearance. In this respect, rotations over 30° begin to be very difficult to solve (See Figure 3.6(b)).

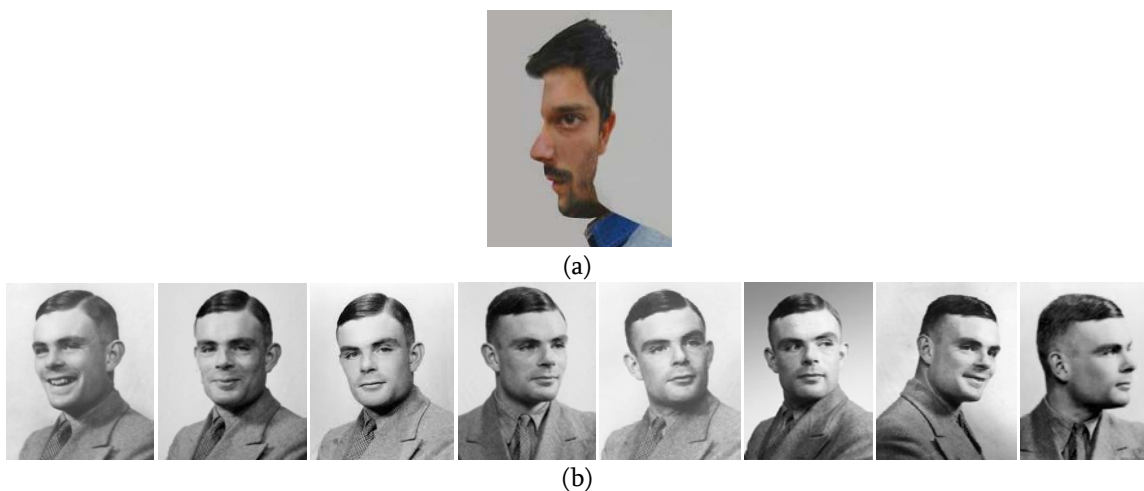


Figure 3.6: (a) Curious frontal and profile face. (b) Full set of different faces of the same subject (Alan Turing) with different pose taking along the same session; Different levels of occlusions are given in rotations over 30° and profile views.

Lighting Conditions. Since the radiance sensed by a camera at a given face location is proportional to the product of face reflectance (also known as *albedo* as we will further see in the next Chapter) and incident light, the effect of such lighting on face images can lead to one of the strongest distortions in facial appearance. Thus, the overall set of lighting conditions over a given face (quality, changes in location and intensity of the lighting source and other related pattern distortions) introduces diverse linear and nonlinear changes to the image content. The problem is aggravated by the fact all lighting parameters are difficult to control indoors and fully uncontrollable outdoors. Actually, a very careful study led by Moses et al. [Mos94] concluded that the variability of the faces of the same subject taken with different illumination conditions can be much larger than the variability among different individuals captured with same neutral expression. The changes in illumination on face images are dramatically illustrated in Figure 3.7.

Additionally, strong lighting conditions may produce specularities, changes in color and self-shadowing. This last effect is due to the hypothesis of a *lambertian surface* reflectance (further discussed in Section 4.2.3) is violated for human face [Bel97, Zou05].

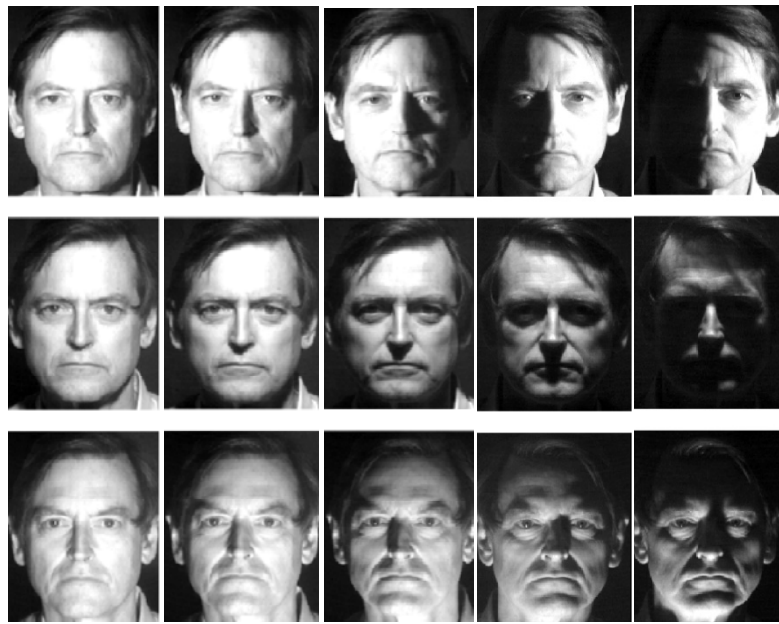


Figure 3.7: Faces of the same subject under varying light conditions [Harvard face DDBB].

As stated at the beginning of this section, irrespective to the described intrinsic and extrinsic factors, other important aspects contribute to harm the final performance of the FR system. *Face detection* and *segmentation* are fundamental issues, which are often prior steps within recognition systems, are known to be key problems in building automated systems that perform FR. While the first one is intended to detect whether a face exists in the acquired image, the second one is focused in extracting the locations in images where presence of faces is known in advance [Yan01, Jai04]. This task becomes easily highly difficult when dealing with difficult scenarios such as those with similar object and background colors and/or variable background conditions. Furthermore, faces are active three-dimensional objects whereas face images are 2D projections of a three-dimensional world. This fact result in another important constraint in 2D FR due to the high degree of variability that exhibits the 3D real structure of a face when represented in a 2D way (such as a photograph) even, in frontal views. Fingerprint acquisition is simpler, because we are always getting the ridge and valley pattern over the sensor surface. Thus, there is not zoom/ panning nor 3D rotations, nor shadows nor objects blocking (hair, glasses, etc.) [Fau05a]. Figure 3.8 tries to show this last consideration.



Figure 3.8: A simultaneously image with a real 3D subject and its 2D projected face (In this picture they are both actually, 2D subjects).

For concluding this first part, the key technical challenges when dealing with this faces for FR tasks are provided:

- Faces to be classified may change smoothness or even strongly with time (i.e. surgery, accidents...) or/and between recording sessions (i.e. lighting conditions, supervised/non supervised acquisition process, camera viewpoints, different scenarios ...).
- Faces, as complex and variable 3D structures, strongly difficult the development of the computational models.
- FR approaches suffers seriously when presented with face images that have pose and illumination variations.
- The variability of the facial patterns with time, leads the need to refresh the facial databases every certain period of time.
- Fail to discriminate monozygotic (identical) twin⁴ brothers.
- Difficulty of gathering a high number of training examples per person.

The above mentioned conditions are challenging when high performance in classification is desired. Classifiers that have high performance on a certain characteristic might behave poorly because of its structure in another situation.

3.3 Face Recognition Technology

Face recognition is the process by which the brain interprets and identifies and/or verifies human faces. A FR system by its side refers to a computer vision application for automatically recognizing an individual from images by matching probe images against the according set of previously recorded images, available in a DDBB. In this effect, for a computer, a human face is just a map of pixels of different gray (or color) levels, and accordantly with this computerized point of view, identify (or verify) a person, will imply to represent his facial image in an intelligent way, as a feature vector of reasonably low dimension and high discriminating power. Developing this representation is the main challenge for automated FR researchers and developers.

Along Sub-section 3.3.2 we will make an overview of the current state of the art major techniques in FR in the visible spectrum, while in following section we will specifically address the concerned literature of this issue over the infrared spectrum. Additionally, before going in detail towards Face Recognition techniques (FRT), some facts of how the problem is performed by humans will be also provided as a starting point, in order to better understand the complexity of FR. This task will contribute to the fitter correspondence among the cognitive-physical implementation of the human beings and the computational models [Sin06]. An interesting reflection about the hardness of this key issue is pointed out by Copeland and Proudfoot in their report “*Alan Turing's Forgotten Ideas in Computer Science*”[Cop99]:

⁴By definition, identical twins cannot be distinguished based on DNA, hand geometry or face. On the contrary, they can be positive discriminated by means of their fingerprints, iris and retina ocular patterns.

We can understand digital computers are high performance number crunchers. Ask them to predict a rocket's trajectory or calculate the financial benefits for a large multinational corporation, and they will give us the answer in seconds. But seemingly simple actions that people routinely perform, such as recognizing a face or reading handwriting, have been devilishly tricky to program. Perhaps the human brain has a natural ability for such tasks that standard computers lack. Scientists have thus been investigating computers modeled more closely on the human brain.

3.3.1 Face Recognition by Humans

Babies may identify their mother's within half an hour after birth and mirror self-recognizing between 18 and 21 months later [Nie04], fully finalizing the overall FR learning process during the teenage stage. Human beings achieve *perceptual constancy* –the correlation of all the different appearances, the transforms of objects– very early, in the first months of life. This amazing human ability⁵ achieved smoothly and unconsciously, constitutes a huge learning task [Wen99] that has fascinated philosophers and scientists for centuries such as Aristotle and Darwin. Currently FR has also become an interesting issue for both neuroscientists and engineers dealing with pattern recognition. However, although many research efforts have been made till date, the enormous complexity of such FR activity that human beings perform every day is scarcely realized. In this sub-section we appropriately review the major interesting aspects of this issue.

A long standing focus of research in psychophysics, physiology and psychology fields reveals that the human visual cortex perceives by decomposing the visual stimuli in a number of frequency bands which are also dependent of the spatial resolution. Experimental results in [Nac75] indicate that these frequency bands have the approximate bandwidth of an octave. The similarities between the mechanisms with which human visual system treat the respective visual stimuli and the preprocessing techniques such as the wavelet transform (that split the signal into various spatial frequency bands), justify the use of this latter approach in FRT [The06]. On the other hand, research in human perception and memory centers on the importance of the *average* or *prototype* in guiding recognition and categorization of stimuli. The theory is that categories of objects, including faces, are organized around a prototype or average. The idea is that the closer an item is to the category prototype, the easier is to recognize as an exemplar of the category. However, in biometric applications, the goal is not to detect a face in an image [Zha06]. The authors of [Lig79] found that faces rated as “typical” were recognized less accurately than faces rated as “unusual”; unusual features in these faces makes the person less confusable with other faces, and somehow or other “more like themselves” [Zha06]. We would expect that the recognizability of individual faces should be

⁵The referred perceptual capabilities are not only human properties. Animals are also capable to recognize individuals. Animals as dolphins are even capable of recognizing themselves when they see themselves as explained in [Phi09]: Ric O'Barry reports the case of Flipper, a dolphin in captivity famous for being the actor of the flipper serial movie in the sixties. He could recognize himself among other dolphins when he saw the serie in a TV specially adapted for him. Penguins by their side, fully exceeds human identification abilities: Their females are capable of recognizing their baby animals without never have seen them before are even among thousands of individuals.

predicted by the density of faces in the neighbouring face space. We might also expect that the face space should be more dense in the center near the average. The space should become progressively less dense as we move away from the average. If a computationally-based face space approximates the similarity space humans employ for face processing, we might expect that “typical” faces would be near the center of the space and that unusual or distinctive faces be far from the center. It follows, therefore, that computational models of FR will not perform equally well for all faces. These systems should, like humans, make more errors on typical faces than on unusual faces.

Extremely related with such discussion, is also a well known fact that human beings are almost unable to distinguish faces of other races. For instance, for a Westerner, all the oriental faces look the same, but what really induces the *cross-race effect* (CRE) is still a mystery. M. Bernstein et al. point out in [Ber07] the origin of the effect may arise from our tendency to categorize people into in-groups and out-groups based on social categories such as race, social class, hobbies, etc. Whatever the reason is, this fact can lead e.g. to reiterative eye-witness misidentifications in police issues, which is an important problem [Rut01]. Who don't remember the funny expression said by a Chinese in a joke context for exonerating: *All Chinese look alike...?*

Apart from cope with different races, some other everyday situations might drive to poor FR performance such as when dealing not directly with faces but also with 2D photographs. In [Bar83], a British anthropologist named Nigel Barley reported an interestingly situation, when he lived with a strangely neglected tribe of the North of Cameroon called Dowayos:

" ... In the end I managed to lay my hands on some postcards depicting African fauna. I had at least a lion and a leopard and showed them to people to see if they could spot the difference. Alas, they could not. The reason lay not in their classification of animals but rather in the fact that they could not identify photographs. It is a fact that we tend to forget in the West that people have to learn to be able to see photographs. We are exposed to them from birth so that, for us, there is no difficulty in identifying faces or objects from all sorts of angles, in differing light and even with distorting lenses. Dowayos have no such tradition of visual art; theirs is limited to bands of geometric designs. Nowadays, of course, Dowayo children experience images through schoolbooks or identity cards; by law, all Dowayos must carry an identity card with their photograph on it. Inspection of the cards shows that often pictures of one Dowayo served for several different people. Presumably the officials are not much better at recognizing photographs than Dowayos... The point was that men could not tell the difference between the male and the female outlines. I put this down simply to my bad drawing, until I tried using photographs of lions and leopards. Old men would stare at the cards, which were perfectly clear, turn them in all manner of directions, and then they say something like "don't know this man"...

In addition, some experiences described by neurologists about people with sight restoration after long-term blindness, are also really interesting [Sac96]:

“...an English surgeon removed the cataracts from the eyes of a thirteen-year-old boy born blind. Despite his high intelligence and youth, the boy encountered profound difficulties with the simplest visual perceptions. He had no idea of distance, space or size. And he was bizarrely confused by drawings and paintings, by the idea of a 2D representation of reality. As he had anticipated, he was able to make sense of what he saw only gradually and insofar as he was able to connect visual experiences with tactile ones.

...[Observing a face]: He had no idea what he was seeing. There was light, there was movement, there was color, all mixed up, all meaningless, a blur. Then out of the blur came a voice that said, well? Then, and only then, he said, did he finally realize that his chaos of light and shadow was a face...sometimes he would get confused by his own shadow (the whole concept of shadows, of objects blocking light, was puzzling to him) and come to stop, or trip, or try to step over it... Moving objects presented a special problem, for their appearance changed constantly. Even his dog, he told me, looked so different at different times that he wondered if it was the same dog”.

For the rest of us, born sighted, is difficult to imagine such confusion. For us, born with full complement of senses, and correlating these, one with the other, create a sight world from the start, a world of visual objects and concepts and meanings. When we open our eyes each morning, it is upon a world we have spent a lifetime learning to see. We make our world through incessant experience, categorization, memory, and reconnection.

On the other hand, many cognition problems may also drastically reduce this natural capability in performing faces recognition. The acquired or hereditary disease called *prosopagnosia* (also often called *Face Blindness*) reveals this inability to recognize familiar and unfamiliar human faces [Sac87]. An extreme case of this disease is done by patients who hardly recognize himself or even, people who do not recognize themselves at all⁶ [Beh05]. By contrast, some prosopagnosic patients seem to recognize facial expressions due to emotions. Interestingly and according to the neurologists, another kind of patients who suffer from organic brain syndrome does poorly at facial expression analysis but can perform FR quite well.

For finalizing this first sub-section is also important to report that contemporary studies in cognitive psychology have suggested that humans learn to recognize objects (e.g. faces) using positive examples without the need for negative examples [Lee99]. This corresponds to a *generative* method. The statistical PR analogous version exists, known as *generative* (or *informative*) approach, such as e.g. Hidden Markov Models (HMM), where an estimation of a probability distribution is pursued. Accordantly to this, may be the other existing statistical approach known as *discriminative* approach that concentrates in founding a decision boundary for discriminating between client / impostors (such as SVMs) could not be the best choice for FR from the perspective of modeling the human performance.

⁶An expert in paintings contacted with Picasso in order to verify the authorship of some paintings apparently of him, but the master determined they were all false copies. When the expert explained to Picasso they were all certified Picasso ones, the master just concluded by saying: *“But I also paint fake ones”*.

Bearing in mind these and other considerations related with FR by humans might reveal some possible effective FR strategies. In fact, early research into automated FR led by Bledsoe [Ble66] was inspired by the ability of humans to recognize people from photographs (hopefully was not Downyos ones). Actually, this project was labeled as *man-machine* since human beings extracted the coordinates of a set of concerned features from the photographs and computers performed the recognition task. Currently, what biometrics are trying to do is to automate this amazing recognition process we perform every day in a natural and easy way, and hopefully to get away from human errors. However, it is an achievement that even the largest supercomputers cannot yet match with a performance comparable to humans, both in accuracy and speed.

3.3.2 Holistic Approaches: Projections, Transformations and Classifiers based Techniques.

The existing approaches for FR may be classified into two categories according to feature extraction for pattern modeling: *Holistic* and *Analytic ones*. The holistic approaches consider the global properties of the pattern, whereas the second one considers a set of geometrical features of the face. The following outlines both approaches.

- a) *Holistic (Statistical or appearance-based)* approaches consider the image as a high-dimension vector, where each pixel is mapped to a component of a vector (see Figure 3.9) in a multidimensional space. Due to the high dimensionality of vectors some vector-dimension reduction algorithm must be used. Typically the Karhunen Loeve expansion applied with a simplified algorithm known as *eigenfaces* [Tur91] is the most prominent global approach⁷.

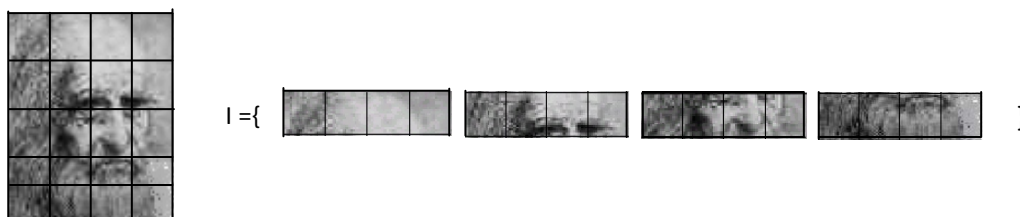


Figure 3.9: Facial image rearranged as a vector. Each pixel responds to a coordinate in the high-dimensional image space.

- b) *Analytic (or geometry-feature-based)* methods are focused on detecting the position and relationship between face parts, such as eyes, nose, mouth, etc., and the extracted parameters are measures of textures, shapes, sizes, etc. of these regions. The set of all normalized size and distance measurements constitute the final feature vectors that will be used for a direct recognition approach. Areas and features that are subject to modification such as facial hair and hairstyle are not taken into account. In such kind of methods a reliable detection of facial features is fundamental to the success of the

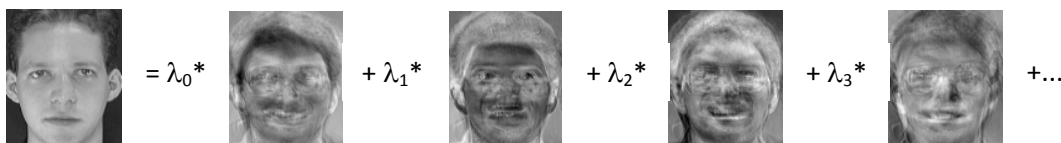
⁷Holistic feature vectors are often also used to represent both two-dimensional (2D) and 3D face images.

overall system; so, they cannot manage with strongly occluded faces. A number of early FR algorithms are based on this approach [Kan73, Man92] and since then many methods have been developed to detect facial features [Zha03].

Thus, Kanade's and Turk and Pentland's approaches reflect these two extremes in solving this problem. Nevertheless template-matching approaches appear to have better robustness and achieve better results as concluded Gutta et al. in their paper about benchmark studies on FR [Gut95] and also previously Brunelli and Poggio in their comparative between features versus templates powerful [Bru93]. In addition, since is difficult to locate and extract features of a facial thermogram (especially eyes), approaches using geometrical features are not usually extended to thermal face images [Sel01]. Thus only statistical approaches as common choices will be further discussed. Thus, the following is a brief introduction of most prominent visible FR holistic algorithms.

As mentioned in the above a) paragraph, high information redundancy presented in images results inefficient, when these images are used directly for recognition. In this respect, in order to keep the number of features reasonably small alleviating this way classification tasks and the amount of required runtime, statistical features extraction has been widely driven by algebraic linear methods such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA), among the most representative ones.

The Eigenface method of Turk and Pentland [Tur91] is one of the main FR methods applied in the literature which is based on the Principal Component Analysis. Their study was motivated by the earlier work of Sirovich and Kirby [Sir87] who invented eigenfaces in 1987 as a way of providing a low dimensional representation for human faces. Eigenfaces are linear approximations to face space that represent faces as a linear deviation from a mean or average face. The proposed eigenface system projects face images onto a lower feature space that spans the significant variations among known face images by using PCA performed by the Karhunen-Loève Transform (KLT). The significant features are known as eigenfaces, because they are the eigenvectors (-the principal components-) of the set of images. The projection operation characterizes any individual face by a weighted sum of the eigenface features, and so to recognize a particular face it is only necessary to compare these weights to those of known individuals. The geometric interpretation is that each face is approximated by a linear combination of eigenvectors of the new low dimensional image subspace, called *face space*, which corresponds to a subspace of the original image space. Figure 3.10 shows the geometric interpretation of an original image projected onto the new face space.



$$\text{Target Face} = \lambda_0 * \text{Mean Face} + \lambda_1 * \text{Eigenface}_1 + \lambda_2 * \text{Eigenface}_2 + \lambda_3 * \text{Eigenface}_3 + \dots$$

Figure 3.10: Face image expressed as a linear combination of EigenFaces.

After PCA, many other algorithms have been proposed for further improvements. The first alternative approach to PCA is *Fisher's Linear Discriminant* (FLD), first developed by Robert Fisher in 1936 for taxonomic classification [Fis36] and better known as *Linear Discriminant Analysis* (LDA). This transform computes the eigenvectors of the product $S_W^{-1}S_B$ (being S_W^{-1} the inverse of the within-class scatter matrix, and S_B , and the between-class scatter matrix), and selects the basis vectors better suited for increasing the separability among classes by maximizing the ratio expressed in (3.1):

$$\frac{|\Phi^T S_B \Phi|}{|\Phi^T S_W \Phi|} \quad (3.1)$$

where Φ is the linear subspace and S_B and S_W are defined as:

$$S_B = \sum_{i=1}^m N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (3.2)$$

$$S_W = \sum_{i=1}^m \sum_{x \in X_i} (x - \bar{x}_i)(x - \bar{x}_i)^T \quad (3.3)$$

where m is the number of classes in the database, N_i is the number of samples of class i , \bar{x}_i is the average class i , and \bar{x} is the mean of all classes. As such, the main difference between these both similar mathematical discussed approaches is while PCA focuses on maximizes the total scatter across all classes (inter and intraclass), producing a set of *Most Expressive Features* (MEF), LDA specifically attempts to minimize intraclass scatter and maximize interclass scatter producing the consequent set of *Most Discriminating Features* (MDF) [Hal99]. With this regard, although the PCA projection is well-suited to object representation, features selected are not necessarily good for discriminating among classes [Swe96, Hal99]. The usefulness of finding such projections can be seen in Figure 3.11, where an example of a two dimensional classification problem is given.

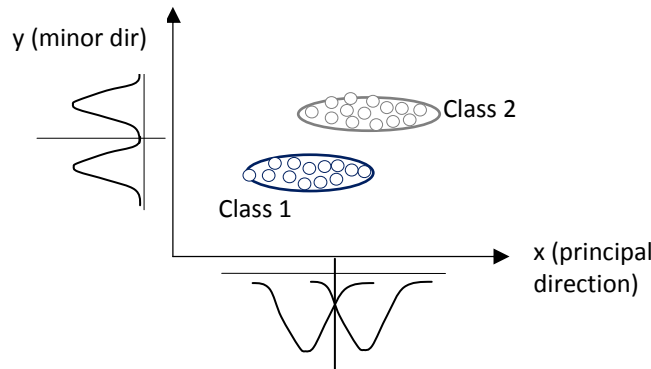


Figure 3.11: Classical graphical interpretation [Bis95] where the projection on the fisher direction, which is vertical, clearly shows the clustered structure of the data, whereas the projection on the PC (horizontal) contains most of the energy of the signal (the maximum variance of the data is in this direction), fails to classify this structure.

However, LDA copes with a difficulty when applied to FR problem: the within-class scatter matrix, becomes singular⁸ when dealing with high dimensional vectors which is the case of faces. In order to overcome such constraint, most methods such as the suggested by Belhumeur et al. in [Bel97] first perform a dimensionality reduction using PCA for avoiding the singularity of S_w matrix and then apply LDA to further reduce the dimensionality. The recognition is then accomplished by a Nearest Neighbor (NN) classifier. This method has been called by the authors as *Fisherfaces*. Nevertheless, when applying PCA in first place some discriminative features may be replaced. Then, in order to avoid this new shortcoming, many other exist. In [Che00, Yu01] the Direct LDA (D-LDA) method is applied that consists in simultaneously diagonalize S_B and S_w matrices, firstly removing S_w null space (without discriminative properties), but retaining S_B null space.

Lu et al. propose a new approach in [Lu03a] called DF-LDA that combines D-LDA approach with the strengths of the fractional-step LDA (F-LDA) method (which is not usually used in high dimensional problems) achieving high performance as will be specifically reported in Table 3.2.

Etemad and Chellappa deal with the scarcity of the learning samples [Ete97] by adding new samples from the original given dataset, by computing the mirror image of all the original faces and also performing a set of new noisy but of reasonable versions of available examples.

For closing LDA approaches review, just to point out that such LDA based systems (also PCA ones) assume the existence of an optimal projection that projects the face images to distinct non-overlapping regions in the reduced subspace where each of these regions corresponds to a unique subject. However, in reality, that assumption may not necessarily be true since images of different people may frequently map to the same region in the face space and, thus, the regions corresponding to different individuals may not always be disjoint. Some authors are recently introduced some variants in order to deal with the concerned problems. A local technique based on LDA, called *Locality Sensitive Discriminant Analysis* (LSDA) is implemented in [Cai07]: LSDA overcome LDA limitations when little samples are available.

Accordingly to the overall above discussion, is not strange that first range of variable exploratory multivariate techniques is still dominated by essentially algebraic second order methods based on the Singular Value Decomposition (SVD) and nonparametric variants of such methods. Indeed, such methods have been widely used for dimensionality reduction and feature extraction by some famous commercial FR products along the last years. Nevertheless, as second order statistics provide only partial information on the statistics of both general images and facial ones, it might become also necessary to incorporate higher order statistics [Liu00]. More recent approaches in appearance-based FR methods include more sophisticated learning methods. *Independent Component Analysis* (ICA) [Bar98, Bar02], for instance, also

⁸The degeneration of scatter matrices is caused by the so-called “*small sample size*” (SSS) problem, where the number of samples images in the learning set, N , is much smaller than the dimension (number of pixels) of the face images.

considers the higher order statistics. ICA refers to another multivariate statistical technique that has the particularity of not requiring any transformation on incoming data and array manifold, but it utilizes only the statistical independence of the incoming data. This advantage is so-called blindness, and the concerned separation problem is known as *blind signal separation* (BSS). From a statistical point of view, this expansion achieves non-Gaussian but independent coefficients, whereas in the PCA case, the coefficients finally achieved was Gaussian (due to Gaussian distribution assumption) and independent. Therefore, ICA is more appropriate for non-Gaussian distributions [Has01] since they do not rely on the second-order of the data [Jai00]. Inspired by the fact that most of the useful information might be involved in the high-order relationships, authors in [Bar02] give an answer to the problem of FR by means of this transform. As major drawbacks of ICAs in general terms are the training time (much larger compared with PCA) and that the independent components obtained are not sorted by relevance [The06].

The search for such an optimal basis leads also to the *Projection Pursuit Regression* (PPR) approach, firstly developed by Friedman and Tukey in 1974 [Fri74], as possible candidate for universal approximation. This data transformation technique exploits the simple basic idea of finding interesting projections of a multidimensional data onto a line or a plane [Fri87, Jon87]. This nice transform requires the design of an index⁹, namely *projection index*, which is assigned to each computed subspace to indicate how well the projection reveals the structure of the data, so, the best subspaces could be found by optimizing this index. In this effect, for different interests in the structure of the data, we can define different projection index and thus find different projections, e.g. PCA projection for the maximal variances of the data (and factor analyzer projection for the maximal correlation among the data), LDA projection for the minimal intraclass variations and maximal interclass variations and ICA projection for minimal statistical dependence of the component of non-Gaussian representation [Tu03]. Therefore this last data transformation technique may be viewed as a generalization of the preceding described methods in some sense. In [Rod10a], authors have taken the above specific idea and have built a new collaborative feature extraction method based on the PP index based on the weighted sum of the most relevant state of the art FR indices, resulting in competitive results in terms of generalization performance and dimensionality reduction. Nevertheless, when the number of projections taken is arbitrarily large, most time becomes impractical to explore possible interesting projections exhaustibly, being more useful for prediction tasks [Has01].

Notwithstanding, due to all the above linear techniques project the data linearly from a high dimensional image space into a lower face subspace, they are unable to preserve the nonlinear and nonconvex variations of the different images of same individual introduced in Section 3.1, necessary to distinguish between different people [Li04]. This fact causes the decrease of the power of linear methods to perform reliable FR systems, especially when lacking representative training data. Thus, researchers in FR have also resorted to kernel methods as a feasible way to exploit such as non linear properties of faces. Then, linear subspaces can extended to a NL domain by mapping the image space into a potentially infinite feature space implemented by kernel methods. Kernel versions of PCA (Kernel PCA; KPCA) [Sch99] and LDA (Kernel LCA;

⁹Usually, this index is some measure of non-Gaussianity such as differential entropy.

KLCA) [Bau00] are based on such principle. Thus, the basic idea of kernel PCA will be first to project original images into a new high dimension face space by using a NL function (a Mercer kernel¹⁰ such as pth-order polynomial or a gaussian kernel) and then perform a linear PCA in the mapped space [Jai00]. Kernel LDA follows an analogous idea. KPCA and KLDA extended methods are behind several FR methods [Kim02, Yan00, Mik99]. In [Bac02], explores kernel methods on ICA approach. For more detailed information related with kernel methods and properties see [Bau02].

An also known technique to handle nonlinearity in FR applications is the statistical Local Feature Analysis (LFA) which was proposed by Penev and Atick in the 1990's [Pen96]. This method systematically extracts a set of local building blocks (or local features) of an original set of faces that can be combined in different ways to produce another set of new computed faces. It is a NL approach in the sense that a subset of all features is chosen for a given face and everything else is set to zero.

As has been seen along the section, the value of the overall set of the above discussed transforms and their variants is undisputed and has become a standard in FR literature. However, irrespective of the nature of the different reported approaches, all share two important shortcomings:

- They do not use fixed basis vectors (*data dependent*). This implies that the extracted features depend on the training image, and then, is statistically more probable that the set of projection vectors be too fitted to the training images that have been used to extract them, and some generalization problems can appear when dealing with new users and test images not used during training.
- They are suboptimal due to the concerned computational burden and memory requirements. Although nowadays, with the improvements on computational speed and memory capacities, it is possible to compute the concerned images transformations directly, their requirements are still important.

Moreover, all these methods deals with features extracted from the space domain. In this respect, is interesting to state that transforms that work in the spatial frequency domain (often called as *correlation filters* or *correlation PR*) offer another way to extract the meaningful information while appropriately managing the above restrictions; They are data independent transforms and their suboptimality with respect to decorrelation property is most often compensated by their low computational requirements. In addition there are additional advantages. Information packing introduced by such kind of transformations is much better in most cases and some transformations may offers of shift-invariance [Kum06]. In this second part of the section we will be concerned with these set of transformation based systems.

¹⁰Mercer Kernel refers to a Kernel function which satisfies Mercer's theorem, allowing computing dot product in the new extended space from the dot product in the original space.

Ahmed et al. first introduced the Discrete Cosine Transform (DCT) in the early seventies. Since then, several variants have been proposed. DCT is closely related to the Discrete Fourier Transform (DFT). It is an invertible and separable linear transformation; that is, the two-dimensional transform (2D-DCT) is equivalent to a one-dimensional DCT performed along a single dimension followed by a one-dimensional DCT in the other one [Str99]. The DCT tends to concentrate information, making it useful for image compression applications, dimensionality reduction, etc. Additionally, it is also well known that DCT transform converges to KLT when the block size to be transformed is large (>64 components) [Jai89], and this is the case of FR. In fact, the most popular image compression standards JPEG and MPEG are based on DCT [Son99]. Two following equations define 2D-DCT:

$$X[k, l] = \frac{2}{N} c_k c_l \sum_{m=0}^{a-1} \sum_{n=0}^{b-1} x[m, n] \cos \left[\frac{(2m+1)k\pi}{2N} \right] \cos \left[\frac{(2n+1)l\pi}{2N} \right] \quad (3.4)$$

where, in equation (3.4) :

$$c_k, c_l = \begin{cases} \sqrt{\left(\frac{1}{2}\right)} & \text{to } k=0, l=0 \\ 1 & \text{to } k=1, 2, \dots, a-1 \text{ and } l=1, 2, \dots, b-1 \end{cases} \quad (3.5)$$

In [Pod96] authors uses DCT approach for FR purpose. The system is based on matching face to a map of invariant facial attributes associated with specific areas of the face. In [McC07], authors modeled two set of DCT based features vectors to represent two forms of face variation (intersubject and intrasubject ones), in an analogous way as LDA approach.

In an attempt to reduce even more, computational requirements, Faúndez-Zanuy, Roure, Espinosa-Duró and Ortega [Fau07a] propose a low-complexity face verification system based on the Walsh-Hadamard Transform (WHT). It is a fast transform that does not require any multiplication in the transform calculations because it only contains ± 1 values. The final system developed can be easily implemented on a fixed point processor because no decimals are produced using additions and subtractions, and offers a good compromise between computational burden and verification rates, revealing that it is competitive with the state-of-the-art statistical approaches to FR. Table 3.1 compares the computational burden of KLT, DCT and WHT [Jai89], and provides the concerned execution times.

Transform	Basis function computation Execution t	Image transformation Execution t
KLT	$O(N^3)$ to solve 2 $N \times N$ matrix eigenvalue problems 347,78s	$2N^2$ multiplications 0,23s
DCT	0 0	$N^2 \log_2(N)$ multiplications 0,0031s
WHT	0 0	$N^2 \log_2(N)$ additions or subtract. 0,0003s

Table 3.1: Computational burden of KLT, DCT and WHT for images of size $N \times N$ and concerned execution time using a Pentium 4 processor at 3GHz.

The interest to use the discrete wavelet transformed domain for multiscale representation of the image data [Mal98, Bur08] are also moved to FR field. Wavelets are families of basis functions generated by dilations and translations of a basis *wavelet*¹¹. The two-dimensional Discrete Wavelet Transform (2D-DWT) is thus a multi-resolution decomposition of the function (image intensity) in terms of these basis functions. By decomposing an image using WT, resolutions of the subband images are consequently reduced. Assuming H_0 to be the (ideal) low-pass filter and H_1 the high-pass filter, the four frequency bands that are formed by the concerned decomposition are illustrated in Figure 3.12. Filtering process is recursively applied on each generated low pass frequency subband. This leads to a number of versions with an hierarchy of resolutions. This decomposition is known as *Multiresolution Decomposition*.

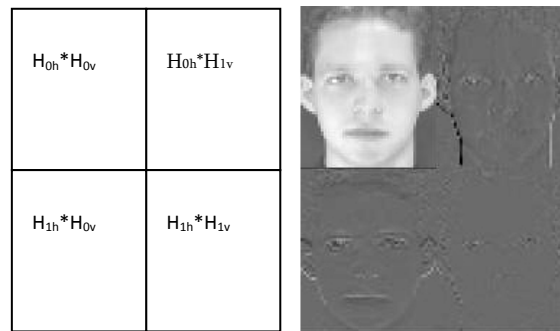


Figure 3.12: The first-level WT decomposition of a face (and the four generated subbands). Note that $H_{0h} * H_{1v}$ area records vertical features such as the outline of the face, whereas $H_{1h} * H_{0v}$ area fix changes of the image along horizontal direction, such as mouth, eyes and eyebrows.

In [Ete97] authors suggest to apply LDA to wavelet transforms of face images and extract the most discriminant vectors of each transform component and combine multiscale classification results by using a proposed method of soft-decision integration.

Feng et al. apply multiresolution techniques in [Fen00] in order to obtain a different wavelet subbands. Subsequently, they extract PCA features from these subbands (the final subband used for eigenfaces projection is a midrange frequency). This performance was carried out over YALE database. In [Ma06], a similar idea is exploited over ORL database.

Authors in [Eke04] decompose each training face into multi-subbands for extracting their PCA or ICA projections, and then exploit these multiple channels by fusing their information for improved recognition. Finally, they perform and compare three fusion approaches (fusion at the subband data level, fusion at feature extraction level, and fusion at the decision level at the subband channel level).

¹¹The wavelet-based approach provides simultaneous local information in both space domain and frequency domain.

Espinosa-Duró and Monte-Moreno explore in [Esp08] the use of wavelets families implemented by using linear-phase filters such as biorthogonal ones, in order to preserve the phase (orthogonal filters such as Daubechies and Haar wavelet families can have non linear phase, except trivial variations on Haar; the lack of this property produces distortions in the processed image). In addition, such kind of filters have the extra advantage of preserving the location of spatial details. Classification is then solved by means of a MLP NNET. Results reveal similar recognition accuracies when compared with classical PCA method.

Imtiaz and Fattah design a DWT as a feature extractor in [Imt11], for exploiting local spatial variations in a face image, obtaining outstanding results with two different databases. The authors, instead of considering the entire face image, an entropy-based local band selection criterion is developed, which selects high-informative horizontal segments from the face image.

All the above discussed spatial frequency domain approaches are good candidates for solving feature extraction tasks and have been used extensively in various FR problems. However as has been already discussed in Section 3.3.1, WT offers an extra advantage, which in some cases can be beneficially exploited: *Its multiresolution properties conform to the way perception is achieved by humans, through their hearing and visual systems* [The06].

An alternative to the reduction methods as stated, might be by dealing with feature vectors as sampled version of the initial feature vectors, by for instance, randomly eliminating some of the features. In this effect, the study carried out by Chawla and Hunter [Cha05] proved the viability of performing training by using high quality downsampled facial models. In brief, this approach proposes PCA methodology as a feature extractor, and the ensemble techniques of subsampling and random subspace method (RSM) are then applied in order to perform the classification stage. A special item of this approach is that depending of the experiment, the tuned or the complete face space is used.

The last, simplest but also promising dimensionality reduction method applied to the FR problem is the so-called *sparse representation*, developed by Yang et al. [Yan07]. The authors exploit an easy sparse representation method by randomly downsampling initial face image, being the computed random matrix, data independent. Then, if a reasonable number of pixels from anywhere of the image is available, a highly accurate way of FR will be achieved, even when eyes, nose and mouth are obscured or distorted. In addition, the differences in performance between different features become irrelevant once the dimension of the face space is large enough.

The overall above discussed methods in this sub-section, might be seen as powerful feature extractor systems, where classification task is solved by computing Euclidean distances among modelled and tested faces. However, this is just one of the available possibilities to the designers: Classification task might also be accomplished by NL classifiers as neural networks and SVMs.

Neural networks¹² (NNET) architectures provide a new suite of NL algorithms for feature extraction (using hidden layers) and classification by means on Feed Forward FF networks: Multilayer Perceptron (MLP) and Radial-Basis Function (RBF) networks [Bis95]. In addition, existing feature extraction and classification algorithms can be also mapped on neural network for efficient (hardware) implementation [Jai00]. By contrast, they have the drawback of high computational cost of classification and overfitting. NNETs are behind much research work in FR field [Tem99, Law97].

Nevertheless, some aspects should be taken into account when dealing with images: For high dimensionality inputs (when feature vectors have considerable size), the solution based on hyperplanes (MLP) should be discarded because the parameters are extremely sensitive to spurious correlations, and noise [Has01]. This fact is aggravated when the number of training examples is scarce. Also in [Bau89] is shown that under quite general conditions the generalization error of a multilayer neural net is proportional to the ratio between the number of weights and the number of examples. The bound is valid for hyperplane type of classifiers, which justifies the use of a local classifier such as the RBF, instead of a solution based on hyperplanes. (However, since radial-basis functions (RBF) networks use fixed Gaussian radial-basis functions with their exponent based on the euclidean distance matrix, they operate by constructing hyperspheres around the centers of the basis functions and therefore also suffer from the curse of dimensionality [Hay94]). Moreover, although RBF networks may be computationally more demanding than MLP (requires more neurons), they usually be fitted to the trained data with less time.

Support Vector Machines (SVM) firstly developed by Cortes and Vapnik [Cor95], are known to be very effective NL classification methods for 2-class discrimination purposes. SVM algorithm finds the hyperplane leaving the largest possible fraction of points of the same class on the same side, while maximizing the distance either class from the hyperplane¹³ [Cri00]. This is the reason because SVM behave like successful machine learning (ML) algorithms applied to the problem of face verification [Jon99]. In this case the system has just to deal with two kind of events: either the person claiming a given identity is the one who he claims to be, or not, and so, the system will have to decide between two unique possibilities. An interesting approach in this sense is [Car02] where the authors compare the performance of both MLP and SVM classifiers to resolve this particular problem of face verification. On the contrary, in the case of FR, the use of SVM is tricky in the sense that originally these classifiers solve a binary classification problem, so the multiclass problem has to be reduced to a set of binary problems [Pla00]. Even though SVM deals very well with high dimensionality input patterns, the extension to multiclass classification (one-against-all framework) is neither intuitive nor easy [Cri00] since reformulating the multiclass problem to a larger binary problem introduces

¹²NNET approaches have been used in FR generally in a geometrical local feature based manner, but there are also some methods where neural networks are applied holistically.

¹³ The obtained hyperplane is called the *Optimal Separating Hyperplane* (OSH). The high dimensionality of the feature spaces makes the OSHs very effective decision surfaces, while the recognition stage is reduced to decide on which side of an OSH lies a given point in feature space [Cri00].

restrictions into the classifiers and might lead the performance of the algorithm to a borderline. In [Phi99] the author reinterprets the problem of FR as a problem in difference space, which models dissimilarities between two facial images. In difference space, FR is redefined as a two class problem, being the two classes, differences between faces of the same subject and differences between faces of different individuals. An useful technique that allows extending in a natural way a binary classifier to a multiclass is the Error Correcting Output Coding (ECOC). ECOC refers to an approach firstly developed for channel coding that transforms a k-class supervised learning method into a large number of two class supervised learning. Authors in [Zor09] also exploit such technique to use SVM-based classification algorithm to the problem of recognition of facial expressions (that involves a lesser number of classes than the FR problem). On the other hand, Kittler et al. explore in [Kit01] this coding method jointly with MLP architectures, where their outputs define the ECOC feature vector.

For closing such NL architecture review and in a like manner to NNETs, the major disadvantages of SVMs are the high computational burden required when increasing the number of training samples, both during training and the test phases and overfitting.

The last research line reported in this section is related to the aggregation of classifiers in order to control the influence of the bias and the variance of the classification error. An example of this is the classic Bagging algorithm developed by Breiman [Bre96], which consists of training classifiers (in his case CART's) by random sampling with replacement of the training database. Thus, any of this random sampling is used to generate independent bootstrap replicates where the size of the subset is the same as that of the original set. Then, this method can be understood as an example of a resampling method. In [Lu03b], Lu and Jain describe other interesting resampling techniques in order to span several subsets of samples from the original training face dataset. In this case, a classic classifier based on LDA is built on each of the generated subsets. Notice that in the LDA approach, both intra and interclass information are used, so the sampling strategy is not randomly sample the whole training set, but is randomly sampling within each class.

Table 3.2 illustrates a layered approach of the solutions available and recognition rates reported as a summary of the current state of the art techniques in the visible spectrum. The reference to each work appears in the first column.

<i>Representative Work</i>	<i>Face DDBB</i>	<i>Strategy</i>	<i>Recognition performance (%)</i>
Turk and Pentland [Tur91]	Collected by the authors DDBB ¹⁴	Eigenfaces	from 20 to 100 as function of the used subset.
P. N. Belhumeur et al. [Bel97]	YALE	Eigenfaces	70,6
K. Etemad and R. Chellappa [Ete97]	A mixture of ORL(40) +FERET+ set of modified samples = 60 Subjects ;2000 faces	Fisherfaces LDA	99,4 99,2
P. J. Phillips [Phi99]	A subset of FERET (400 frontal faces; 200 subjects)	SVM	77
G. F. Feng et al. [Fen00]	YALE	DWT + SVM	85,45
J. Kittler et al. [Kit01]	XM2VTS	MLP + ECOC	93,25
A. M. Martínez [Mar01]	AR	Eigenfaces Fisherfaces	70 88
H. Yu and J. Yang [Yu01]	ORL	D-LDA	90,8
J. Lu et al. [Lu03a]	ORL	DF-LDA	96
M. S. Bartlett et al. [Bar02]	FERET	ICA	89
J. Tu and T. Huang [Tu03]	A subset of FERET (256 faces; 64 people)	PCA Projection Pursuit	82,8 92,2
X. Lu and A. Jain [Lu03b]	A mixture of ORL(40) +Yale (15) +AR(120) +Own DDBB = 206Subjects ;2060 faces	LDA + Resampling	88,7
C. Travieso et al. [Tra04]	ORL	DWT + SVM	98,9
N. V. Chawla and K. W. Hunter [Cha05]	A subset of FERET (462) +a subset of ND (138) = 600	PCA +RSM + NN Subspace reductions: 10% 25% 40% 50% & 90%	45 63 69 72 76
M. Faúndez et al. [Fau07]	FERET	DCT + NN WHT + NN	73,1 66,5
A. Y. Yang et al. [Yan07]	Extended YALE B	Sparse Representation	92,1

Table 3.2: Reviewed FR techniques and related recognition performance.¹⁴ 2500 face images; 16 subjects; 3 head orientations; 3 head sizes and 3 lighting conditions.[Tur91].

3.4 Infrared Approaches

As we have seen in Section 3.2, 2D FR systems must address face images degraded by many factors, intrinsic and extrinsic ones, some of which are extremely difficult to solve. Irrespective of difficulty, several FR algorithms have been proposed to deal with these challenges over the last years. Although significant advances have been made in this direction and high recognition rates over the constrained conditions of the experiments have been achieved, reports of performance of commercial FR systems in real-world scenarios reveals that a big gap exists between laboratory and real conditions [Zha06a]. Recognizing faces reliably across changes in pose and illumination has proved to be the most difficult problem to solve. A recent large scale evaluation of commercial face recognition system called the *Facial Recognition Vendor Test* 2002 (FRVT) showed in [Phi03] that FR and verification accuracy deteriorated significantly when there were differences in pose and lighting between images used for enrollment and matching [Geo01], and that errors increased as the lapsed time between enrollment and recognition increased.

FR based only on faces in the visible spectrum has shown difficulties in performing consistently in strong lighting conditions. In order to minimize this influence in lesser constrained environments, research in FR has been biased towards IR imagery (NIR and Thermal ones, which will be looked at in a much greater depth in the following chapter). While sacrificing color of the skin, eyes and hair, IR images provide detailed missing information in the optical window [Jai99]. Compared with conventional broadband images, this purpose can improve FR accuracy in several conditions.

On the other hand, infrared approaches do not significantly differ from FR approaches in the visible spectrum, being majorly appearance-based approaches like eigenfaces or fisherfaces [Akh08a]. Nevertheless, while a considerable amount of experiments exist that reveal the usefulness of Near-IR for FR, it is evident from the literature that no much research has been conducted by using thermal spectrum.

Additionally, it is widely known that human beings utilize a large amount of additional and contextual information in performing FR. Without this contextual information is highly questionable whether the face itself is really enough for recognizing with a reasonable level of performance [Cla94]. In this respect and in analogous way than people does, promising fusion approaches at different level (sensor, feature, score and decision as discussed in Section 2.4) has also been taken into account in FRT in order to enhance final recognition performance under different illumination conditions. Image fusion can be understood as a combination that extracts complementary and redundant information of different images from different spectra and fuses them into one more useful image [Gom01]. Thus, fusion methods can exceed FR performance beyond that of either (VIS & IR based methods) acting alone, representing a feasible help to visible imaging in an attempt to develop more robust solutions in this subject.

Since several authors have also recently suggested non-conventional imaging modalities such as *multispectral*¹⁵ imaging (which deal with several images at discrete and somewhat narrow bands covering the spectrum from the VIS to the LWIR range) these techniques will be also briefly reviewed.

3.4.1 Face Recognition in the Near Infrared Spectrum

FR with the so called Near Infrared (NIR), which refers to the closest part of the Infrared spectrum to the visible spectrum in the range of 700-1000nm, presents progress made in automatic techniques over the last two decades. In this sense, NIR images have an effectiveness checked for detecting and recognizing faces in areas where identification is critical, such as outdoor environments with variable lighting conditions [Bow06]. One of the main benefits of using NIR spectrum is that people in the scene are unaware that they are being illuminated by the system, increasing camouflage capabilities in facial recognition. However, this fact also implies an inherent drawback: NIR imagery require an specific illumination system (whereas recognition in the VIS spectrum may works with standard indoor illumination as well as with sunlight). Another advantage is related with human eyes physiology that provides acquisition of black eyes in such spectrum, as it will be further described in Section 4.4.2. Such property facilitates eyes detection task when dealing with simultaneous VIS & NIR acquisition systems, by simply using the difference image. Similar eyes detector algorithm is used in [Qui02] for 3D face pose estimation and tracking. Actually, authors in [Pav00] show that face detection task in VIS & NIR dual-band fusion systems can segment human faces much more accurately than traditional FR visible systems. By contrast, although NIR images are more insensitive to visible illumination, they may be also variant to invisible light [Zou05] due to in the same way that in visible case, they are also formed by the reflection of a certain kind of light (irrespective whether it is visible or invisible).

As already pointed at the beginning of the Section 3.4, many of the techniques used in NIR FR are inspired from their visible counterparts.

J. Wilder et al. report in [Wil96] initial efforts in determining whether NIR imagery provides a viable alternative to visible in the search for a robust, practical identification system. Authors extracted NIR images from a set of sequences acquired with a special purpose videocamera. Based on recognition results, authors conclude that both visible and IR imagery perform similarly across eigenfaces.

¹⁵In [Kos08] the author provides the following definition for the term MULTISPECTRAL:

A multispectral image is a collection of several monochrome images of the same scene, each of the taken with additional receptors sensitive to other frequencies of the visible light, or to frequencies beyond visible light, like the IR region of the EM spectrum. Each image is referred to as a band or a channel. There is no common agreement yet on the definition of the term hyperspectral image, (also known as imaging spectroscopy). However, the term is commonly used for images with more than a hundred bands. While “multi” in multispectral means many spectral bands, the “hyper” means over as in more than many and refers to the large number of measured wavelengths bands.

Espinosa-Duró presents in [Esp04a] a related work in the study comparing the effectiveness of visible and NIR imagery for recognizing faces over a first NIR database approach, projected over the eigenface space. In our case, images were acquired by using a broadband conventional visible camera which also provides sensitivity in the NIR spectrum as will be appropriately discussed in Section 4.3.1. Preliminary FR results point to same conclusions.

Another holistic technique that can be used for dimensionality reduction of NIR images is the DCT. S. Zhao et al. propose in [Zha05] the use of DCT lowest frequencies and a SVM classifier for recognition task, obtaining promising results. Authors also contribute with an interesting alignment and face detection method based on localizing eyes as a first step.

In [Zou06] authors explore NIR face identification over the LDA subspace and a SVM as classifier and conclude that performance results in a consistent approach to remove the illumination effects.

Furthermore, and as already mentioned at the beginning of the section, some authors have used image fusion, performing the recognition over this fused image at sensor (image) level. In [Hiz09] a synchronized visible and NIR facial database is presented, and the authors notice the stability of the performance of all the tested algorithms on IR images upon illumination variation, and the improvement in performance that results from the fusion of these two different kinds of images. Actually, in [Rag11] authors study the fusion of visible and near infrared images and finds slightly better accuracies when fusing images than applying other fusion levels. Moreover, image fusion techniques have also usually been assisted by means of a widely used wavelet based methods. Thus, different research groups have studied the concept of image fusion on WT between different spectral images from different approaches.

On the other hand, the potential of using multispectral images over the NIR spectrum for FR purposes is investigated by Pan et al. in [Pan03, Pan04, Pan05]. Once at the specific database was collected, authors demonstrate that spectral images of faces acquired in the concerned range was feasible for recognizing subjects across the eigenface algorithm under different poses and expressions. In addition, results reported indicate that the local spectral properties of human tissue are nearly invariant under different poses and expressions. However, their recognition performance was not compared with that of the broadband images acquired with conventional cameras in the visible spectrum.

To appropriately establish the concerned comparison, Chang et Kong acquire in [Che08] multispectral imaging in the visible and NIR spectra and conclude that both provide a new approach to separating the color information and the illumination in the image. In [Cha09] almost same research team insists in the use of multispectral imaging as alternative means to conventional broadband monochrome or color imaging sensors. In this new research work, authors address a new method that automatically specifies the optimal spectral range for multispectral faces images in the visible spectrum, according to given illuminations by the introduction of a distribution separation measure and the selection of the optimal frequency band by ranking the computed separation measures, whereas fused image is builded using the Haar wavelet based pixel-level fusion.

It is also interesting to briefly cite the approach of Gomez et al. in [Gom01], where authors extend the wavelet based method for data fusion among both multispectral and hyperspectral images.

3.4.2 Face Recognition in the Thermal Spectrum

As already pointed in the introduction chapter, FR based on thermal imaging, known as *Thermal Face Recognition* (TFR), has attracted considerable attention in the last few years, emerging as both an alternative and a complementary modality for FR in the visible spectrum but not much research work can be still found in literature, majorly due to not affordable and high performance thermal imagers availability.

While visible and near-IR images are based on the acquisition of the illumination reflected by the face, thermal images (TIR) are formed thanks to the measurement of heat emission. We are certainly producing infrared emissions, and it has been demonstrated that they are useful for biometric recognition, with some advantages such as the independence to illumination conditions (been fully operative even in darkness conditions which is still a challenge in the visible spectrum) [Soc01], robustness against disguises [Pav00] and the ability to discriminate between twin-brothers, which are all difficult tasks when using visible images. Nevertheless, thermal faces are inherently less discriminative and are heavily subject to change due to differences in body temperature caused by physical exercise or ambient conditions (specially, currents of cold or warm air), seasons and emotions of the subjects, even revealing different moods and anxiety in some cases [Heo05]. In addition, glasses and sunglasses becomes a stronger limitation in the thermal spectrum, fully occluding eyes area, resulting in loss of useful information for FR. Fortunately, this can be solved by detecting and segmenting the area on a face that is blocked by eyeglasses and replace the thermal image with average thermal eyes. The following part of the section gives an overview of existing approaches.

Prokoski provides in 2000 [Pro00] a consistent starting point in thermal face recognition by reporting an accurate current status and future trends. Facial thermogram's minimum required biometric properties are also assured since then. Previously, same author reported in [Pro92] that TFR first used MWIR cameras. At that time, cooled LWIR technology was very expensive. By the late 1990s, although uncooled thermal cameras in the LWIR became more affordable and accurate (even though their sensitive was still about ten times lower than uncooled MWIR cameras), MWIR technology still discerned more detail of the human faces. By contrast, we are currently living a turning point in TFR technologies mainly because of two prominent factors:

- Uncooled microbolometer LWIR cameras are rapidly approaching to almost one half of the sensitivity of cooled MWIR.
- Faces radiate more in the LWIR range, as we will appropriately discuss in Section 4.5.2.

These factors lead for the first time that the most affordable TIR imaging technology (i.e. LWIR), is also the most appropriate for recognizing human faces.

Socolinsky and Selinger also contribute in this field with a valuable line of investigation. In [Sel01], referenced authors deal with VIS and LWIR own single session face DDBB and perform a comprehensive comparison of classical and state-of-the-art appearance-based FR approaches (PCA, LDA, ICA and LFA) applied to visible and LWIR imagery. Going in depth with such work, regardless of what algorithm is used, performance on LWIR images has got better recognition accuracy than those on visible images over all experiments. On this regard, is important to mention that prior to experimentation, images used were previously subsampled by a factor of 10 in each dimension and it is possible that visible imagery may lose the advantage of relatively higher resolution comparing to thermal imagery.

In [Soc02] same authors perform a comparative analysis of FR performance with VIS & TIR imagery across the well-known techniques, PCA, LCA, LFA and ICA, reporting interesting results. It is also worth mentioning their following study reported in [Soc03] where the most comprehensive analysis to date on same session TFR is given. We also strongly recommend the reading of the tutorial referenced [Kon04].

In a new work [Soc04a] same authors move towards dealing with multisession databases in operational scenarios. In these more challenger conditions they observe that whereas multi-session thermal FR system under controlled indoor illumination was statistically poorer than visible recognition with two standard algorithms, significance was substantially reduced with an algorithm more specifically tuned to thermal images. On the other hand outdoor recognition performance was worse for both modalities, with a sharper degradation for visible images regardless of the algorithm. In the same year 2004, authors examine in [Soc04b] the influence of time-lapse on thermal images and conclude that recognition based on thermal images does not imply an inferior performance than visible ones when several weeks have elapsed between enrollment and testing. Authors in [Che03a, Che03b] include the impact of the short term (minutes) and longer term (weeks) change in facial thermogram appearance. Research on time-lapse recognition has been fundamental in determining under what conditions thermal systems can and should be used. Recent studies also focused on facial thermogram influence over time, leaded by Farokhi et al. in 2012 [Far12], give support to such previous conclusion once proven the reasonable reproducibility of the skin temperature patterns. The battery of experiments was performed over the University of Notre Dame time lapse thermal face database (2 years). This conclusion is very important because almost all experimental results reported in the literature are in fact same-session scenarios but not time-lapse scenarios.

Bebis et al. address their research on the sensitivity of thermal IR to facial occlusions caused by glasses [Beb06] and conclude that recognition performance in the thermal spectrum degrades seriously when eyeglasses are present in the test set but not in the training set and viceversa.

In [Jia04] Jiang answers the question of distinguishing frontal from non-frontal face views with the assumption that at any time, only one person is found in the scene and no any other heat-emitting object is present, by developing a system based on the distance from centroid (DFC) method.

Che et al. propose in [Che05] a semiautomatic system for registering multiple 3D termogram views and integrate them into one model.

In a similar way than in visible and NIR images, some researchers have considered the combined use of visual and thermal IR imagery constituting another viable means of improving the performance of TFR. This is challenging because the former is greatly affected by variation in illumination, while the latter frequently contains occlusions due to eye-wear and is inherently less discriminative.

In [Kwo05, Sin08] concerned authors perform the recognition over the visual and thermal fused image, while in [Nea07] decision fusion is fully exploited. In [Abi04] authors design and discuss these two fusion techniques (sensor and decision fusion), establishing interesting conclusions. In addition, an algorithm to detect and remove eyeglasses in the thermal images using ellipse fitting is also proposed, being the detected eyeglass regions then replaced by an average eye template. In [Che03b] the authors propose a score-based strategy and report that the fusion of visible and TIR images outperforms the individual spectrum recognition. Some other papers have also studied a score combination [Pop10, Aran20] obtaining encouraging results. Actually, in [Buy10] authors fully discuss all the fusion levels studies the sensor, feature and score level and conclude that data fusion at score level outperforms the other ones when combining visible and thermal images.

Arandjelovic et al. in [Ara06] propose an algorithm for combining the similarity scores in visual and thermal spectra emphasizing the presence of prescription glasses. Authors also examine the effects of preprocessing of data in each domain, obtaining a recognition rate of 97% in the best performance on the IRIS DDBB.

Moreover, and in analogous way that when dealing VIS and NIR images data fusion, wavelet based method's are also commonly used toward VIS and TIR image fusion.

Liu, Zhou and Wang propose in [Liu08] a novel wavelet-based method and visible and TIR images for FR. Authors firstly decompose original VIS and TIR images with a wavelet transform to fourth level in an attempt to search for the subband that is more insensitive to the variation in expression and in illumination. Once isolated, the fisherfaces method is then applied to the low frequency sub-image.

A related work to the previous one is performed by Bhowmik et al. [Bho10]. Authors present a comparative study on fusion of VIS and TIR images using different both Haar and Daubechies (db2) wavelet transforms. Here, coefficients of DWT once computed separately from both visual and thermal images are then fused. Next, inverse DWT is performed in order to obtain fused face image. This computed image will be subsequently projected onto an eigenspace in order to reduce components. Finally, modeled image will be classified by means of a MLP. The resulting system was tested on IRIS database and was shown to outperform the VIS images over direct eigenface method.

In [Bou07] a new step is done. Authors firstly decompose VIS and TIR images using Haar wavelets and then the data fusion is computed by obtaining a weighted combination of corresponding DWT coefficients of the both spectra. Afterwards, a score fusion is applied over the three concerned scores (VIS, TIR and wavelet fused) based on the average or on the highest matching score.

For concluding this section, new recent trends in this field, far the visible counterparts techniques already discussed will also be briefly discussed below.

Based on the idea already suggested by Prosoky and Riedel in 1999 of using physiological information extracted from high temperature regions in thermal face images [Pro99], Buddharaju et al. are the first to present a novel framework for TFR based on capturing facial physiological patterns [Bud06, Bud07]. Authors localize the superficial blood vessel network by using mathematic morphology NL techniques. In a similar manner that a fingerprint is encoded [Esp02], the ridge bifurcations of the thinning vascular network, called by the authors as *Thermal Minutia Points*, constitute the modeled face. The encouraging experimental reported results demonstrate the feasibility of the new physiological framework and open the way for further methodological and experimental research in the area. Later works leaded by Akhloufi and Bendada [Akh08b, Akh08c] insist in the idea of extracting different physiological information from facial thermogram as a new way of performing promising high TFR performance.

More recently, same authors introduced in [Akh10] another new framework for TFR by using 3D imaging and texture descriptors, achieving encouraging first results over Equinox database. The best result was obtained in the short wave infrared spectrum (reported in the following Chapter) using non linear dimensionality reduction techniques.

Table 3.3 summarizes the most relevant work of FR when dealing with IR images.

<i>Representative Work</i>	<i>Face DDBB</i>	<i>Strategy</i>	<i>Recognition performance (%)</i>		
			VIS	IR	Fusion
Socolinsky and Selinger [Soc04a]	Own DDBB:	PCA	81,54	58,89	87,87
	385 subjects ; 4 sessions	LCA	94,98	73,92	97,36
	VIS 640x480	EQUINOX Comercial algorithm	97,05	93,93	98,40
	LWIR 320x240				
Bebis et al. [Beb06]	EQUINOX	Fused (wavelet domain)	-	-	91
		Fused (eigenspace domain)	-	-	83
Chen et al. [Che03b]	NOTRE DAME	PCA + Score Fusion	75	73	90
	(VIS & TIR)+ UQUINOX				
S Zhao and R. Grigat [Zha05]	Own DDBB:	DCT + SVM	-	79,6	-
	10 subjects (25 images per subject) NIR 320x240				
O. Arandjelovic et al. [Ara06]	IRIS	PCA w/o glasses detection			90
	(VIS & TIR)	PCAw glasses detection			97
M. K. Bhowmik et al. [Bho10]	IRIS	Fusion Haar DWT + PCA	-	-	87
	(VIS & TIR)	Fusion db2 DWT + PCA			91,5
Sajad Farokhi et al. [Far12]	NOTRE DAME	PCA	76,23	61,92	83,24
	(VIS & TIR)	Zernike Moments + NN(Euclidean)	87,24	73,23	92,37
		ZMs + NN (TSR)	91,35	76,93	95,23

Table 3.3: Reviewed FR techniques in the IR spectrum and related recognition performance.

3.5 Face DataBases

In Section 3.2.1 a variety of different intrinsic and extrinsic factors that result in variations in facial appearance has been reported. The development of robust algorithms to these variations requires data sets benchmark of sufficient size specially designed to evaluate them, which include carefully controlled variations of these factors. In this sense, the choice of an appropriate DDBB based on the property to be tested will be desired. The most well known face data sets are FERET, AT&T (former ORL dataset), AR, M2VTS, XM2VTS and EQUINOX [Equi09], which are briefly described in Tables 3.4 and 3.5 [Gro04]. Furthermore, when benchmarking an algorithm, the use of standard databases in order to be able to directly compare the results between different researchers techniques, will be also highly recommendable. However, researchers usually test their systems with their own sets, which are not available to other researchers making the comparison itself a complex task.

While there are many DDBBs in use currently, we are especially interested in DDBBs where the nature of the spectrum and/or the illumination issues has been specifically taken into account. In this section we review the most prominent public domain face DDBBs. Tables 3.4 and 3.5 overview these publicly available data sets categorized into two groups: Those databases

that deal with faces in the VIS spectrum and those that deal with images in overall IR spectrum. In italics, literal expressions, but not detailed ones from the developers, has been collected. When available the relation between males (coded as “*m*”) and female (coded as “*f*”), has been also detailed. Blank cells correspond to non-available settings (either because the underlying measurement was continuous or the set was not controlled during acquisition process).

Notice that although several databases exist in the visible spectrum and less that simultaneously acquire visible and NIR or VIS and thermal images, there is not any existing multisession database containing VIS, NIR and THIR information simultaneously. In Chapter 6 we will describe the new database acquired with three different sensors in order to acquire these last three mentioned spectra under different illumination conditions.

Facial DDBB	# of Subjects	#Faces	#Sessions t between Sessions	Resolution Bit Depth	Facial Expressions	Pose	Illumination	Others
AR Purdue CVC	126 70m/56f	3276 126x26	2 2 weeks	768x576 24 bits; RGB	High variability Occlusions (sun glasses, scarfs)	Frontal (some tolerant)	Frontal	
AT&T Lab. (former ORL)	40 35m /5f	400 40x10	24 months	92x112 8 bits; grey levels	Neutral	Frontal (some tolerant)	<i>varying lighting</i>	
BANCA	52 26m /26f		12 3 months	720x756				Multi-modal DDBB (face & speech)
FERET US Army FacE REcognition Technology	≈3000 m/f	7562	15 23 months	256x384 8 bits; grey levels	<i>No any restriction imposed</i> (Between neutral and smiling)	<i>At least, two frontal images</i>		Diversity across gender
HARVART Harvart Robotics Lab]	10 m/f			193x254		Frontal	77-84: 0° 30° 45° 60° 75°	Highlight: large variations in illumination
ND (NotreDame) Univ of Notre Dame	>300	>15000	10/13 10/13 weeks	1600x1200 2272x1704 Color; JPG	2: Neutral and smiling		a) 2 side focus b) 2 side focus + frontal focus	
MANCHESTER Univ. of Manchester	30 23m /7f		3/54 weeks		Varialibity Strong Occlusions (glasses, objects...)			Ethnic origins mixed
XM2VTS (Extended M2VTS) Univ of Surrey	295	9440 295x32	4 5 months	720x576	4x2 head rotations 4x6 Speech: <i>In the frontal- view images, subjects read a specific text</i>	Frontal and some Rotation (Hor & Ver)		Multi-modal DDBB (face & speech) 3D models of 293 are available as well
YALE	15	165 15x11		320x243 8 bits; grey levels	High variability	Frontal , 12° and 24°		
YALE B	10	5760 10x576 [5760+9 0=5850]		640x480		9	64 lighting conditions	Highlight: large variations in illumination and pose

Table 3.4: Overview of the public domain Face DDBBs in the VIS spectrum.

Facial DDBB	# of Subjects	#Faces	#Sessions t between Sessions	Resolution Bit Depth	Pose, Facial Expressions, Details (Glasses)	Illumination	Spectrum Range
CMU	54	3510 54x65	5 6 weeks	640x480		4 : Three lamps individually and then combined	VIS-NIR (0,45-1,1 μ m) ; BW of 650nm Hyperspectral [Range=10nm] (650/10 = 65 Images) ; Time lapsus of 8s
EQUINOX	91	822	1 -	320x240	3 Facial expressions: smiling, frowning and surprising. Occlusions (eyeglasses)	3: Frontal-L-R	VIS (0,4-0,7 μ m) ; LWIR (8-12 μ m) For some subjects, additional SWIR (0,9-1,7 μ m) and MWIR (3-5 μ m) are also acquired.
TERRAVIC	20 19m/1f		1 -	320x240	3 different poses 1 (Neutral expression) Occlusions (eyeglasses and hat) Indoor / Outdoor	-	TIR
IRIS Imaging, Robotics & Intelligent Systems Lab.	30 28m/2f	3058 1529 VIS 1529 LWIR 30x(176-250)	1 -	320x240 RGB 320x240	11 different poses 3 Facial expressions: smiling, anger and surprising.	6	VIS LWIR
IRIS Imaging, Robotics & Intelligent Systems Lab.	82 62m/20f ≠ ethnici- cities	2624 VIS Multispectral _{vis}	10 From August 2005 to March 2006			3 sources: halogen, fluorescent and outdoor	VIS Multispectral _{vis} (0,4-0,72 μ m ; 25 bands)
ND (NotreDame) time-lapse (Collection X1)	82	2292 VIS 2292 LWIR	From 2002 to 2004		Frontal Without glasses		VIS LWIR

Table 3.5: Overview of the public domain Face DDBBs in different spectral ranges. Two additional rows (*facial details –glasses- and spectrum range*) have been included as important properties for analyzing IR database approaches. Special features where our face database obtains good marks has been highlighted in the header of the table.

Chapter 4

Visible, Near-Infrared and Thermal Face Imaging

*Dans les yeux d'un jeune brille la flamme.
Dans les yeux d'un vieux brille la lumière.
-En els ulls del jove crema la flama. En els del vell brilla la llum-*
Víctor Hugo.

This chapter covers some general aspects related with images to deal with. With the exception of the first two sections of introductory nature, the following three sections share the same structure, and are devoted to issues surrounding sensing and digital image nature in the different spectra (visible, near-infrared and thermal-infrared), as well as the concerning facial images in such spectra. A review of the most relevant acquisition technologies and details of the technical aspects will be also given.

4.1 Introduction

Although audio-visual human systems have several well-know limitations, artificial sensors can measure information beyond human limits. In contrast to speaker recognition infrasounds and ultrasounds are not directly applicable, for instance, to speaker recognition due to the impossibility of human beings to generate sounds in these frequencies, this is not the case with image signals beyond the visible spectrum for FR due to human beings can reflect body images in the reflected VIS and NIR spectrums and emit thermal radiation in thermal band of the infrared spectrum (TIR and referred as TH in our experiments) [Esp11].

4.2 Background Fundamentals

This section provides a basic understanding of the concerned electromagnetic (EM) spectrum and the underlying atmospheric behavior, thermal radiation and heat transfer, as well as the image model and the photometric sensor models. In addition, the equations described along the section will provide the basic framework for dealing with the different acquisition systems required in this dissertation.

4.2.1 Electromagnetic Spectrum and Atmospheric Influence

The portion of the EM spectrum visible for the human eye roughly ranges approximately from 300nm to 700nm when measured in terms of daylight conditions, being not a flat response and showing a maximum sensitivity at 555nm. This is also called *photopic curve* and matches with the CIE (Commission Internationale de l'Éclairage) standard curve used in the CIE 1931 color space. This curve shifts itself towards shorter wavelengths in darkness conditions due to the *Purkinje effect*, where it becomes the *scotopic curve* which has a peak luminance sensitivity at 510nm.

While visible spectrum comprises a narrow portion of the spectrum (400nm), IR spectrum comprises a broad range from 700nm to 1mm, being a large region of EM waves. The Near Infrared (NIR) window lies just outside of the human response window, and the medium Infrared (MIR) and far IR (FIR) are far beyond the human response region. Many species can see wavelengths that fall outside the visible spectrum. Bees and many other insects can see light in the ultraviolet (UV), which helps them find nectar in flowers. Plant species that depend on insect pollination may owe reproductive success to their appearance in UV light, rather than how colourful they appear to us. Birds can also see into the UV region nearest to the optical window (300-400nm), and some have sex-dependent markings on their plumage, which are only visible in the UV range [Cut97, Jam07].

An especially interesting sub-band of FIR spectrum lies from 3 μ m to 14 μ m, called Thermal Infrared (TIR), which humans experience every day in the form of heat or thermal radiation. This special band of the spectrum presents two important windows called Mid-Wave IR (MWIR) comprised in the range between 3 and 5 μ m, and Long-Wave IR (LWIR), the second thermal windows that lies in the range between 8 μ m and 14 μ m. Between them there is a band which is blocked by contamination due to solar reflectance and water vapour absorption¹. Figure 4.1(a) shows the EM spectrum paying special attention to the overall existing infrared sub-bands, also known as *Atmospheric Transmission Windows* (c). These bands define the IR channels that are usable at technologic level and useful for imaging. The intermediate graph (b) depicts the atmospheric transmittance in the band of IR.

¹Water has two resonance frequencies, in the IR spectrum and in the microwave band.

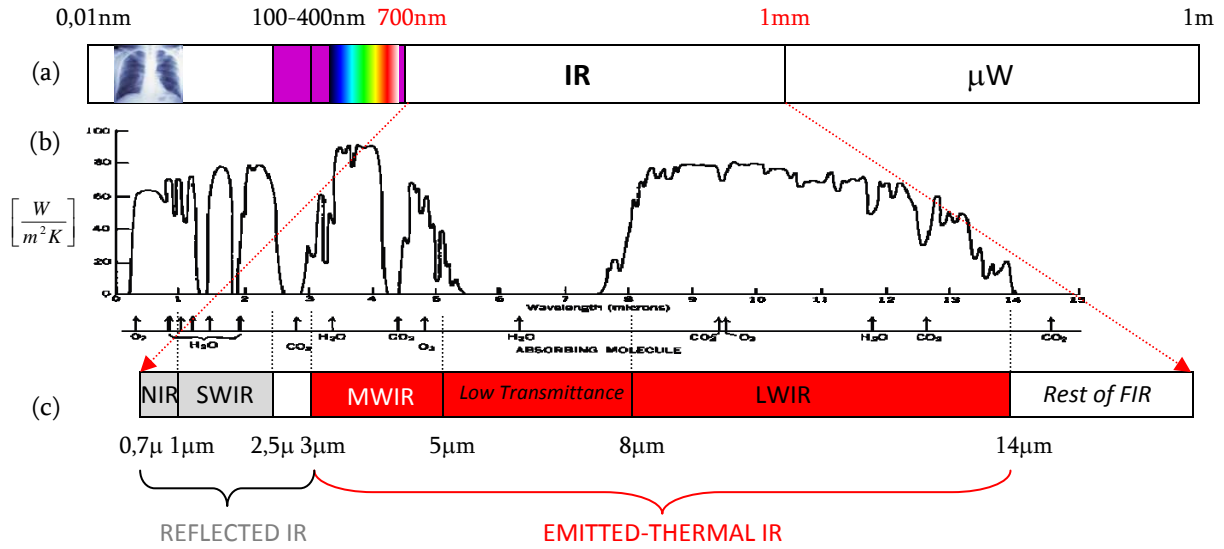


Figure 4.1: (a) EM spectrum as function of the wavelength. (b) Atmospheric Transmittance in the region of the IR spectrum. Note the atmosphere strongly absorbs between 5 and 8μ due to water molecules in the atmosphere. (c) IR Channels.

Whereas most of the IR range is not useful for transmission because it is blocked by the atmosphere (from 14μm to 1mm and also the low transmittance bands depicted from 700nm to 14μm and showed in Figure 4.1(b)), EM waves with shorter wavelength radiations as VIS and UV, e.g., are highly sensitive to the so-called *Rayleigh scattering* [Der51] that appears when particles present in the atmosphere are much smaller than the scattered photons. This effect is particularly severe in the blue end of the visible spectrum causing important levels of degradation over the remotely sensed images². Therefore the degree of Rayleigh scattering that an EM wave undergoes is function of its wavelength and the dimension of the particle, being the respective intensity of the scattered light computed by the well-known *Rayleigh Law*, defined by the following analytic expression:

$$I = I_0 \frac{1}{\lambda^4} \frac{(1 + \cos^2 \theta)}{2R^2} (2\pi)^4 \left(\frac{n^2 - 1}{n^2 + 2} \right)^2 \left(\frac{d}{2} \right)^6 \quad (4.1)$$

where R is the distance to the particle, θ is the scattering angle, n is the refractive index of the particle, and d is the diameter of the particle. Rayleigh Law applies to particles that are small with respect to wavelength of light, and that are optically "soft" (i.e. with a refractive index close to 1).

²Rayleigh scattering of sunlight into the atmosphere is also the reason why the sky is blue.

4.2.2 Principles of Infrared Thermometry

The most frequently measured physical property throughout history has been the time and the second one, the temperature, being this last one also a good indicator for the status of physical systems and one of the major used measures to characterize the molecular behavior of any material. However, in practice some difficulties exist, being the most important, the no possibility to obtain it in a direct way, having to relate through another variable [McG86]. An usually used indirect variable is the EM radiated energy by objects and this is also the key basic principle of temperature measurement by means of infrared radiation, and consequently, of the thermal imaging of *thermography* as is also called. The really large steps in the story of thermography started in 1800, when Sir William Herschel discovered the infrared radiation by accident while studying the colors of the stars [Vol10]. Second, Max K. Planck established one of the basis of the fundamental physics, when concluding that every object with a temperature above absolute zero *radiates electromagnetic energy³, principally in the infrared spectrum (thermal radiation⁴)*. The spectral radiant distribution of the referred radiation is described by the Planck's distribution function for perfect emitter (blackbody) radiation [Bar00] can be written as:

$$M(T, \lambda) = \frac{2\pi hc^2}{\lambda^5} \left(\frac{1}{e^{\frac{hc}{\lambda kT}} - 1} \right) \left[\frac{W}{m^2} \right] \quad (4.2)$$

where h is the Plank's constant ($h=6,623.10^{-24}$ J/Hz), c is the light speed in the vacuum, λ is the wavelength of the radiated EM and T denotes the absolute temperature of the blackbody given in Kelvin (K). The respective radiance is $L_\lambda(T)=M_\lambda(T)/\pi$. Such postulate is regarded as the birth of quantum physics. For more information about the development and properties about Planck's function, the interested reader is referred, among many others, to Eisberg and Resnick books [Eis76].

By integrating the above expression over all wavelengths, *Stefan-Boltzmann Law* is obtained:

$$M(T) = \int_0^\infty M(T, \lambda) d\lambda = \sigma T^4 \quad (4.3)$$

where σ is the Stefan-Boltzmann constant ($5,67.10^{-8}$ W/m² K⁴) and T is the absolute temperature of the blackbody. In addition, the net amount of intensity of the radiation which is emitted by any object different to the blackbody, also depends on the radiation features of the surface material of the measuring real body. The *Emissivity⁵* (ϵ) is a physical property of materials that describes how efficiently it radiates and ranges between 0 and 1. The blackbody is the ideal radiation source and has an emissivity equal to one.

³ The assertion does not take into account dark matter.

⁴ *Thermal radiation* refers to the transmission of heat by means of electromagnetic waves.

⁵ *Emissivity* is analogous to the notion of albedo used in the computer vision literature.

The greybody is a body having an emissivity less than one and constant for all wavelengths, while the non-grey body (or color body) is a body whose emissivity changes with wavelength. Consequently, for a greybody, Stefan-Boltzmann Law, takes the specific form:

$$M(T) = \varepsilon \sigma T^4 \quad (4.4)$$

Figure 4.2 graphically shows the analytic expression described in (4.2):

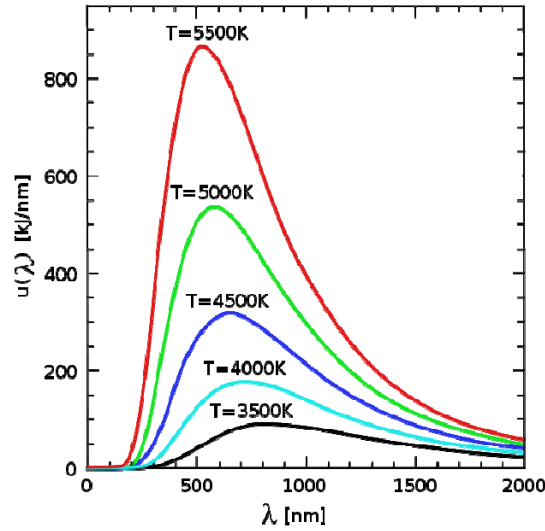


Figure 4.2: Illustration of the Plank's Law.

Two conclusions can be directly extracted from the graphic of the Figure 4.2: the amount of radiation emitted by an object increases with temperature, and the peak shifts to shorter wavelengths as the temperature of the object increases. The *Wien's displacement Law*, named in honor of physicist Wilhelm Wien, obeys the mathematical expression of this observation and computes the wavelength corresponding to the maximum energy (4.5). This value can be computed by differentiating the expression (4.2) and equating to zero:

$$\lambda_{\max} = \frac{2897,8}{T} \quad (4.5)$$

This expression can also be viewed as the wavelength corresponding to the maximum energy, multiplied by the absolute temperature of the body is equal to a constant, equal to 2897,8.

Thermal infrared cameras that will be appropriately discussed in the last section of this chapter, detects heat transfer by the EM radiation process stated in this section [Bar00] for measuring the surface temperature of the objects. Nonetheless, in physics, one usually distinguishes two additional known heat transfer's modes, called *radiation and convection*. The underlying physical processes that involve conduction and convection are very similar; therefore, the distinction is rather artificial. Conduction refers to the heat transfer in a solid or fluid which is

at rest, while convection refers to the heat transfer between a solid and a fluid which is in motion, when there is a temperature difference between them.

Formulate the general heat transfer (also called *energy flow* in thermodynamics) problem considering all such contributions, involves solving the equations that describe the phenomenon characteristics, which result from the principles of conservation of mass, momentum and energy, being the referred analysis beyond the scope of this thesis. Interested readers can see [Inc96]. Our discussion in this chapter regarding such additional heat transfers modes, hence is limited to very simple example such as those of the everyday life, where the three heat transfer modes are simultaneously given: we refer to the classic cup of coffee. As depicted in Figure 4.3, the plastic coffee cup's surroundings are about 24°C (see point 1) while the temperature of the coffee is considerably higher: In such situation, the conduction heat transfer phenomenon (also called *diffusion*), becomes evident, as states the second law of thermodynamics, while gradient temperature is obtained by observing the difference in temperature between points 1 and 2.

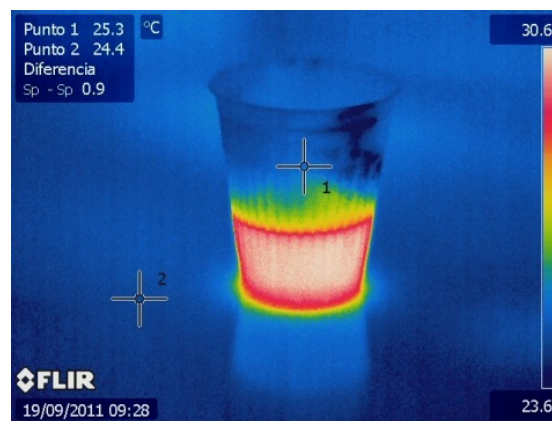


Figure 4.3: A cup of hot coffee and the concerned three forms of heat transfer. Yellow zones marks heat transfer by conduction while the difference of temperatures among points 1 and 2 is due to the convection and radiation heat transmissions (coffee was previously removed by a spoon). [Image acquired with a FLIR SC620 thermal imager; resolution of 640x480 and NETD <40mK].

4.2.3 Face Image Model

As described in Section 3.3, a picture of a human face is modeled as a bidimensional array of pixels of different gray levels. Thus, in order to recognize an individual using his face it has to represent a human face in a way that the FR system can use. In a more generalized way, and taking into account that infrared has the same properties as visible light regarding reflection, refraction and transmission for the image in visible and reflected IR sub-bands, the light arriving at a camera can be understood and easily modeled as a bidimensional light-intensity function, represented by $I(x,y)$ [Gon87]:

$$I(x, y) = L(x, y)\rho(x, y) \quad (4.6)$$

where $L(x, y)$ is the *illumination* or amount of source light falling on the surface of the object and $\rho(x, y)$ is the *reflectance* of the surface that the light is leaving, also known as *albedo*⁶, and refers to the emissivity intrinsic property of the objects (-faces). In a more generalized way, and taking into account the atmospheric influence in the EM waves transmission process, as discussed in Section 4.2.1, equation (4.6) can be newly written in a more generalized way, as follows [Kos08, Cha08]:

$$I(x, y) = L(x, y)\rho(x, y)\tau(x, y) \quad (4.7)$$

where $\tau(x, y)$ is the *transmittance* or *transmission* of the medium (atmosphere). Equally, for thermal faces images, and considering that the same laws that govern the transfer of light also govern the radiant transfer of heat, the expression (4.7) can be reformulated as:

$$I(x, y) = E(x, y)\tau(x, y) \quad (4.8)$$

where $E(x, y)$ denotes the EM emission of the object, whose spectrum is $E(\lambda)$.

Moreover, the modeling of image formation commonly used, generally requires the hypothesis of a *lambertian surface* reflectance (after Johan Lambert, who first formalized the idea). The surface of an object is considered lambertian (or also, *ideal diffuse surface*) when reflects light equally in all directions when illuminated⁷. This means that the illuminated surface has the same brightness independently of the direction of observation [For03]. Consider a point p on a Lambertian surface illuminated by a point light source at infinity. Let $\mathbf{s} \in \mathbb{R}^3$ be a column vector signifying the product of the light source intensity with the unit vector for the light source direction. When the surface is viewed by a camera, the resulting image intensity of the point p (x, y) is given by:

$$E(p) = \rho(p)\mathbf{n}(p)^T \mathbf{s} \quad (4.9)$$

where $\mathbf{n}(p)$ is the unit inward normal vector to the surface at the point p , and $\rho(x, y)$ is the albedo of the surface at p . This shows that the image intensity of the point p is linear on $\mathbf{S} \in \mathbb{R}^3$. [Bel97]. When dealing with the special case of faces, the assumption of lambertian reflection is also given. Under this hypothesis, the set of images of a given face acquired under all possible illumination conditions, but fixed pose, lie in a 3D linear subspace of the high-dimensional image space, as demonstrated. However, and as pointed in Section 3.2, since they

⁶The more efficient is a body in reflecting energy of a given wavelength (more reflectance albedo) the less efficient it is in thermally emitting energy at that same wavelength respective to its temperature (less thermal albedo).

⁷ In computer graphics, Lambertian reflection is often used as a model for diffuse reflection.

are not truly Lambertian surfaces without shadowing, and indeed produce self-shadowing depending on direction of the illumination, images will deviate from this linear subspace.

Other FR techniques recreate the shape of the face in 3 dimensions. The 3D model is known to deal with the correspondence problem when images showing faces in different poses as well as some illumination problems [Vet97]. First approach, seeks to infer the 3D structure of the face from one or more 2D images of the same face. More sophisticated approaches, uses three dimensional devices that record facial images in three dimensions providing more information than two dimensional ones, which enables recognition from different angles and thus enhancing performance. The overall system provided with fine algorithms allows learn a pose-invariant shape of the faces. For detailed information on the appropriate processes in 3D modelling and consequent recognition techniques, see [Hal99].

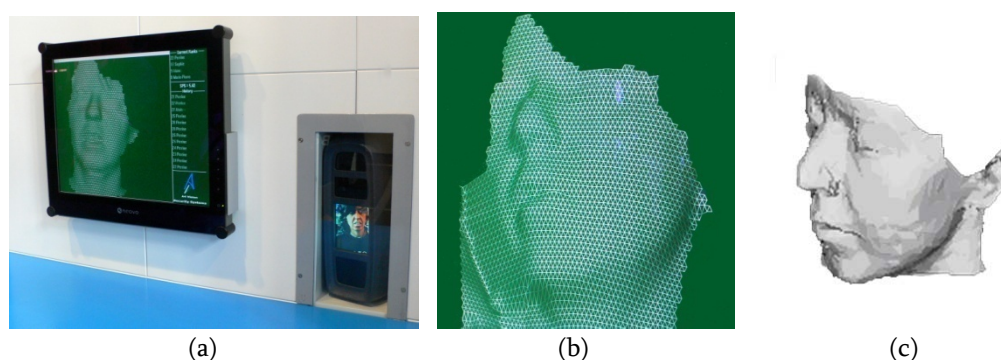


Figure 4.4: 3D face acquisition. (a) The author just standing in front of the screen to acquire and see 3D acquired face. (b) Generating the 3D face image. (c) Rendered image.

4.2.4 Photometric and Thermal Sensor Models

One of the main parts of any image acquisition system is the sensor and it is also one of the most sensitive parts. Due to properties of the FR algorithms strongly depend on the characteristics of the acquired image as pointed out in Section 3.2, appropriate sensors with low sensibility against noise are desirable for the success of recognition systems.

For image sensing, two different feasible detectors can be distinguished according to the different operating principle and consequently, two different models can be found: *passive* and *active*. In the former group, image signal is obtained by the *reflecting* light of the scene (human face in our case), when illuminated with the light source (UV-VIS-NIR-SWIR) of the referred vision system, while the second ones collect the *emitted* EM energy in the MWIR and LWIR ranges⁸ as illustrates Figure 4.5.

⁸This is the reason why NIR and SWIR sub-bands, are sometimes called *Reflected IR* while MWIR and LWIR are referred to as *Emitted or Thermal IR*.

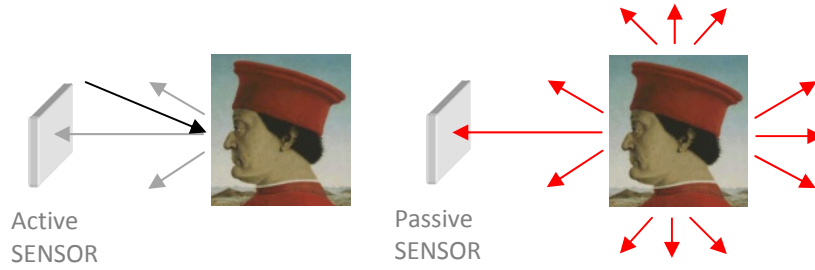


Figure 4.5: Basic operative principle of the active and passive sensors. Notice that the passive sensor senses the energy that is directly radiated by objects.

Since, according to the black body radiation law, EM radiation, mainly IR, is emitted by all objects, it will be possible to sense live beings and warm objects with or without visible illumination in a similar way that some animals are capable to do it. An interesting example of living beings that can detect such kind of thermal information are some kind of snakes. Some snakes can “see” radiant heat at wavelengths between 5 and 30 μm to a degree of accuracy such that a blind rattlesnake can detect warm blooded animals and then target vulnerable body parts of the prey at which it strikes [Kar91]. It was previously thought that the organs evolved primarily as prey detectors, but recent evidence suggests that they might be used in thermoregulation and predator detection, making it a more general-purpose sensory organ than it was firstly supposed [Kro94]. Second and last illustrating example is the Ethiopian monkey gelada, a primate that depicts a strange bright red patch of skin on its chest. This exaggerated feature has received various interpretations, being the last one, an effective way of exhibition of strength and virility (when more heat it emits, more excess energy it has) serving as a sexual ornament.

The spectral response function $I_p(x,y)$ of a photonic sensor in a wavelength range (from λ_{\min} to λ_{\max}) can be represented as [For03]:

$$I_{\text{photosensor}}(\lambda) = \int_{\lambda_{\min}}^{\lambda_{\max}} S(\lambda) L(x, y, \lambda) \rho(x, y, \lambda) \tau(x, y, \lambda) \quad (4.9)$$

The functions $\rho(x,y,\lambda)$ denotes spectral reflectance of the object, $L(x,y,\lambda)$ is spectral power distribution of the illumination, and $S(\lambda)$ is spectral sensitivity of the photodetector and $\tau(x,y,\lambda)$ denotes spectral transmittance of the atmosphere.

For passive thermal sensors, (4.10) is reformulated as:

$$I_{\text{thermal}}(\lambda) = \int_{\lambda_{\min}}^{\lambda_{\max}} S(\lambda) E(x, y, \lambda) \tau(x, y, \lambda) \quad (4.10)$$

4.3 Visible Imaging

The first part of the image processing chain is acquisition and it is one of the most sensitive parts. As introduced in Section 3.2 the properties of the FR algorithms strongly depend on the characteristics of the acquired image (face), especially, resolution, spectral sensitivity and noise image sensor concerned parameters and lighting conditions (majorly, direction and quality of the light). This section will review the most prominent related concepts, while in the next chapter will properly discuss the focusing problem.

4.3.1 Acquisition Systems in the Visible Spectrum

Whatever the system (for acquiring single still images or video sequences), all share the same basic idea, based on dedicated light source and the proposed camera provided with the relevant image detector positioned at the focal plane of the optical lens, being majorly a Couple Charge Device (CCD) or a CMOS sensor. They both are conventional broadband monochrome⁹ sensors formed by an array of photosensitive elements that follows the *photoelectric effect* (mainly photodiodes due to their linearity) over a slightly p-doped or intrinsic layer of silicon. The number of these photosensitive cells determines the resolution of the device. Although the sensitivities of both image sensors approaches to the photopic curve in the VIS spectrum, almost all silicon sensors also provide light response in UV to about 330nm and NIR band to about 1000nm (See Figure 4.6).

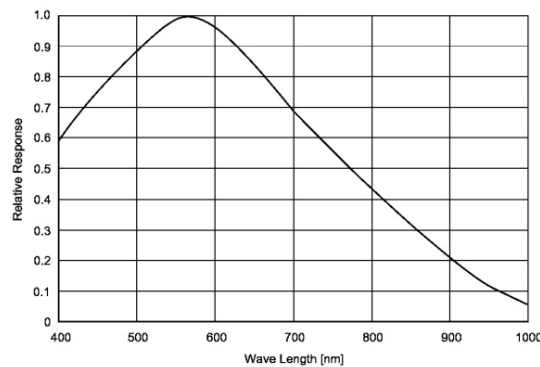


Figure 4.6: Spectral sensitivity response of a general purpose CCD. Note the cut effect of the UV filter over 400nm and the sensitivity showed beyond both sides of the visible spectrum.

⁹The described sensors detect light intensity, not color information. In order to also provide color sensitivity, essentially, two architectures exist. An additional built-in color filter array (CFA) over the sensor (largely a Bayer mosaic mask) in one CCD (1CCD) color sensor approach is used while for higher performance, three-CCD approach (3CCD) is used, arranged with three different fully R, G and B filters (each one assembly over the surface of each of the three CCDs) to respond to each color and a dichroic beam splitter prism embedded to divide the incoming light beam into the three required output beams [Kos08].

Note that in order to minimize UV&IR contamination of VIS images; most manufacturers cover their silicon sensors with *Internal UV and IR Cut Filters* (IUVCF & IIRCF).

As just pointed, the illumination system is the second required element to build the full acquisition system for detecting visible (and NIR) light and the most critic part of the vision system. This is more evident when illuminating objects with complex forms and/or with high reflectance levels, which is the case of human faces. As discussed in Section 3.2, lighting conditions in FR might introduce considerable variations due to natural or artificial sources of light, whose presence or absence affects the distribution of shadows.

Lighting parameters directly related to the performance of the FR systems design are direction, quality and spectral power distribution of the light. The right choice of these involved parameters is one of the keys to high quality image acquisition and accurate image analysis. Hence, optimized illumination system will reduce both, programming and image analysis computer burden. Figure 4.7 depicts different acquisition results as function of the standard lighting directions as well as the different quality (-nature-) of the light.

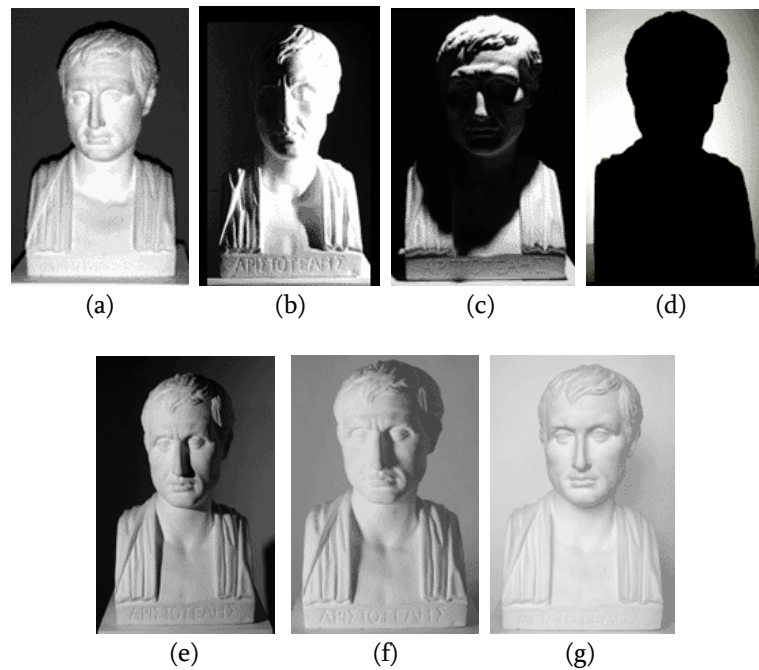


Figure 4.7: Direction and quality of the light: 1st row: frontal (a), lateral (b), zenithal (c) and background light (d). 2nd row: hard (e), diffuse (f) and soft (g) light.

4.4 Near-Infrared Imaging

In the same way than their familiar visible images, NIR and SWIR bands have a similar behavior. The image is obtained by the reflecting light of the scene, which means that, in analogous way than in visible systems, some external light source with components in the NIR or SWIR reflected-IR bands are required [Gra79]. Furthermore, both NIR and SWIR systems can take advantage of the light provided by the sunlight, moonlight, starlight, and also by a phenomenon called *nightglow* or background radiance, which provides a steady light harvested from the whole night sky with illuminance levels that oscillate from 0,1lx to 10⁴lx relaying on how covered by clouds is the sky [Gon10], skipping in these cases of the need of additional lighting systems.

The use of cameras in the reflected-IR spectrum, originally conceived for military purposes, has been opened to a wide range of applications since second half of the last century. Most typical ones encompass from night vision, near-to-zero lux video-recording, surveillance cameras, machine vision and scientific photography and its use in police investigations [Der54] among the most important ones. Furthermore, since each material type, is, in general, uniquely characterized by its corresponding spectral signature (or reflectance spectra) in the NIR (and SWIR) spectrum, many other applications using such property are fastly emerging, e.g, as for distinguishing different types of plastics (working in the range between 1000nm 1600nm), for discriminating (and separate) glass in function of its color in recycle plants, or even, in crop inspection methods to assert the optimum maturation point and yield of different variety of fruit by using remote IR systems (sometimes driven by drones) are currently being introduced in the agriculture sector.

4.4.1 Acquisition Systems in the Near-Infrared Spectrum

As has been seen in Section 4.3.1, CCD and CMOS sensors used for visible light cameras based on silicon technology are also sensitive to the NIR¹⁰, so they are also useful for near-IR imaging. In addition, when higher performances are required, enhanced CCD sensors provided with HAD CCD chips (a kind of CCD sensor that drastically improves light efficiency by including a basic structure of *Hole Accumulation Diode* for optimal NIR sensitivity) are used. Despite of its high sensibility, the camera is resistant against bright light sources.

Nevertheless, IR light focuses in a slightly different focus respect to the visible light. Thus, ordinary lens not perfectly match visible and NIR light in the same focus point due to a phenomenon known as *chromatic distortion* (that will be discussed in detail in the following Chapter), resulting in blurred images, reduced contrast and overall lower image quality. In this regard, straightforward approaches with angular objectives and the corresponding large DOFs,

¹⁰CCDs approaches with reduced infrared and red response also exist [Tos85]. This approach manufactured on an n-wafer, also reduces smear and dark current.

usually solves the problem. Additionally, setting high f numbers to concentrate light beam, are also accomplished. Following chapter will address such particular solution.

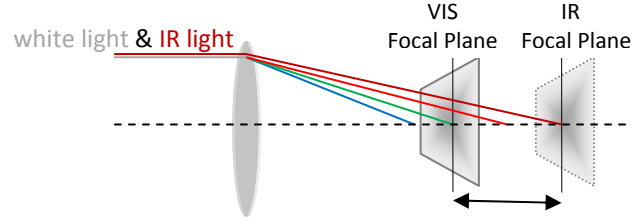


Figure 4.8: NIR focus shift by the focal plane difference among VIS and NIR spectrums.

For more demanding designs, IR corrected lenses yield better performance in cameras that are sensitive to both kinds of light allowing the use of maximum infrared light, while providing clear focus. Figure 4.9 shows a sample of different performance as function of the lens deployed.



Figure 4.9: A comparative of two night vision surveillance cameras provided with a standard lens and an IR corrected lens. (a) Scene acquired in a low-light situation with the standard lens. (b) Same scene and lighting conditions, acquired with the IR corrected lens. (Courtesy of Pelco).

By eliminating the problem of IR focus shift, IR corrected lenses deliver superior image quality in a variety of lighting and camera environments, even better than visible images due to chromatic distortion is lower in NIR where the wavelengths are bigger than in VIS spectrum.

The band gap energy of a silicon CCD is 1.26eV. This energy corresponds to that of light with a wavelength of 1000nm (1 μ m), following the analytic solid-state physics principle given by the expression (4.11). Thus, this analytic expression determines this so-called *cut-off wavelength*:

$$\lambda = \frac{1240}{E_G[eV]}[nm] \quad (4.11)$$

Light having a wavelength longer than this wavelength passes right through the chip as silicon becomes transparent [Pas08] and CCDs constructed from silicon become insensitive. Thus, for detecting IR light over 1000nm other materials should be considered. InGaAs is well adapted to the end of sensibility of silicon; images from InGaAs sensors are comparable to visible images in resolution and detail. Germanium (Ge) by his side offers maximums of sensibility in the range between 1,3 and 1,6 μ m, while silicon-germanium (SiGe) alloys also works well in such spectral bands [Liu05]. Despite of especial situations pointed at the beginning of this section, some additional lighting source is largely required in both environments, indoors and outdoors at night. Arrays of IR LEDs (IREDs), can be appropriate for short distances but to achieve good performances at medium and large distances (over tens of meters or more) dedicated illumination will be required¹¹.

Because NIR and SWIR have a wavelength longer than visible light, energy in these bands is less scattered by atmospheric obscurants (particles suspended in the atmosphere such as fog, dust and smoke) due to the Rayleigh scattering, as described in Section 4.2.1, which involves an extra advantage when dealing with such frequencies. Additionally, due to the lighting required is invisible to humans, users may not be aware that they have been illuminated (-and recognized-) by a FR system. Figure 4.10(a) and (b) depicts the behavior of visible and SWIR remote sensing taking into account this effect, while (c) and (d) exploits NIR capabilities in night vision.

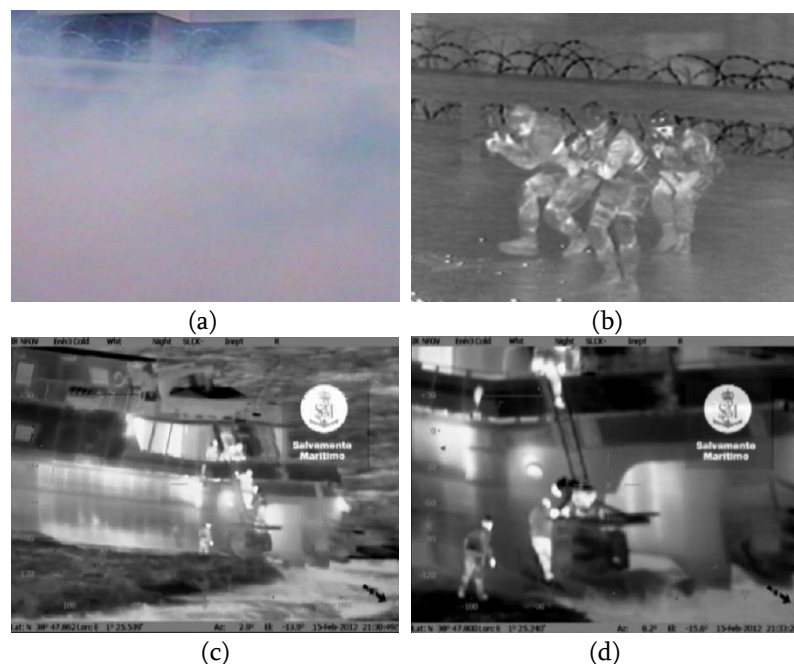


Figure 4.10: Some special applications: Behaviour against scattering: Same scene taken in VIS (a) and SWIR (b) [© FLIR]. (c),(d) Rescue tasks at night near Formentera island, the last winter taking advantage of the nightglow phenomenon [© Salvamento Marítimo].

¹¹Typical medium to long-range systems employ a focused beam from a laser or specialized spotlight.

4.4.2 Near-Infrared Faces

As pointed out in Section 4.2.3 every face of a human being has a characteristic albedo (the particular way of reflecting light), which is function of the color and the texture properties of the skin. Black people for instance, absorb a significantly greater portion of the incoming light than white one [Bey94]. In the same way, tanned skins increase their own albedos. Besides, images acquired in such band are capable to acquire subsurface information of the subject's in zones with low levels of adipose tissue. Figure 4.11 shows a comparative of the same scene acquired with the same camera (sensitive to both, visible and Near-IR spectrum) and provided with a standard lens (not an infrared corrected one). Note that picture in NIR spectrum (b) is slightly out of focus due to the chromatic aberration. Due to the lighting required is invisible to humans, users may not be aware that they have been illuminated (-and recognized-) by a FR system.

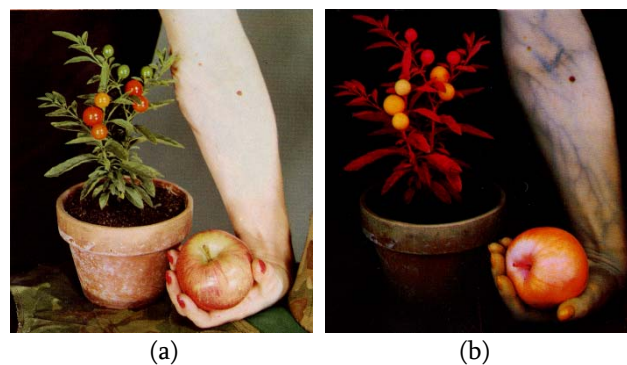


Figure 4.11: Same scene taken in VIS (a) and NIR (b) spectrums with the concerned films (in the second case, an especially sensitized to the region between 700 and 1200nm film, has been used). Strong differences in human skin reproduction can be found, as well as the fruit color reproduction. Extracted from [Gra79]. In addition, veins map may be also distinguished in (b) due to NIR deepness ability.

Human eyes have a high proportion of water compared with dermis and epidermis. In particular, sclera contains much blood and hemoglobin in blood almost fully absorbs all the NIR radiation due to the proximity with the resonance frequency in the microwave range (appears black in the Figure 4.12 (a) [Dau06]. This property is used as an easy eye detector mechanism (The eyes are easy to detect in the image difference between both visible and NIR images simultaneously taken). By contrast, the most part of vertebrate animals are provided of the so-called *tappetum lucidum*, a special reflective texture located in the bottom of the eyes that reflects lost light again to the retinal area, highly increasing their night vision skills, and producing the phenomenon of eye-shine [Oll04].

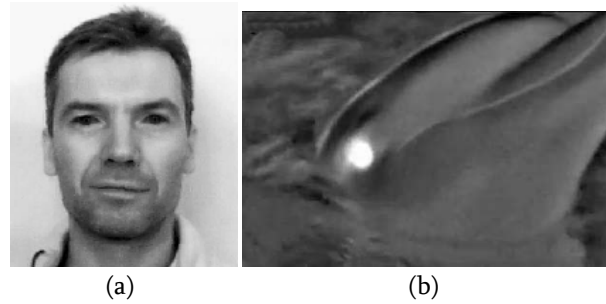


Figure 4.12: Eyes of a human and a dolphin illuminated with NIR light. (a) Sclera appears often as dark as the iris. (b) Strong reflection perceived in a dolphin's eye due to the presence of the referred biologic reflector *tappetum lucidum* system; Frame extracted from *The Cove* documentary [Psi09].

Hair is another specific part, which reveals a lot of additional information, specially related with the use of cosmetic products such grease, tints, etc, due to its different way of reflecting. Figure 4.13 shows two people with and without using tint over the hair, which is not appreciated in the visible spectrum but it is in the NIR one.



Figure 4.13: Different results obtained when acquiring hair in different spectra. Additionally, note that while both sweaters are white (a) and black (c), both NIR images (b) and (d) depict them in white colour due the powerful absorption property of the black surfaces.

4.5 Thermal Infrared Imaging

An imaging system that operates in MWIR and LWIR bands is able to sense the energy that is radiated directly from objects in the scene, and transforms it to a thermal infrared imaging or *thermography*, as is also known. As showed in Figure 4.2, the amount of radiation emitted by an object increases with temperature; this relationship will be used in thermography, to see

such variations in temperature. Afterwards, heat patterns or temperature in objects are inferred from changes in the dedicated sensors. Thus, as an object gets hotter, it radiates more energy and appear brighter to a thermal IR camera. The object's emissivity is another important factor that drastically contributes to determine how bright an object appears in the screen of the camera [Chr00]. Furthermore, an additional false color using an standard color palette (*iron*, *rainbow*, *blue-red*,...etc) assures a significantly better interpretation of the thermography measurement. In this effect, "iron" color palette is the most commonly used, because it is more intuitive for humans due to the colors code follow the same change of color than the iron when gheathed. Thus, white and orange areas are the warmest and the dark ones are the coldest. This thermal IR imaging gives us a different view of the visible world as well as information that we could not get from a VIS picture.

Additionally and in analogous way than in visible spectrum, a thermal contrast is also required in order to obtain good quality thermal images. Thermal contrast is the change in signal for a change in target temperature [Fli08]. Figure 4.14 shows an example of a scene with a high visible contrast and a poor thermal contrast, resulting in an extremely low quality thermal image.

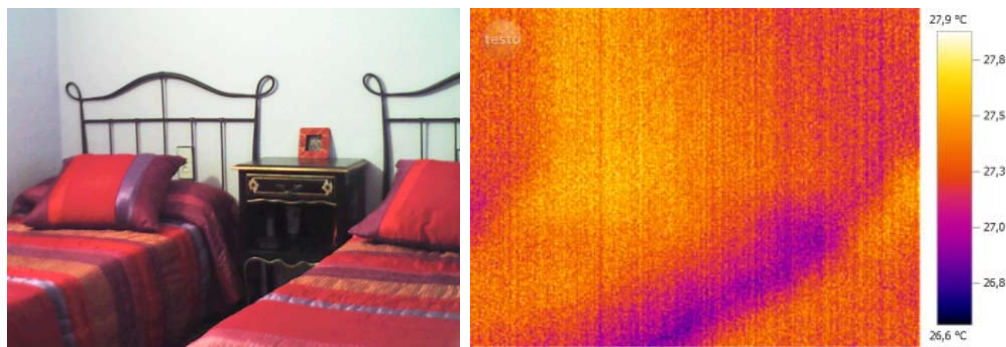


Figure 4.14: A sample of a simultaneously scene taken in visible and thermal IR spectrum.

Typical applications of this extremely powerful technology comprise high end scientific research and development, medical and veterinary imaging (febrile temperature control¹² and detection of a width range of diseases, among the most important ones), materials science, preventive and predictive maintenance and quality control of electrical, mechanical and industrial processes, energy conservation, building inspection (identifying discontinuities of insulation, thermal bridges and air leakage paths) and defense, among the most important ones. Figure 4.15 shows a sample of many of these applications.

¹²Since the outbreak of SARS (Severe Acute Respiratory Syndrome), in Southeast Asia in 2003, it had become clear that thermography could be used to identify people with a high temperature who may be among a number of travellers, especially those arriving or departing from ports and airports [Rin08].

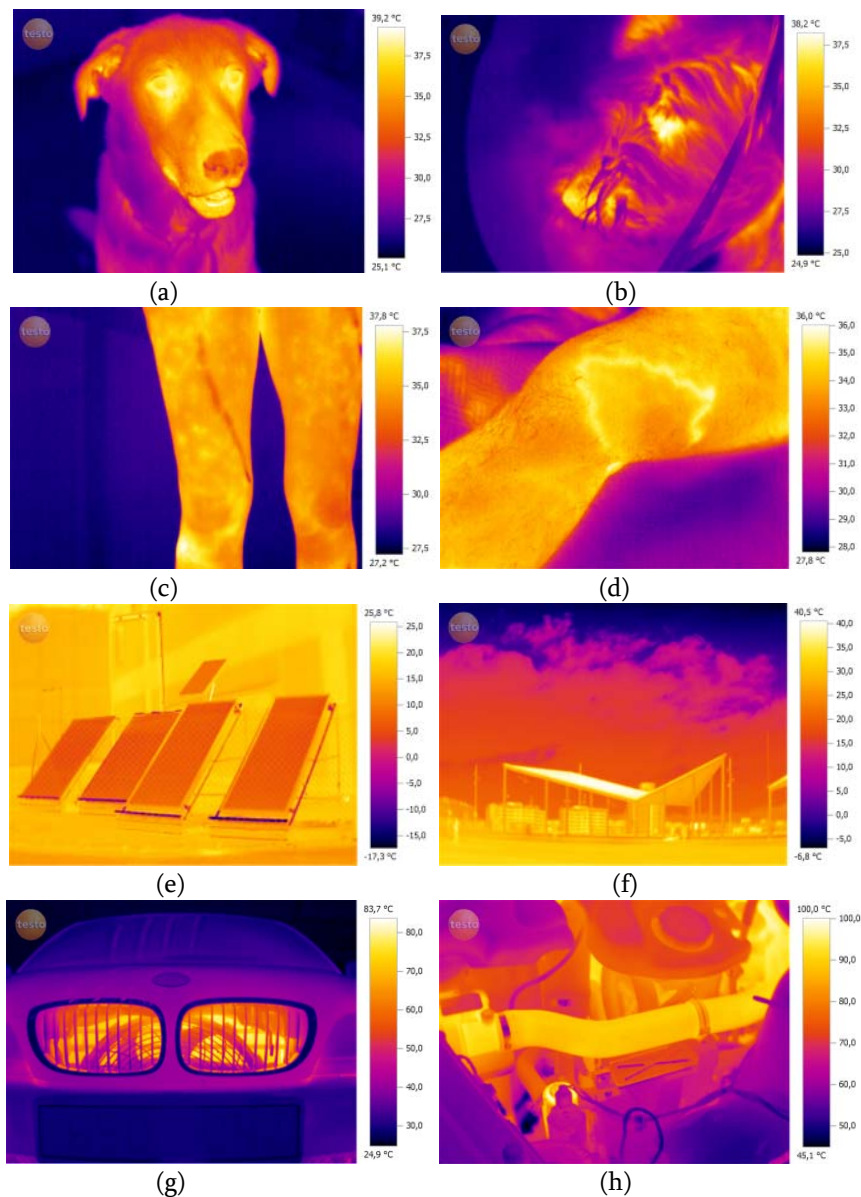


Figure 4.15: A sample of some applications in thermography. 1st row and 2nd rows: Veterinary & medical support: A couple of healthy (a) and ill (b) dogs (Tango and Fosca). (Notice that nose in dogs behaves as a thermal regulator and is also reliable indicator of the health of the animal). Non invasive detection of cold (c) and warm (d) pathologic veins. Last two rows: Inspection of some devices: solar cells at CREA (Centre de Recerca d'Energies Alternatives) (e) and Forum space in Barcelona (f) and a radiator (g) and the engine (h) of a car. [Images acquired with the TESTO 882 thermal imager; resolution of 320x240 and NETD <60mK].

4.5.1 Acquisition Systems in the Thermal Spectrum

Thermal IR cameras, also known as *thermal imagers*¹³ [Chr00], are provided of fully passive detectors (see Section 4.2.4), requiring no any external illumination as happened in the two previous cases. Furthermore, since over than three orders of magnitude separate visible photon energies of the thermal photon ones, detection techniques will change accordingly. For this reason, special purpose sensors based on *FPA* (Focal Plane Array; consisting of an array of one kind of IR sensing detector at the focal plane) are required, which are normally highly-priced. Two kinds of IR detectors depending of the type of the sensor being incorporated in FPA, exists: *thermal* detectors based on *microbolometers*, and *quantum* (or *photon*) detectors based on the photoelectric effect [Fli08, Bar00].

First ones are generally uncooled sensors (UFPA; Uncooled FPA), specially designed to work in the LWIR band where the uncooled detection is easy (are mostly temperature stabilized by a Peltier element), while second ones concerns to enhanced cooled¹⁴ sensors, cooled to cryogenic temperatures which are highly most sensitive to discriminate small temperature differences in scene temperature [Chr00] and are generally designed to acquire images in both, MWIR or LWIR bands. These second thermal imagers tend to be expensive and are strongly limited in complexity. Therefore, heat patterns or temperature in objects are inferred from changes in the resistivity of sensors based on microbolometers or analyzing the photoelectric activity in quantum detectors [Fli08, Bar00]. Table 4.1 summarizes most important properties of both discussed sensors.

As discussed at the beginning of this section, thermal sensors depend on the temperature and emissivity of the target, as well as intrinsic properties of the sensors being the most relevant spatial resolution, measured in pixels and thermal sensitivity, also known as NETD (Noise Equivalent Temperature Difference), measured in miliKelvins.

¹³If we make a review of literature dealing with infrared technology, we will find that there are several different additional terms used as synonyms of the original term of “*thermal infrared camera*”: *thermal imager*, *infrared camera*, *thermograph*, *thermal camera*, *FLIR* (*Forward Looking IR*), *thermovision*, *thermal imaging camera* and *thermal video system*, among the most important ones.

¹⁴Cooled sensors are packaged with the so-called *crycoolers*, majorly Stirling cycle coolers or Joule Thompson coolers that decline the temperature of the sensor to cryogenic levels in order to reduce the dark current and therefore, the thermal noise to negligible levels.

Detector	Material	Pros	Cons
Thermal: <i>Microbolometer</i>	<ul style="list-style-type: none"> • Metal: VOX¹⁵ • SC: amorphous Silicon 	<ul style="list-style-type: none"> • Operable in ambient T • Low cost • No cooling requirements 	<ul style="list-style-type: none"> • Low sensitivity (30mK is an achievement) • Slow response (time ct ~10ms)
Quantum: <i>Photodetector</i>	<ul style="list-style-type: none"> • InSb [MWIR] • MCT¹⁶ [MWIR and LWIR] 	<ul style="list-style-type: none"> • High sensitivity (lower than 30mK) • Fast response 	<ul style="list-style-type: none"> • High priced • Sensitivity wavelength dependence (Increase when decreasing in wavelength) • Cooling requirements

Table 4.1: Properties of the FPA detectors.

Thermal imagers are usually provided of an additional camera in order to simultaneously acquire same scene in visible and thermal spectra. However, an effect called *parallax error*¹⁷ may appear between images taken in both spectrums, at small working distances between camera and scene, in several cameras. (Both TESTO cameras used in this research work exhibit this drawback, while both FLIR cameras used for specific tests provide parallax-free images).

The optical reason is that, both, thermal and visible cameras have different spatial positions, viewing objects from slightly different angles and providing the concerned two different field of views (FOV) resulting in small coincidence areas (see Figure 4.17) and also producing problems of alignment when fusing both spectral regimes. This error highly increase when reducing the distance between the scene and the camera as can be seen in the sequence of images taken from different distances showed in Figure 4.16.

¹⁵VOX (Vanadium Oxide).

¹⁶MCT refers to the HgCdTe ternary alloy (Mercury Cadmium Telluride).

¹⁷This parallax effect is a photographic legacy problem, when dealing with twin-lens reflex cameras (TLR) [Jac00] as well as in compact cameras provided with direct visors.

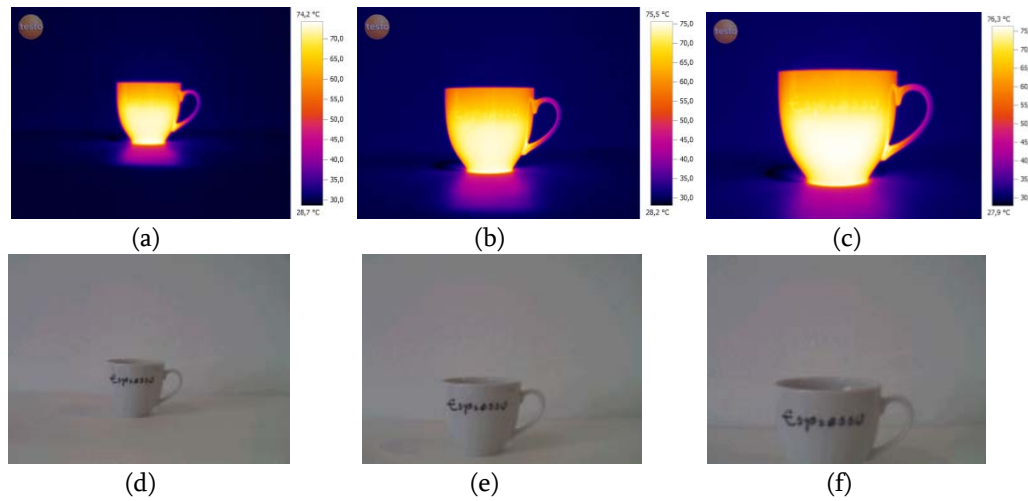


Figure 4.16: Same scene simultaneously taken in visible and thermal IR spectrum by a thermal imager, from three different distances: 1st row: TIR images taken at a distance of: 40cm (a), 20cm (b) and 10cm (c). 2nd row: VIS images taken at a distance of: 40cm (d), 20cm (e) and 10cm (f). [The minimum focusing distance of the used objective was of 10cm]. Notice that the inside of the cup can be appreciated with the image taken with the visible camera as a result of the referred effect.

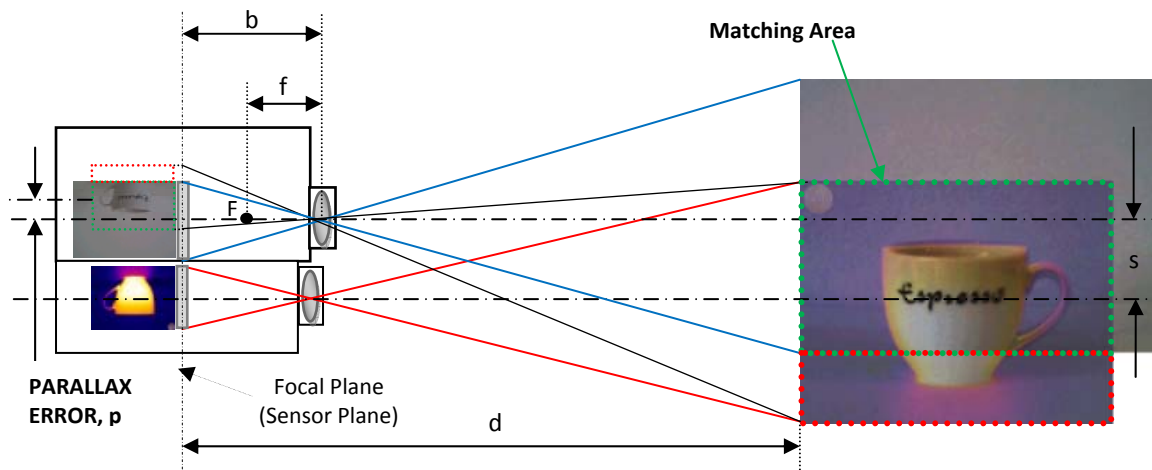


Figure 4.17: Optical geometric description of parallax error for the worst case discussed in Figure 4.16: (c) and (f) images.

The parallax error, p , can be therefore analytically expressed as:

$$p = \frac{sb}{d} \quad (4.12)$$

And taken into account the following equality:

$$\frac{1}{f} = \frac{1}{d} + \frac{1}{b} \quad (4.13)$$

then, we can obtain a more general expression of the parallax error including all the involved parameters as follows:

$$p = \frac{s}{\left(\frac{d}{f} - 1\right)} \quad (4.14)$$

where, s is the displacement, or *shift* between both involved optical axes (OA), b is the lens-to-image distance

Optical axis: Horizontal axis perpendicular to the sensor plane.

You may found all the parameters defined are graphically described in Figure 4.17.

4.5.2 Thermal Infrared Faces

Briton Robert Boyle was the first who discover in the XVII century, the body temperature was approximately constant over the environment, currently being accepted as normal, the temperature of 37°C (310K, 98,6°F)¹⁸. Such heat-preservation regardless of the external environment, called *endothermy* or *homeothermy* is due to a thermoregulation process that mainly consists in the interaction of three thermal regulation systems: skin blood flow (SBF) regulation, perspiration and thermogenesis [Kak02]. In particular, SBF kind of regulation plays an important role in the facial zone of the human body. A recent study carried out by Bergman and Casadevall [Ber10] hypothesizes that the poorly understood origin of endothermy in mammals would be to prevent certain infections (for every degree the temperature rises, the body rejects 6% of existing microbes). Specifically, they analyzed the tradeoff between the metabolic costs required to maintain a body temperature and the benefit gained by creating a thermal exclusion zone that protects against environmental microbes such as fungi. Their result yielded an optimum at 36,7°C; beyond such temperature, benefits do not outweigh the metabolic cost of maintaining thermal equilibrium, which coincides with reality.

It is easy to compute by means of the Wien's displacement law, analytically described in equation (4.5), that human body, at a normal body temperature of 37°C (310K), radiates

¹⁸Temperatures of 38°C and above located on the face, are considered to be febrile.

energy in the in the second window of the thermal spectrum with a maximum peak emission at 9,35 μ m. Since the human body emits this kind of energy as heat, it follows to harvest this energy by means of thermography technology from a biometric perspective. In this respect, the average emissivity of human facial skin needs to be given due to their greybody nature. Fortunately, temperature of humans remains almost constant, not producing variations in its emissivity as other material discussed in Section 4.2.2 that have emissivity levels that vary with its wavelength and therefore, with its temperature. Therefore, the estimate of such parameter in the MWIR and LWIR thermal sub-bands is given by the expressions (4.15) and (4.16):

$$\varepsilon_{MWIR}^{skin} > \frac{skin_{meanEnergy}}{\int_{\lambda=3\mu m}^{\lambda=5\mu m} W(\lambda, 310K) d\lambda} = 0,91 \quad (4.15)$$

$$\varepsilon_{LWIR}^{skin} > \frac{skin_{meanEnergy}}{\int_{\lambda=8\mu m}^{\lambda=14\mu m} W(\lambda, 310K) d\lambda} = 0,97 \quad (4.16)$$

According to the values given, FR in the thermal IR favors the LWIR band, since LWIR emissivity (and emission) is much higher than that in the MWIR. Then, and taken the emissivity of 0,97 for the LWIR case, we can easily compute according to Boltzmann law, the total radiated energy by a human's body, resulting to be about 450W at a distance of one meter, which is certainly a considerable emission of radiation.

Thermal image of a face, which is usually called *facial thermogram*, concerns to the skin surface temperature distribution produced by the underlying vascular system in the human face when heat passes through the facial tissue and is emitted from the skin [Hon98] and can be currently acquired by a thermal imager. This distribution is considered a unique facial signature as pointed out in Section 3.4.2. Thus, while general temperature can rise in thermal images, the relative difference between different portions of the face remains similar, because the hottest point will always be related to the vein positions, and these remain the same. Nevertheless, reproducibility of this thermal distribution of the face as well as the rest of the human body is not completely assured as is described in [Zap08] due to both, measurement environment and physiological variability of the blood flow. Similarly, the evaluation of the measurement results of a human thermal model developed by Kakuta et al., [Kak02] suggest that the skin surface temperature distribution does not change in the forehead -and neck-, as it does in the rest of the face. The study point out that chest and abdomen are also invariant zones in terms of thermal distribution.

Adipose tissue also plays an important role acting as an insulator, due to its low thermal conductivity. In this respect, face is considered again a good thermal pattern emitter due to its relatively free of adipose tissue compared with other parts of the human being more sensible to gain and/or lose weight. In particular, the forehead, the temples and the frontal zone of the nose are particular good zones in terms of radiating thermal energy. Figure 4.18 showcases the differences in the face shape that can be seen to the naked eye when the subject undergoes a

considerable increase of weight. As expected, the incremented volume at zones previously detailed is negligible.

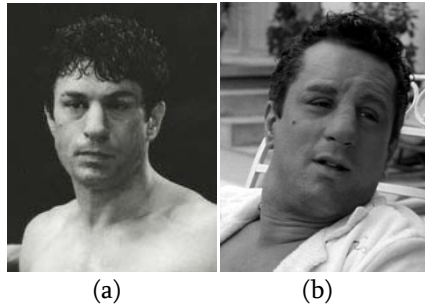


Figure 4.18: The actor Robert de Niro in different moments of the film *Bull Ranging* (He gained weight 27Kg to play with great effect the boxer Jack LaMotta). Minimum differences can be appreciated in forehead-temples and frontal zone of the nose among both images.

On the other hand and as major drawbacks, apart from the low discriminability and reproducibility of the facial thermogram, and the concerned with thermal imagers' technical issues already discussed, include the following:

- Thermal cameras are still expensive, and some minimum background and experience is required by the users. Particularly, it is more difficult to focus an image than with a visible camera.
- Thermal calibration is required in ambient temperature (or activity level may change thermal characteristics).
- Effects of thermal losses (convective, radiative and conductive) affect thermal contrasts.
- The human specific parts, as nose, ears, hands and lower extremities [Uem88], are thermoregulators parts of the human body, often being colder than other parts of the body due to less blood flow and more convective cooling. This fact produces a high variance as function of the temperature of the surrounding and consequently providing minimum information as well as mouth and eyes areas, as can be found in literature. In this last respect, we have experimentally checked with our most sensitive available camera (TESTO 882; 60mK) that the temperature is not exactly the same in all parts of the eye's surface, having a slight thermal variation along the eye's surface, as is graphically reported in Figure 4.19.

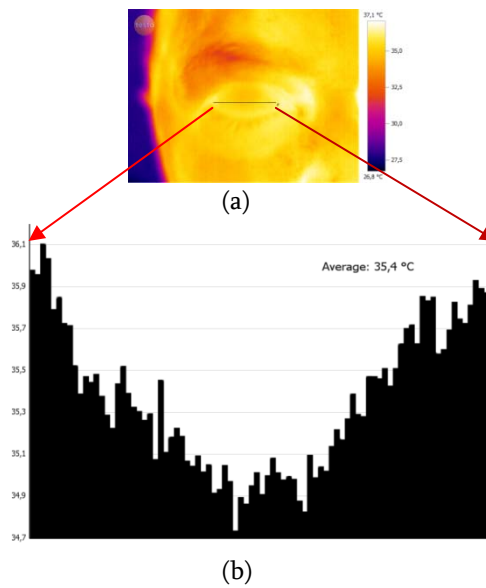


Figure 4.19: Temperature profile of the eye of the author's brother. (a) Thermogram of an eye. (b) Temperature profile graph. A Maximum of 36,1°C and a minimum of 34,7°C are detected, while the average temperature is 35,4°C.

- Glasses¹⁹ exhibit a fully different behavior as function of the spectrum, being transparent from the VIS to the NIR spectrum and fully opaque beyond 3 μ m approximately as shown in Figures 4.20 (a), 4.20 (b) and 4.20(c) [Vol10]. Beyond this wavelength, the behaviour of any kind of glass is fully opaque in a similar way that the blocking effect of sunglasses in visible spectrum, becoming the great last limitation [Esp12]. Equally, contact lenses, behave in a similar way than glasses, as shown in Figure 4.20(d). Due to the same reason TFR is not appropriate for recognition of vehicle occupants (because of glass and also due to speed).

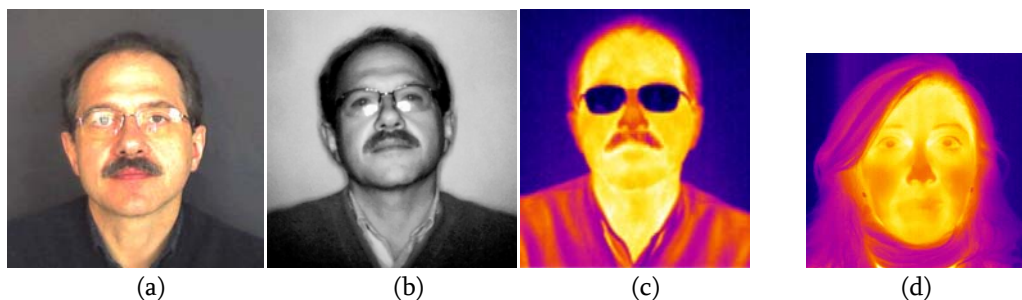


Figure 4.20: A set of different images of an author's colleague, wearing glasses in VIS (a), NIR (b) and TIR (c) spectrums. Facial thermogram of the author's sister wearing contact lenses (d).

¹⁹IR transmission of glass is function of its composition and thickness, behaving inversely proportional with respect to this second factor.

Nevertheless, although glasses are considered an undesirable blocked object when dealing with thermal images, it is also true that glasses produces undesirable reflections in VIS and NIR spectrums, especially when they are not provided with antireflection glass, as can be appreciated in Figures 4.20(b) and 4.20(c). This is one of the main reasons that have lead to finally reject the inclusion of people wearing glasses in our acquired multispectral data set, as is further discussed in Chapter 6. Sunglasses by its side will produce the same negative opaque behavior in all the three analyzed band ranges.

Irrespective to the above considerations, and, recovering the initial discussion addressed in the introduction Chapter, technical aspects related with thermal IR cameras, especially spatial resolution and even more, thermal sensitivity, become critics when dealing with the acquisition of human faces, being inoperative NETD values larger than 100mK for such purpose²⁰. Figure 4.21 shows a visual comparative between same subject taken with the two available thermal imagers.

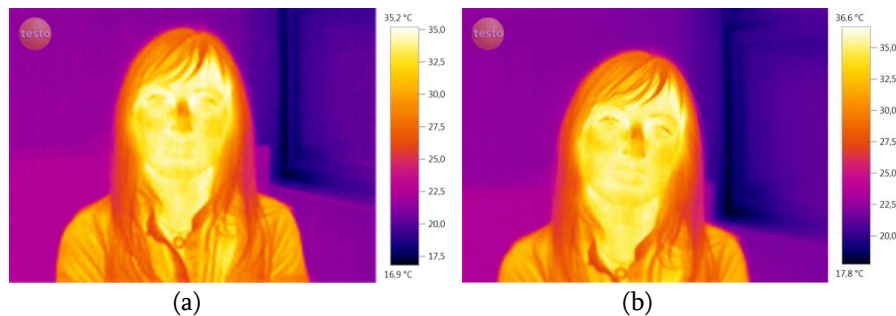


Figure 4.21: Same subject taken with two different thermal imagers with a spatial resolution of 160x120 pixels and NETD of 100mK (a) and 60mK (b). In the second sample, a better degree of detail is perceived (see both fringes, e.g). Additionally, note that due to hair has a lower emissivity than skin, it appears darker than skin even when both have exactly the same temperature.

Despite facial thermograms are due to the map of arteries and veins, they cannot be appreciated in current LWIR images. Vasculature of a face will be seen in detail in the next generation of high resolution thermal IR imagery, being a valuable information to assist new trends in Thermal Face Recognition (TFR) systems and might be a promising direction in this field [Pro00]. Figure 4.20 shows a current thermal image of a human where the jugular vein is already sensed.

²⁰The nerve endings in human skin can respond to temperature difference as little as 9mK [Flu09].

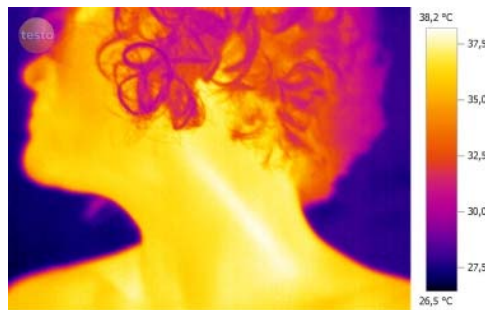


Figure 4.22: Thermogram of a human being, appreciating the jugular in the neck.

Notwithstanding, even if the equipment accuracy is relatively lower than that of a visible image, a TFR system can still achieve satisfactory performance. With regard to its interesting properties concerning biometric applications, a few of which are outlined here:

a) About Facial Thermograms:

- Due to the fact that the acquisition system is fully passive, it does not carry any health risks.
- They are not affected by illumination, shadows, etc. In fact, they can perfectly be acquired in fully darkness (and, from another point of view, no any external illumination system is required).

b) About the inherent biometric authenticator properties:

- Are more difficult to fake than visible ones, because artifacts, makeup, disguises and other specific devices cannot imitate the flow of blood through the veins, which is directly related with the heat emission.
- They are also robust to plastic surgery, which does not reroute the vein distribution under the skin.
- They have the benefit of providing an aliveness detection method without the need of any additional mechanism and/or image processing.
- Thermal images can differentiate twin brothers because heat emission is related to the veins distribution under the skin and this is different for each person, even for twins.

c) And specifically in the FR systems:

- Due to its high temperature, a higher thermal contrast against the background is easily achieved, making easier face detection task in not-controlled environments and specially, in fully darkness conditions, being this goal a big challenge in visible images as pointed in the third Chapter.
- If both facial images are simultaneously available (in the visible and in the thermal spectrum), and considering the parallax error negligible, it is possible to use the thermal

information as a consistent “mask” in pixel-level fusion methods in order to easily detect faces in the visible spectrum [Akh10].

- In the same case than when operating with NIR images, as discussed in Section 3.4.1 the users are not aware that they have been illuminated by the system.
- Due to their large wavelengths, this light is not affected by low levels of obscurants in the atmosphere like fog, dust and smoke.

Chapter 5

On the Relevance of Focusing in Thermal Imaging

Sorprenderse, extrañarse, es comenzar a entender.
José Ortega y Gasset.

This Chapter is devoted to the study of thermal image focusing. Section one addresses the problem of focusing images in the visible and the thermal spectra, providing a basic understanding of the optical principles underlying. Second section briefly outlines the different methods for focusing images manually or automatically. In section three, an overview of the most well known mathematical tools to assess the quality of an unknown visible image degraded by blurring is provided as well as the description of a new thermographic image database suitable for the analysis of automatic focus measures. Related experiments and conclusions are also included. The section ends with the extension of some methods for analyzing the focus of thermal images, being one of the contributions of this thesis. Finally, last section of the chapter deals with the analysis of the contribution of temperature of the objects when focusing thermal images and concentrates the results to the facial thermograms focusing problem.

5.1 The Focusing Problem

Because of the lack of suitable introductory treatments about the focusing of thermal images, we propose to convey as much physical understanding as possible using only the minimum required amount of mathematics.

The phenomenon of *refraction* is vital for its application to optical. When a light beam obliquely traverses a transparent material to penetrate another, undergoes a change of direction known as refraction. [Fon90]. In the seventeenth century, Isaac Newton demonstrated in his

treatise on Optics that the index of refraction of materials depends on the visible light. The same holds for IR radiation. Thus, and as discussed in Section 4.4, due to visible and infrared lights have the same reflection, refraction and transmission properties, lenses for thermal imagers will be designed in a similar to those of a visible cameras.

Therefore, and as well as in visible image acquisition systems, an optical system capable of focusing all rays of light from a point in the object plane to the same point in the focal plane is desired, the same goal is also desired when dealing with infrared images, in order to get clear and focused images in infrared spectrum. However, as is well known in optics, optical properties usually depend on wavelength, leading to major kind of lens *aberrations* (imperfections in the optical formula of a lens that put a stop to perfect convergence). Maxwell established how fundamental the problem of aberrations is in mid-eighteenth century. He proved that no optical system can produce ideal imaging at all focal depths¹, because such a system would necessarily violate the basic mechanisms of reflection and refraction [Ng06]. Additionally, deviations due to *diffraction* become more present in infrared light, due to the longer wavelength, drastically reducing the capability to focus IR images. As a result, IR light from the desired point is blurred over a spot on the image plane, reducing contrast and resolution. In this section we will specifically discuss this problem beyond the visible and the NIR spectrum, since referred aspect in the NIR spectrum has been appropriately discussed in Section 4.4.1.

We will firstly consider the *achromatic aberration*, as the most critical aberration aspect when dealing with broadband spectrum images (which is the case of MWIR and LWIR thermal images). Secondly, and not least, we will focus on diffraction effects, highly dependent with the wavelength, which also spread the image, even if the optical system is free of any kind of aberration (chromatic, spherical, astigmatism...).

- a. *Chromatic aberration*. Chromatic aberration [Jac00] is an undesirable optical effect that promotes the inability of the lens to focus all different wavelengths of the beam light (-all colors in visible light-) at the same focal point [Zak93]. This effect, sometimes also called *achromatism* or *chromatic distortion*, is due to the *spread dispersion phenomenon* concerning the *refractive index* variation with wavelength². Normal lens shows normal dispersion, that is, the index of refraction n , decreases with increasing wavelength. Thus, the light beam with longer wavelength is refracted less than the shorter wavelength one. This behavior produces a set of different focal points, as can be seen in Figure 5.1.

¹Focal Depth (FD) or Depth of Focus, is defined to be the tolerance of the film's displacement within the camera, without altering the definition of the image of a flat object.

²Both of these phenomena occur because all optical signals have a finite spectral width, and different spectral components will propagate at different speeds. One cause of this velocity difference is that the index of refraction of the lens core is different for different wavelengths. This is called *material dispersion* and it is the dominant source of chromatic dispersion in optical lens.

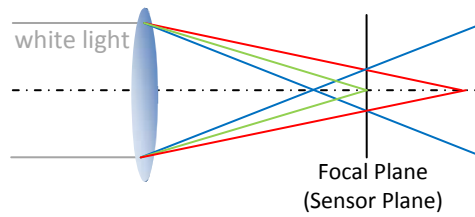


Figure 5.1: Optical representation of chromatic aberration. Note that longer wavelengths focus behind the focal plane, while smaller ones, focuses before the focal plane.

Chromatic aberration is generally broken into two categories: Axial chromatic aberration present in most of normal and high focal length lenses, which manifest itself as blurring and lateral chromatic aberration which manifest itself as geometric distortions and is mainly present in angular lenses [Bou92]. In order to reduce the impact of such aberrations, especially the first one, special low and extra-low dispersion glasses (ED) exist, which present a small variation of refractive index with wavelength. Additionally, a number of high performance optical lenses also exist such as *achromatic* and *apochromatic* ones, for most demanding imaging corrections. For a review of chromatic aberration and related issues in lenses design, see [Lai91, Sla80].

In any case, the correction tasks in the VIS spectrum are reasonable to achieve due to the short range of wavelengths to deal with. This is not the case in the MWIR and LWIR operating ranges, where thermal IR sensors measure simultaneously over broadband wavelength. Thus, while the change is 400nm between the violet and red end of the EM spectrum, in both the MWIR and LWIR spectra the wavelength ranges are 2000nm and 6000nm, respectively. Nevertheless, although the broadband, another constraint concerned with the coupling between large wavelength and low refractive index will be a challenging in itself beyond of the chromatic aberration correction. To promote the required refraction to deviate and to accurately converge any IR light comprised in the range, an IR transparent material with high refraction index is required for the design of lenses that might not otherwise be possible [Gre07].

Germanium (Ge) material has the highest index of refraction of any of the existing IR transmitting materials (around 4,0 from 2-14 μ m, while normal values for focusing visible images ranges from 1,5 to 2) and also offers a low dispersion, attending both constraints. Germanium³ also provides more than the required transmission in the desired spectral

³Germanium is subject to called *thermal runaway* effect, meaning that the hotter it gets, the more the absorption increases. Pronounced transmission degradation starts at about 100°C and begins rapidly degrading between 200 and 300°C, resulting in possible failure of the lens. Thus, must important weakness of this solution in thermal scenario, is the narrow temperature operating range which is the case of face acquisition.

band as showed in Figure 5.2. Another useful high performance IR transmissive material is ZnSe.

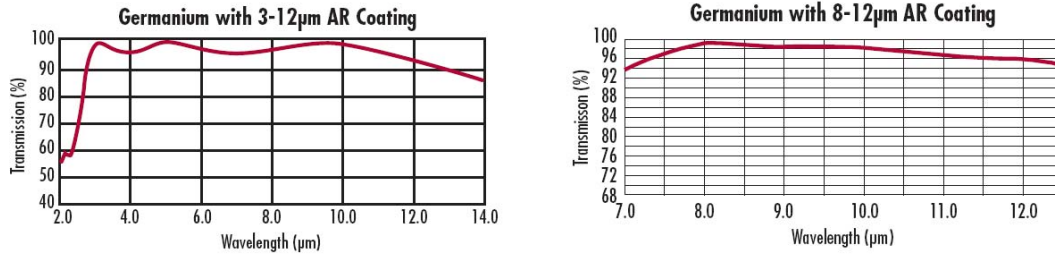


Figure 5.2: Germanium transmission curves. (© Edmund Optics).

- b. *Diffraction effect.* Diffraction is an optical phenomenon, due to the wave nature of light, which can limit the total resolution of any image acquisition process. Usually, light propagates in straight lines through air. However, this behavior is only valid when the wavelength of the light is much smaller than the size of the structure through which it passes. For smaller structures, such a gap or a small hole, which is the case of camera's aperture, light beams will suffer a diffraction effect [Zak93, Ng06] caused by a slight bending of light when it passes through such singular structures. Figure 5.3 exhibit the referred phenomenon:

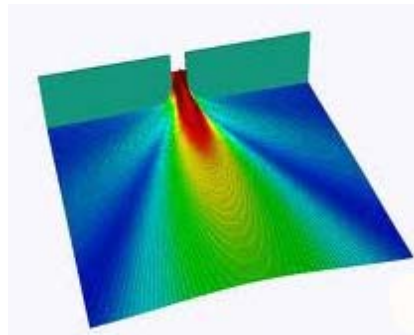


Figure 5.3: Wave after diffraction through a gap.

Due to this effect, any image formed by a perfect optical lens of a point of light, not correspond to a point, but to a circle called *Airy disc* in honor to its discoverer, George Airy (not to be confused with the *circle of confusion*), and determines maximum blur allowable by the optical system [Fon90]. Furthermore, the diameter of this circle will be used to define the theoretical maximum *spatial resolution* of the sensor and will be given by the following equation:

$$d = 2,44\lambda \frac{\nu}{D} \quad (5.1)$$

where λ is the light wavelength, ν is the distance from the image to the lens and D is the effective aperture diameter. The Equation (5.1) can be generalized as (5.2) when the system is working slightly far of the minimum focal distance, and N being, the lens f number. Then:

$$d = 2,44\lambda N \quad (5.2)$$

Solutions in this field only consider diffraction limited optical systems that provide a measure of the diffraction limit of a system. Knowing this limit can help to avoid any subsequent softening. Table 5.1 depicts the resulting *pixel size* designs using the expression (5.2) for three different f -numbers computed for the boundary frequencies in each thermal band.

	MWIR	Low Trans	LWIR	
f/ λ	3 μm	5 μm	8 μm	14 μm
$f/1$	7,32	12,2	19,52	34,16
$f/2$	14,68	24,4	39,04	68,32
$f/4$	29,28	48,8	78,08	136,64

Table 5.1: Resulting maximum pixel sizes given in μm^2 . Special cases of LWIR and $f/1$ and $f/2$ have been highlighted. (Thermal infrared range has been included in order to better understand the involved results).

Typical available pixel sizes for MWIR and LWIR range from 20 to 50 μm^2 (-understanding squared pixels-), while less than 2 μm^2 may be found for visible spectrum.

Two main conclusions can be extracted from the above equations and the results depicted in Table 5.1:

- Diffraction increases (-and spatial resolution decreases-) when the wavelength increases: This first conclusion clearly reveals the size of the sensing elements and the diffraction blur are two of the major differences between THIR and visible images systems due to the radically different wavelength values.
- Diffraction also increases when increasing f -number: Said in another way: *A big f -number results in a larger optics blur due to diffraction*. Thus, this second conclusion leads any IR vision system to be provided with lenses with *low f -numbers* (implying large apertures). Additionally, low sensibilities also forces thermal imagers to work with low f -numbers in order to collect enough thermal energy. In this sense, fixed apertures near to $f/1$ in thermal cameras with a sensibility of 100mK have traditionally

been required⁴ [Gre07]. Be aware on the other hand that is not difficult to see visible cameras with $f/22$ and over (or even $f/64$ in optical bancs as the used by Ansel Adams in the first half of the 20th century!).

This second restriction seriously determines a closely related parameter: the *depth of field* (hereafter DOF) -the range of distances that appears acceptably sharp in the resulting image-. As is well known in optics, DOF depends on three mainly parameters, in order to arrange any possible focusing situation: Aperture ($f/$ number), Focus distance and Focal Length (FL), being the last one negligible compared two first ones. In this sense, low f -numbers constraint reduces blur of objects away from the object plane, but also results in shallows DOFs, as depicted in Figure 5.4. By contrast, setting high f -numbers implies larger DOFs as well as concentrating the light beam forcing it to pass through the lens to the center, where the curvature is lesser and reducing chromatic aberration just discussed [Jac00].

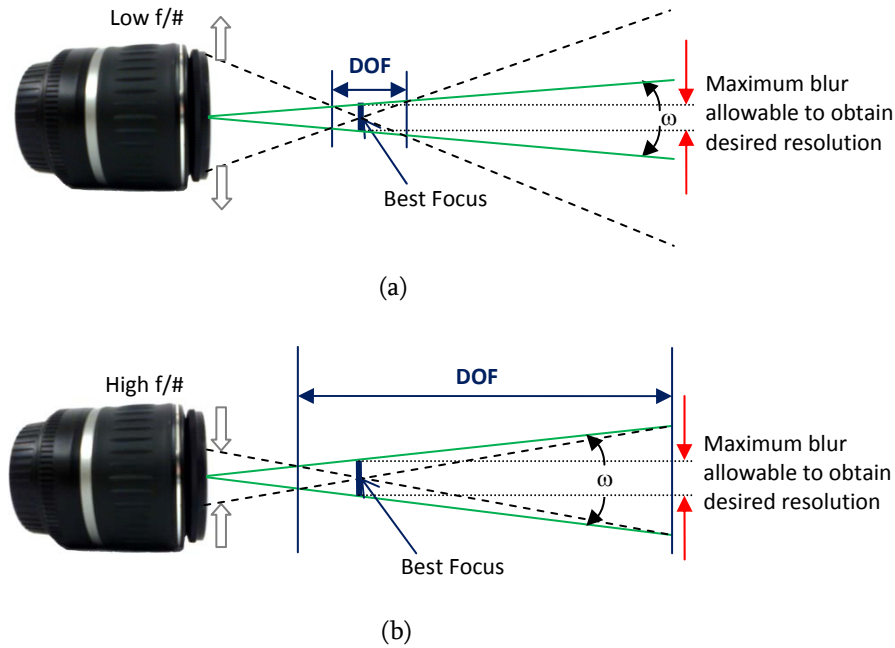


Figure 5.4: DOF as function of $f/$ number: (a) High $f/$ (small aperture) involves large DOFs, while (b) Low $f/$ (large aperture) involves short DOFs.

Finally and assuming that a diffraction limited system is used, an approached expression of the DOF can be given as follows:

⁴Future more accuracy sensitivities will play an important role in order to alleviate diffraction problem: Take into account that high sensitivities, will allow reduced f numbers (also increasing depth of field) to collect the same energy.

$$DOF = \frac{D^2}{4\lambda} \quad (5.3)$$

Note that this approached expression is only function of the aperture diameter and the wavelength. Once evaluated, another important conclusion can be extracted: *DOF also decreases when increasing wavelength*. This is the last major difference between VIS and IR acquisition systems and uncovers why thermal imagers may not focus all planes of the acquired scene. Thus, thermal imagers will require higher distances to scenes than VIS and NIR vision systems in an attempt to fully focus the overall scene. The challenge becomes more difficult to solve when more than one object located at different distances in scene demands to be focused, especially due to the constraints in *depth of field* design. This assertion has also led us to collaborate in parallel in carrying out a multifocusing thermal image fusion proposal in [Ben12] in an attempt to provide a novel solution to this last described problem and be able to extend the solution to multiple faces. An additional discussion of the referred work will be provided in the future research Section of the last Chapter of this dissertation.

5.2 Focusing Approaches

Several approaches exist for focusing an image, manually or automatically. In the former, called Manual Focus (M) modality, the user must trim the optics till he obtains a satisfactory result. Trimming can be fully manual or helped by an engine that adjusts the focus. In each case, the user must decide the optimal focus position. This process can be tedious and complicated for users with sight problems (such as myopia) or lack of sufficient skill. In addition, typical digital camera screens do not provide enough resolution to determine fully whether the image is focused or blurred.

Automatic focusing (or Autofocus-AF as is also known) systems use electronic analysis to perform the task without user intervention. Automatic focusing can be split into two main categories. *Active systems* rely on an integrated sensor. Normally, the sensing equipment will emit an IR or ultra-sound signal and wait for its reflection. The distance between object of interest and imaging equipment is calculated from the time between signal transmission and the receipt of the reflected signal. Trigonometric triangulation is required for IR based systems. On the basis of the estimated distance, the camera parameters such as the lens position and aperture setting are adjusted accordingly in order to attain the best focus. *Passive systems* perform a set of focusing measurements on the acquired image. This can be carried out in the spatial or frequency domains. Spatial domain techniques require fewer operations and are more suitable for real time applications. The procedure is quite straightforward in the frequency domain because focused images contain sharp edges, which are associated with high frequency content. Table 5.2 summarizes current focusing techniques.

There was no thermal camera which incorporates an automatic focusing facility at the moment when analysed and published our study [Fau11]. In our case, the TESTO 880-3 and TESTO 882 thermal cameras used for the experiments, incorporate an engine able to move the focus, but the trim must be performed visually (motorized manual focus, as described in Table 5.2). A

focused image can be obtained by making small adjustments to the focusing knob until edges appear visually sharp.

Focus	Kind	Domain	Pros	Cons
M	Manual/ motorized	Spatial	<ul style="list-style-type: none"> • Cheap • User can select the focused object inside the scene. 	<ul style="list-style-type: none"> • Tedious • Complicated for several users
	Active	Temporal	<ul style="list-style-type: none"> • Can work in complete darkness 	<ul style="list-style-type: none"> • Requires and additional transmitter-emitter (infrared or ultrasonic) system. • Some objects tend to absorb the transmitted signal energy. • There is distance-to-subject limitation (6m). • A source of IR light from an open flame (birthday cake candles, for instance) can confuse the IR sensor. • May fail to focus a subject that is very close to the camera.
AF	Passive	Spatial	<ul style="list-style-type: none"> • Suitable for real time. • There is no distance-to-subject limitation. 	<ul style="list-style-type: none"> • Requires good illumination and image contrast.
		Frequency	<ul style="list-style-type: none"> • High accuracy. • There is no distance-to-subject limitation. 	<ul style="list-style-type: none"> • Requires good illumination and image contrast. • High computational complexity.

Table 5.2: Summary of kinds of focus.

5.3 On the Focusing of Thermal Images

While there is a considerable amount of work reported on focusing visible image, e.g. [Nay94, Cre07, Gam06, Lee01, Mor07, Wee07; Hua07, Kro87, Sub98], we are not aware of previous similar studies on thermal imagery. For this reason it is important to create robust, objective criteria to evaluate if a given thermal image is focus or not [Fau11]. This section describes several focusing measures, presents the database specially created for this analysis and summarizes experimental results and main conclusions. Using this database we evaluate the usefulness of six focus measures with the goal to determine the optimal focus position. Experimental results reveal that an accurate automatic detection of optimal focus position is possible, even with a low computational burden.

5.3.1 Focus Measures

We have evaluated several measures suitable for automatic focusing. The study of the focusing of thermal images is especially challenging for the following two main reasons [Fau11]:

1. It is harder to manually focus a thermal image than a visible one, mainly because most operators are not used to seeing thermal images. This means that some skill and habituation in using thermal cameras is required.
2. Imaging cameras usually incorporate a small screen with little resolution. Although a human operator can consider that a given image is focused, sometimes it appears blurred when visualized in a larger screen (e.g. once transferred to a personal computer).

In addition, a typical focus measure should satisfy these requirements [Hua07]:

1. It should be independent of image content. However, if the image contains a large amount of fine detail, it is generally easier to focus.
2. It should be monotonic with respect to blur. If we move away from the optimal focus position, the focus measure should decrease monotonically. Typically this will happen when moving the focus in both directions (left and right).
3. The focus measure should be unimodal, that is, it should have one and only one maximum value. While this is simple for “flat” scenes, this cannot be true for scenes with objects at different focal distances. For instance, if the nearer object is focused, the most distant will be blurred and vice versa. However, computational photography methods are able to combine multifocus images.
4. Large variations in the value of the degree of blurring. This will permit a sharp peak (maximum focus value).
5. Minimal computation complexity. For real time image acquisition, focusing should be conducted as quickly as possible.
6. Robust to noise: the maximum focus value should be stable and unique in the presence of noise. In this aspect, is important to emphasize, as discussed in Section 4.4 that NIR and visible images are sensitive to illumination conditions. If illumination is not sufficient, the image is noisy.

In order to obtain the most suitable focus measure for a thermal image, we have performed a set of experiments with several images and alternative measures. A human operator considers that a thermal image is in focus when it presents the highest amount of details and sharpness. Not surprisingly, the operation instructions of commercial thermographic cameras assert that a thermal image is in focus when edges within the image appear sharp. Thus, focus measures should be similar to those used for visible images, which are mainly sharpness measures. The following sections describe the measures we have used.

- a. *Variance*: A very simple measure is the variance of the image. Blurred images have smaller variance than focused ones.
- b. *Energy of the image gradient*: The energy of image gradient (EOG) is based on the vertical and horizontal gradients of the image, and is obtained as:

$$EOG = \sum_{x=1}^{M-1} \sum_{y=1}^{N-1} (f_x^2 + f_y^2) \quad (5.4)$$

- c. *Tenengrad*: This measure [Kro87] is based on the gradient magnitude from the Sobel operator:

$$Tenengrad = \sum_{x=2}^{M-1} \sum_{y=2}^{N-1} [\nabla S(x, y)]^2 \quad \text{for } \nabla S(x, y) > T, \quad (5.5)$$

where T is a discrimination threshold value, and $\nabla S(x, y)$ is the Sobel gradient magnitude value.

- d. *Energy of Laplacian of the image*: Energy of Laplacian (EOL) can be computed as [Sub98]:

$$EOL = \sum_{x=1}^{M-1} \sum_{y=1}^{N-1} (f_{xx} + f_{yy})^2 \quad (5.6)$$

where

$$\begin{aligned} f_{xx} + f_{yy} = & -I(x-1, y-1) - 4I(x-1, y) - I(x-1, y+1) - 4I(x, y-1) + 20I(x, y) \\ & - 4I(x, y+1) - I(x+1, y-1) - 4I(x+1, y) - I(x+1, y+1) \end{aligned} \quad (5.7)$$

- e. *Sum-modified Laplacian*: [Nay94] noted that, in the case of the Laplacian, the second derivatives in the x- and y-directions can have opposite signs and tend to cancel each other. Therefore, he proposed the sum modified Laplacian (SML), which can be obtained by means of:

$$SML = \sum_{i=x-W}^{x+W} \sum_{j=y-W}^{y+W} \nabla_{ML}^2 f(i, j) \quad \text{for } \nabla_{ML}^2 f(i, j) \geq T, \quad (5.8)$$

where T is a discrimination threshold value and:

$$\begin{aligned} \nabla_{ML}^2 f(x, y) = & |2I(x, y) - I(x - step, y) - I(x + step, y)| \\ & + |2I(x, y) - I(x, y - step) - I(x, y + step)| \end{aligned} \quad (5.9)$$

In order to accommodate for possible variations in the size of texture elements, Nayar and Nakagawa used a variable spacing (step) between the pixels to compute ML. The parameter W determines the window size used to compute the focus measure.

f. Crete et al. To be independent from any edge detector and to be able to predict any type of blur annoyance, Crete et al. [Cre07] proposed an approach, which is not based on transient characteristics but on the discrimination between different levels of blur perceptible on the same picture. The algorithm for the no-reference blur measurement can be described by these formulas:

$$h_v = \frac{1}{9} [111111111] \quad (5.10),$$

$$h_h = \frac{1}{9} [111111111]^T \quad (5.11),$$

$$b_v = h_v * I(x, y) \quad (5.12),$$

$$b_h = h_h * I(x, y) \quad (5.13),$$

where h_v and h_h are the impulse responses of horizontal and vertical low-pass filters which are used to make the blurred version of the image $I(x, y)$. In the next step the absolute difference images $Di_v(x, y)$ and $Db_h(x, y)$:

$$Di_v(x, y) = \text{abs}(I(x, y) - I(x - 1, y)) \quad \text{for } x = 1, 2, \dots, M - 1, \quad y = 1, 2, \dots, N - 1, \quad (5.14),$$

$$Di_h(x, y) = \text{abs}(I(x, y) - I(x, y - 1)) \quad \text{for } x = 1, 2, \dots, N - 1, \quad y = 1, 2, \dots, M - 1, \quad (5.15),$$

$$Db_v(x, y) = \text{abs}(b_v(x, y) - b_v(x - 1, y)) \quad \text{for } x = 1, 2, \dots, M - 1, \quad y = 1, 2, \dots, N - 1, \quad (5.16),$$

$$Db_h(x, y) = \text{abs}(b_v(x, y) - b_v(x, y - 1)) \quad \text{for } x = 1, 2, \dots, N - 1, \quad y = 1, 2, \dots, N - 1, \quad (5.17),$$

Then the variation V_v and V_h of neighboring pixels is analyzed:

$$V_v = \max(0, Di_v(x, y) - Db_v(x, y)) \quad \text{for } x = 1, 2, \dots, M - 1, \quad y = 1, 2, \dots, N - 1, \quad (5.18),$$

$$V_h = \max(0, Di_h(x, y) - Db_h(x, y)) \quad \text{for } x = 1, 2, \dots, M - 1, \quad y = 1, 2, \dots, N - 1, \quad (5.19)$$

If the variation is high, then the initial image is sharp; on the other hand, if the variation is low, the initial image $I(x, y)$ is blurred. In the next step the sum of coefficients $Di_v(x, y)$, $Di_h(x, y)$, $V_v(x, y)$ and $V_h(x, y)$ is calculated in order to compare the variations from the initial picture:

$$Si_v = \sum_{x, y=1}^{M-1, N-1} Di_v(x, y) \quad (5.20),$$

$$Si_h = \sum_{x, y=1}^{M-1, N-1} Di_h(x, y) \quad (5.21),$$

$$SV_v = \sum_{x, y=1}^{M-1, N-1} V_v(x, y) \quad (5.22),$$

$$SV_h = \sum_{x, y=1}^{M-1, N-1} V_h(x, y) \quad (5.23)$$

The vertical Bi_v and horizontal Bi_h blur values from range 0 to 1 are calculated according to equations:

$$Bi_v = \frac{Si_v - SV_v}{Si_v} \quad (5.24),$$

$$Bi_h = \frac{Si_h - SV_h}{Si_h} \quad (5.25)$$

This algorithm is designed to calculate the blur value, but for our purpose the value describing the sharpness is more useful. This value S can be obtained easily according to formula:

$$S = 1 - \max(Bi_v, Bi_h) \quad (5.26)$$

This implies that the sharper images will have the value S closer to 1 and the blurred images closer to 0.

5.3.2 Materials and Methods

In order to analyze the focus of a thermal image using the focus measures described in the previous sub-section, we have constructed several databases. Using these databases, we have evaluated the focusing measure for each image and plotted it against the focus position.

Acquisition System. Thermal images have been acquired using a thermographic camera TESTO 880-3, equipped with a silicon uncooled microbolometer detector (UFPA) with a resolution of 160x120 pixels and a NETD of 100mK and a removable germanium optical lens. The key technical characteristics of this optical lens are summarized as follows:

- FL: 10mm
- Fixed Aperture: f/1,0
- FOV: 32°x24°
- IFOV: 3,5mrad (Geometric Resolution)
- Closest Focusing Distance: 10cm

Database description. The overall database consists of ten image sets detailed in [Fau11]. In each set, the camera acquires one image of the scene at each lens position. In our case we have manually moved the lens in 1mm steps which provides a total of 96 positions. Thus, each set consists of 96 different images of the one scene. For this purpose, we have attached a millimeter tape to the objective (as showed in Figure 5.8 (b)), and used a stable tripod in order to acquire the same scene for each scene position. We have developed a program to control the thermographic camera from a laptop. This program shows the focus measure values in order to facilitate the image acquisition. The image is stored in *.bmp file format. This program is freely available, as well as the database. For the specific purpose of the document, we will describe two of the most representative subsets: The Heater (Static scene) and the Face ones (related with our particular goal), which allows to obtain the same reported conclusions in [Fau11] when evaluating the full dataset.

- Heater: This subset consists of a single set of images of a heater. This scene contains a large amount of detail because the metallic parts are warmer than the spaces between. Figure 5.5(b), shows the best focused image of this database.
- Face: This subset consists of a single set of images of a human face. This sequence contains a scene that is not fully static because of involuntary physical movement. Figure 5.5(e), shows the best focused image of this set.

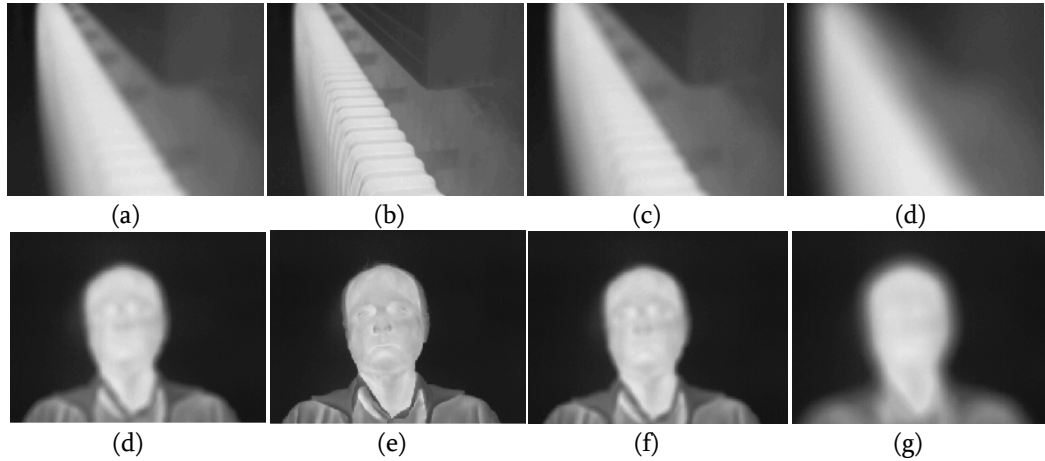


Figure 5.5: Example of different focus for the two subsets described.

5.3.3 Experimental Results and Conclusions

In this section we present the values of the different focus measures described in Subsection 5.3.1. Early results revealed that the variance was not a reliable focusing measure for some image sets so it was discarded. Figures 5.6 and 5.7 shows the focus values described in using the two subsets finally chosen.

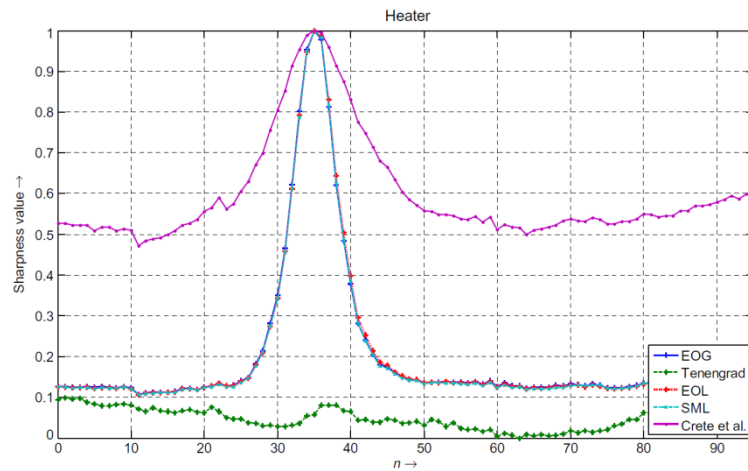


Figure 5.6: Focusing measures obtained with the Heater database.

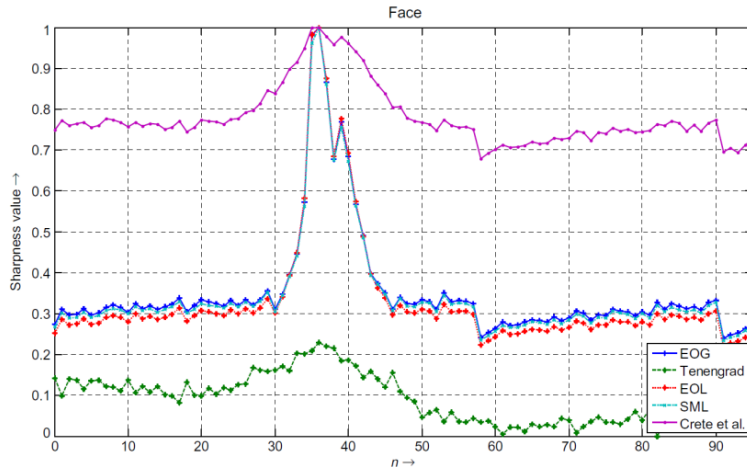


Figure 5.7: Focusing measures obtained with the Face database.

The experimental results in Figure 5.6 reveal that all the methods provide the same focus position with the exception of the Tenengrad and Crete et al. algorithms, while Figure 5.7 shows a clear peak for the EOG, EOL and SML measures. So we do not find any special difficulty in focusing a thermal image of a human body part (face and hand). On the other hand, Tenengrad method fails to provide an accurate peak location, which implies that it is not possible to find an accurate focus using this measure.

Computational time has been obtained with optimized (we have avoided loops) MATLAB algorithms. This was tested on a 32-bit MATLAB 7.4.0.287 (R2007a). MATLAB was used on a laptop with an Intel Core 2 Duo 2,4 GHz, 4 GB RAM and operating Microsoft Windows Vista Home Premium. Table 5.3 summarizes the computational time for each algorithm. As can be seen, all of them are reasonably fast.

Database	<i>Method</i>				
	EOG	Tenengrad	EOL	SML	Crete et al.
HEATER	0,67	2,88	1,71	1,81	2,59
FACE	0,67	2,94	1,70	1,75	2,53

Table 5.3: Computational time (ms) for each method.

We have presented a set of measurements and have reached the following conclusions (More experimented results with other databases can be found in [Fau11]. They lead as to the same conclusions):

- It is possible to automatically focus a thermal image: some operators can find an optimum focus position that matches the human decision. Among the measures we have used, only EOG, EOL and sum-modified Laplacian (SML) offer good performance in both scenarios. These measures provide an accurate and sharp peak which clearly identifies the optimal focus position.

- We have observed that Tenengrad operator was unable to provide an accurate peak in both image sets. We do not want to conclude that this algorithm fails when applied to thermal images. We think that this may be due to the low spatial resolution of our thermal images (160 x120) which may not be adequate for this operator.
- In general, the simplest operators to compute provide the best experimental results. Probably the more sophisticated algorithms require more statistical information (more pixels) in order to provide better results.
- Considering computational issues, EOG performs the analysis in less than 0,7ms, when programmed in MATLAB. This implies that this operator is suitable for obtaining a fast automatic focus.

5.4 Contribution of the Temperature of the Objects to the Problem of Thermal Imaging Focusing

When focusing an image, DOF, aperture and distance from the camera to the object, must be taking into account, both, in visible and in infrared spectrum. Our experiments reveal that in addition, the focusing problem in thermal spectrum is also hardly dependent of the temperature of the object itself (and/or the scene).

In a similar way than in the visible case, where polychromatic images were defocused due to achromatic aberration, as deeply discussed in Section 5.1, temperature of the objects (or the involved wavelengths) produce a parallel chromatic aberration in the thermal spectrum. In this sense, we will focus on how problematic may be this effect, and more specially, we will aim whether temperature of the humans is specially focused or defocused with respect the focal plane of the thermal imagers [Esp12]. According to results presented in the previous section, sum-modified Laplacian (SML), will be selected in order to analyze this last relevant aspect.

5.4.1 Materials and Methods

To address this last issue of analyzing the effect of chromatic aberration in the thermal spectrum when dealing with human faces, we have built two additional special purpose databases.

Acquisition System. Thermal images of the first scene have been acquired using a second thermographic camera available: The TESTO 882. The main differences between TESTO 882 and the previous one (TESTO 880-3) are summarized as follows:

- UPFA of 320x240 pixels and a NETD of 60mK
- Related to the optical lens:
 - Fixed Aperture: f/0,95
 - FOV: 32°x23°
 - IFOV: 1,7mrad

Database description. The database consists of eight image subsets of the same scene showed in Figure 5.9(a) at eight different fixed temperatures of the bulb. Thus, each subset consists of 96 different images of the scene taken at an ambient temperature of 20 degrees. Bulb has been chosen in order to firstly approach the human face (It is a similar 3D object with the possibility to depict the same temperature than a human body). In this respect, bulb's temperature has been regulated by means of its current by using the proposed dimmer D1KS 220-240V 50-60Hz 1000W, obtaining a final span of temperatures near to 40°C (from 40°C to 80°C). Bulbs have the extra advantage of being fixed, compared with humans, avoiding involuntary physical movements such as eye blinking, breathing, etc. Bulbs also have a constant texture in the thermal spectrum as showed in Figure 5.8 due to the opaque behavior of the glass, as appropriately discussed in Section 4.5.2.

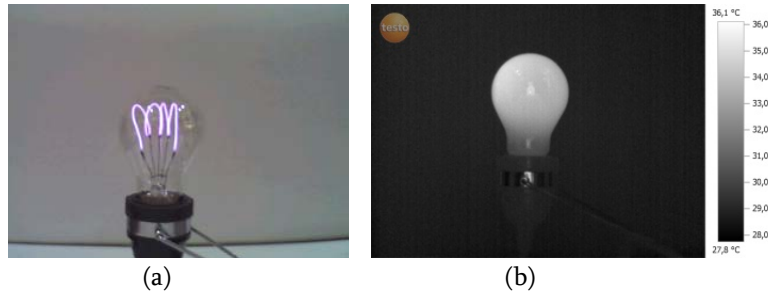


Figure 5.8: Used bulb in both Visible (a) and Thermal spectra (b).

In each set, the camera acquires one image at each lens position following the same procedure defined in Section 5.3.2. Thus, we have again manually moved the lens in 1mm steps which provides a total of 96 positions. For this purpose, we have attached a millimeter tape to the objective, as showed in Figure 5.9(b). The final performed database consists of $8 \times 96 = 768$ thermal images. Figure 5.10 shows a sample of the best focused image of each subset.



Figure 5.9: (a) Scenario. (b) TESTO 882 thermal imager with the stepping ring manual adapter.

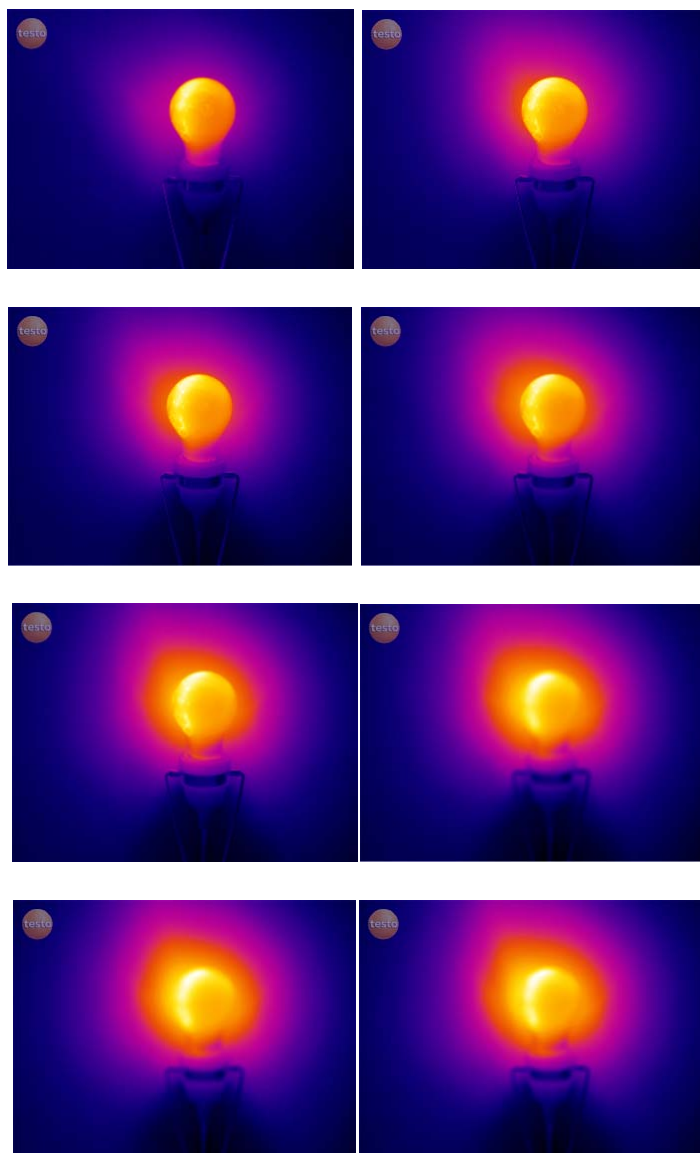


Figure 5.10: Best image of each set, at the eight evaluated temperatures.

5.4.2 Experimental Results and Conclusions

In this section we present the experimental results with the full database described in Section 5.4.1. On the one hand, is direct to obtain by means of the Wien's displacement law, that the bulb's temperature range of 40°C implies a second span of wavelengths of 1048nm that is more than two and a half times the fully visible spectrum, producing optical distortions due to the achromatic aberration. On the other hand, and taking into account that the temperature of the bulb's surrounding is about 20°C while the temperature and the concerned gradient is increasing in each subset, the conduction heat transfer phenomenon (also called *diffusion*), becomes more evident, as states the second law of thermodynamics. This involves solving the equations that describe the phenomenon characteristics, which result from the principles of conservation of mass, momentum and energy [Icr96], being the referred analysis beyond the scope of this research, as just discussed in Section 4.2.2.

We have determined and analytically verify, as sampled in Figure 5.11, that the larger the conduction heat transfer, the smaller the sharpness, (producing as in the analyzed case, a glare around the hotter bulbs that dramatically contributes to the defocus of themselves). This will be sufficient for the ultimate purpose of making evidence that chromatic aberration can be considered negligible when dealing with facial thermograms (with temperatures near to 37°), being such features reliable and robust for biometric recognition purposes.

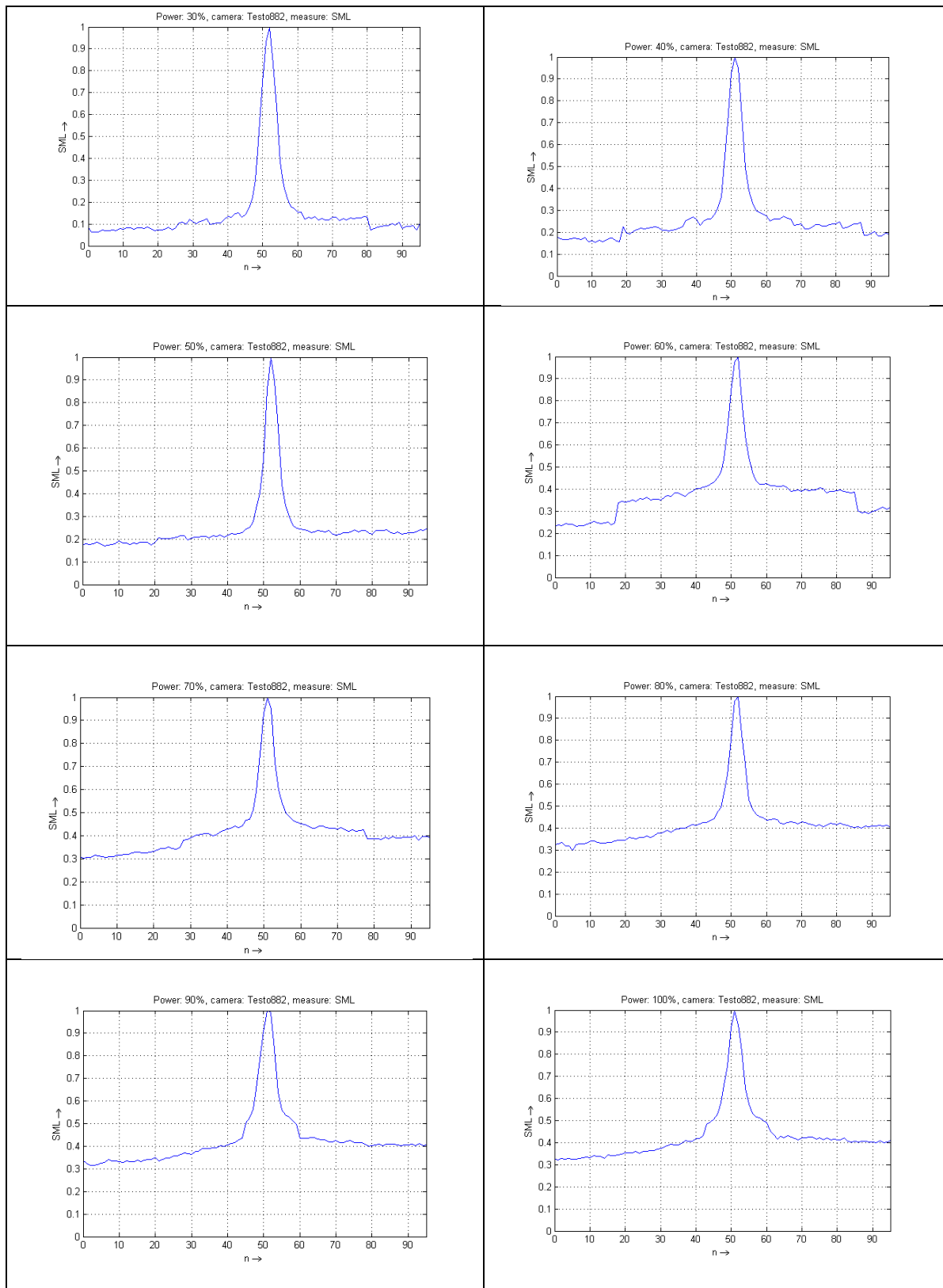


Figure 5.11: Focusing measures of each of the 96 images per subset. The images that depict more heater bulbs depict less sharper peaks. Therefore, higher temperature, involves worse focusing results.

It is necessary to comment that once accepted the paper [Fau11], we were aware of an existing new thermal camera (FLIR SC660) provided with a novel integrated AF system. The related AF system is based on a passive AF method called *contrast detection AF*, where the image sensor itself works as a contrast sensor, and do not necessarily has multiple discrete AF sensors (which is required when approaching the phase detection second passive method of AF). Since the camera does not know the distance to the object, two tests contrast with different focus will be required. Final processing approach compares the computed histograms and establishes that the higher the level of peaks in the histogram the more focused is the related image. Figure 5.12 shows the computed histograms of a set of images of the face database previously analyzed, depicting this property.

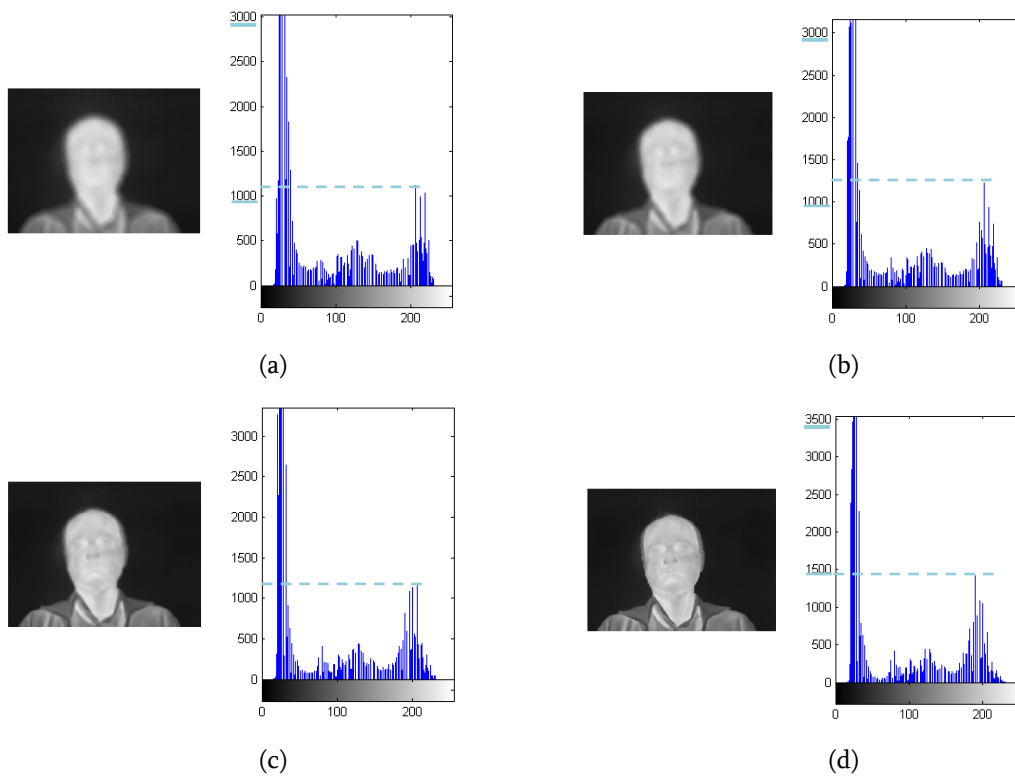


Figure 5.12: A sequence of four thermal faces with different levels of blurriness and their related histograms. Three defocused faces (a,b,c) and a last focused one (d) and the involved histograms.

Chapter 6

Multispectral Face Database

*If you want to make an apple pie from scratch,
you must first create the Universe.*

Carl Sagan

This Chapter is mainly a technical description of the novel database for human faces specially developed for this research work, constituting one of the contributions of this dissertation. The database is novel as it for the first time systematically records human faces in the three selected frequency bands. The description and some recognition experiments have been published in [Esp12].

6.1 Why a new Multispectral Database is required

The limits of human performance do not necessarily define upper bounds on what is achievable. Specialized identification systems, such as those based on novel sensors, may exceed human performance in particular settings [Sin06]. For this reason, it is interesting to perform automatic experiments with images acquired with different sensors. On the other hand, computer based systems can go beyond human limitations because they can “see” beyond cognitive limits. The basic idea behind the design of the **CARL**¹ (CAtalán Ray Light) face Database is guarantee a working solution for experimentation and statistical performance evaluations in such cases, based on both, the single features, as well as the combination of them

¹The name of the DDBB is also a tribute to Carl Sagan, for teaching us to see the Universe...*one-vers...*, to M. Karl E.L. Planck, by enter the universe of the atom and set the starting point of quantum physics and finally, to Karl-el, or better, to Superman, the common name by which people know him, without whom we would never have believed that a man not only could see beyond the visible spectrum, even when took off his glasses, but could also put the interests of others ahead of his own ones.

in more demanding biometric approaches. In addition, three different illumination environments have been performed in order to extend the algorithms assessment capabilities for dealing with the challenging of FR in less restricted lighting conditions.

As advanced in Section 3.5, although several databases exist that simultaneously acquire visible and near infrared [Hiz09] or visible and thermal images [Soc04], we are not aware of an existing multi-session database containing VIS, NIR and thermal information simultaneously. Analyzing the Table 3.5 in a more detail, only the performance of the most well referenced Multispectral NIST/EQUINOX database [Equi09] came close to our research interests although it collects even a more extensive set of faces in the visible, SWIR, MWIR and LWIR infrared spectral bands. However, irrespective of the known reliable properties, EQUINOX database has also a set of disadvantages, most of them are summarized below:

- It has been acquired in just one session, whereas, as is pointed by Jain et al. [Jai04] a multisession performance is highly recommended when dealing with variant biometric traits, which is the case of facial thermograms.
- MWIR technology is highly expensive and involves cooling requirements. In addition, suffers important levels of contamination due to the solar reflectance during daylight. Then, it is not possible to extend FR problem in daylight outdoor scenarios, which is a dramatic limitation for thermal imaging FR daylight approaches.
- SWIR technology also requires expensive devices. By contrast, and as discussed in Section 4.4.1 acquisition in SWIR range has the advantage to see through smoke and fog, which involves a more invariant behavior to weather outdoor conditions.
- Placing acquisition systems is in row, causing some rotations among different acquired faces depending on the camera used.

For all the above reasons, except for the latter isolated consideration, although collecting the performed database means a resource-intensive task that has also meant the user's involvement, it has been considered a compulsory requirement for the advancement of this research field.

CARL database includes a total amount of 7.380 pictures of 41 different subjects. Forty-five different faces (15 of each of the three sensors) of each individual under three different illumination conditions in each session have been collected, and the process has been repeated four times, in order to cope with the level of variability of the faces in each spectrum. Next sections describe the details of this new acquired database as well as technical problems and how they have been overcome. The experiments reported in Chapter 9 have been performed using this new database, whereas the experiments included in Chapter 7, deals with a subset of this database.

6.2 Previous Design decisions

This section describes the previous considerations that have been taken into account in order to implement a useful database to be able to exploit both, the FR powerful beyond the visible spectrum as well as the mathematical analysis techniques to analyze and take advantage of the expected complementary information among the images in different spectra.

6.2.1 Frequency Bands of the Multispectral System

Many Imaging technologies beyond the visible spectrum have been used to capture a body part (XR, magnetic resonance, IR, ultrasounds, etc). In any case, the purpose of these biometric measures is related to health, rather than security application, such as biometrics. In addition, not all of them could be used for biometric purposes due to their high level of intrusiveness [Mor09]. Likewise, although cameras in the UV spectrum also exist² [Esp06], they have been also discarded due to the high energy of the photons in this part of the spectrum is harmful for human beings. Furthermore, EM waves are strongly affected by Rayleigh scattering effect in this part of spectrum, as discussed in Section 4.2.1, drastically reducing outdoor capabilities in medium and large recognition distances. In this sense, the more adequate imaging technologies for our purpose have been VIS, NIR and LWIR (coded as TH). Figure 6.1 shows a sample of the three kinds of pictures acquired in the three spectrums: NIR, Thermal and Visible. The main reason for selecting such bands can be summarized as follows:

- VIS:
 - Is the most appropriate spectrum for trial balances.
 - In addition, commercial thermal imagers provides both spectrum images (VIS & TH) in a simultaneously way. Thus, no any other device is required for acquiring both spectrums.
- NIR:
 - Easy and inexpensive device.
 - Reduced device size (allow to be embedded very near to the other sensor (Thermal imager), as can be seen in Figure 6.2), in order to minimize the point of view between different cameras.
- LWIR (TH):
 - Current devices are affordable (The cost of these systems has dropped by more than a factor of ten over the past decade).
 - Thermal energy levels of human beings are centered in this range as described in Section 4.5.2.
 - Cooled systems are not required.

²UV camera technology is supported by silicon based sensors, whereas due to common glass is opaque to ultraviolet wavelengths, usually quartz, is used instead, for the appropriate UV lenses performing.

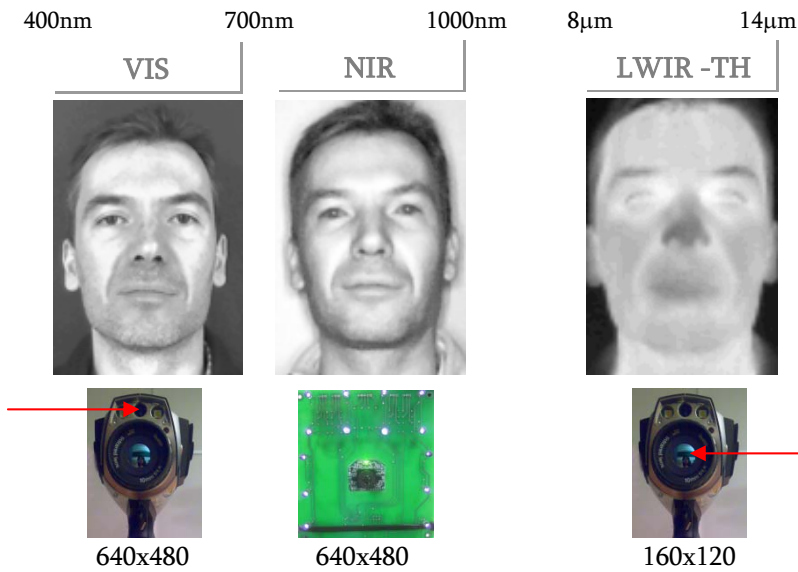


Figure 6.1: A sample of a subject acquired in the three spectral bands. Co-registered VIS/LWIR thermal imagery. (Images are depicted once segmented).

6.2.2 Sensors Arrangement

In order to appropriately distribute the set of image sensors to finally achieve a similar field of view (FOV), a negligible parallax error, and a set of almost equal facial poses (horizontal rotations) irrespective of the vision system used, a number of preliminary considerations were taken into account:

- Related with the TESTO Thermal (and Visible) camera:
 - FOV of thermal camera, by using a lens of a focal length (FL) of 10mm, and considering the dimensions of the thermal sensor, is equal to 32°x24°.
 - The inherent parallax error between visible and LWIR images occurs when working distances are shorter than 40cm, as addressed in Section 4.5.1.
- Related with the customized NIR camera:
 - Although webcams are provided with angular lenses, no more technical information about the focal length and the corresponding FOV are available.
- Related with the relationship between both acquisition systems:
 - Vertical alignment is chosen (producing a slightly “low angle” (“contrapicada” in Spanish)) NIR images (See Figure 6.1). Such selection is preferable than horizontal alignment approach, used in EQUINOX e.g, due to it provides undesirable rotations in face pose, depending of the camera.

Once computed the thermal imager settings, the initial optimal computed distance between the face of the user and the tripod that holds the sensors was 110cm, while after manually adjusting in order to synchronize both FOV of thermal camera and NIR camera, the final chosen distance was fixed to 135cm.

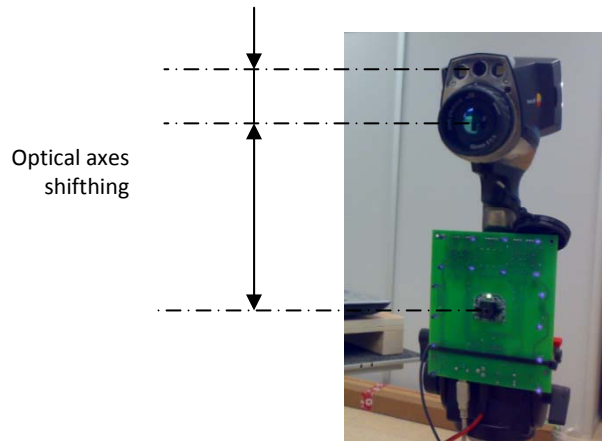


Figure 6.2: Hardware components of the multispectral imaging system mounted on a tripod.

A reliable embedded solution for acquiring frontal views minimizing the difference between angles of acquisition among the different sensors has been accomplished with both sensors mounted over a same tripod in order to better test algorithm performance. In order to appropriately fix the tripod along the multiple sessions (and rest of the elements of the overall scenario) we have used markings on the ground.

6.2.3 Further Considerations

A set of the last additional considerations to close the setting database parameters was the following:

- Related with the temperature range of the thermal imager: Due to camera TESTO automatically autoscale each image as shown in Figure 6.3, such condition forced us to directly use .bmt files, in a similar way than raw image format for standard images, despite of such format has an interpolated resolution³ (320x240 pixels instead of the originals 160x120). This consideration led us to downsampling again interpolated samples in order to recover the original resolution. Once the file was completely preprocessed, the image in VIS spectrum was extracted, the temperature matrix was stored to MATLAB *.mat file and also transformed to grayscale image using a fixed scale and stored to *.bmp format.

³TESTO 880-3 thermal imager acquires images with original resolution just saving as BMP format.

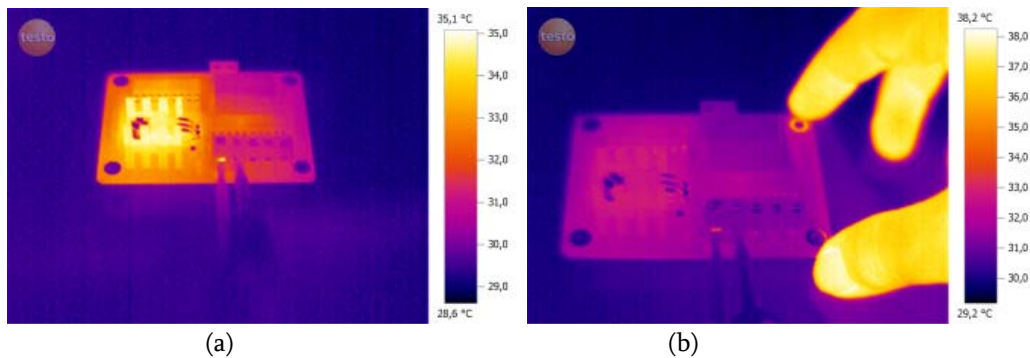


Figure 6.3: A sample of the autorange process automatically carried out by the thermal camera (This process is function of the maximum and minimum temperatures of the scene. In the case shown: (a) PCB with radiator. (b) Same PCB handled by the author's hand. Also notice the different pseudocolors used to code the radiator despite of it has the same temperature in both images.

- **Background Treatment:** A background screen using a special stand kit which supports a roll of matt black paper was also designed. It is important to point out that this matt black background is mandatory behind the user in order to avoid undesirable thermal reflections from the operator, due to its well-known extra low albedo. This smooth background also facilitates the segmentation of the visible and NIR images.
- **Glasses Treatment:** Although the most part of literature treats the presence of eyeglasses in thermal face images as a big weakness that blocks large portions of thermal energy and resulting in loss of valuable information on and around the eyes [Abi04] it is also true that wearing glasses, is a problem in visible and NIR reflected spectrums, despite not in the same intensity, as pointed is Section 4.5.2. However, because our focusing goals are to investigate the viability of thermal face images for FR and to analyze the redundant versus complementary information between face images in different spectrums, we have considered better to deal with faces in a rawer version, finally disesteeming the inclusion of people wearing glasses.



Figure 6.4: Implications of Wearing glasses in the visible spectrum. Clark Kent (a) hiding his true identity of Superman (b) just using a big glasses. Reflections in (a) can be additionally appreciated.

6.3 Acquisition Scenario

The image acquisition system is based on both thermal and webcam camera. The main function of the image sensors is to convert the viewed scene in a bidimensional array of data values. In our case, these values will be both, reflected VIS and IR data and emitted IR data depending of the final sensor used. We have used two different cameras provided with different sensors. The first one is a webcam provided with a simple CMOS image sensor and the second one is the TESTO 880-3 thermographic camera provided with two different sensors to encompass visible and LWIR spectral bands. In the following sections we describe these image acquisition devices as well as the illumination system, the acquisition protocol and the database concerned features.

6.3.1 Visible and Thermal Acquisition System

Visible (VIS) and thermal infrared images have been both acquired using a thermographic camera TESTO 880-3, equipped with a silicon uncooled microbolometer detector with a spectral sensitivity range from 8 to 14 μ m and provided with a germanium optical lens, and an approximate cost of 8.000€ (in 2009). This thermal imager also integrates a 640x480 resolution AF visible camera. The key technical characteristics are summarized below:

- Sensor:
 - Type: UFPA, temperature stabilized.
 - Resolution: 160x120 pixels (320x240 pixels interpolated)
 - Spectral Sensitivity: 8 to 14 μ m
 - Thermal Sensitivity (NETD): <100mK at 30°C
 - Operating Temperature Range: -20÷350°C
- Removable angular optical lens:
 - FL: 10mm
 - Fixed Aperture: f/1,0
 - FOV: 32°x24°
 - IFOV: 3,5mrad (Geometric Resolution)
 - Closest Focusing Distance: 10cm

6.3.2 Near-Infrared Acquisition System

For the acquisition in the Near- Infrared (NIR) spectrum, a commercial customized Logitech Quickcam messenger E2500 has been used. This inexpensive webcam with a cost around 30€, is provided with a Silicon based CMOS solid-state image sensor and a fixed focus lens embedded in the same single module, providing a reasonable optical performance. The sensor has a sensibility to the overall visible spectrum and the half part of the NIR (until 1.000nm approximately) and a still picture maximum resolution of 640x480 for NIR images; this has been the final resolution selected for our experiments. The default IR optical filter of this camera (IR cutoff Filter or IRCF) has been replaced by a couple of Kodak daylight filters for IR

interspersed between lens and sensor. They both have similar spectrum responses as showed in Figure 6.5 and are coded as wratten filter 87 FS4-518 and 87C FS4-519, respectively.

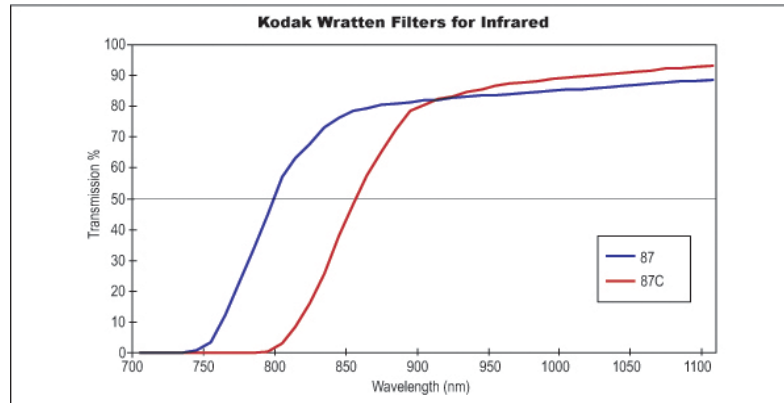


Figure 6.5: Spectral sensitivity of the two visible opaque IR filters, specifically matched to our application.

Regarding blurring images due to NIR focus shift phenomenon described in Section 4.4.1, the standard angular lens assembled in the webcam used almost fully compensate the effect, providing large DOF levels and delivering acceptable quality images in any particular lighting situation, as can be appreciated in Figure 6.12.

After considering the preliminary tests of the first inspired approach, developed in our previous research work [Esp04a] and motivated by its simplicity and consistency (In Figure 6.6(a) the referenced device can be seen), a more robust solution based on the same idea and developed in the context of the TEC2006-13141-C03/01/TCM Biopass Project (a first prototype of it can be found in [Mor08]) has been used as the final solution. It is a special purpose printed circuit board (PCB) provided with a set of 16 IREDs with a range of spectral emission from 820 to 1.000nm placed with an inverted U shape in order to provide the required illumination. The final number of diodes chosen responds to a trade-off among power consuming and correct facial lighting. In the following section, a developed graphic user interface (GUI) to control this second acquisition system, will be appropriately pointed.

Meanwhile, Figure 6.7 closes this section showing an individual in front of the overall acquisition system (LWIR-VIS & NIR spectra).

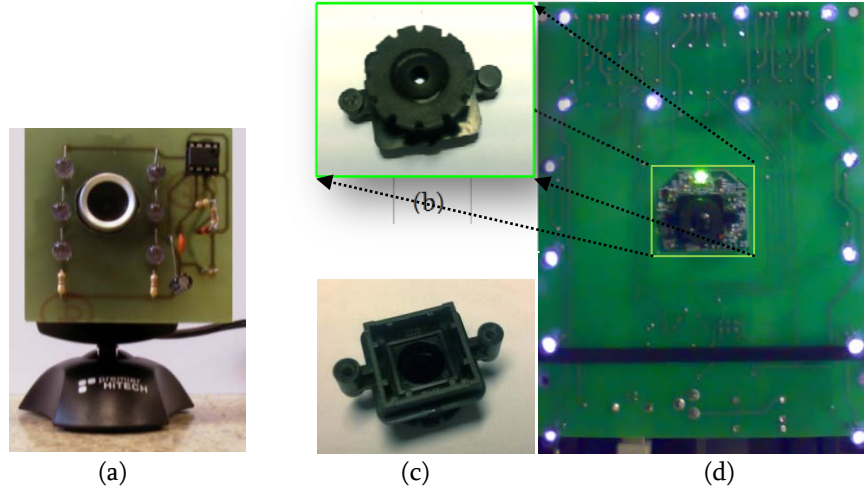


Figure 6.6: Webcam cameras and PCB boards for infrared illumination. (a) First approach using a general purpose USB CMOS webcam. Final approach: Top (b) and bottom view (c) of the mini web camera packaged in a light-blocking, IR transmissive plastic. (d) Final approach performed built onto a PCB. Note the array of 16 IREDs powered by the system, embedded onto the camera to illuminate the face. (Notice that the IREDs can be seen lighting. This is because the camera with which the photo has been taken is provided with a broadband sensor capable of acquiring this fraction of the NIR spectrum).



Figure 6.7: A subject acquired by the system. The thermal imager (a) simultaneously acquires thermal and visible images in dual mode, whereas the embedded NIR acquisition system must be connected to a laptop (b) to acquire the NIR image. The distance between camera and user has been chosen as a tradeoff between having a low parallax error and take advantage of the maximum sensing area of all the sensors.

6.3.3 Lighting Conditions

The selected room is provided with both natural and artificial light. In each recording session the images have been acquired under three different illumination conditions, as follows:

- 1) *Natural illumination* [Coded as NA]: Windows are open and daylight (the full color spectrum) enters the room and provides the necessary illumination for illuminating the scene. Obviously this illumination is not constant along days (due to weather conditions) and it also varies in function of the different hours of the day, providing a set of different quality, direction and intensity of the natural lighting.
- 2) *Infrared Illumination* [Coded as IR]: As is described in the previous section, IR illumination is provided by an array of high directional IREDs, which is an appropriate solution for short-range illumination. PCB around the webcam is turned on and the remaining sources of light are disconnected. The designed GUI (see Figure 6.8) has been developed in order to set properly the IRED's intensity level. This camera software has been extended to also set the image involved image acquisition parameters (*exposure*, *gamma* and *brightness*) which are useful during the adjustment. Appropriate exposure time and analog gain applied, ensures a reasonable quality image in every particular operating condition. Additionally, it is also possible to manually fully optimize them.

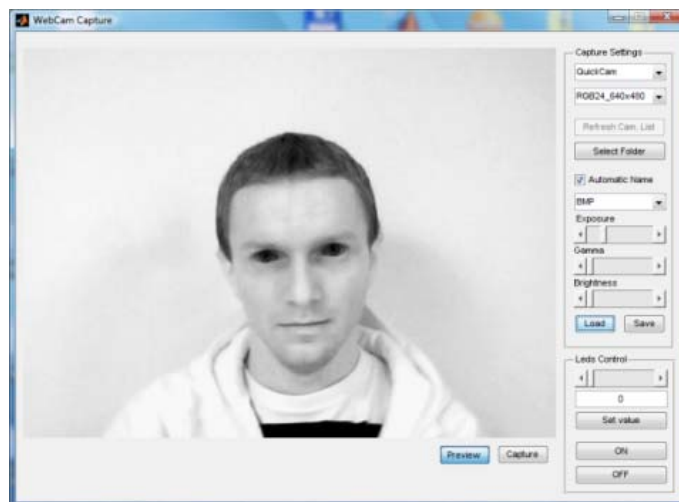
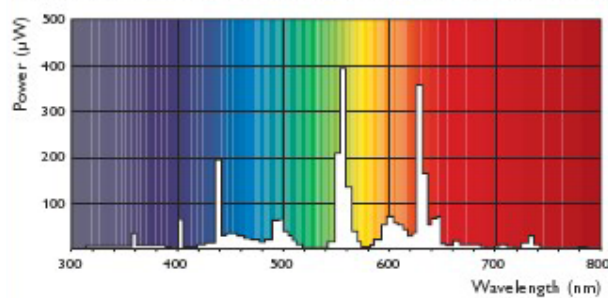


Figure 6.8 : GUI developed to control the IR illumination and camera settings.

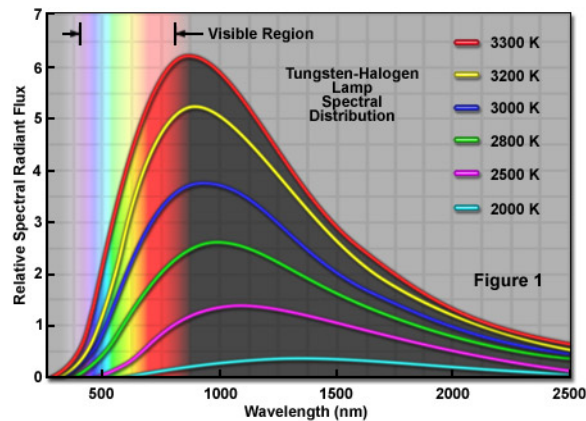
- 3) *Artificial Illumination* [Coded as AR]: The provided equipment used for illumination is the following:
 - A dominant light source approached by a set of 9 PHILIPS TL-D 58W/840 cool white *fluorescents* lamps uniformly distributed in order to produce the required base illumination of the scene.

- A second pair of IANIRO Lilliput lights fitting 650W-3400K *tungsten halogen lamps* have also been used in order to fill and smooth the well-known discontinuous fluorescent spectral emission (see Figure 6.9(a)) and to provide an additional IR portion of light. Figure 7.9(b) shows the related portion of spectral emission in this band emitted by a set of different color temperature halogen bulbs. At the beginning, high pair of power focus produced important dark shadows over the users' face. In order to solve this drawback, we had finally used a LEE 3ND 209 Filter to minimize the referred effect. This neutral density gel reduces light without affecting color balance.

Spectral energy distrubition MASTER TL-D Secura/840



(a)



(b)

Figure 6.9: Spectral power distribution of the different lighting systems employed. (a) Related fluorescent curve. Notice only certain wavelengths (the spikes) are strongly present. The rest come close to zero power. (© Philips). (b) Related tungsten- halogen light sources curves. Notice the strong presence of “warm” wavelengths in all cases. (© Zeiss)

A couple of halogen focus disposed 30° away from the frontal direction and about 3m away from the user, match the artificial light of the room. Figure 6.10 shows the overall scenario performed.

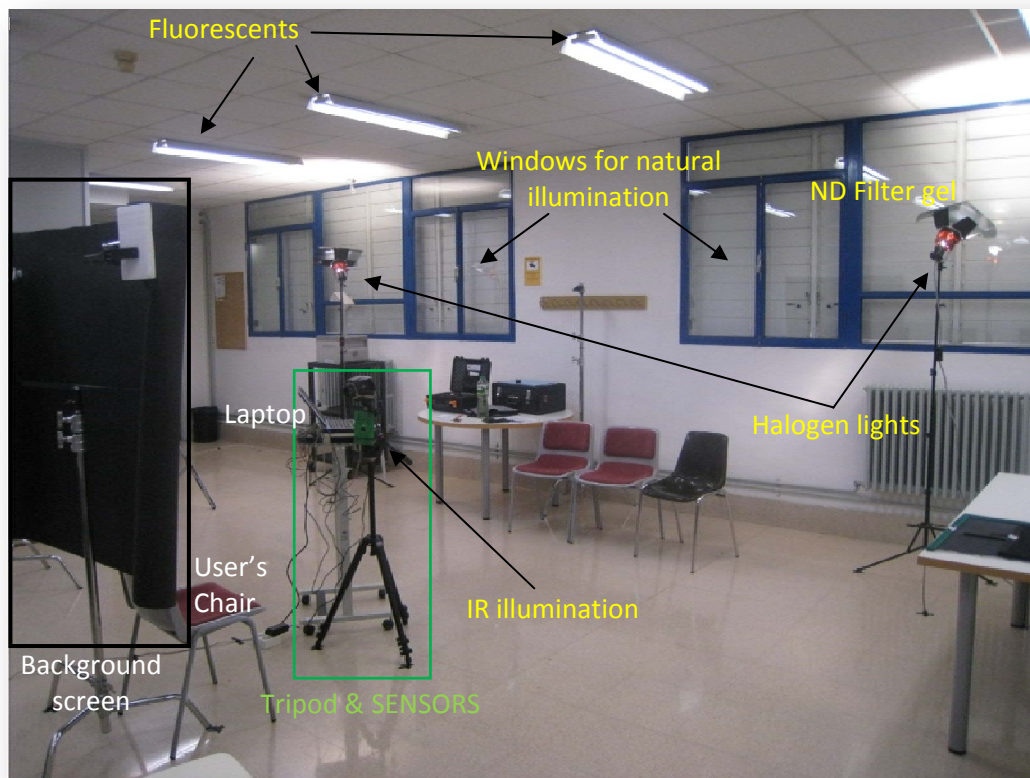


Figure 6.10: Overall multispectral face acquisition scenario.

6.3.4 Acquisition Protocol

In this section the acquisition procedure established during the development of the CARL Face Database is described in detail. The whole process of face acquisition has been accomplished under the assistance of an operator in a supervised way (J. Mekiska or the Author). Each user has been recorded in four different acquisition sessions performed between November of 2009 and January 2010. In this sense, distinctive changes in the haircut and/or facial hair of some subjects may be appreciated. The acquisitions have been done in the whole day from 9 AM to 5 PM, because it was getting dark after 5 PM. The average time required for the full acquisition process of a skilled user has been around 10 minutes, being 15m for a non skilled one. The whole set of users were acquired in two days per session. The time slot between each session is shown in Figure 6.11.

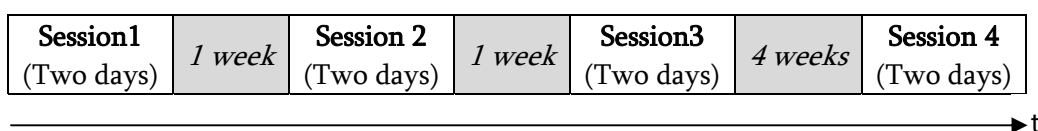


Figure 6.11: Full Acquisition plan.

In each illumination condition five different frontal snapshots are acquired. During the acquisition process, the user is required to look straight at the same place and held the head relatively steady in order to reduce defocusing additional problems. No keeping neutral facial expression is required. Thus, a different facial expressions have been collected (smiling/non smiling, open-closed and blinking eyes...etc). People wearing glasses were asked to remove them before acquisition. No any other physical restriction has been taken into account in order to acquire a face image.

In order to try to fastly modify the temperature of the subjects and to reduce the correlation between consecutive acquisitions of the same session, between a couple of snapshots the user is asked to stand up, make a loop to the room, including one step which corresponds to the portion of the room close to the blackboard, and sit down again. It is worth to mention that thermal camera was able to detect a temperature increase due to this additional physical exercise.

6.3.5 Database Features

Final database consists of 41 people. A special effort in enrolling a considerable number of females has been done, obtaining a final relation of 32 males and 9 females. Each individual contributed in four acquisition sessions and provided five different snapshots in three different illumination conditions and under three image sensors. Thus, in each session, 45 (5x3x3) acquisitions per individual have been carried out. This implies a total of: $45 \times 4 \times 41 = 7.380$ images, grouped in folders as shown in Figure 6.12.

The images in NIR spectrum are stored in lossless *.bmp files. The developed acquisition software, automatically named, accordantly with the protocol, the acquired files. The images from thermal camera were firstly stored to *.bmt format provided by TESTO company (this format includes VIS image, temperature matrix and metadata describing for example the outside humidity, temperature range etc) and then were appropriately preprocessed to solve autorange and resolution issues, as discussed in Section 6.2.3.

In order to normalize all the images to the same size and remove the background a new face segmentation algorithm for thermal images has been developed [Mek10]. All the faces have been segmented following the process described in Section 8.2.1 and consequently resized to 100x145 pixels using bicubic interpolation.

Each file in database has an 8-letter code name. The meaning of each letter is described in Table 6.1.

Letter position	1-2	3-4	5	6-7	8
Meaning	Personal ID	Session number	Sensor	Illumination	Sample
Possible values	01-41	S1-S4	C – visible I – NIR T – Thermal	NA – NATural IR – INfrared AR – ARtificial	1-5

Table 6.1: Meaning of the file code name.

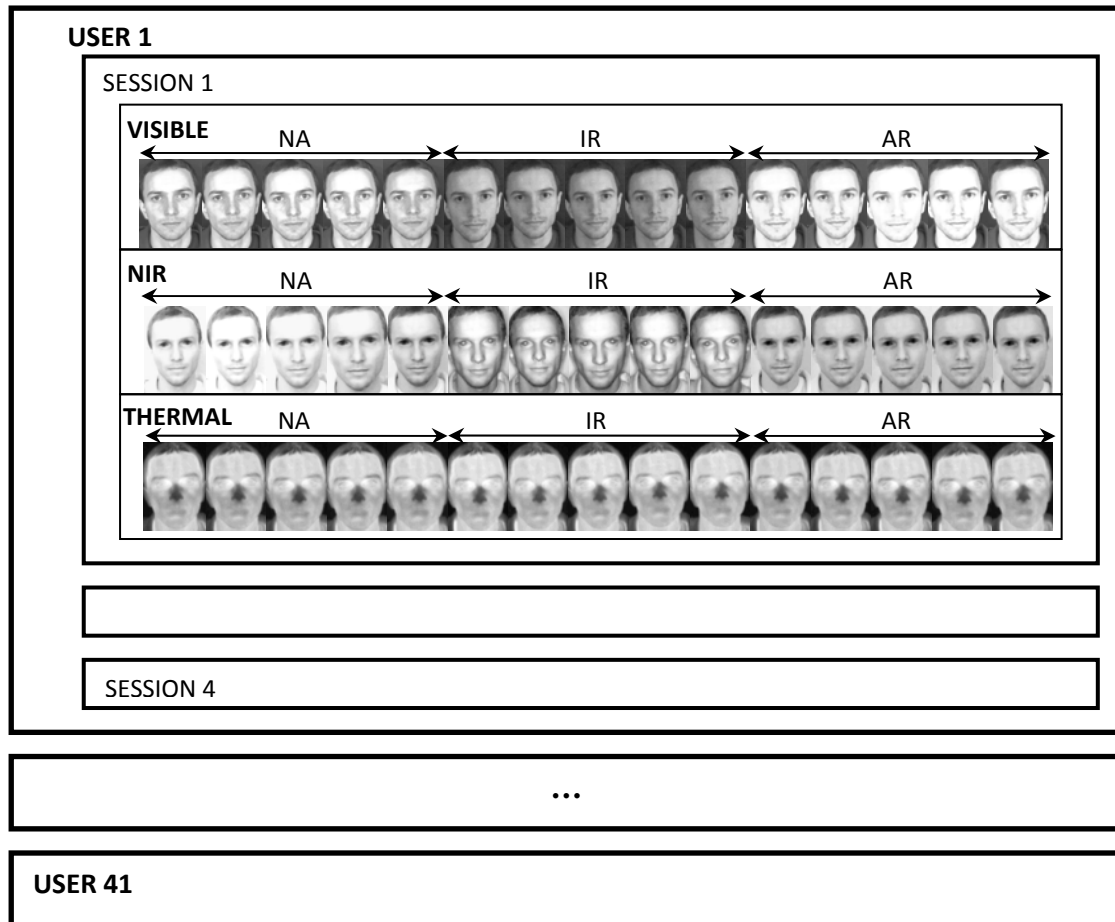


Figure 6.12: Database structure. For each user there are four sessions and each session contains three kinds of sensors and three different illuminations per sensor.

Chapter 7

Information Analysis of Multispectral Images

*We know accurately only when we know little.
With knowledge, doubt increases.*
Goethe.

In this chapter, the images of CARL multispectral face database are analyzed from an information theory point of view in order to explore the redundancy between several spectral bands and to evaluate the power of data fusion. The first part of the chapter gives a straightforward introduction to the related mathematical tools and introduces the basic analytic expressions. The second one proposes a new criterion based on the Fisher score for the case of mutual information, which allows evaluating the usefulness of different sensor combinations for data fusion and cross-sensor recognition. The chapter ends with a set of experimental results and some relevant conclusions. General contents of this chapter have been published in Issues [Esp10, Esp11].

7.1 Introduction

As deeply discussed in Chapter 4, a new generation of enhanced image sensors can perform acquisition in different wavelengths, obtaining information beyond our limits. *What would happen if we were able to overcome our limitations? Is there any additional not redundant information provided by such sensors in order to get a better knowledge of our environment?* And more especially: *Would we able to use this additional information for biometric purposes?*

For answering such set of opening questions, a criterion for analyzing concerned images from different spectra, has been addressed from an information theory point of view [Esp10]. In

addition, a criterion for pairwise combination of information from different sensors in order to decide how a given pair of sensors is useful for our purpose is also proposed [Esp11].

7.2 Background on Information Theory

In order to more easily understand the following sections, a straightforward background to Information Theory will be previously given. More detailed and complete information of the topic can be found in [Cov91, Mac03] among other references.

Information theory is a framework in which one can set up different features related to perception and which allows for the quantification of the phenomena that one wants to model. From an information theory [Sha48] point of view we can establish these following measurements:

- a. *Entropy*: Considering that the random variable X consists of several events x , which occur with probability $p(x)$, the entropy $H(X)$ can be calculated according to equation:

$$H(X) = -\sum_{x \in X} p(x) \log_2(p(x)) \quad (7.1)$$

Entropy measures the information contained in a message (in our case will be an Image) as opposed to the portion of the message that is determined (or predictable).

- b. *Conditional Entropy*: The conditional entropy (or *equivocation* in communication theory) quantifies the remaining entropy (i.e. *uncertainty*) of a random variable Y given that the value of a second random variable X is known. It is referred to as the entropy of Y conditional on X , and is defined as:

$$H(X | Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2(p(x | y)) \quad (7.2)$$

- c. *Joint Entropy*: The joint entropy of two random variables X and Y measures how much entropy is contained in a joint system of these two random variables. It is defined as:

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2(p(x, y)) \quad (7.3)$$

- d. *Mutual information*: The mutual information (MI) $I(X; Y)$ of two random variables X and Y (representing two images in our case) is defined as:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (7.4)$$

where $p(x, y)$ is the joint probability distribution function of X and Y , and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively. Intuitively, mutual

information measures the information that X and Y share: it measures how much knowing one of these variables reduces our uncertainty about the other. For example, if X and Y are independent, then knowing X does not give any information about Y and vice versa, so their mutual information is zero. At the other extreme, if X and Y are identical then all information conveyed by X is shared with Y : knowing X determines the value of Y and vice versa. As a result, in the case of identity, the mutual information is the same as the uncertainty contained in Y (or X) alone, namely the entropy of Y .

Mutual information quantifies the dependence between the joint distribution of X and Y and what the joint distribution would be if X and Y were independent. Mutual information is a measure of dependence in the following sense: $I(X;Y)=0$ if and only if X and Y are independent random variables. This is easy to see in one direction: if X and Y are independent, then $p(x,y)=p(x)p(y)$ and therefore:

$$\begin{aligned} I(X;Y) &= \sum_{y \in Y} \sum_{x \in X} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) \\ &= \sum_{y \in Y} \sum_{x \in X} p(x,y) \log_2(1) = 0 \end{aligned} \quad (7.5)$$

Mutual information can also be expressed as:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \quad (7.6)$$

The diagram of Figure 7.1 helps to explain the relation between these different four definitions.

- e. *Normalized 2D cross-correlation*: Cross-correlation is a measure of similarity of two signals as a function of a time-lag applied to one of them. It is possible to calculate a two-dimensional cross-correlation $c(u,v)$ according to the formula [Lew01]:

$$c(u,v) = \sum_{x \in X} \sum_{y \in Y} f(x,y) t(x-u, y-v) \quad (7.7)$$

where f is the image and t is the feature positioned at u,v . The disadvantage of the two-dimensional cross-correlation is that the range of $c(u,v)$ is dependent on the size of t and it is not invariant to changes caused by changing of illumination conditions across the image sequence. These disadvantages are suppressed by normalization of image and feature to unit length which is used in normalized 2D cross-correlation $\gamma(u,v)$ defined by:

$$\gamma(u,v) = \frac{\sum_{x,y} [f(x,y) - \bar{f}_{u,v}] [t(x-u, y-v) - \bar{t}]}{\left\{ \sum_{x,y} [f(x,y) - \bar{f}_{u,v}]^2 \sum_{x,y} [t(x-u, y-v) - \bar{t}]^2 \right\}^{0.5}} \quad (7.8)$$

where \bar{t} is the mean of feature and $\bar{f}(u, v)$ is the mean of $f(x, y)$ in the region under the feature. This measure is interesting because we should take into account that information theory measurements do not take care of pixel position. They consider the statistical properties of the pixels regardless of their distribution inside the image. This is not the case with cross-correlation that certainly considers pixel positions.

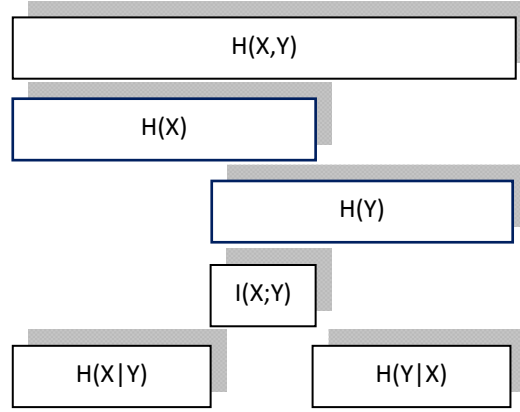


Figure 7.1: Relation between different measurements.

7.3 Our Proposal: Information Theory-Based Fisher Score

Our objective is twofold: On the one hand, we want to evaluate the usefulness of different sensor combination in order to decide how a given pair of sensors is useful for our purpose. On the other hand, we also want to analyze the possibility for crossed-sensor recognition (matching of images acquired in different spectral bands).

The criterion that we propose is related to Linsker's idea of the principle of maximum information processing [Lin88, Lin86]. This idea was proposed as a criterion for modeling the self-organization in two-layer networks that process sensory signals. In our case, we are not explicitly interested in finding a self-organizing network, but on the ways of using information from different sensory sources. The uses we envisage in this dissertation are also common to the sensory perception done by humans and animals, i.e. recognition of different individuals, fusion of information coming from different senses and recognition based on mixed (cross) information from different senses. In these cases, the fusion of information coming from different senses provides a better knowledge of the environment. For instance, we would not be able to locate the precedence of a sound if we had only one ear, we would not have any deepness sensation with a single eye and when we look and touch an object we have better knowledge about it. The criterion known as *infomax* [Lin88, Lin86] explains, in part, how neurobiological structures are formed; therefore, it is a good candidate as a criterion to decide how to combine information originated at different senses.

The criterion proposed is a generalization of the Fisher score for the case of mutual information, which is measured as the ratio of the interclass information to the intraclass. The

proposed score measures the behavior of a pair of sensors either when they are used in combination or when they are used to discriminate between classes. The motivation for the score that we propose comes from the fact that, in pattern recognition, we are interested on features of the data that yield a small intraclass variation and higher interclass variation. When this criterion is used to design a linear classifier, the resulting classifier is known as Fisher discriminant as discussed in Section 3.3.2. A transformation that maximizes this score also maximizes the recognition rate in the case of two classes that follow a Gaussian distribution with the same covariance matrix. For instance, in a biometric recognition application, we are interested on a small variation between different snapshots of the same person. This will make it feasible to obtain a stable model for a given person. This score can be generalized and be used for selecting sensors, with the criterion that the score is maximized. This is one of the ideas that we will follow in this work. Another idea is to use mutual information instead of the standard deviation as a notion of variability. We will use the score in order to determine the sensor combinations that are candidates to either data fusion or interoperability¹.

In discriminant analysis of statistics, within-class, between-class and mixture scatter matrices are used to formulate criteria for class separability [Fuk90]. A within-class scatter matrix (S_w) shows the scatter of samples around their respective class expected vectors. On the other hand, a between-class scatter matrix (S_b) is the scatter of the expected vectors around the mixture mean. The mixture scatter matrix (S_m) is the covariance matrix of all samples regardless of their class assignments. In order to formulate a score for class separability, the information contained in the matrices is summarized as a scalar number. In the case of the Fisher discriminant, the objective is to find a projection that maximizes this ratio, i.e. to maximize the between-class scatter and simultaneously minimize the within-class scatter. Other uses of this score might be to find a combination of sensors that maximize this score. The score used to design linear classifiers has a general form that relates the product of the inverse of the class scatter to a matrix that measures the between-class scatter. In the case of features distributed as multivariate Gaussians, the resulting solution is reduced to a generalized Eigenvalue problem. This score is the solution of an optimization problem that maximizes a Rayleigh quotient and is known as *Fisher score*. The general form of the score can be summarized as:

$$J_1 = \text{tr}(S_2^{-1} S_1) \quad (7.9)$$

where S_1 and S_2 are one of S_b , S_w or S_m . In the case of a general multivariate distribution of the data, the solution is not so simple. Nevertheless, it makes sense to use a measure similar to the scatter matrices. The initial idea of Fisher was to quantify the dispersion of the data, and a natural measure of dispersion in the case of multivariate Gaussians is the covariance matrix. An alternative in the case of a general distribution might be the number of bits needed to code a random variable. Therefore, the number of bits needed to code a random variable captures the idea of uncertainty that is present in the concept of standard deviation. As a tool for

¹Interoperability of different sensors implies the recognition of a same object when the test image and the model for this object have not been obtained with the same kind of sensor. An example of this situation is a surveillance camera acquiring nocturnal infrared images and the matching of these sets of images with a database of visible images of suspicious people.

quantifying the number of bits needed to code a random variable is the *Shannon's entropy* [Cov91].

Thus, redefining the Fisher score, it will give information about the performance of pairwise combination of sensors when measured intraclass or measured between classes. Therefore, we will compute a ratio related to the uncertainty between sensors, i.e. how the number of bits needed to code the information of a sensor increases or decreases when the other sensor is known. This measure of the pairwise performance is given by the following ratios:

- i. The ratio between mutual information between-classes and the mutual information intraclass.
- ii. The ratio between normalized inter- and intraclass mutual information, which would take into account the different scales of both mutual information.
- iii. The maximum of the cross-correlation between sensors for the intra- and interclass cases.

As our objective is twofold, as discussed at the beginning of this section, we would like to evaluate the proposed measures in relation to several properties that should reflect how the measure or index captures the information between sensors, and this information is reflected in the intra/interclass dispersion.

From a theoretical point of view a desirable list of properties is the following:

- High mutual information and cross-correlation between images of a given person acquired with the same sensor. This implies that different acquisitions of the same person show small variability (they are almost identical, highly redundant). In this case, it would be easy to obtain a model for a person, due to the stability of different samples that belong to a specific individual.
- Low mutual information and correlation between images acquired with both the same and a different sensor when they belong to different people. This implies that different acquisitions of different people show large variability (they are different and not redundant). In this case, it would be easy to discriminate between different samples that belong to different individuals.
- Low mutual information and correlation between images of a given person acquired with a different sensor. This implies complementary information between images. In this case, it is worth combining information of both images in order to improve recognition accuracies.
- High mutual information and correlation between images of a given person acquired with a different sensor. This implies the possibility of properly matching images of the same person but acquired with different sensors. An example of this situation is the identification of people inside a black list (whose model consists of a “classical” visible photo) in a recording at night obtained with an infrared surveillance camera. Clearly this desirable property is opposed to the previous one.

Table 7.1 summarizes these main desirable properties. In order to obtain good recognition rates, the next ones must be simultaneously fulfilled:

- Same sensor: (1) and (2) provide good recognition rates for images acquired with the same sensor.
- Different sensor: (4) and (5) provide good cross-sensor recognition.
- Different sensor: (3) and (5) provide good possibilities to enhance recognition accuracies by means of data fusion.

#	Sensor	Person	$I(X;Y)$	Implications
Type 1	Same	Same	High	Low intraclass variation: easy to obtain a model
Type 2	Same	Different	Low	High interclass variation: easy to discriminate people
Type 3	Different	Same	Low	Data fusion can improve recognition accuracy
Type 4			High	Possibility to perform cross sensor recognition
Type 5	Different	Different	Low	High interclass variation: easy to discriminate people

Table 7.1: Desirable values for mutual information (and correlation) in all the possible combinations and their implication.

In order to sum up these aspects, we propose the following scores:

1. *Ratio of the mutual information in the interclass case to the intraclass case.* This ratio measures the number of bits needed to code one sensor after the other one is known; therefore, it measures the increase in the number of possible states needed to code the information of one sensor when the information of the other sensor is known in the case of the interclass.

$$R_1 = \frac{I_{inter}(X;Y)}{I_{intra}(X;Y)} \quad (7.10)$$

2. Ratio of the maximal cross-correlation between sensors for the interclass case to the intraclass case. This score is justified because while Information Theory based on memory less source entropy considers the image histogram and neglects the spatial position of a given pixel value, this is not the case with cross-correlation, which certainly takes into account the spatial distribution of different pixel values. For this reason, we consider it interesting to check this kind of measure in addition to entropy. Worth mentioning is that Information Theory permits the consideration of relations between consecutive pixels by means of conditional entropy (suitable for memory sources, where the knowledge of previous pixel values reduces the uncertainty to know the next pixel value). Nevertheless, the problem for conditional entropy estimation is that a large amount of pixels conditioned by previous possible values are necessary and normally this is not possible due to the limited amount of data. Thus, a measure that captures these aspects is the ratio of the maximum values of the cross-correlation.

$$R_2 = \frac{\max\{\gamma_{inter}(X;Y)\}}{\max\{\gamma_{intra}(X;Y)\}} \quad (7.11)$$

3. *Normalized ratio of the mutual information.* Measure R_1 could be criticized in the sense that the measures should be normalized in order to capture the fact that the number of bits needed to code the case of the intraclass is different from the number of bits needed to capture the case of the interclass. This is due to the fact that the latter has to deal with a higher number of possibilities. The normalization was done with the joint entropy that captures the different number of states in both situations.

$$R_3 = \frac{\left(\frac{I_{inter}(X;Y)}{H_{inter}(X,Y)} \right)}{\left(\frac{I_{intra}(X;Y)}{H_{intra}(X,Y)} \right)} \quad (7.12)$$

Note equation R_3 is similar to R_1 but instead of considering the absolute value of $I(X;Y)$ it is considered the relative portion to the joint entropy $H(X,Y)$.

Taking into account these proposed ratios, the uses of different combinations of sensors can be summarized in Table 7.2.

<i>Sensor</i>	<i>Ratio value (i.e. R_1, R_2 or R_3)</i>	<i>Theoretical usefulness</i>
Same	High	Worst for biometric recognition
	Low	Best for biometric recognition
Different	High	Best for data fusion
	Low	Best for cross sensor recognition

Table 7.2: Implications of different ratio values for combining information from sensors.

7.4 Experimental Results

We have studied the relations between several pairs of images (TH=thermal, NIR= Near Infrared, VIS= Visible) from the perspective of Information Theory. The experiments have been performed over a subset of the CARL database that consists of 10 different users and 5 different images of each person acquired with three different sensors. Thus, a total amount of $10 \times 5 \times 3 = 150$ images have been used. Table 7.3 shows the experimental entropies for each kind of image alone whereas Table 7.4 depicts the joint, mutual and maximum cross-correlation results for intraclass situation. These results have been obtained averaging the overall previously detailed subset. Figure 7.2 represents an example for a specific person.

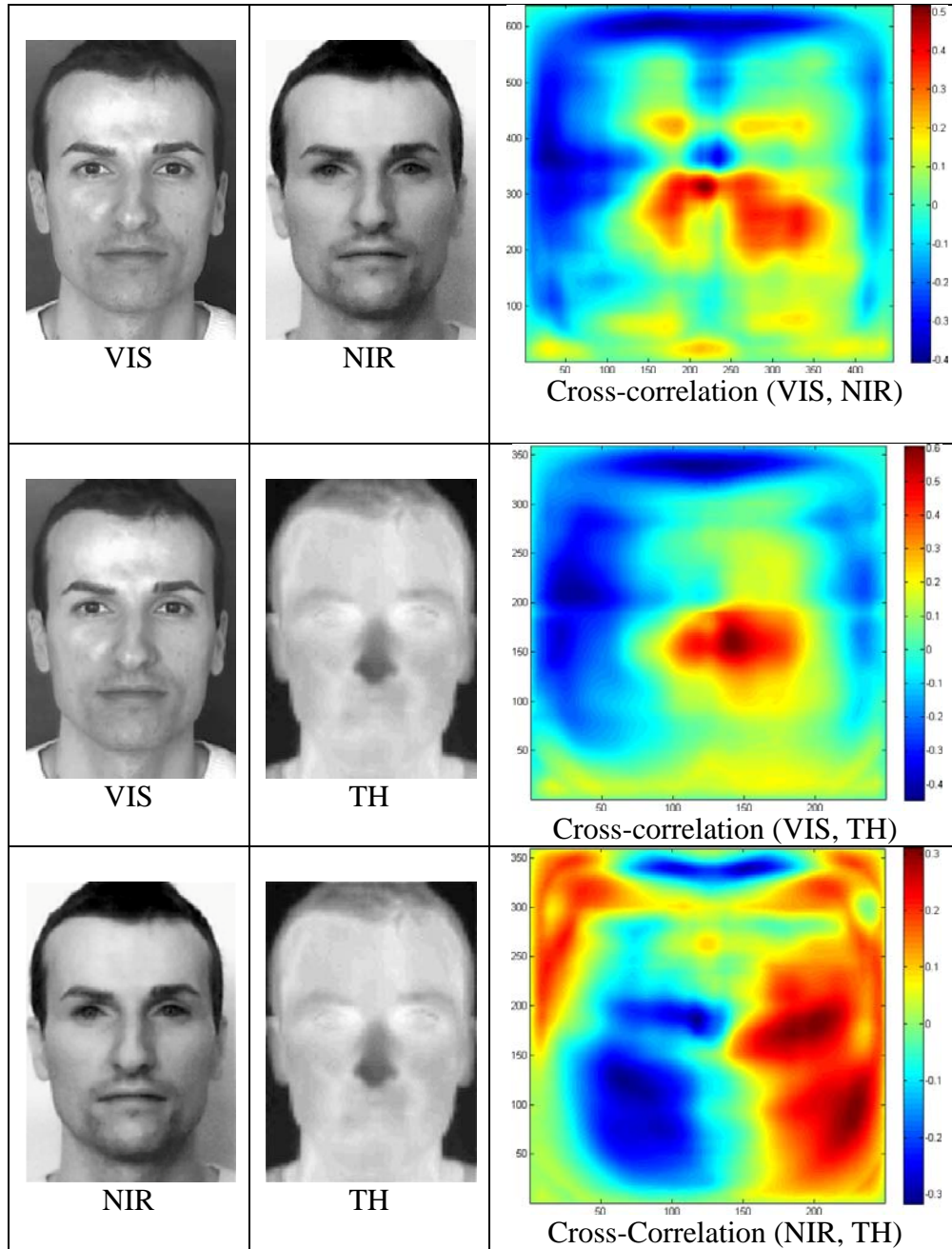


Figure 7.2: TH, NIR, VIS and cross-correlation for person number 1.

	VIS		NIR		TH	
	Mean	std	Mean	std	Mean	std
H(X)	6,73	0,93	7,04	0,40	4,81	0,48

Table 7.3: Experimental entropies for a single image (averaged for 10 people, 5 different images per person) for Visible, Near-Infrared and Thermal images. As shown, NIR images have the largest amount of information, followed by VIS and TH.

Conditions	H(X,Y)		I(X;Y)		H(X/Y)		H(Y/X)		max($\gamma(X,Y)$)		min($\gamma(X,Y)$)	
	Mean	std	Mean	std	mean	std	mean	std	mean	std	mean	std
X=VIS, Y=NIR	12,20	0,84	1,55	0,43	5,18	0,87	5,48	0,54	0,58	0,09	-0,46	0,09
X=VIS, Y=TH	10,66	0,49	0,89	0,28	5,85	0,98	3,92	0,67	0,62	0,09	-0,46	0,08
X=NIR, Y=TH	10,87	0,35	0,99	0,47	6,05	0,51	3,83	0,35	0,49	0,12	-0,52	0,09
X=VIS, Y=VIS	11,60	1,35	1,69	0,43	4,70	0,91	5,23	0,50	0,74	0,08	-0,46	0,08
X=NIR, Y=NIR	12,06	0,59	2,10	0,45	5,03	0,43	4,93	0,39	0,71	0,12	-0,49	0,12
X=TH, Y=TH	9,42	0,96	1,58	0,33	4,26	0,81	3,58	0,59	0,82	0,05	-0,45	0,05

Table 7.4: Experimental results (averaged for 10 people, 5 different images per experimental result) for Visible, Near-Infrared and thermal images.

On the other hand, Table 7.5 extends the measurements computed in Table 7.4 for both intraclass and interclass situation. The values given in this second table will be used to establish the different performances of the pairwise combination of sensors.

Sensors	Person	I(X;Y)	H(X,Y)	I(X;Y)/H(X,Y)	max($\gamma(X,Y)$)
VIS-VIS	Same	1,69	11,60	0,15	0,74
NIR-NIR		2,10	12,06	0,17	0,71
TH-TH		1,58	9,42	0,17	0,82
VIS-VIS	Different	1,23	12,28	0,10	0,62
NIR-NIR		1,48	12,66	0,12	0,49
TH-TH		1,00	10,15	0,10	0,71
VIS-NIR	Same	1,55	12,20	0,13	0,58
VIS-TH		0,89	10,66	0,08	0,62
NIR-TH		0,99	10,87	0,09	0,49
VIS-NIR	Different	1,32	12,60	0,11	0,56
VIS-TH		0,82	10,79	0,08	0,60
NIR-TH		0,88	10,96	0,08	0,50

Table 7.5: Experimental conclusions using the criterion defined in table 1.

Afterwards, we will use the ratio between the interclass to the intraclass in order to compute the three scores (i.e. R_1 , R_2 , R_3) defined in Section 7.3. These scores will indicate the best use of a given pairwise combination. These scatter plots allow us to define the different uses of the pairwise combinations of sensors. Note that of all the possible combinations of interest

summarized in Table 7.1, for the three bands studied in this paper (i.e. VIS, NIR, and TH), only types 1, 3 and 5 were found relevant. The combinations of interest were the following:

- i) Type 1: Low intraclass variation means that models for each class are easy to obtain, and the preferred combination of sensors for this case is the VIS–VIS and TH–TH.
- ii) Type 3: Data fusion that allows for improvements in recognition accuracy. The preferred combinations for all three measurements are the NIR–TH and VIS–TH.
- iii) Type 5- High interclass variation: people easy to discriminate. This type complements type 2. The preferred combination is VIS–NIR.

In Figure 7.3, is depicted that the different combinations of sensors generate two clusters, one related to the combination of the same kind of sensor, which is better for recognition tasks, and one related to the combination of different kinds of sensors, which is better for data fusion. Note that the three scores are coherent in the sense that they give similar values for each pairwise combination of sensors. It should be noted that the difference between R_1 and R_3 was related to a normalization process within the joint entropy, which does not introduce a difference between the ratios, just a scale. One can conclude from the results that, for the separation of applications, the normalization to the joint entropy did not introduce additional improvements. On the other hand, Figure 7.4 plots R_1 versus R_3 . In addition, Table 7.6 provides the computed rates and comments used for building Figure 7.3 and 7.4.

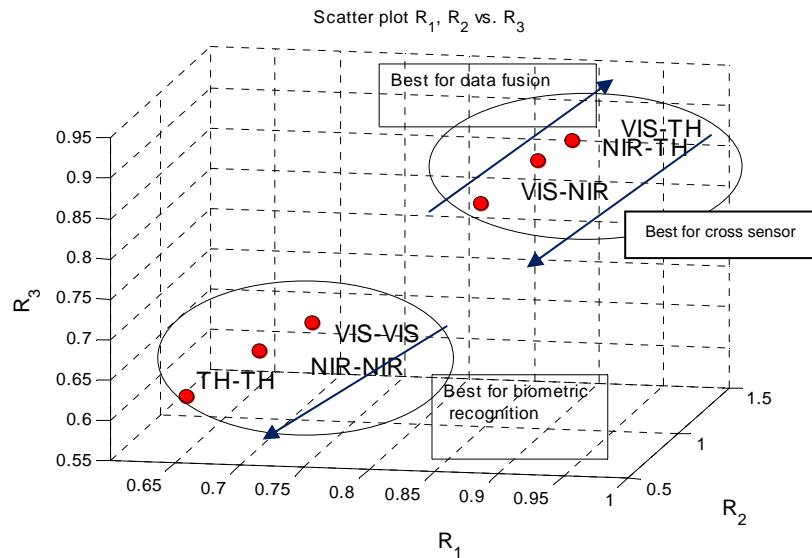
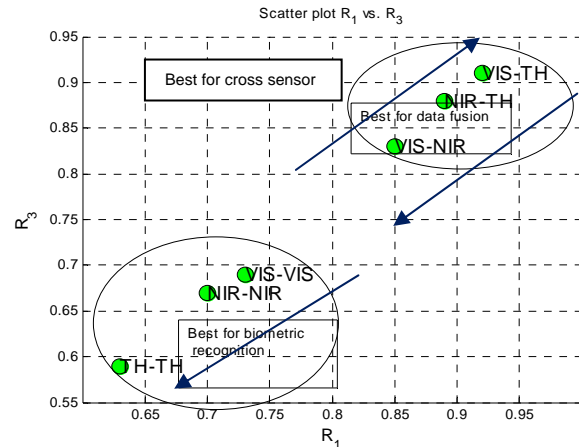


Figure 7.3: Usefulness of each combination.


 Figure 7.4: Comparison of R_1 and R_3 .

Sensors	R_1	R_2	R_3	comment
VIS-NIR	0,85	0,97	0,83	Best combination for sensor fusion
VIS-TH	0,92	0,98	0,91	Best cross-sensor recognition according to $I(X;Y)$ and normalized $I(X;Y)$
NIR-TH	0,89	1,02	0,88	Best cross-sensor recognition according to $\gamma(X,Y)$
VIS-VIS	0,73	0,84	0,69	Best sensor according to $I(X;Y)$ and normalized $I(X;Y)$
NIR-NIR	0,70	0,69	0,67	
TH-TH	0,63	0,87	0,59	Best sensor according to $\gamma(X,Y)$

Table 7.6: Experimental ratios and conclusions.

7.5 Conclusions

Based on the experimental results presented in Tables 7.3 and 7.4, we can do the following observations, concerning to the image properties:

- NIR images have higher entropy than the other ones and lesser standard variation. In principle, this is a good result (high amount on information and small variation along the 50 evaluated images). We think that this is due to the infrared LED illumination.
- The three kinds of images analyzed (visible, NIR and thermal) are not redundant. Thus, for instance, it is possible to enhance a face recognition system by means of data fusion between biometric classifiers performing on each kind of image. In fact, experimental results of Table 7.4 reveal that the mutual information is low, between 0,89 and 1,55 bit when comparing a pair of images of the same person acquired with a different kinds of sensor and between 1,58 and 2,10 bit when acquired with same kind of sensor. This indicates that more advantage can be taken when combining couples of images acquired with different sensor.
- The mutual information is higher for the NIR-TH couple than for the VIS-TH one. This is reasonable, because NIR-TH are close to each other than VIS-TH.

- The highest normalized cross-correlation for couples of same kind of sensor images is obtained for TH images ($\max(\gamma(X, Y)) = 0,82$). This implies more redundancy between different acquisitions of same person for this kind of images. This can be interpreted as stability (less variation), which makes them interesting for pattern recognition applications such as biometric face recognition based on thermal images. This higher stability can also be observed when looking at mutual information, because a value of 1,58 bit in images of entropy 4,81 is higher redundancy than 2,10 in images of entropy 7,04 (32,8% in front of 29,8%).

Comparing the desirable properties of Table 7.5 and 7.6, we can establish the following conclusions:

- Best combination for sensor fusion should be obtained with VIS and TH.
- Best cross-sensor recognition (training with one sensor and recognition with a different one) will be obtained with the pair VIS–NIR. This is an interesting finding because in real scenario applications during the night, one would normally have only the NIR source, while the training database most probably will have samples in the visible range.
- Best recognition will be obtained with NIR images (according to R_2) and thermal images (according to R_1 and R_3). Worth mentioning is the fact that R_2 is related to cross-correlation, while R_1 and R_3 are based on entropies. Thus, we can interpret that NIR images will produce better recognition rates than other spectral ranges when using recognition algorithms based on correlation between images. On the other hand, thermal images will produce better results when using histogram-based methods such as [Bre03], because entropy computation requires the image histogram information.

Therefore, research results obtained are relevant and give support to the first notion pointed at the beginning of this chapter, that images in different spectral bands are rich in not redundant but complementary information, and consequently useful for data fusion. This encourages us to continue our aim of working with different spectral bands for FR, as will be done and will be fully described in the following chapter.

Chapter 8

Proposed Face Recognition Approach

Any sufficiently advanced technology is indistinguishable from magic.
Arthur Clarke, *Profiles of the Future*. 1962.

This chapter is devoted to the implemented recognition system, which exploits the strength of multispectral information for FR, and constitutes on the main contributions of this dissertation. First section presents the system overview of our proposal, whereas in section two a much more in-depth view is given, detailing the principles and procedures for the design of each stage: face segmentation and normalization procedure, feature extraction, feature selection, matching process and score fusion. A description of the overall system has been published in [Esp12], whereas the proposed new criterion for face segmentation of facial thermal images has been published in [Mek10].

8.1 System Overview

The perspective considered to approach the concerned system is derived from the discussion presented in the overall preceding chapters, in an attempt to deal the issues directly related with the special requirements of FR when dealing with faces simultaneously taken in different spectra.

Following list presents the different proposed steps in order to carry out the goal pursued: (i) multispectral face acquisition, (ii) face segmentation and normalization, (iii) feature extraction, (iv) feature selection, (v) classification and (vi) fusion. We will focus on the last five ones due to the first singular image acquisition process, has already been fully discussed in Chapter 6.

Face segmentation, as pointed in Section 3.2, is an important preprocessing step for FR that consists in detecting the face in an image (previously known its existence), enabling subsequent background removal. Nevertheless, due to the inexistence of specific segmentation algorithms for dealing with thermal images, a new algorithm has been required and accordingly, developed. The referred design will be fully discussed in the following Section 8.2.1. In addition, a normalization process of all the images of the concerned database to the same size for compensating remaining zoom and translation problems has also been carried out.

As has been widely discussed along the overall document, the case of face recognition has the particularity that the dimensionality of the images is extremely high. Thus, once the segmented faces are available, we mainly focused on the study and design of the feature extractor block, ranging from simple to complex statistical modelling in order to better represent the underlying data structure. Taking into account that we were looking for a low-complexity FR system, we have finally proposed the use of the discrete cosine transformed domain in order to compute the feature extraction. Thus, the obtained output result will be managed as a *pattern vector*, whose components will be computed one at a time to allow the quickest possible response. Since this moment, images (faces) will be seen as the set of coefficients in the DCT transformed domain. Nevertheless, even after transforming high-dimensional images into much lower dimensional pattern vectors, a selection method for detecting the most relevant coefficients will be also helpful in order to enhance final performance. In this sense, a feature selection algorithm focused on optimizing low intra-class and high inter-class variations, has been subsequently designed and applied.

On the other hand, the issue of the classification task will specifically perform the recognition process and will provide the corresponding matching scores. This module has been carried out by means of a simple and yet powerful pattern matching based on distances computation by using a fractional distance. In order to preserve the user privacy and to avoid new user enrollment drawbacks, a *generative (informative)* model has been suggested as estimation criteria for adjusting such classifier block. Note that feature extractor and matcher blocks have been implemented in an easy way in order to make feasible the required overall final complex fusion system and to alleviate computational burden and time requirements. In this sense, empirical results reported in Chapter 9, strongly support the proposed idea in almost all conditions.

Finally, a fine fusion method for appropriately dealing the available complementary information of each spectrum, as reported in Chapter 7, has been carried out, in order to better improve the performance of any of the three individual recognition systems. This module has been formulated at matching score level, and both a fixed and a trained technique to combine the set of matching scores have been exploited in order to establish the final identification decision. The architecture of the overall system is shown in Figure 8.1.

To the best of our knowledge, this is the first system that deals with facial images simultaneously taken in the VIS, NIR and TIR spectra and under different illumination conditions, as appropriately discussed in Chapter 6.

8.2 The Proposed System

This section provides the system description and the implementation details of the finally designed stages: segmentation, feature extraction and selection, pattern matching and fusion. The concerned theories were firstly addressed in Sections 2.4, 2.5.2 and 3.2 and subsequently more deeply explained in Section 3.3.2.

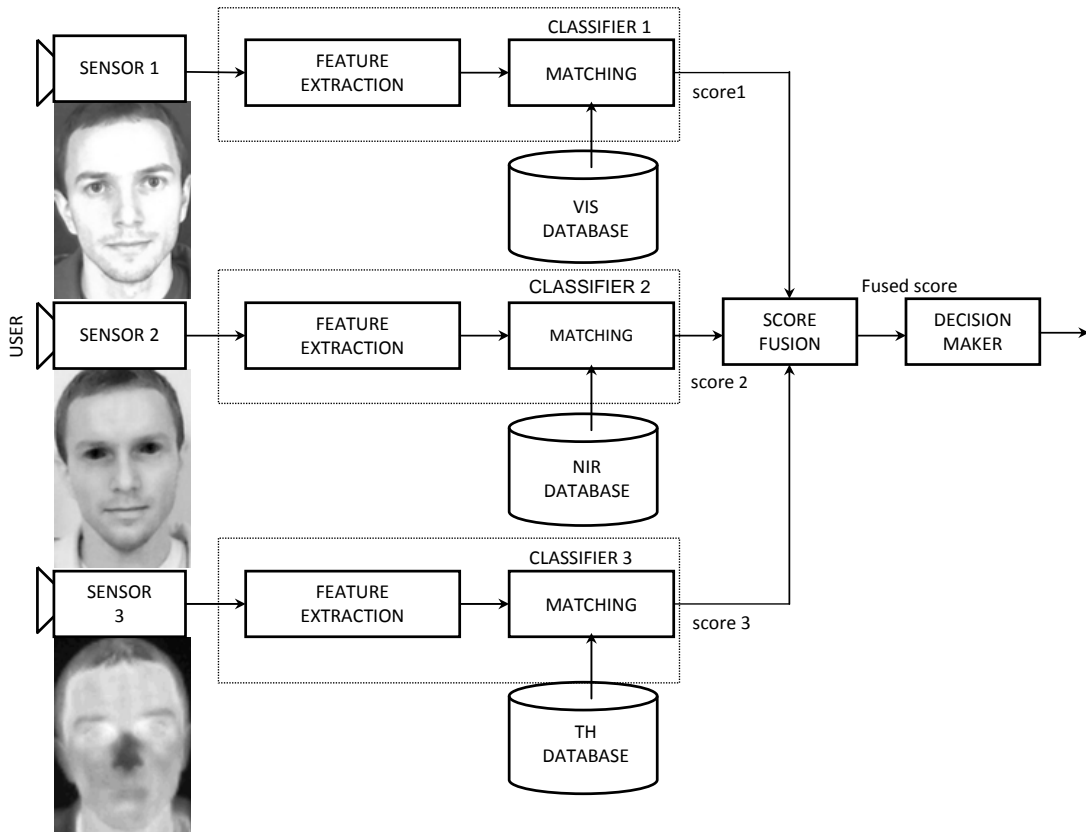


Figure 8.1: Overall purposed processing system. Multi-sensor fusion at score level.

8.2.1 Face Segmentation Algorithm

The success of the system is directly related with the prior step to FR, the accurate detection of human faces, becoming one of the most important involved processes. When faces are located rightly in the image, the subsequent recognition process will constitute a feasible task.

While this topic has attracted a considerable attention when dealing with VIS images, a large amount of topics remains unsolved for thermal images. We have used a Viola and Jones face detector for face segmentation of the VIS & NIR images [Vio01]. However, according to our experiments reported in Chapter 9, this algorithm is unable to correctly segment thermal images due to their different properties. Therefore, a new fast and efficient face segmentation algorithm for thermal images has been proposed [Mek10].

The goal is to detect the coordinates of the rectangle $(x_1, y_1 ; x_2, y_2)$ that contain the face in a simple and efficient way, as follows:

1. The image is binarized, where the threshold T is chosen by the Otsu's method [Ots79]. If $I(x,y) < T$ then $I(x,y) = 0$, otherwise, $I(x,y) = 1$.
2. The vertical and horizontal (VH, V= Vertical, H= Horizontal) projections are calculated.
3. The first border of the vertical profile is marked as y_1 (see Figure 8.2(b)).
4. $h/2$ is the length from the y_1 to the lower part of image.
5. The part of face from y_1 to $h/2$ is selected and then positions x_1 and x_2 are estimated as the left and right limits of this portion.
6. The lower part of the face is detected by means of: $y_2 = y_1 + (x_2 - x_1) * 13/9$

Note that the described algorithm does not need any training procedure. Thus, the accuracy and duration of face segmentation is thus not dependent on the amount of training data, which is also an extra advantage of this method. The steps of the VH projection are schematically visualized on Figure 8.2 using a sample thermal image, on the top of the image. Last image shows a green rectangle or box with the outcome of the detection process. Thus, each green box will correspond to an image fragment that the system will consider as a face.

8.2.2 Feature Extraction Algorithm

The goal of this block is to reduce the dimensionality of the vectors in order to simplify the complexity of the classifier and to improve recognition accuracy. In order to compare the identification rates using the full battery of different sensors and illumination conditions, rather than use complex cutting edge techniques, we have decided to use the well known feature extraction method based on Discrete Cosine Transform (DCT). The reasons to take such approach are summarized as follows:

- According to our previous experiments published in [Fau07a, Fau07b], DCT approach outperforms the well known eigenfaces [Tur91] algorithm with lesser computation burden and time consuming (see Table 3.1). This fact provides real time applications, even on a low-cost processor.
- Although DCT is similar to the DFT, DCT has two advantages when compared with DFT:
 - The application of the DCT to an image (real data), produces a real result, whereas DFT produces complex values.
 - The high-intensity coefficients of the DCT transform are concentrated in a smaller area¹ than for DFT [The06].

¹The size of such area depends on the energy compaction properties of the respective transform.

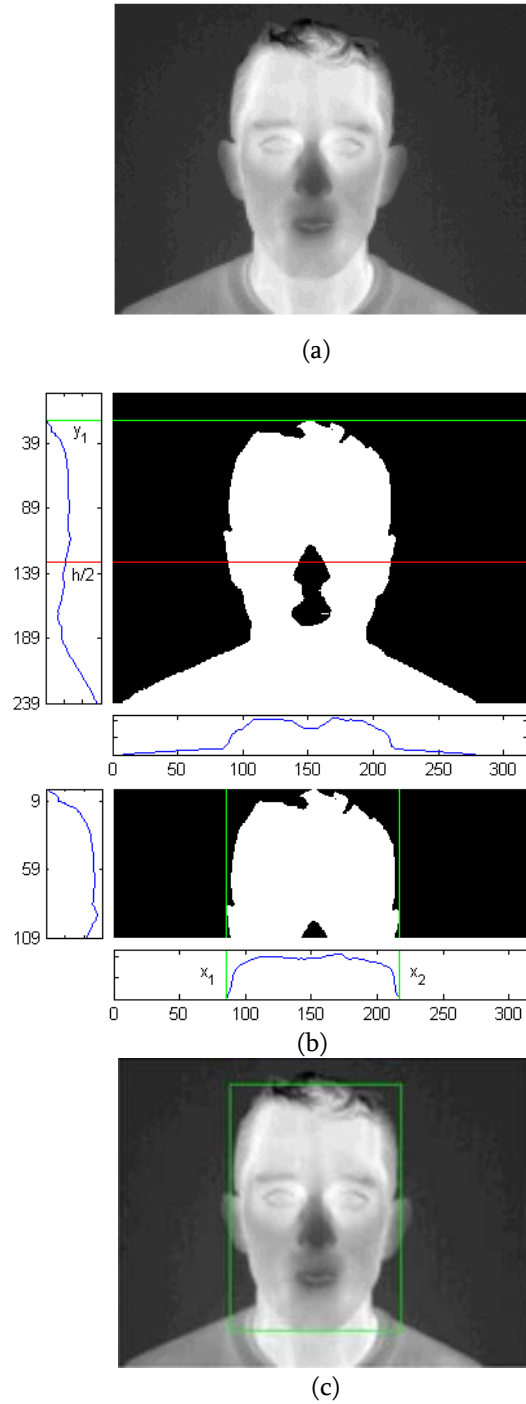


Figure 8.2: Thermal Face Segmentation algorithm steps for a sample image.

- Is a data independent transform (basis functions are known in advance, which means is not necessary to find any projection vector set). Therefore, no generalization problem can appear when dealing with new users and images not used during enrollment.
- Their properties and behaviour are well understood and documented in the available state of the art. This last reason is crucial to extract conclusions from results related to the different nature of the images in the different considered lighting conditions.
- In addition:
 - Feature vectors are fixed-length. (Fusion at feature extraction level will be easier feasible).
 - Is possible to apply filters on the transformed domain defining a zonal mask² as the array of the form defined in (8.1) [Fau 07a] and multiply the transformed image by the zonal mask, which takes the unity value in the zone to be retained and zero on the zone to be discarded.

$$m(f_1, f_2) = \begin{cases} 1, & f_1, f_2 \in I_t \\ 0, & \text{otherwise} \end{cases} \quad (8.1)$$

This same kind of solution is also preferred in image coding algorithms (JPEG, MPEG, etc.) that use DCT instead of KLT, because it is a fast transform that requires real operations and it is a near optimal substitute for the KL transform of highly correlated images, and has excellent energy compactation for images. Moreover, is important to emphasize as discussed in Section 3.3.2, that DCT asymptotically tends to the KLT when the block size sample is large small ($N > 64$), which is our FR addressed problem, because we deal with final images of 14.500 components.

Given a face image, the first step is to perform a two dimensional DCT, discussed in Section 3.3.2, which provides an image of the same size but with most of the energy compacted in the low frequency bands (upper left corner).

The original signal is converted to the frequency domain by applying the cosine function for different frequencies. After the original signal has been transformed, its DCT coefficients reflect the importance of the frequencies that are present in it. The very first coefficient refers to the signal's lowest frequency, and usually carries the majority of the relevant (the most representative) information from the original signal. The last coefficients represent the component of the signal with the higher frequencies. These coefficients generally represent greater image detail or fine image information, and are usually noisier. We have used `dct2` (two dimensional DCT) defined in equations (3.4) and (3.5) in order to compute the concerned

²In image coding it is usual to define the zonal mask taking into account the transformed coefficients with largest variances. (Whereas in image coding the goal is to reduce the amount of bits without appreciably sacrificing the quality of the reconstructed image, in image recognition the number of bits is not so important).

transformation. On the other hand, Figure 8.3 summarizes the process to obtain a feature vector from a DCT transformed image.

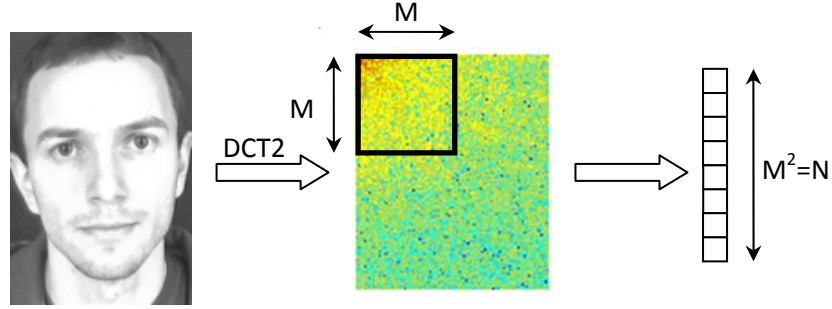


Figure 8.3: Feature extraction after transforming by the DCT, and selecting a subset of the low frequency components.

Mainly, the process consists of preserving the low frequency components around the $X[0,0]$ coefficient (DC coefficient) of the transformed domain, which at the same time have more energy. In earlier versions of our system, we produced low-dimensional feature vectors re-ordered in this way.

8.2.3 Feature Selection Algorithm

As has already been stated in the preceding Section 8.1, although feature extraction accomplishes with the reduction of the dimensionality of the feature vector by truncating the coefficients to a smaller window of lower frequency coefficients, a feature selection algorithm will be also required in order to better improve the accuracy of the overall system. Later experiments also revealed that a slightly higher performance could be delivered by modifying the generation of the feature vector. In this sense, feature vector coefficients have been appropriately chosen by following the *fisher discriminant*. The goal is to pick up those frequencies that yield a low intra-class variation and high inter-class variation based on identifying and maximizing independence relations [Esp12].

On the other hand, those frequencies that provide a high variance for inter and intra-class distributions should be discarded. Thus, the fully process for obtaining the final feature vector from face image will consist on applying the 2D-DCT immediately afterward the concerned frequency selection mechanism. The notation is the following one:

- P is the number of people inside the database.
- F is the number of images per person in the training subset.
- $i_{p,f}(x,y)$ is the luminance of a face image f that belongs to person p , where $p = 1, \dots, P; f = 1, \dots, F$
- $I_{p,f}(f_1, f_2) = \text{transform}\{i_{p,f}(x, y)\}$ is the 2D-DCT transformed image.

- $m(f_1, f_2) = \frac{1}{P \times F} \sum_{p=1}^P \sum_{f=1}^F I_{p,f}(f_1, f_2)$ is the *average* of each frequency obtained from the whole training subset images.
- $m_p(f_1, f_2) = \frac{1}{F} \sum_{f=1}^F I_{p,f}(f_1, f_2) \quad \forall p=1, \dots, P$ is the *average* of each frequency for each person p .
- $\sigma_p^2(f_1, f_2) = \frac{1}{F} \sum_{f=1}^F (I_{p,f}(f_1, f_2) - m_p(f_1, f_2))^2$, $\forall p=1, \dots, P$ is the *variance* of each frequency for each person p .
- $\sigma^2(f_1, f_2) = \frac{1}{P \times F} \sum_{p=1}^P \sum_{f=1}^F (I_{p,f}(f_1, f_2) - m(f_1, f_2))^2$, $\forall p=1, \dots, P$ is the *variance* of each frequency evaluated over the whole training subset.
- $\sigma_{\text{intra}}^2(f_1, f_2) = \sum_{p=1}^P \sigma_p^2(f_1, f_2)$ is the average of the variance of each frequency for each person.
- $\sigma_{\text{inter}}^2(f_1, f_2) = \sigma^2(f_1, f_2)$.

We will use the following measure, which is the *Fisher discriminant*:

$$M_1(f_1, f_2) = \frac{|m_{\text{intra}}(f_1, f_2) - m_{\text{inter}}(f_1, f_2)|}{\sqrt{\sigma_{\text{intra}}^2(f_1, f_2) + \sigma_{\text{inter}}^2(f_1, f_2)}} \quad (8.2)$$

It is interesting to point out that this procedure is similar to the threshold coding used in transform image coding [Jai89]. Nevertheless, we are using a *discriminability criteria*, instead of a *representability criteria*, which is only based on energy (the higher the frequency coefficient value, the higher its importance). Figure 8.4 shows an example of the M_1 ratio defined in (8.2) obtained from visible images of session 1. It is important to point out that feature selection has been done using only training samples. Thus, we have not selected frequencies using testing samples, which would provide better results, but unrealistic because feature selection must be done a priori, before classifying samples.

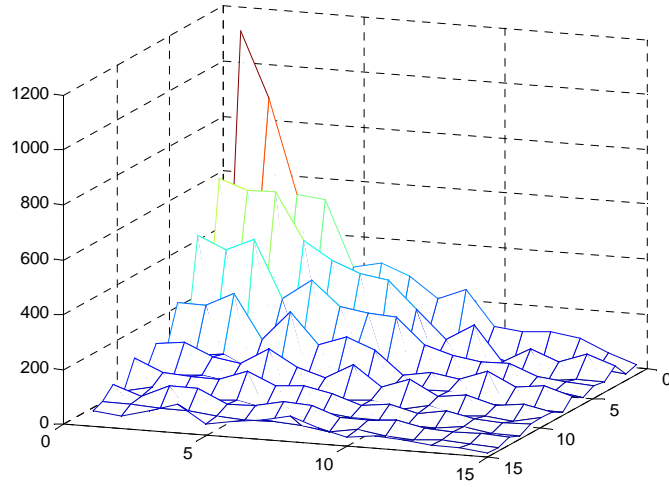


Figure 8.4: 15x15 first coefficients M_1 ratio for visible images of session 1. It is evident that the highest discriminability power is around the low frequency portion (upper left corner).

8.2.4 Classification

All facial images in different spectra are projected to a lower dimensional feature space as described in the former sub-section, and therefore, each pattern and template (training and testing feature vectors) will be represented by a feature vector of dimension N .

To accomplish the following classification task we have applied a simple yet robust classifier based on computing distance between training and testing feature vectors using a *fractional distance*. It is modelled as follows:

$$d(\vec{x}, \vec{y}) = \left(\sum_{i=1}^N (|x_i - y_i|)^p \right)^{1/p} \quad (8.3)$$

where i is the feature vector component.

For $p=2$ the equation corresponds to the Euclidean distance. When data are high dimensional, however, the euclidean distances and other *Minkowsky norms* (p -norm with p being an integer number, i.e, $p = 1; 2; \dots$) seem to concentrate and, so, all the distances between pairs of data elements seem to be very similar [Fra07]. Therefore, the relevance of those distances has been questioned in the past, and fractional p -norms (Minkowski-like norms with an exponent p less than one) were introduced to fight the concentration phenomenon. In our case, we have used $p=0.5$.

With regard to the number of coefficients (vector dimension) we have experimentally selected it by trial and error, selecting windows of 1×1 , 2×2 , 3×3 , ..., $M \times M$, where the frequency coefficients have been previously ordered using the strategy defined in the preceding Sub-section.

We have used a simple method because the experiments are quite time consuming. For each feature vector dimension we have executed the algorithm, we have studied hundreds of feature vector dimensions for each condition, and this implies thousands of executions. If using a more sophisticated method this would imply, probably, to train a complex algorithm for each studied feature vector dimension. This would be impractical from the computational burden point of view. In fact, in its current version, we required several weeks to work out the whole experimental section.

In addition, we were looking for a method with few parameters because a more complex algorithm can require fine tuning, and this fine tuning could be different for each spectral band. Thus, in this case, it would be difficult to know if one spectral band provides better results due to different tuning or to the frequency itself. Our suggested method is so simple and effective that we did not required any fine tuning.

On the other hand, we have developed the classification task from a *generative* point of view, as a matter of fact, as human brain does [Lee99], instead of a *discriminative* one (see Section 3.3.1), mainly due to the following reasons: In biometrics, *discriminative* approaches are usually a one-vs-all approach and correspond to training the classifier to differentiate one user from the others. This means that the algorithm requires samples from the given user but also samples belonging to the other ones. Sometimes this can raise privacy concerns. On the other hand, when a new user must be added into the BS, the whole system must be retrained, which is time consuming and can be a drawback for a real operating system that must enroll new users (or remove users) quite frequently. Thus, in our system, when adding a new user, it will be no necessary to re-train the models of the other users (we compute the model of a specific user without requiring samples from the other users). A more detailed discussion about generative vs discriminative models can be found in [Rub97, Fab08, Yan01].

8.2.5 Fusion Method

As already discussed along the document, Near-IR and thermal images contain complementary information [Esp11] to visible images. This opens a large amount of possibilities of data fusion for enhancing pattern recognition applications.

The image produced by employing fusion method will consolidate evidences obtained from three handled sensors, providing with this way more detailed and reliable information which will assist us in constructing a more efficient FR system.

As previously described in Section 2.4 and subsequently, reviewed in Section 3.4, data fusion takes into account information at different levels (sensors, feature, score and decision) in order to improve the final recognition accuracies. In our approach, the final fusion step has been formulated as a score matching problem. In this sense, the final solution will be viewed as a combination of the scores provided by each of the three matchers. Actually, the overall system will fit to a multi-sensor fusion architecture at score level, due to it takes into account three different sources of information (sensors in different spectra) and every matcher just provides a

distance measure or a similarity measure between the input features and the models stored on the database.

Before score fusion, normalization must be done when the scores provided by different classifiers do not lie in the same range. In our case, we experimentally found that this normalization is not necessary because the three studied classifiers gave similar range.

After the normalization procedure, fusion stage is finally applied. In order to establish the relevance of each classifier, a combination scheme based on the weighted averaging has been proposed, where weights can be both fixed or trained, as discussed in Section 2.4. Equations (8.4) and (8.5) describe computed training rules when combining two and three classifiers respectively:

$$O = \omega_1 o_1 + \omega_2 o_2 = \omega_1 o_1 + (1 - \omega_1) o_2 \quad (8.4)$$

$$O = \omega_1 o_1 + \omega_2 o_2 + (1 - \omega_1 - \omega_2) o_3 \quad (8.5)$$

It is important to take into account that in our case the purpose of the trained rule is to evaluate the weights assigned to each classifier, rather than to maximize the identification rate. In fact, trained rules should be done with a development set different than the test set. Otherwise the experimental results are unrealistic. Nevertheless, these optimistic results are well accepted by the scientific community. The same situation occurs, for instance, in biometric verification performance measured by means of EER. EER corresponds to a posteriori threshold setup, which is an optimistic (unrealistic) situation. In “real systems” the threshold must be setup a priori and then the errors should be computed with testing samples not used for threshold computation. This implies some degradation on experimental results, but it is well accepted this unrealistic (optimistic) situation.

Chapter 9

Experiments and Results

Seeing is believing, but feeling is being sure.

John Ray

In this chapter, a large set of experimental results regarding the designed recognition system in different conditions is appropriately reported, demonstrating the competitiveness of our approach. Last section provides main conclusions.

9.1 Introduction

The hypothesis of our research work has been intrinsically formulated at different levels along the document. In broad lines, we can postulate it as follows:

The use of multispectral images assisted by fusion methods appears to be a very useful strategy for representing faces taking advantage of complementary meaningful information of each spectrum, relieving with this way the strong dependence of lighting conditions in current FR performance.

The aim of this chapter is to develop an experimental framework in order to explore the goodness of the hypothesis as well as the capabilities of the proposed system that will give statistically significant conclusions [Box78] about the properties of our proposal. The set of experiments have been designed to assess the contribution of each factor (kind of sensor, illumination conditions, pre-processing steps, number of coefficients, training and testing sub datasets, weighting factors, combination of classifiers, etc.) to the overall effectiveness of the system. Notice that the number of comparisons can be extremely high, so the experimental design should be done carefully.

For the face segmentation experiments we have used a specific subset of the CARL face database appropriately described along the Chapter 6 whereas for the rest on the experiments we have used the full database, which is large enough to both train the system and evaluate its performance.

9.2 Experimental Results

All the reported data below comes from an accurate real study conducted via programs written in Matlab, with the exception of Viola Jones algorithm which was written in *.cpp and then compiled to *.mex32.

Moreover, CARL database comprises a total amount of 41 users; so from the identification point of view, we are facing a 41 class problem.

9.2.1 Experimental Results of the Face Segmentation Preprocessing Step

This subsection provides the data results once applied the VH projection algorithm presented in Section 8.2.1, over both visible and thermal images. In addition and for valid comparison purposes, we have also applied the classical competitive state of the art Viola and Jones algorithm [Vio01] used for visible images over thermal ones. As known, this second algorithm requires some training. Hence, the experiment contains three different scenarios regarding the procedure of training Haar cascade used.

During the training procedure two Haar cascades were trained: one for VIS images and the other one for TH images. These scenarios are described below:

- i) Scenario 1 (SC1): 900 negative (background) and 1800 positive (containing face) training images. 1800 positive images were divided into three parts containing 600 images. First part was acquired under artificial illumination (AR), second under near-infrared (IR) and third under natural (NA) illumination.
- ii) Scenario 2 (SC2): 300 negative and 600 positive training images. These 600 positive images were acquired under artificial illumination.
- iii) Scenario 3 (SC3): 1400 negative and 1800+2400 positive training images. 1800 positive images were divided into three parts containing 600 images. First part was acquired under AR illumination, second under NIR and third under NA illumination. Next 2400 positive images were acquired under artificial illumination.

The Haar cascades were trained on frontal faces. In case of TH images, the training faces were segmented by VH projection and then manually checked or edited. In case of VIS images, the training faces were segmented by another Haar cascade trained on 7000 positive images. These faces were again manually checked or edited.

During the evaluation we were interested in successful detection rate (SDR) and time needed to segment the face. The SDR has been defined as $100 \cdot (N_c / N_a) [\%]$, where N_c is number of correctly detected faces and N_a is number of all faces.

Criterion of successful detection applied has been the following: (i) the face must have contained at least browns, both eyes and whole lips. (ii) The face contained maximally bottom of head and in the lower parts the part of neck. Examples of badly and correctly detected faces can be seen on Figure 9.1.

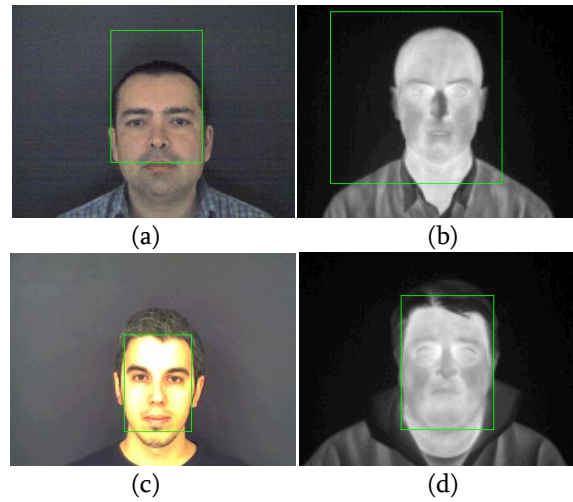


Figure 9.1: 1st row: Examples of badly detected face in VIS (a) and TH spectrum (b) using Viola and Jones. 2nd row: Examples of correctly detected faces in VIS and TH spectrums.

The following constitute the set of results of the algorithm developed for all three scenarios. As stated at the beginning of this section, these results have been compared with Viola and Jones algorithm that will provide the baseline on which to make comparisons.

<i>Scenario</i>		<i>AR</i>	<i>IR</i>	<i>NA</i>
SC1	VIS VH projection	95%	55%	83%
	Viola-Jones	51%	48%	46%
	TH VH projection	100%	100%	100%
	Viola-Jones	93%	96%	97%
SC2	VIS VH projection	95%	55%	83%
	Viola-Jones	40%	42%	40%
	TH VH projection	100%	100%	100%
	Viola-Jones	69%	71%	66%
SC3	VIS VH projection	95%	55%	83%
	Viola-Jones	62%	72%	55%
	TH VH projection	100%	100%	100%
	Viola-Jones	95%	93%	96%

Table 9.1: Successful detection rates.

<i>Scenario</i>			<i>d [s]</i>	<i>mean(d) [ms]</i>	<i>var(d) [ms]</i>
SC1	VIS	VH projection	0,740	2,311	0,012
		Viola-Jones	10,068	31,463	0,503
	TH	VH projection	0,578	1,806	0,066
		Viola-Jones	5,808	18,149	0,208
SC2	VIS	VH projection	0,754	2,356	0,011
		Viola-Jones	9,971	31,159	0,553
	TH	VH projection	0,547	1,708	0,044
		Viola-Jones	5,756	17,987	0,184
SC3	VIS	VH projection	0,767	2,397	0,011
		Viola-Jones	10,438	32,618	0,524
	TH	VH projection	0,450	1,407	0,019
		Viola-Jones	6,440	20,125	0,240

Table 9.2: Detection time of 220 images under artificial illumination.

<i>Scenario</i>			<i>d [s]</i>	<i>mean(d) [ms]</i>	<i>var(d) [ms]</i>
SC1	VIS	VH projection	0,646	2,019	0,002
		Viola-Jones	10,097	31,552	0,531
	TH	VH projection	0,435	1,358	0,001
		Viola-Jones	5,726	17,894	0,185
SC2	VIS	VH projection	0,674	2,106	0,003
		Viola-Jones	9,548	29,838	0,411
	TH	VH projection	0,432	1,350	0,001
		Viola-Jones	5,884	18,388	0,227
SC3	VIS	VH projection	0,677	2,114	0,002
		Viola-Jones	10,348	32,339	0,549
	TH	VH projection	0,371	1,158	0,001
		Viola-Jones	6,197	19,365	0,172

Table 9.3: Detection time of 220 images under IR illumination.

<i>Scenario</i>			<i>d [s]</i>	<i>mean(d) [ms]</i>	<i>var(d) [ms]</i>
SC1	VIS	VH projection	0,675	2,108	0,002
		Viola-Jones	10,178	31,807	0,528
	TH	VH projection	0,430	1,345	0,001
		Viola-Jones	5,800	18,125	0,203
SC2	VIS	VH projection	0,692	2,163	0,003
		Viola-Jones	9,886	30,895	0,539
	TH	VH projection	0,438	1,368	0,001
		Viola-Jones	5,702	17,819	0,165
SC3	VIS	VH projection	0,697	2,180	0,003
		Viola-Jones	10,404	32,512	0,524
	TH	VH projection	0,376	1,175	0,001
		Viola-Jones	6,231	19,472	0,175

Table 9.4: Detection time of 220 images under natural illumination.

Experimental results reveal that VH projection outperform the Viola-Jones algorithm for the three kinds of image illumination in both spectrums (VIS and TH) according to computational burden (Tables 9.2 to 9.4) as well as SDR (Table 9.1). Only for IR and SC3 Viola and Jones provides higher SDR (see Table 9.1). Nevertheless we should emphasize that classical Haar cascade training requires a higher number of face images than we have used. We could not use more due to the limitation of the used databases.

In addition and as has been broadly mentioned along the document, thermal images are not affected by the illumination. This property is also good for the detection rate of Viola-Jones, because if the images are acquired under three different kinds of illumination, we can theoretically use for training three times less TH images than VIS images to obtain the same SDR. If we compare SDR of Viola-Jones algorithm using haar cascade trained on 600 TH images (SC2) and SDR of Viola-Jones using Haar cascade trained on 1800 VIS images (SC1) affected by illumination, the results in case of SC2 are still better.

Figure 9.2 shows an example of several segmented faces (over different illumination conditions). In case of first and third row, our proposed algorithm was used. In case of second and fourth row, we applied the Viola-Jones algorithm. It is also interesting to comment the possibility to use the combination of VH projection and Viola-Jones algorithm. This combination has been also used in this research work. VH projection can be used to automatically create a large thermal training database for Viola-Jones.

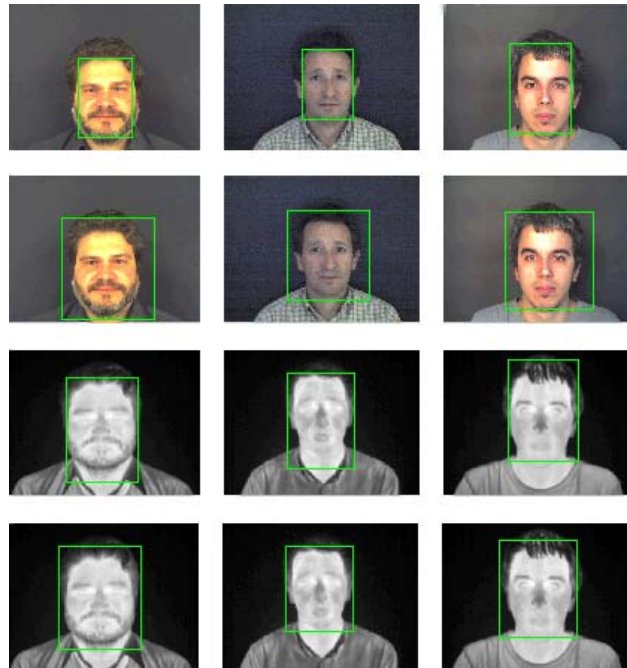


Figure 9.2: Some examples of pictures and the resulting face segmentation using our proposed algorithm (rows 1 and 3) and Viola-Jones algorithm (rows 2 and 4). Pictures in column 1 is acquired under AR illumination, in column 2 under IR illumination and in column 3 under NA illumination.

9.2.2 Experimental Results with Different Illumination Conditions

During the first phase of the analysis, initial performance evaluations have been made to assess the ability of every sensor in an individual way in function of the different illumination conditions.

In this section we have compared the identification rates for the visible (VIS), Near infrared (NIR) and thermal (TH) sensors for natural (NA, Figure 9.3), artificial (AR, Figure 9.4) and infrared (IR, Figure 9.5) light. These experimental results have been obtained training with session 1 & 2 and testing with session 3 as function of M (see Figure 8.3). Thus, the number of selected coefficients for each point in these plots is $M^2=N$.

These figures reveal several general interesting facts:

- Feature selection is indeed important, because a too large number of coefficients decreases the identification rate.
- Different sensors provide a different number of optimal feature dimension M^2 .

Specifically:

For natural light, Figure 9.3 shows that:

- The NIR sensor provides lower identification rates than visible and thermal, which provide similar rates. In addition, the optimal feature vector size is more critical, because identification rates drop quickly when moving away from the optimal point.
- The TH sensor requires a lesser amount of coefficients to reach the highest identification rate, and the identification rate drops slower than for visible sensor.

For artificial light, Figure 9.4 shows that:

- All the sensors provide nearly similar results, although the visible sensor outperforms the other ones.
- Optimal feature vector size selection is lesser critical for the VIS sensor than for the other ones because a large range of M^2 values produce the highest achievable identification rate.

For Near-IR light, Figure 9.5 shows that:

- The NIR sensor provides the best behavior and the VIS sensor fails to provide a reasonable identification rate. This makes sense considering that infrared illumination in the proposed scenario for a visible sensor is equivalent to an almost dark scene.
- TH and NIR provide similar behavior, although TH sensor results drop faster beyond the optimal value.

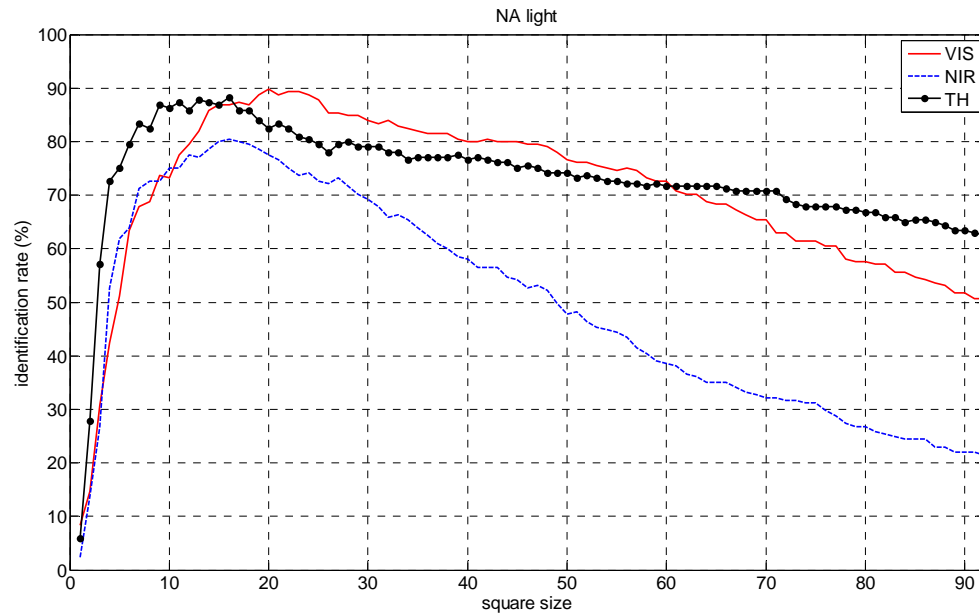


Figure 9.3: Identification rate as function of the square size (N) of selected coefficients for visible (VIS), near infrared (NIR) and thermal (TH) sensors for natural (NA) illumination.

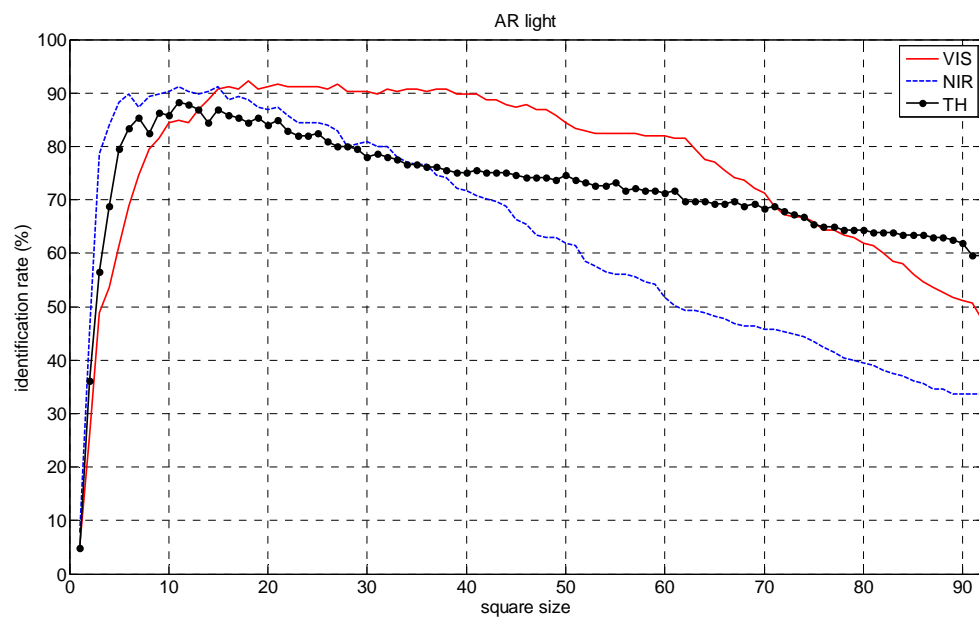


Figure 9.4: Identification rate as function of the square size of selected coefficients for VIS, NIR and TH sensors for artificial (AR) illumination.

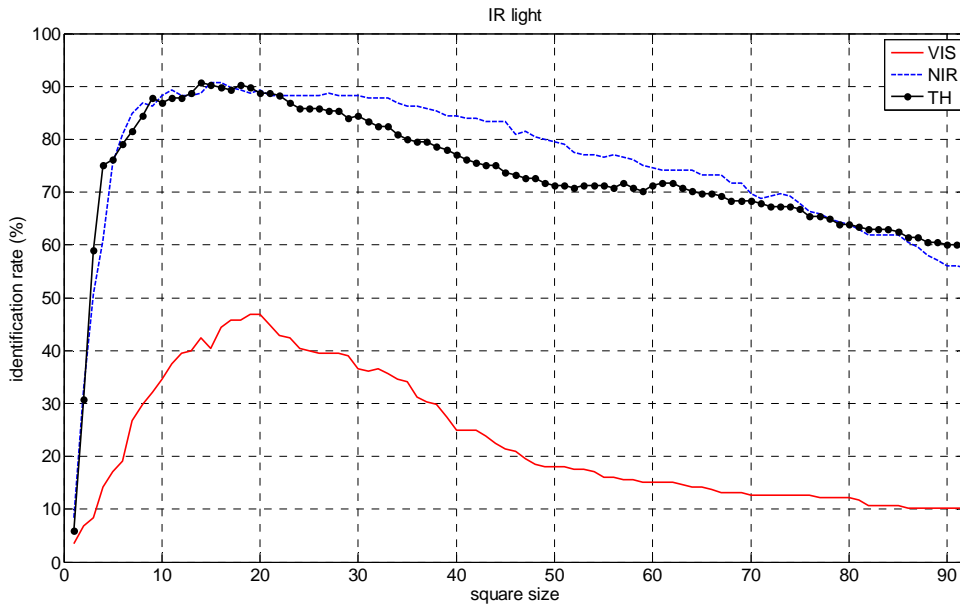


Figure 9.5: Identification rate as function of the square size of selected coefficients for VIS, NIR and TH sensors for near infrared (IR) illumination.

9.2.3 Experimental Results for a Specific Sensor

In this section we have compared the identification rates for a specific sensor regarding the different illuminations. We have studied the VIS sensor (Figure 9.6), the NIR (Figure 9.7) and the TH (Figure 9.8) for natural (NA), artificial (AR) and infrared (IR) illumination.

Figure 9.6 reveals that:

- VIS sensor performs better with artificial illumination. This makes sense because the variation along acquisition sessions is smaller than when using natural light, which varies from day to day.
- Optimal feature selection value is more critical when using natural light when compared to artificial light.
- VIS sensor fails when using NIR illumination. This is due to the acquisition conditions for this scenario, which is almost dark for a visible sensor.

Figure 9.7 reveals that:

- IR sensor performs similarly well with AR and IR illumination, and around 10% worse when evaluated with natural light. This can be due to the larger variability when analyzing faces with natural light.
- Feature selection is less critical when using IR illumination. This is reasonable considering that NIR sensors should perform optimally with IR illumination.

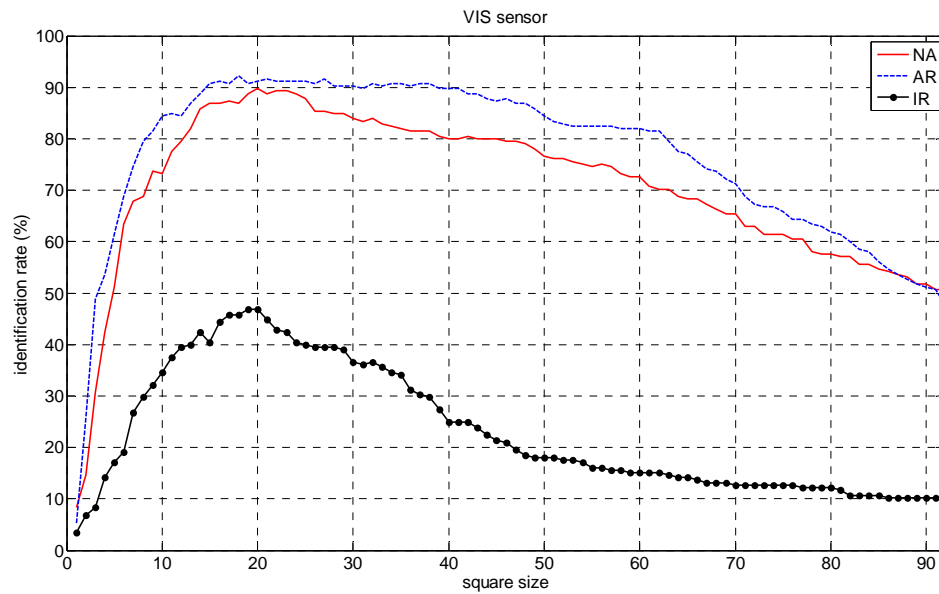


Figure 9.6: Identification rate as function of the square size of selected coefficients for Visible sensor and natural (NA), artificial (AR) and near infrared (NIR) illumination.

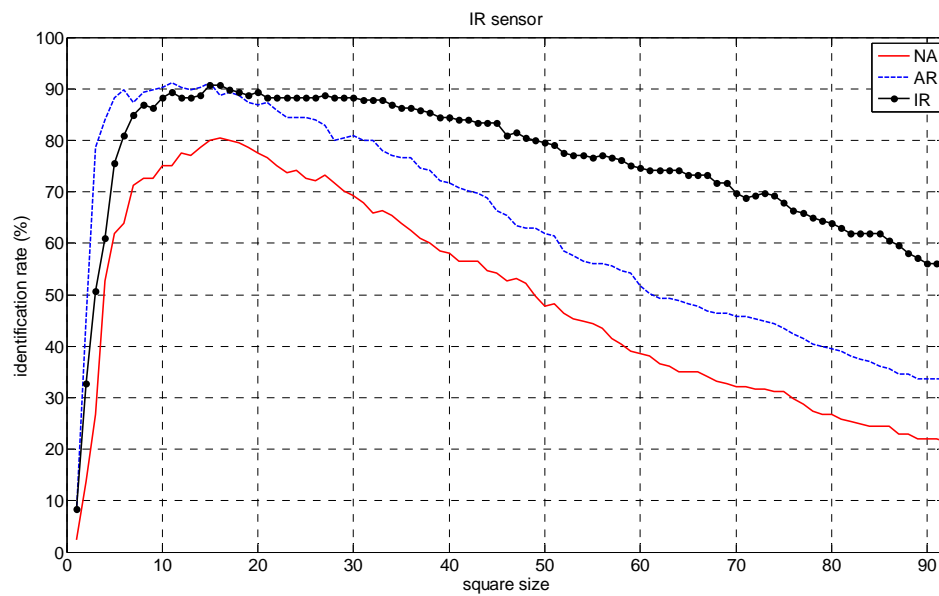


Figure 9.7: Identification rate as function of square size of selected coefficients for Infrared sensor and NA, AR and NIR illumination.

While Figure 9.8 shows an expected conclusion:

- TH sensor performs almost the same with all the studied illuminations. This is reasonable considering that thermal cameras do not measure the light reflection on the face, as previously discussed along Chapter 4. They measure the heat emission of the body. In fact, they could perfectly work in fully darkness because the illumination is irrelevant.

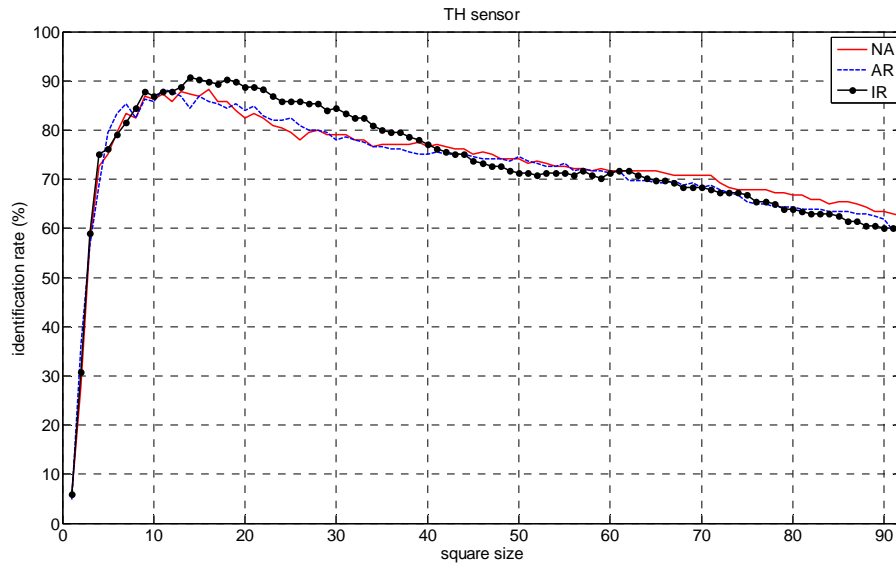


Figure 9.8: Identification rate as function of the square size of selected coefficients for Thermal sensor and NA, AR and NIR illumination.

It is important to point out that although there are small variations between the three studied illumination conditions, they are not due to the illumination. The motivation is the inherent variability of the acquired subject from day to day and acquisition to acquisition. If the subject would be an inanimate object with a fix temperature along the different acquisitions, the behaviour shown in Figure 9.5 would be the same under the three illuminations. However, a human being cannot fulfill this property.

Table 9.5 summarizes the optimal results and the optimal feature vector dimension (when evaluated from $1 \times 1, 2 \times 2, \dots, M \times M$) for different sensors and illumination conditions. This table reveals similar identification rates for all the sensors, although the thermal one requires a lower number of coefficients. In addition the visible sensor provides low identification rates when using IR illumination for the reasons previously commented.

Sensor	<i>Illumination</i>					
	NA		IR		AR	
	Identification	Coefficients	Identification	Coefficients	Identification	Coefficients
VIS	89,76	20x20	46,83	19x19	92,20	18x18
NIR	80,49	16x16	90,73	15x15	91,22	11x11
TH	88,29	16x16	90,73	14x14	88,29	11x11

Table 9.5: Optimal results for VIS, NIR and TH sensor under NA, IR and AR illumination conditions. Experimental conditions are the same of previous Figures 9.3 to 9.8.

9.2.4 Experimental Results in Mismatch Conditions

Using the setup of previous sections we have studied the identification rates in function of the different illumination conditions for training and testing. Table 9.6 shows the experimental results when using 20x20 coefficients for VIS, NIR and TH respectively. The models have been computed using session 1 and 2 and the testing has been done with session 3 and 4 separately.

Although it is possible to trade-off an optimal feature vector dimension for each scenario we decided to select a fix window size of 20x20 coefficients. According to previous plots (Figures 9.3 to 9.8) this tends to benefit the identification rates of the VIS sensor. Nevertheless the goal of this table is to study the mismatch illumination effect between training and testing conditions, rather than to find the highest identification rate for each scenario.

Due to the bad results obtained specially when using IR illumination we have decided to use some normalization procedure. The image has been normalized previous to apply the DCT2. The normalization procedure maps the values in intensity image to new values such that 1% of data is saturated at low and high intensities of the image. This increases the contrast of the normalized image. Thus, Table 9.6 includes experimental results with and without normalization.

On the other hand, we trained with sessions 1 and 2 and tested with sessions 3 and 4. Thus, the experimental results are affected by the time evolution. Nevertheless we have applied a feature selection defined in Section 8.2.3. Thus, stability along time is achieved by means of feature selection algorithm defined in Section 8.2.3, which is different for each spectral band. Comparing the experimental results of testing sessions 3 and 4 we can speculate that the stability of the different frequency bands over long periods of time is reasonably good, because there is a minor degradation when comparing session 4 and 3.

			<i>Test</i>					
			NA		IR		AR	
Sensor	Normalization	Train	3	4	3	4	3	4
VIS	NO	NA 1&2	89,8	84,4	51,2	60	85,9	82,9
	YES	NA 1&2	90,2	83,9	80,5	79	86,8	83,9
	NO	IR 1&2	85,9	82,4	46,8	61	92,2	85,9
	YES	IR 1&2	86,8	83,9	90,2	91,7	90,7	86,3
	NO	AR 1&2	90,2	82	57,6	60	91,2	87,3
	YES	AR 1&2	89,8	81,5	85,4	84,9	91,7	89,8
NIR	NO	NA 1&2	77,6	89,8	19	21	79	85,4
	YES	NA 1&2	82	93,7	44,4	38	85,4	87,8
	NO	IR 1&2	70,7	54,1	89,3	89,3	63	61,5
	YES	IR 1&2	74,1	57,1	94,1	95,1	62	54,6
	NO	AR 1&2	78,5	81	25	26,3	86,8	88,8
	YES	AR 1&2	85,4	86,3	49,3	49,3	89,8	88,3
TH	NO	NA 1&2	82,4	80,5	82	78	83,4	77,1
	YES	NA 1&2	82,4	80	82,9	79,5	84,4	78,5
	NO	IR 1&2	82,4	78	88,8	78	84,9	75,1
	YES	IR 1&2	82,4	76,6	88,3	78,5	84,9	75,1
	NO	AR 1&2	81,5	76,1	82,4	80	83,9	81
	NO	AR 1&2	82	78,5	82	79,5	84,9	81

Table 9.6: Identification rates (%) under different illuminations, sensors and normalization conditions for testing sessions 3 and 4, labeled on the table as 3 and 4. Best results are marked on bold face.

Table 9.6 reveals the following aspects:

- IR sensor provides the best result, which is 94,1% identification rate. This experimental result is in agreement with our previous paper [Esp11] conclusion, because NIR images have higher entropy than the other ones.
- Looking at the standard deviation (*std*) and mean value (*m*) of the experimental results of Table 9.6 for an specific sensor we obtain: $m=81,6$ and $std=12,4$ for visible sensor, $m=69,6$ and $std=22,8$ for NIR sensor and $m=80,9$ and $std=3,2$ for Thermal sensor. Thus, thermal image recognition rates are more stable than the other sensors.
- Image normalization is important for the case of illumination mismatch when using the visible and NIR sensor, and less important for the thermal one.

9.2.5 Experimental Results Using Multi-Sensor Score Fusion

In this section, a broad set of different fusion solutions discussed in Section 8.2.5, in function of the type of combination rule and the concerned weighting values, is given.

Table 9.7 shows the identification rates under different training and testing conditions for a fixed rule using the same weight for all the classifiers.

			<i>Test</i>					
			NA		IR		AR	
Sensors	Normalization	Train	3	4	3	4	3	4
VIS&NIR	NO	NA 1&2	91,7	96,1	51,2	58	91,7	92,7
	YES	NA 1&2	94,1	96,6	82	81,5	94,6	94,6
	NO	IR 1&2	93,7	90,7	91,7	92,7	91,7	91,7
	YES	IR 1&2	93,7	89,8	97,6	98	96,1	91,2
	NO	AR 1&2	92,7	91,7	49,3	52,2	95,1	94,6
	YES	AR 1&2	96,1	93,7	86,8	82,4	97,07	96,1
VIS&TH	NO	NA 1&2	91,7	87,3	85,9	82,9	95,1	88,8
	YES	NA 1&2	94,1	88,3	91,7	87,8	95,1	88,3
	NO	IR 1&2	96,1	89,3	90,2	84,4	96,6	90,2
	YES	IR 1&2	96,6	86,8	97,1	94,1	98	90,7
	NO	AR 1&2	95,6	89,3	83,4	82,9	98	93,7
	YES	AR 1&2	95,6	89,3	93,2	91,2	98	93,7
NIR&TH	NO	NA 1&2	91,7	96,1	85,9	74,1	95,1	96,6
	YES	NA 1&2	94,1	97,6	91,7	82	97,1	97,1
	NO	IR 1&2	96,6	89,3	96,1	96,6	92,2	87,8
	YES	IR 1&2	94,6	84,4	98,5	99	90,2	85,4
	NO	AR 1&2	93,7	94,6	77,1	67,8	97,1	93,7
	YES	AR 1&2	96,1	96,1	84,4	80,5	98,5	93,7
VIS&NIR&TH	NO	NA 1&2	94,6	97,1	82,9	76,6	97,1	96,1
	YES	NA 1&2	94,6	97,6	94,1	91,7	97,6	97,1
	NO	IR 1&2	98,5	95,1	98	97,6	98,5	97,1
	YES	IR 1&2	98	94,6	98,5	100	99,5	96,6
	NO	AR 1&2	97,1	95,1	80,5	74,6	99,5	97,1
	YES	AR 1&2	98,5	98,5	93,7	91,7	99,5	98

Table 9.7: Identification rate for the combination of two and three sensors under different illumination conditions (NA=Natural, IR=Infrared, AR=Artificial).

When combining two classifiers using a trained rule, a trial and error procedure must be done to set up the optimal value of the weighting factor. Figure 9.9 shows the identification rates as function of the weighting factor *alpha* (described as ω_i in the general case as seen in Equation 8.4), where the combination function is:

$$O = \alpha * d_{VIS} + (1 - \alpha) d_{NIR} \quad (9.1)$$

It is interesting to point out that for $\alpha=1$ the combination consists of the visible classifier distance alone d_{VIS} , whereas $\alpha=0$ fully removes the effect of the visible classifier, being the classification based on NIR sensor distance alone d_{NIR} . Thus, for $\alpha=0$ we obtain 89,8% identification rate and for $\alpha=1$, 84,4%. In the middle, there is an area that provides higher recognition rates (up to 95,6%) due to the combination of distances.

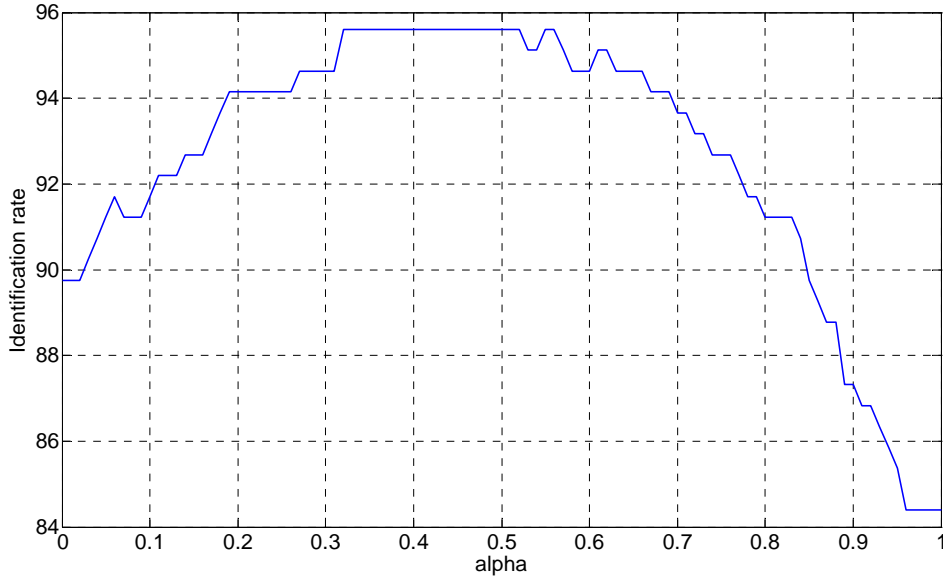


Figure 9.9: Trained rule combining VIS and NIR classifiers for NA illumination for training and testing, session 4.

When combining three classifiers the generalized procedure described in Equation (8.5) will be rewritten as:

$$d = \alpha * d_{VIS} + \beta * d_{NIR} + (1 - \alpha - \beta) d_{TH} \quad (9.2)$$

In this case in order to combine their outputs in the right way, we should trade-off two parameters and the graphical representation will become a three dimensional plot, such as the one shown in Figure 9.10. This three dimensional plot is not very informative due to the limitations of three dimensional representations and an alternative is to represent its contour plot. A contour plot is the level curves of the bidimensional matrix formed by giving values to the two parameters α and β . For the sake of simplicity only a few level curves are plot, as well as a black dot that indicates the highest value.

Some interesting remarks about this kind of plot are the following:

- In fact, the addition of the three weighting factors should be one. However, in order to avoid discontinuities and sudden gradients, we have filled up a whole matrix with $\alpha, \beta \in [0,1]$ using increments of 0,01. Thus, 100 values have been worked out for each variable.

- $\alpha=100$ implies $\beta=0$. Thus, the combined system consists of the visible sensor alone.
- $\beta=100$ implies $\alpha=0$. Thus, the combined system consists of the near infrared sensor alone.
- $\alpha=\beta=0$ implies that the combined system consists of the thermal sensor alone.
- $\alpha=\beta=33$ implies that the three system are equally weighted in the averaged distance computation.
- α and β adjustments on the diagonal line depicted in each of the Figures 9.11 and 12 imply that the thermal sensor is not used. The closest is the optimal point to this line, the lesser the weight of the thermal system. Adjustment points far from this diagonal imply a strong weight on the thermal system.

Observing the 18 plots of Figures 9.11 and 9.12 it can be observed that the three systems are almost equally important in the weighting process. There is only one exception, which is the second plot of Figure 9.10. In this case $\alpha=33$, $\beta=0$. Thus, NIR images are ignored and TH images are weighted two times more than visible ones. This is reasonable considering the identification rates of each sensor alone (see Table 9.6: VIS=60%, NIR=21%, TH=78%). Using these optimal combination values the identification rate reaches 84,9%.

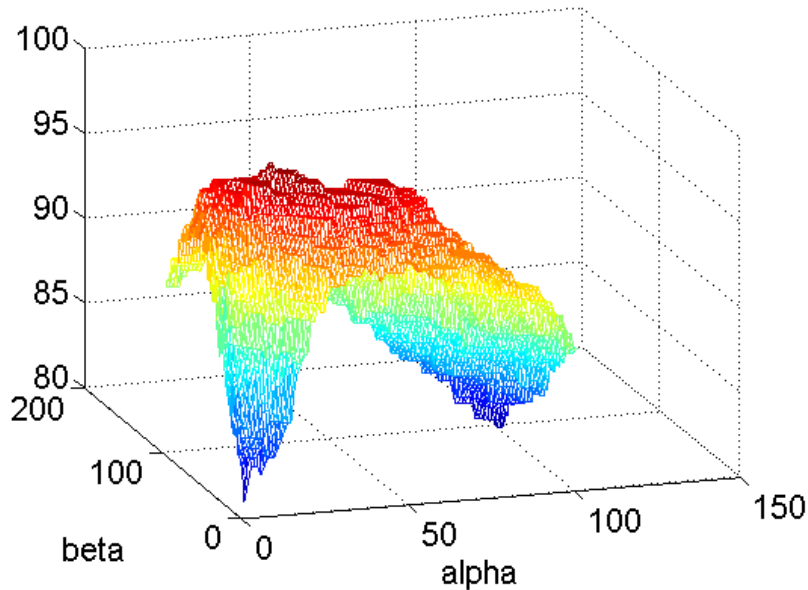


Figure 9.10: Example of trained rule identification rates combining three classifiers.

Figure 9.11 shows the contour plots as well as the maximum identification rate for the VIS, NIR and TH combination from top down and left to right for the following training and testing illumination conditions: NA-NA, NA-IR, NA-AR, IR-NA, IR-IR, IR-AR and AR-NA, AR-IR, AR-AR for session 4 and unnormalized feature vectors. Figure 9.12 represents the experimental results under the same illumination conditions for the normalized feature vectors case.

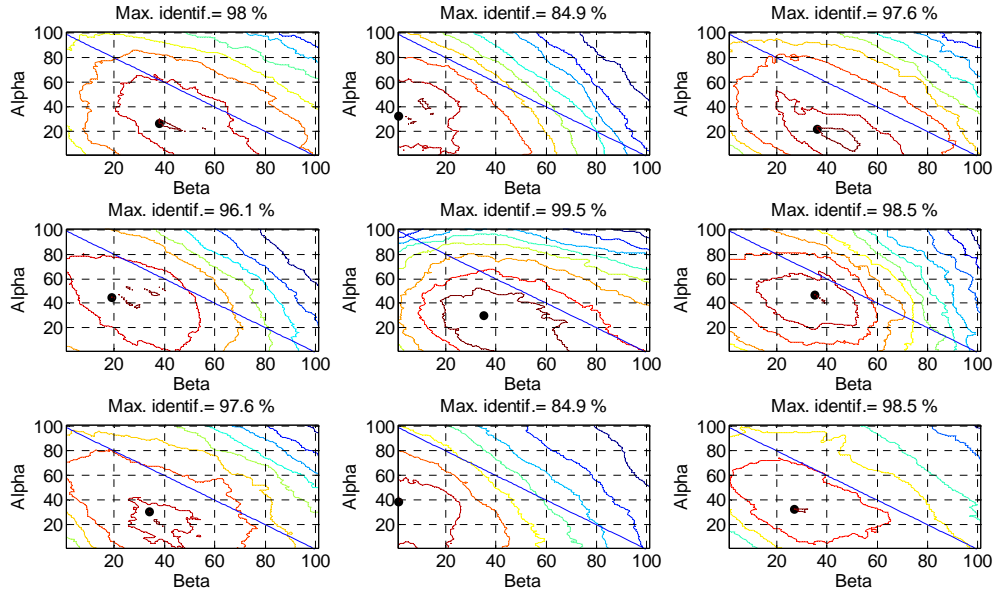


Figure 9.11: Contour plots when combining VIS, NIR and TH sensors under the following training and testing illumination conditions: NA-NA, NA-IR, NA-AR, IR-NA, IR-IR, IR-AR and AR-NA, AR-IR, AR-AR for session 4 and unnormalized feature vectors.

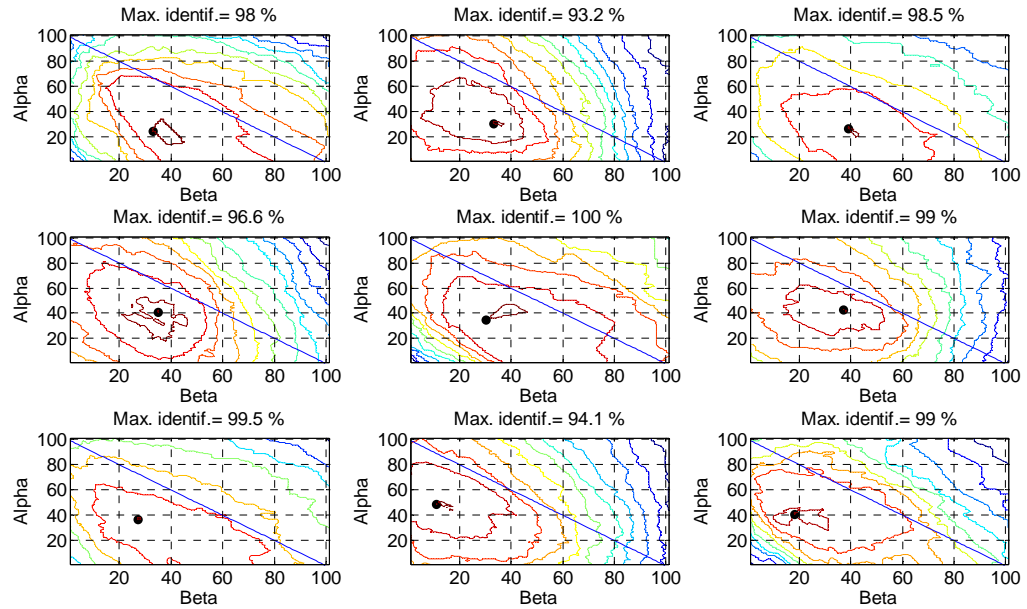


Figure 9.12: Contour plots when combining VIS, NIR and TH sensors under the following training and testing illumination conditions: NA-NA, NA-IR, NA-AR, IR-NA, IR-IR, IR-AR and AR-NA, AR-IR, AR-AR for session 4 and normalized feature vectors.

9.3 Conclusions

The encouraging results obtained strongly support the highlight of our hypothesis: The use of fused multispectral images results in a significant promise for dealing with the FR problem.

This last section of the chapter provides the main conclusions about first of all, face segmentation and afterwards, about face recognition, when using a single sensor and when fusing two or three sensors.

Related with the segmentation stage, the following facts worth to notice:

- Classical methods applied to visible images fail for segmenting thermal faces due to their peculiarity properties.
- VH projection algorithm produces accurate segmentations.
- As was already mentioned in Section 8.2.1, the advantage of VH projection is that we can leave out the training procedure.
- In case of segmentation based on Viola-Jones algorithm, a large amount of training data is needed to achieve good detection rate. In the reported experimental results, we have trained the Haar cascades by max. 4200 images, but for obtaining good results, 7000 images and more would be better. Thus, the disadvantage of this algorithm is also the detection time, which is dependent on the amount of training data. Segmentation using Haar cascade trained on many thousands images can be then more than 50 times slower than in case of VH projection.
- The disadvantage of VH projection is the fact that it can only detects one face on the image whereas Viola-Jones algorithm can detect more than one face. (Therefore VH projection is suitable for segmentation of databases containing a large number of single faces).

Related with the recognition stage when using a single sensor, the highlights are:

- The three studied sensors can provide good identification rates. This conclusion is particularly satisfying in that it suggests that these simple measurements capture perceptually important high-level information.
- The highest identification rate has been obtained for the NIR sensor under NIR illumination conditions.
- Visible sensor suffers seriously when presented with NIR lighting conditions, as expected.
- Thermal sensor is more stable along different illumination mismatch, as expected, and it also provide good enough identification rates, and requires smaller number of coefficients. In addition, optimal feature selection is less critical than for the other sensors.
- On average, visible sensor provides higher identification rates.

Related with the recognition stage when fusing two or three sensors, the set of highlights are:

- The combination noticeably improves the identification rates. The best system alone provides a 95,1% identification rate, and the combined system **reaches the 100%** in a particular scenario.
- In general the three sensors are almost equally important, because a quite balanced weighting factor is obtained by exhaustive trial and error of the whole set of weighting combinations.
- Normalized feature vectors always outperform the un-normalized system for the trained combination rule, and are slightly worse in 3 of 18 cases for the fixed combination rule.
- When studying the three sensors simultaneously we have not found any couple of redundant sensors. Fusion process has been shown to be very effective in taking advantage of available complementary information of the three spectral bands.
- The combined system is more robust with respect to illumination mismatch.

Chapter 10

Conclusions and Future Research

L'Utopie d'aujourd'hui est la vérité de demain.
-La Utopia d'avui és la veritat del demà-
Víctor Hugo

This last chapter summarizes our work and establishes a discussion of the main contributions advantages and limitations of our approach, suggesting some directions for future research. For concluding, a brief summary of open issues is finally provided.

10.1 Conclusions

Face recognition has demanded a lot of attention though-out the years, leading to a wide range of different currently available biometric solutions such as law enforcement support and security and surveillance applications. Modern FR technology has reached identifications rates greater than 90% for larger databases with well-controlled pose and illumination conditions. Nevertheless, FR is highly degraded when more demanding conditions are required. As has been widely reported along the document, illumination of the scene (-faces-) is one of the most longstanding problems related to recognize faces in an accurate way.

In this dissertation, a new FR approach has been introduced based on relatively novel sensors that gives a solution to the pervasive problem of FR when lighting conditions are less restrictive, achieving good results in all testing settings. The system has been tested over the face database specifically developed for our purpose, the multispectral (VIS-NIR-TIR) CARL DDBB, segmented with a new proposed face segmentation algorithm, resulting in a good baseline not only for exploiting the capabilities of the system in the well defined conditions but also, to strong support from information theory point of view, that valuable information for FR is contained in all explored frequency bands.

The rest of the designing decisions have been the following:

The DCT approach has been addressed for feature extraction, resulting in a good compromise between computational complexity reduction and accuracy. On the other hand, feature vector coefficients have been chosen by means of a developed feature selection algorithm that accurately looks for low intra-class variation and high inter-class variation. Afterwards, a distance calculation by using a fractional distance performs the template matching algorithm for classification. The advantage of having faces classified in these different three frequency bands have been finally exploited by means of data fusion at score matching level performed by both fixed and trained combination strategies.

In addition, the focusing problem in thermal images has been also addressed, firstly, for the more general case and then for the case of facial thermograms, from both the theoretical and the practical point of view. In this respect, and in order to analyze the quality of the thermal face images degraded by blurring, an appropriate algorithm has also been suggested.

Experimental results reveal that fusion outperform the performance, and has been also found that the three studied spectral bands contribute in a nearly equal proportion to a combined system. Recognition over 98% has been achieved in some conditions. These results represent a new step in providing a robust matching across changes in illumination, and also a new advance in the pursued consolidation of FR systems as competitive biometric modalities in practical scenarios. Figure 10.1 draws an outline of the overall performed system.

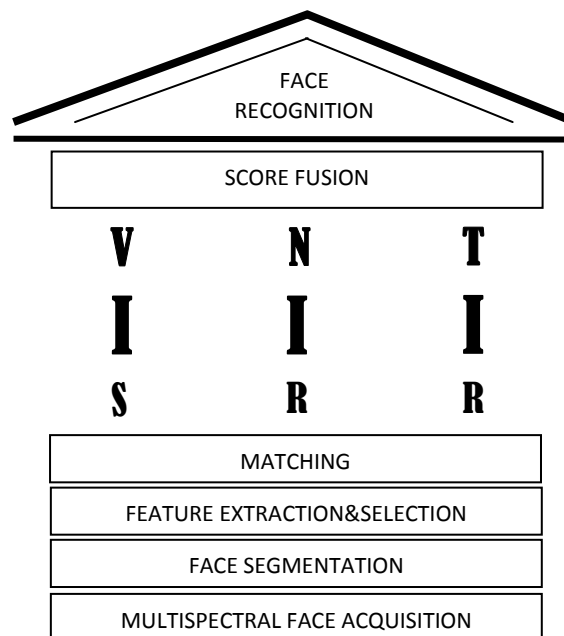


Figure 10.1: Overview of the Face Recognition performed system.

The key technical aspects and conclusions of the proposed solution are summarized below.

a) About the images properties:

- We have mathematically demonstrated that visible, near-IR and thermal images contain a small amount of redundancy. Actually, experimental results reveal that the mutual information is below 1,55 bits when comparing a pair of images of the same person acquired with different kind of sensor.
- Based on information theory measurements we conclude that the best spectral band combination always contains the thermal image, whereas the best combination for crossed-sensor recognition is VIS and NIR.
- The observation that thermal face images (or facial thermograms) contain complementary information to visible faces images opens a large amount of possibilities of data fusion for enhancing pattern recognition applications.
- NIR acquisition system minimizes influence of environmental lighting on face images. Nevertheless, near-IR images, (that belongs to a specific range of the reflected IR), are weakly affected by illumination due to their analogous properties to those of the visible light.
- Both NIR and thermal FR approaches, optimize the acquisition in camouflaged way, due to the involved light radiation is invisible in the NIR case and nonexistent for the case of thermal imaging.
- Thermal sensor is more stable along different illumination mismatch, as expected, and it also provides good enough identification rates, and requires smaller number of coefficients. In addition, optimal feature selection is less critical than for the other sensors.
- Thermal imaging features are not only invariant to the lighting conditions, but also essential for providing a robust matching across changes in illumination.
- If both visible and thermal facial images are simultaneously acquired and in reasonable working distances where the parallax effect discussed in Section 4.5.1 is negligible, it is possible to use the thermal information as a reliable mask in pixel-level fusion approaches for visible face detection, even when thermal image quality is not appropriate for recognition purposes.
- Thermal imaging FR approaches are also less sensitive to plastic surgery, disguises, and are also capable to distinguish between twin brothers. In addition, due to their large wavelengths TFR systems are less affected to atmospheric scattering, can see through fog, dust and smoke.

b) About the Focusing problem in thermal images:

- It has been analytically and experimentally demonstrated that a direct relation exists between the temperature of the objects to focus in the thermal spectrum and its blur level, in a similar way than the well-known chromatic aberration in the visible spectrum.

- The above relation can be considered negligible when dealing with facial thermograms (with temperatures near to 37°), being such features reliable and robust for biometric recognition purposes.
- c) About the overall system:
- Experimental results reported in Chapter 9 reveal that, when segmenting the visible and thermal face subset, the proposed face segmentation algorithm is more than 10 times faster, whereas the accuracy of face segmentation in thermal images is also higher than in case of competitive state of the art Viola and Jones method.
 - Results reported in Chapter 9 reveal that fusion of the three spectrums increases system performance, even when using a simple feature extractor and classifier. In six of the nine studied scenarios, identification rates higher or equal to 98% when using a trained combination rule have been obtained, being two cases of nine when using a fixed rule.
 - Multimodal systems are generally more complex and expensive than unimodal solutions, but it is not exactly true in the case of using thermal cameras, because of they incorporate two different sensors in one device, which make easier the final multimodal system implementation.

For closing this section, and as personal final conclusion, I can say that having worked in this research field, biometrics in general and thermal face recognition in particular, has been an interesting, stimulant and fruitful experience along these years.

10.2 Future Research

There are undoubtedly numerous technical, algorithmic and computational considerations that can arise to a greater extend the performance of the presented research work. This section briefly addresses some aspects that need to be faced, as follows:

- To extend the system for working under more uncontrolled illumination and thermal environments, where their conditions have a strong influence on skin surface temperature and then propose:
 - i. To exploit the adaptative combination strategy for score matching, in which, instant time becomes a factor that strongly contributes in determining the relevance of each classifier. This strategy will be useful, for instance, when system detects a low or side illumination scene, weighting more the thermal score, as well as when change in ambient temperature becomes relevant. Hence the weights will need to be adjusted to reflect more visible and NIR information in the fusion.
 - ii. To explore algorithms for eliminating abnormal areas, beyond the nose, and compute the discriminability power of the remaining more persistence parts.
- To address the problem of wearing glasses in thermal spectrum images, using standard algorithms, such as Hough Transform and/or other specifically designed to detect and replace eyeglasses.

- To compute the degree of thermal asymmetry of nose area due to its thermoregulator powerful properties, and depending on the relevance of the results, propose a parallel detection and removing system for extracting such part.
- To explore the powerful of updating each subject's model every time he is presented to the facial biometric system, especially in thermal images, because they are more subjected to variations in short time scales. Alternatively, a temperature normalization can also be studied.
- To achieve an easier-to-use acquisition system is also desirable.
- To extend the overall solution to less restricted outdoor environments.

Aside from the stated future tasks, we are currently contributing in the generation of a multifocusing thermal images system [Ben12] in order to overcome the inherent low DOF that limits future FR of multiple subjects in different focal planes, as has been deeply discussed in Chapter 5. This work has been accepted to *Pattern Recognition Letters* and has just passed main evaluation steps and it is just pending minor revisions. In Figure 10.2 a diagram of the proposed goal is showcased.

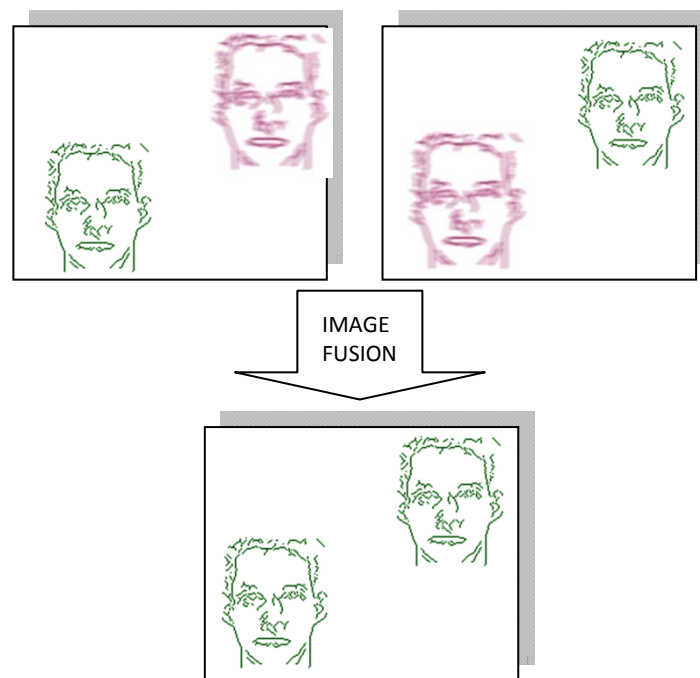


Figure 10.2: Multifocusing thermal images proposed goal. (color code: focused people in green and blur people in red).

In this sense, it would be also interesting to extent current VH projection as face segmentation method, due to the developed version is not a feasible method for selecting more than one face in the same picture. This limitation can be in the further work cancelled by an extension, which will firstly select the areas with potential faces and then will apply VH algorithm which

will extract these faces from the areas. This extension can be based on image binarization with consequent split of scene to areas according to the detected objects.

Finally, and for concluding this last section of this dissertation, we want to briefly highlight the more opened future FR directions. As is widely known, the emphasis in this research field is on providing robust solutions that do not make restrictive assumptions regarding the observer or its environment. In this sense, the ultimate goal in such field is usually viewed as the ability to simultaneously recognize different faces of freely moving people acquired from less restricted camera viewpoints, in real time, with high performance and in environments where the lighting conditions are fully uncontrolled. Figure 10.3 exhibits a sample that collects many of such assumptions. Understanding the recognition of human faces under this unrestricted point of view remains an open challenge.



Figure 10.3: A very poor quality frame from a camouflaged surveillance video, registering two people of different races (the Author and an eastern woman) in the SF Chinatown neighborhood.

It is clear that an enormous amount of work is still required before facial thermography can be established as a key facial recognition technology. Nonetheless, future advancements in thermal imaging sensors technology, fueled by even more fine machine learning and fusion algorithms and the increasing availability processing power, can led to a breakthrough in the context of FR, extending the pointed ultimate goal, even, to full darkness environments. Thus, all this technology at our disposal will allow us not only detecting the presence of a person in anywhere at any time of day, but also identify who is. Hopefully for good.

References

Either write something worth reading or do something worth writing.
Benjamin Franklin

- [Abi04] B. ABIDI, B. HUQ and M. ABIDI, *Fusion of Visual, Thermal, and Range as a Solution to Illumination and Pose Restrictions in Face Recognition*. 38st IEEE-ICCST International Carnahan Conference on Security Technology. pp 325-330. Albuquerque. USA. October 2004. ISBN 0-7803-8506-3.
- [Akh08a] M. A. AKHLOUFI and A. BENDADA, *Thermal Faceprint: A new thermal face signature extraction for infrared face recognition*. Proceedings of the 5th Canadian Conference on Computer and Robot Vision (CRV 2008). Windsor, Canada. May 2008.
- [Akh08b] M. A. AKHLOUFI and A. BENDADA, *Infrared Face Recognition Using Distance Transforms*. Proceedings of the 5th International Conference on Image and Vision Computing (ICIVC08), Vol. 30, pp 160-163. Paris, France. July, 2008.
- [Akh08c] M. A. AKHLOUFI, A. BENDADA and J.C. BATSALE, *State of the Art in Infrared Face Recognition*. Quantitative Infrared Thermography (QIRT) Journal. Vol. 5, Issue 9, pp. 3-26, 2008.
- [Akh10] M. A. AKHLOUFI and A. BENDADA, *Infrared Face Recognition Using Texture Descriptors*. Proceedings of SPIE 7661, Thermosense XXXII, 766109. May 2010. DOI:10.1117/12.849764.
- [Ara06] O. ARANDJELOVIC, R. HAMMOUD and R. CIPOLLA, *Multi-Sensory Face Biometric Fusion (for Personal Identification)*. Computer Vision and Pattern Recognition Workshop CVPRW'06. Pp 128 June 2006. ISBN: 0-7695-2646-2.
- [Ara10] O. ARANDJELOVIC, R. HAMMOUD and R. CIPOLLA, *Thermal and Reflectance Based Personal Identification Methodology Under Variable Illumination*. Pattern Recognition Issue 43, pp 1801–1813. 2010.
- [Ash00] J. ASHBOURN, *Biometrics. Advanced Identity Verification. The Complete Guide*. Ed. Springer. 2000.
- [Asp68] J. R. J. Van ASPEREN de BOER, *Infrared Reflectography: A Method for the Examination of Paintings*. Applied Optics. Vol. 7, Issue 9, pp 1711-1714. 1968.
- [Auc00] K. AUCOIN, *Face Forward*, Ed. Little Brown and Company. October 2000.

-
- [Bac02] F. R. BACH and M. I. JORDAN, *Kernel Independent Component Analysis*. Journal of Machine Learning Research. Vol. 3, pp 1-48. 2002.
- [Bar86] N. BARLEY, *The Innocent Anthropologist*, Waveland Press, Inc., Prospect Heights, Illinois, 1983.
- [Bar00] W. R. BARRON, *Principles of Infrared Thermometry*. Helmers Publishing, INC. 174 Concord St. Peterborough, NH 03458. 2000.
- [Bar98] M. S. BARTLETT and T. J. SEJNOWSKI, *Independent Component Representations for Face Recognition*. Proceedings of the Third IEEE International Conference on Human Vision and Electronic Imaging III. Vol. 3299, pp 528-539. 1998.
- [Bar02] M. S. BARTLETT J. R. MOVELLAN and T. J. SEJNOWSKI, *Face Recognition by Independent Component Analysis*. IEEE Transactions on Neural Networks. Vol. 13, Issue 6 pp 1450-1464. 2002.
- [Bau89] E. B. BAUM and D. HAUSSLER, *What Size Net Gives Valid Generalization?* Neural Computation, Vol. 1, Issue 1, pp 151-160. 1989.
- [Bau00] G. BAUDAT and F. ANOUAR, *Generalized Discriminant Analysis Using a Kernel Approach*. Neural Computation Num. 12 pp 2385-2404. 2000.
- [Bau02] G. BAUDAT and F. ANOUAR, *Learning with Kernels*. MIT Press. 2002.
- [Bea01] C. BEAVAN, *Fingerprints: The Origins of Crime Detection and the Murder Case That Launched Forensic Science*. Ed. Hyperion; 1st edition. 2001. ISBN-10: 0786866071.
- [Beh05] M. BEHRMANN, G. AVIDAN, J.J. MAROTTA and R. KIMCHI, *Detailed Exploration of Face Related Processing in Congenital Prosopagnosia: Behavioral Findings*. Journal of Cognitive Neuroscience, 17, pp 1130-1149. Massachusetts Institute of Technology. 2005.
- [Bel96] P. N. BELHUMEUR and D. J. KRIEGMAN, *What is the Set of Images of an Object under all Possible Lighting Conditions*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1996.
- [Bel97] P. N. BELHUMEUR, J. P. HESPANHA, and D. J. KRIEGMAN, *Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection*. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 19, Issue 7, pp 711-720. 1997.
- [Ben12] R. BENES, P. DVORAK, M. FAÚNDEZ, V. ESPINOSA-DURÓ and J. MEKYSKA, *Multi-Focus Thermal Image Fusion*. (Accepted in Pattern Recognition Letters).
- [Ber07] M. J. BERNSTEIN, S.G. YOUNG and K. HUGENBERG, *The Cross-Category Effect: Mere Social Categorization is Sufficient to Elicit an Own-Group bias in Face Recognition*. Psychological Science. Vol. 18, Issue 8, pp 706-12. August 2007.
- [Ber10] A. BERGMAN and A. CASADEBALL, *Mammalian Endothermy Optimally Restricts Fungi and Metabolic Costs*. mBio. Vol. 1, Issue 5. November 2010. DOI: 10.1128/mBio.00212-10.
- [Bho10] M. K. BHOWMIK, D. BHATTACHARJEE, M. NASIPUPI, D. K. BASU and M. KUNDU, *Fusion of Wavelet Coefficients from Visual and Thermal Face Images for Human Face Recognition: A Comparative Study*. International Journal of Image Processing (IJIP). Vol. 4, Issue 1. 2010.
- [Bis95] C. M. BISHOP, *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford. 1995.
- [Ble66] W. W. BLEDSOE, *The Model Method in Facial Recognition*. Technical Report PRI 15, Panoramic Research, Inc. Palo Alto, California. August 1966.

-
- [Bou92] T. E. BOULT and G. WOLBERG, *Correcting Chromatic Aberration using Image Warping*. Proceedings Image Understanding workshop. Pp 363-377. San Diego, California, 1992.
 - [Bow06] K. BOWYER, *Face Recognition Using 2-D, 3-D, and Infrared: Is Multimodal Better than Multisample?* Proceedings of the IEEE. Vol. 94, Issue 11, pp 2000-2012, November 2006.
 - [Box78] G. E. P. BOX, W.G. HUNTER and J. S. HUNTER, *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*. 1st Edition. Ed. John Wiley & Sons. 1978.
 - [Bre96] L. BREIMAN, *Bagging Predictors*. Machine Learning. Vol. 24, Issue 2, pp 123-140. 1996.
 - [Bre03] M. BRESSAN, D. GUILLAMET and J. VITRIA, *Using an ICA representation of Local Color Histograms for Object Recognition*. Pattern Recognition. Vol. 36, Issue 3, pp 691-701. 2003.
 - [Bru93] R. BRUNELLI and T. POGGIO, *Face Recognition: Features versus Templates*. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 15, Issue 10 pp 1042-1052. October 1993.
 - [Bud06] P. BUDDHARAJU, I. T. PAVLIDIS and P. TSIAMYRTZIS, *Pose-Invariant Physiological Face Recognition in the Thermal Infrared Spectrum*. IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06). pp 53-60, NY, June 2006. ISBN:0-7695-2646-2.
 - [Bud07] P. BUDDHARAJU, I. T. PAVLIDIS, P. TSIAMYRTZIS and M. BAZAKOS, *Physiology-Based Face Recognition in the Thermal Infrared Spectrum*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 29, Issue 4, pp 613-626, April 2007.
 - [Bul03] M. BULMER, *Francis Galton. Pioneer of Heredity and Biometry*. The Johns Hopkins University Press. November 2003.
 - [Bur98] B. BURKE, *The World According to Wavelets. The Story of a Mathematical Technique in the Making*. A. K. Peters, 2nd edition, 1998.
 - [Buy10] P. BUYSENS and M. REVENU, *Fusion Levels of Visible and Infrared Modalities for Face Recognition*. Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), pp 1-6. 2010.
 - [Cai07] D. CAI, X. HE, K. ZHOU, J. HAN and H. BAO, *Locality Sensitive Discriminant Analysis*. Proceedings of the 20th International Joint Conference on Artificial Intelligence, pp 708-713. India, January 2007.
 - [Cal1933] J. CALICO, *Els Mètodes d'Identificació Personal. Monografies Mèdiques*. Num. 74. 1933.
 - [Car02] F. CARDINAUX and S. MARCEL, *Face Verification Using MLP and SVM: A Comparison*. COST 275- The Advent of Biometrics on the Internet, 2002.
 - [Cha05] N.V. CHAWLA and K.W. BOWYER, *Ensembles in Face Recognition: Tackling the extremes of high dimensionality, temporality, and variance in data*. Conference on Systems, Man and Cybernetics, 2005 IEEE International Vol. 3 pp 2346-2351. 10-12 Oct. 2005. ISBN: 0-7803-9298-1.
 - [Cha08] H. CHANG, A. KOSCHAN, M. ABIDI S. G. KONG and C. H. W0N. *Multispectral Visible and Infrared Imaging for Face Recognition*. IEEE Conference Proceedings. July 2008. DOI:10.1109/CVPRW.2008.4563054. ISBN: 978-1-4244-2339.
 - [Cha09] H. CHANG, Y. YAO, A. KOSCHAN, B. ABIDI and M. ABIDI, *Improving Face Recognition via Narrowband Spectral Range selection Using Jeffrey Divergence*. 2009. IEEE Transactions on Information Forensics and Security. Vol. 4, Issue 1. March 2009.

-
- [Che00] L. F. CHEN, H. Y. MARK LIAO, M. T. KO, J. C. LIN and G. J. YU, *A new LDA-based Face Recognition System Which Can Solve the Small Sample Size Problem*. Pattern Recognition. Vol. 33, pp 1713–1726. 2000.
- [Che03] X. CHEN, P. FLYNN and K. BOWYER, *PCA-Based Face Recognition in Infrared Imagery: Baseline and Comparative Studies*. Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03), pp. 127, Nice, France, October, 2003.
- [Che05] X. CHEN, P. FLYNN and K. BOWYER, *Fusion of Infrared and Range Data: Multimodal Face Images*. Lecture Notes in Computer Science-LNCS 3832. Ed. Springer pp 55–63. 2005.
- [Che06] R. CHELLAPPA, P. JONATHON-PHILLIPS and D. REYNOLDS, *Special Issue on Biometrics: Algorithms and Applications*. Proceedings of the IEEE. Vol. 94, Issue 11, pp 1912-1914 November 2006.
- [Chr00] K. CHRZANOWSKI, *Testing Thermal Imagers. Practical Guidebook*. Military University of Thecnology. Warsaw, Poland. 2010.
- [Cla94] R. CLARKE, *Human Identification in Information Systems: Management Challenges and Public Information Issues*. Information Technology and People. Vol. 7, Issue 4, pp 6-37. December 1994.
- [Cop99] J. COPELAND and D. PROUDFOOT, *Alan Turing's Forgotten Ideas in Computer Science*. *Scientific American*. W. Heffer and Sons LTD., Cambrigde. pp 99-103. April, 1999.
- [Cor95] C. CORTES and V. N. VAPNIK, *Support-Vector Networks*. Machine Learning. Vol. 20, Issue 3, pp 273-297. September 1995.
- [Cov91] T. M. COVER and J. A. THOMAS, *Elements of Information Theory*. Wiley, New York, 1991.
- [Cre07] F. CRETE, T. DOLMIERE, P. LADRET, M. NICOLAS, *The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric*. Proceedings of the SPIE Electronic Imaging Symposium Conf. Human Vision and Electronic Imaging, Vol. 6492. San Jose, USA, February 2007. DOI:10.1117/12.702790.
- [Cri00] N. CRISTINIANI and J. SHAW-TAYLOR. *An Introduction to Support Vector Machines and other Kernel-based Learning Methods*. Cambrigde University Press, 2000.
- [Cut97] I. C. CUTHILL et al., *Ultraviolet Vision in Birds*. in Peter J.B. Slater. *Advances in the Study of Behavior*. 29. Oxford, England: Academic Press. pp 161. 1997. ISBN 978-0-12-004529-7.
- [Dau93] J. G. DAUGMAN, *High Confidence Visual Recognition of Persons by a Test of Statistical Independence*, IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol 15, Issue 11, pp 1148–1161, November 1993.
- [Dau94] G. DAUGMAN, *Biometric Personal Identification System Based on Iris Analysis*. Patent number: 5291560 ; Filing date: Jul 15, 1991 ; Issue date: Mar 1, 1994.
- [Dau01] J. G. DAUGMAN, *High Confidence Recognition of Persons by Iris Patterns*. 35th IEEE-ICCST International Carnahan Conference on Security Technology. London. October, 2001. ISBN 7803-6636-0/01.
- [Das94] DASARATHY, *Decision Fusion*. IEEE Computer Society Press. 1994.
- [Der51] M. DÉRIBÉRE, *De l'Ultraviolet a l'Infrarouge*. Paris: Les Éditions Textile et Technique. Series L'Édition textile moderne. 1951.

-
- [Der54] M. DÉRIBÉRE, *Les Applications Pratiques des Rayons Infrarouges*. 3th Edition. Ed. Paris Dunod. 1954.
 - [Dud01] R. O. DUDA, P. E. HART and D. G. STORK. *Pattern Classification*. John Wiley and Sons, Inc., 2nd edition. 2001.
 - [Eis85] R. EISBERG and R. RESNICK, *Quantum Physics of Atoms, Molecules, Solids, Nuclei and Particles*. Ed. Wiley. 2nd Revised edition. 1985.
 - [Ekm78] P. EKMAN and W. FRIESEN, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA, 1978.
 - [Equi09] EQUINOX, *Multispectral Face DDBB*. <http://www.equinoxsensors.com/products/HID.html>.
 - [Esp00] V. ESPINOSA-DURÓ, *Biometric Identification using a Radial Basis Network*. 34th International Carnahan Conference on Security Technology. IEEE-ICCST. pp 47-51. ISBN 0-7803-5965-8. Ottawa. Canada. October 2000.
 - [Esp02] V. ESPINOSA-DURÓ, *Minutiae Detection Algorithm for Fingerprint Recognition*. IEEE AES-Aerospace and Electronics Systems Magazine. Vol. 17, Issue 3, pp 7-10. March 2002. ISSN 0885-8985.
 - [Esp04a] V. ESPINOSA-DURÓ, *Face Recognition using VIS and Near-IR Images: A Comparison*. 8st SCI. MultiConference on Systemics, Cybernetics and Informatics. pp 294-297. Orlando.USA. July 2004. ISBN 980-6560-13-2.
 - [Esp04b] V. ESPINOSA-DURÓ, M. FAÚNDEZ and J. A. ORTEGA, *Face Detection from a Video Camera Sequence*. 38st IEEE-ICCST International Carnahan Conference on Security Technology. pp 318-320. Albuquerque. USA. October 2004. ISBN 0-7803-8506-3.
 - [Esp06] V. ESPINOSA-DURÓ, *Detecció de l'Efecte Corona Mitjançant Sensors de Radiació UV*. XII Jornades de Conferències d'Enginyeria Electrònica (JCEE). Invited speech. Departament d'Enginyeria Electrònica de la UPC, December 2006.
 - [Esp08] V. ESPINOSA-DURÓ, E. MONTE-MORENO, *Face Recognition Approach Based on Wavelet Transform*. 42st International Carnahan Conference on Security Technology. IEEE-ICCST pp 187-190. Prague. Czech Republic. October 2008. ISBN 978-1-4244-1816-9.
 - [Esp10] V. ESPINOSA-DURÓ, M. FAÚNDEZ, J. MEKYSKA and E. MONTE-MORENO, *A Criterion for Analysis of Different Sensor Combinations with an Application to Face Biometrics*. Cognitive Computation. Vol. 2, Issue 3, pp 135-141. September 2010.
 - [Esp11] V. ESPINOSA-DURÓ, M. FAÚNDEZ and J. MEKYSKA, *Beyond Cognitive Signals*. Cognitive Computation. Ed. Springer. Vol. 3 pp 374-381. June 2011.
 - [Esp12] V. ESPINOSA-DURÓ, M. FAÚNDEZ and J. MEKYSKA, *A New Face Database Simultaneously Acquired in Visible, Near Infrared and Thermal Spectrums*. Cognitive Computation. Ed. Springer. July 2012. DOI: 10.1007/s12559-012-9163-2.
 - [Ete97] K. ETEMAD and R. CHELLAPPA, *Discriminant Analysis for Recognition of Human Face Images*. Journal of the Optical Society of America. Vol. 14, Issue 8, pp 1724-1733 August 1997.
 - [Fab08] J. FABREGRAS and M. FAÚNDEZ, *Biometric Face Recognition with Different Training and Testing Databases*. International COST 2102 Conference on Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction. LNAI 5042 pp 46-58. Patras, October 2008.

-
- [Far12] S. FAROKHI, S. M. SHAMUDDIN, J. FLUSSER and U. U. SHEIKH, *Assessment of Time-lapse in Visible and Thermal Face Recognition*. International Journal of Computer and Communication Engineering. Pp 181-186. 2012.
- [Fau05a] M. FAÚNDEZ, V. ESPINOSA-DURÓ and J. A. ORTEGA, *A Low-Cost Webcam&Personal Computer Open Door*. IEEE AES Aerospace and Electronics Systems Magazine. Vol. 20, Issue 11, pp 23-26. November 2005. ISSN: 0885-8985.
- [Fau05b] M. FAÚNDEZ, *Data Fusion in Biometrics*. IEEE Aerospace and Electronic Systems Magazine. Vol. 20, Issue 1, pp 34-38. January 2005.
- [Fau07a] M. FAÚNDEZ, J. ROURE, V. ESPINOSA-DURÓ and J.A ORTEGA, *An Efficient Face Recognition Method in a Transformed Domain*. Pattern Recognition Letters. Vol. 28, Issue 7, pp 854-858. May 2007. ISSN: 0167-8655.
- [Fau07b] M. FAÚNDEZ, V. ESPINOSA-DURÓ, J. A. ORTEGA, *Low Complexity Algorithms for Biometric Recognition*. Chapter in Verbal and Nonverbal Communication Behaviours. Lecture Notes in Computer Science-LNCS 4775. Ed. Springer pp 275-285. 2007. ISBN-13 978-3-540-76441-0.
- [Fau11] M. FAÚNDEZ, J. MEKYSKA and V. ESPINOSA-DURÓ, *On the Focusing of Thermal Images*. Pattern Recognition Letters. Ed. Elsevier. Vol. 32, pp 1548-1557. August 2011.
- [Fen00] G. C. FENG, P. C. YUEN and D. Q. DAI, *Human Face Recognition using PCA on Wavelet Subband*. SPIE Journal of Electronic Imaging. Vol. 9, Issue 2, pp 226-233. April 2000. DOI:10.1117/1.482742.
- [Fis36] R.A. Fisher, *The Use of Multiple Measures in Taxonomic Problems*, Ann. Eugenics. Vol. 7, pp 179-188. 1936.
- [Fli08] FLIR Corporation, *The Ultimate Infrared Handbook for R&D Professionals*. Technical Report, Flir Systems. 2008.
- [Flu09] FLUKE, *Introduction to Thermography Principles*. Technical Report. American Technical Publishers, Inc. 2009.
- [Fon90] J. FONTCUBERTA, *Fotografía: Conceptos y Procedimientos*. Ed. Gustavo Gili. 1990.
- [For03] D.A. FORSYTH and J. PONCE, *Computer Vision, A Modern Approach*. Ed. Prentice Hall. 2003. ISBN 0-13-191193-7.
- [Fra07] D. FRANOIS, V. WERT and VERLEYSEN, *The Concentration of Fractional Distances*, IEEE Transactions on Knowledge and Data Engineering. Vol. 19, Issue 7, pp 873-886. July 2007.
- [Fri74] J. H. FRIEDMAN and J. W. TUKEY, *A Projection Pursuit Algorithm for Exploratory Data Analysis*. IEEE Transactions on Computers C-23 Vol 9 pp 881-890. September 1974. DOI: 10.1109/T-C.1974.224051. ISSN 0018-9340.
- [Fri87] J. H. FRIEDMAN, *Exploratory Projection Pursuit*. Journal of the American Statistical Association, Vol. 82, Issue 392, pp 249-266, 1987.
- [Fuk90] K. FUKUNAGA, *Statistical Pattern Recognition*. 2nd ed. New York: Academic Press; 1990.
- [Gam06] M. GAMADIA and N. KEHTARNAVAZ, *A Real-time Continuous Automatic Focus Algorithm for Digital Cameras*. Proc. IEEE Southwest Symposium on Image Analysis and Interpretation, pp 163-167. 2006.

-
- [Geo01] A. S. GEORGHIADIS, P. N. BELHUMEUR and D. J. KRIEGMAN, *From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 23, Issue 6, pp 643-660. June 2001.
 - [Gom01] R. B. GOMEZ, A. JAZAERI and M. KAFATOS, *Wavelet-Based Hyperspectral and Multispectral Image Fusion*. Proceedings of SPIE, Vol. 4383, pp 36-42, Geo-Spatial Image and Data Exploitation II, William E. Roper; Ed. 2001.
 - [Gon10] J. F. GONZALEZ, *Aplicaciones de los Sistemas de Visión Nocturna en la Navegación Marítima y la Seguridad en la Mar*. PhD Thesis. Ed. UPC. 2010.
 - [Gon87] R. C. GONZALEZ and P. WINTZ, *Digital Image Processing*. 2nd Edition. Ed. Addison-Wesley. 1987.
 - [Gra79] *International Edition of the Encyclopedia of Practical Photography*. Eastman Kodak Company Inc. American Photographic Book Publishing Company Inc. Editions Grammont, S.A. 1979. ISBN 84-345-3949-7.
 - [Gre07] M. W. GRENN, P. PERCONTI, J. VIZGALIS and J. G. PELLEGRINO, *Infrared Camera and Optics for Medical Applications*. CRC Press. 2007. DOI: 10.1201/9781420008340.ch5
 - [Gro04] R. GROSS, *Face Databases; Handbook of Face Recognition*, S.Li, A.Jain, Ed., Springer, 2004. ISBN 0-387-40595.
 - [Gut95] S. GUTTA, J. HUANG, D. SINGH and I. SHAH, *Benchmark Studies on Face Recognition*. Proceedings of the International Workshop on Automatic Face- and Gesture- Recognition, IWAAGR 95, pp 227-231. Zurich, 1995.
 - [Hal99] P. HALLIGAN, G. GORDON, A.L. YUILLE, P. GIBLIN and D. MUMFORD, *Two- and Three- Dimensional Patterns of the Face*. A.K. Peters, Natick, Massachusetts. 1999. ISBN 1-56881-087-3.
 - [Has01] T. HASTIE, R. TIBSHIRANI and J. FRIEDMAN, *The Elements of Statistical Learning*. Springer series in Statistics. Springer, New York, 2001.
 - [Haz03] T. J. HAZEN, E. WEINSTEIN, R. KABIR, A. PARK and B. HEISELE, *Multi-Modal Face and Speaker Identification on a Handheld Device*. Proc. Works. Multimodal user Authentication. Pp 113-120 Santa Barbara, CA 2003.
 - [Her1884] W. J. HERSCHEL, *Fingerprints*. Nature. Vol. 22, pp 77-78. November 1884.
 - [Heo05] J. HEO, M. SAVVIDES and B. V. K. VIJAYAKUMAR, *Performance Evaluation of Face Recognition using Visual and Thermal Imagery with Advanced Correlation Filters*. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005.
 - [Hiz 09] W. HIZEM, L. ALLANO, A. MELLAKH, and B. DORIZZI, *Face Recognition from Synchronized Visible and Near-Infrared Images*. IET Signal Processing Special issue on Biometrics. Vol. 3, Issue 4. Pp 282-288. July 2009.
 - [Hon98] L. HONG, *Automatic Personal Identification Using Fingerprints*. PhD dissertation. Michigan State University. Department of Computer Science. June 1998.
 - [Hua07] W. HUANG and Z. JING, *Evaluation of Focus Measures in multi-Focus Image Fusion*. Pattern Recognition Letters. Vol. 28, pp 493-500. 2007.

-
- [Imt11] H. IMTIAZ and S.A. FATTAH, *A Wavelet-Domain Local Dominant Feature Selection Scheme for Face Recognition*. ISRN Machine Vision. Vol. 2012. 2011. DOI:10.5402/2012/976160.
- [Inc96] F.P. INCROPERA, *Fundamentals of Heat and Mass Transfer*. 4th Edition. John Wiley & Sons, Inc, New York. 1996.
- [Jac00] R. JACOBSON, S. RAY, G. ATTRIDGE and N. AXFORD, *Manual of Photography*, Ninth Edition. Focal press. 2000.
- [Jaf09] R. JAFRI and H. R. ARANNIA, *A Survey of Face Recognition Techniques*. Journal of Information Processing Systems. Vol. 5, Issue 2, pp 41-68. June 2009. DOI: 10.3745/JIPS.2009.5.2.041.
- [Jai89] A. K. JAIN, *Fundamentals of Digital Image Processing*. Prentice Hall. 1989.
- [Jai97] A. JAIN, L. HONG and S. PANKANTI, *An Identity-Authentication System Using Fingerprints*. Proceedings of the IEEE. Vol. 85, Issue 9, pp 1365–1388. September 1997.
- [Jai99] A. JAIN, R. BOLLE and S. PANKANTY, *Biometrics: Personal Identification in Networked Society*. Kluwer Academic Publishers. 1999.
- [Jai00] A. K. JAIN, P. W. DUIN, R., J. MAO, *Statistical Pattern Recognition: A Review*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, Issue 1, pp 4-37. January 2000.
- [Jai01] A. JAIN, S. PRABHAKAR and S. PANKANTI, *Twin Test: On Discriminability of Fingerprints*. Audio and Video-Based Biometric Person Authentication. Lecture Notes in Computer Science, Vol. 2091/2001, pp 211-217. 2001.
- [Jai04] A. JAIN, A. ROSS and S. PRABHAKAR, *An Introduction to Biometric Recognition*. IEEE Transaction on Circuits and Systems for Video Technology, Special Issue on Image-and Video-Based Biometrics. Vol. 14, Issue 1, pp 4-20. January 2004.
- [Jai11] A. K. JAIN, A. A. ROSS and K. NANDAKUMAR, *Introduction to Biometrics*. Ed. Springer Science+Business Media. 2011. ISBN 978-0-387-77325-4.
- [Jam07] B. G. M. JAMIESON, *Reproductive Biology and Phylogeny of Birds*. Charlottesville VA: University of Virginia. p. 128. 2007. ISBN 1578083869.
- [Jia04] L. JIANG, A. YEO, J. NURSALIM, S. WU, X. JIANG and Z. LU, *Frontal Infrared Human Face Detection by Distance From Centroid Method*. In Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing. Pp 41-44. Hong Kong. 2004.
- [Jon87] M. C. JONES and R. SIBSON, *What is Projection Pursuit?* Journal of the Royal Statistical Society, Vol.1, pp 1–37. 1987. DOI:10.2307/2981662.
- [Jon99] K. JONSSON, J. KITTLER, Y. P. LI and J. MATAS, *Support Vector Machines for Face Authentication*. In T. Pridmore and D. Elliman, editors, British Machine Vision Conference. pp 543–553. BMVA Press, 1999.
- [Jou97] P. JOURLIN, J. LUETTIN, D. GENOUD and H. WASSNER, *Integrating Acoustic and Labial Information for Speaker Identification and Verification*. Proc 5th European Conf. Speech Communication Technology. Pp 1603-1606 Rhodes, Greece, 1997.
- [Kar91] K.V. KARDONG and S. P. MACKESSY, *The Strike Behavior of a Congenitally Blind Rattlesnake*. Journal of Herpetology. Vol. 25 pp 208-211. 1991.

-
- [Kim02] K. I. KIM, K. JUNG and H. J. KIM, *Face Recognition using Kernel Principal Component Analysis*. IEEE Signal Processing Letters. Vol. 9, Issue 2, pp 40-42. 2002.
 - [Kon04] S. G. KONG, J. HEO, B. R. ABIDI, J. PAIK and M. A. ABIDI, *Recent Advances Fusion in Visual and Infrared Face Recognition*. Vol. 71, Issue. 2, pp 103-105. 2004.
 - [Kon07] S. G. KONG, J. HEO, F. BOUGHORBEL, Y. ZHENG, B. R. ABIDI, A. KOSCHAN, M. YI and M. A. ABIDI, *Multiscale Fusion of Visible and Thermal IR Images for Illumination-Invariant Face Recognition*. Computer Vision and Image Understanding. Vol. 71, Issue. 2, pp 215-233. 2007.
 - [Kro94] A. R. KROCHMAL, G. S. BAKKEN and T. J. LADUC. *Heat in Evolution's Kitchen: Evolutionary Perspectives on the Function and Origin of the Facial Pit of Pitvipers (Viperidae: Crotalinae)*. The Journal of Experimental Biology. Vol. 207, pp 4231-4238. 1994.
 - [Kak02] N. KAKUTA, S. YOYOYAMA and K. MABUCHI, *Human Thermal Models for Evaluating Infrared Images*. IEEE Engineering in Medicine and Biology. November-Desember 2002. 0739-5175/02.
 - [Kan73] T. KANADE, *Computer Recognition of Human Faces*. Birkhauser. 1973.
 - [Kev04] S. KEVIN ZHOU, *Face Recognition Using More Than One Still Image: What Is More?* Sinobiometrics 2004, LNCS 3338, Ed. Springer-Verlag Berlin Heidelberg. pp 212-223. 2004.
 - [Kit01] J. KITTLER, R. GHADERI, T. WINDEATT and J. MATAS, *Face Identification and Verification via ECOC*. Proceedings of the 3rd International Conference, AVBPA. LNCS, Vol. 2091. Sweden, June 6-8, 2001. ISBN 978-3-540-42216-7
 - [Kos08] KOSCHAN and M. ABIDI, *Digital Color. Image Processing*. Ed. Wiley. 2008. ISBN 978-0-470-14708-5.
 - [Kum06] B. V. K. V. KUMAR, M. SAVVIDES and C. XIE, *Correlation Pattern Recognition for Face Recognition*. Vol. 94, Issue 11. Pp 1963-1975. November 2006.
 - [Kwo05] O. K. KWON and S. G. KONG, *Multiscale Fusion of Visual and Thermal Images for Robust Face Recognition*. IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety. Pp. 112-116. April 2005.
 - [Lai91] M. LAIKIN, *Lens Design*. Marcel Dekker, Inc., NY, 1991.
 - [Law97] S. LAWRENCE, C. L. GILES, A. C. TSOI and A. D. BACK, *Face Recognition: A Convolutional Neural-Network Approach*. IEEE Transactions on Neural Networks. Vol. 8, Núm. 1, pp 98-133. 1997.
 - [Lee99] D. D. LEE and S. SEUNG, *Learning the Parts of Objects by Non-Negative Matrix Factorization*. Nature. Vol. 41, Issue 21, pp 788-791. 1999.
 - [Lee01] J. S. LEE, Y. Y. JUNG and B. B. KIM, S.S KO, *An Advanced Video Camera System with Robust AF, AE, and AWB Control*. IEEE Transactions Consum. Electron. Vol. 47, Issue 3, pp 694-699. 2001.
 - [Lew01] J. P. LEWIS, *Fast Normalized Cross-Correlation*. Document available online at <http://www.idiom.com/~zilla/Papers/nvisionInterface/nip.html>
 - [Li04] S. Z. LI and A. K. JAIN, *Introduction*, Handbook of Face Recognition. S. Z. Li, A. K. Jain. Ed. Springer. 2004. ISBN 0-387-40595.
 - [Lig79] L. LIGHT, F. KAYRA-STUARTT and S. HOLLANDER, *Recognition Memory for Typical and Unusual Faces*. Journal of experimental psychology-Human Learning and Memory. Vol. 5, pp 212-228, 1979.

-
- [Lin86] R. LINSKER, *From Basic Network Principles to Neural Architecture (series)*: Proceedings of the National Academy of Sciences USA 83: 7508–12, 8390–94, pp 8779–83. 1986.
 - [Lin88] R. LINSKER, *Self-Organization in a Perceptual Network*. *Computer*. Vol. 21, Issue 3, pp 105–17. 1988.
 - [Liu00] C. LIU, *Evolutionary Pursuit and its Application to Face Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol. 22, Issue 6, pp 570–582. June 2000.
 - [Liu05] J. LIU, D. D. CANNON, K. WADA, Y. ISHIKAWA, S. JONGTHAMMANURAK, D. T. DANIELSON, J. MICHEL, and L. C. KIMERLING, *Tensile Strained Ge p-i-n Photodetectors on Si Platform for C and L band Telecommunications*. *Applied Physics Letters*. Vol. 87, Issue 1. 2005.
 - [Liu08] D. T. LIU, X. D. ZHOU and C. W. WANG, *Wavelet-Based Multispectral Face Recognition*. *Optoelectronics Letters*. Vol. 4, Issue 5, pp 384–386. September 2008. DOI: 10.1007/s11801-008-8049-8.
 - [Lu02b] J. LU, K. N. PLATANOTIS and A. N. VENETSANOPOULOS, *Face Recognition Using LDA Based Algorithms*. IEEE Transactions on Neural Networks, Vol. 14, Issue 1, pp 195–200. 2003.
 - [Lu03b] X. LU and A. JAIN. *Resampling for Face Recognition*. AVBPA, LNCS 2688, pp 869–877. 2003.
 - [Luk07] M. LUKSCH, *Faceless. The Movie*. Autumn 2007.
 - [Ma06] Y. MA and S-B. LI, *Face Recognition by Combining Eigenface Method with Different Wavelet Subbands*. *Optoelectronics Letters*. Vol. 2, Issue 5, pp 383–385. September 2006.
 - [Mac03] D. J. C. MacKAY, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press 2003.
 - [Mal98] S. MALLAT, *A Wavelet Tour of Signal Processing*. Ed. Elsevier 2nd edition. 1998.
 - [Mal03] D. MALTONI, D. MAIO, A. K. JAIN and S. PRABHAKAR, *Handbook of Fingerprint Recognition*. Ed. Springer. 2003.
 - [Man92] B. S. MANJUNATH, R. CHELLAPPA and C. Von der MALSBERG, *A Feature Based Approach to Face Recognition*, Proc. IEEE CS Conf. Computer Vision and Pattern Recognition. Pp 373–378. 1992.
 - [Man01] T. MANSFIELD, G. KELLY, D. CHANDLER and J. KANE, *Biometric Product Testing Final Report*. Technical Report. Centre for Mathematics and Scientific computing. National Physical Laboratory. Issue 1.0. March 2001.
 - [Mar97] A. MARTIN, G. DODDINGTON, T. KAMM, M. ORDOWSKI and M. PRZYBOCKI, *The DET Curve in Assessment of Detection Performance*. European Speech Processing Conference Eurospeech. Vol. 4, pp1895–1898. 1997.
 - [Mar00] A. M. MARTÍNEZ, *Semantic Access of Frontal Face Images: The Expression-Invariant Problem*. Proceedings IEEE Content Based Access of Images and Video Libraries. Pp 55–59. June 2000.
 - [Mar01] A. M. MARTÍNEZ, *PCA versus LDA*. IEEE Transactions on Pattern analysis in Machine Intelligence. Vol. 23, Issue 2, pp 228–233. February 2001.
 - [Mar02] A. M. MARTÍNEZ, *Recognizing Imprecisely Localized Partially Occluded, and Expression Variant Faces from a Single Sample per Class*. IEEE Transaction on Pattern Analysis and Machine Intelligence Vol. 24, Issue 6, pp 748–763. 2002.

-
- [McC07] C. McCOOL, V. CHANDRAN, S. SRIDHARAN and C. FOOKES, *Modelling Holistic Feature Vectors for Face Verification*. Ed. Elsevier Science. April 2007.
 - [McG86] T. D. McGEE. *Principles and Methods of Temperature Measurement*. Ed. John Wiley & Sons Inc. 1988. ISBN-10: 0471627674.
 - [Mek10] J. MEKISKA, V. ESPINOSA-DURÓ and M. FAÚNDEZ, *Face Segmentation: A Comparison Between Visible and Thermal Images*. 44th IEEE-ICCST International Carnahan Conference on Security Technology. San José, US. October 2010. ISBN 978-1-4244-7401-1.
 - [Mil96] J. MILTON, *Tramp: The Life Of Charlie Chaplin*. Da Capo Press. 1998. ISBN-10: 0306808315.
 - [Mik99] S. MIKA, G. RATSCH, J. WESTON, B. SCHOLKOPF and K. R. MULLER, *Fisher Discriminant Analysis with Kernels*. Proceedings of IEEE Neural Networks for Signal Processing Workshop. 1999.
 - [Mor07] R. L. MORRISON, M. N. DO and D. C. MUNSON, *SAR Image Autofocus by Sharpness Optimization: A Theoretical Study*. IEEE Trans. Image Process. Vol.16, Issue 9, pp 2309–2321. 2007.
 - [Mor08] A. MORALES, M. A. FERRER, C. M. TRAVIESO, J. B. ALONSO, *Comparing Infrared and Visible Illumination for Contact-less Hand Based Biometric Scheme*. 42st IEEE-ICCST International Carnahan Conference on Security Technology. pp 191-197. Prague. Czech Republic. October 2008. ISBN 978-1-4244-1816-9.
 - [Mor09] M. MORENO-MORENO, J. FIERREZ and J. ORTEGA-GARCIA, *Biometrics Beyond the Visible Spectrum: Imaging Technologies and Applications*. BioID_Multicomm09. LNCS. Pp 154-161. October 2009.
 - [Mos94] Y. MOSES, Y. ADINI and S. ULLMAN, *Face Recognition: The Problem of Compensating for Illumination Changes*. Proc. European Conference Computer Vision. Pp 286-296, 1994.
 - [Nac75] J. NACHMAIS and A. WEBER, *Discrimination of simple and Complex Gratings*. Vision Research. Vol. 15 pp 217-223. 1975.
 - [Nan02] S. NANAVATI, M. THIEME and R. NANAVATI, *Biometrics. Identity Verification in a network World*. Ed. Wiley Computer Publishing. 2002.
 - [Nay94] NAYAR and Y. NAKAGAWA, *Shape from Focus*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 16, Issue 8, pp 824-831, August 1994.
 - [Nea07] V. E. NEAGOE, A. D. ROPOT and A. C. MUGIOIU, *Real Time Face Recognition Using Decision Fusion of Neural Classifiers in the Visible and Thermal Infrared Spectrum*. IEEE Conference on Advanced Video and Signal Based Surveillance. Pp 301-306. 2007.
 - [Ng06] R. NG, *Digital Light Field Photography*. PhD Thesis. Department of Computer Science of Stanford University. July 2006.
 - [Nic11] C. NICKEL, C. BUSCH, *Classifying Accelerometer Data via Hidden Markov Models to Authenticate People by the Way the Walk*. 45th IEEE-ICCST International Carnahan Conference on Security Technology. Mataró, Spain. October 2011. ISBN 978-1-4577-0901-2.
 - [Nie04] M. NIELSENA and C. DISSANAYAKEB, *Pretend Play, Mirror Self-Recognition and Imitation: A Longitudinal Investigation Through the Second Year*. Infant Behavior& Development Vol. 27, pp 342–365. 2004.
 - [Nix06] M. NIXON and J. CARTER, *Automatic Recognition by Gait*. Proceedings of the IEEE. Vol. 94, Issue 11, pp 2013–2024. November 2006.

-
- [Oll04] F. J. OLLIVIER, D. A. SAMUENSON, D. E. BROOKS, P. A. LEWIS, M. E. KALLBERG and A. M. KOMÁROMY, *Comparative Morphology of the Tapetum Lucidum among Selected Species. Veterinary Ophthalmology*. Vol. 7, Issue 1, pp 11-22. 2004.
- [Ort03] J. ORTEGA-GARCIA, J. FIERREZ, D. SIMON, J. GONZALEZ, M. FAÚNDEZ, V. ESPINOSA-DURÓ, A. SATUE, I. ERNAEZ, J. J. IGARZA, C. VIVARACHO, D. ESCUDERO, Q. I. MORO, *MCYT Baseline Corpus: A Bimodal Biometric DataBase*. IEE Proceedings. Vision, Image and Signal Processing. Vol. 150, Issue 11, pp 395-401. December 2003. ISSN 1350-245X.
- [Ots79] N. OTSU, *A Threshold Selection Method from Gray-level Histograms*. IEEE Transactions on Systems, Man and Cybernetics, Vol. 9, Issue 1, pp 62-66. 1979.
- [Pan03] Z. PAN, G. HEALEY, M. PRASAD, and B. TROMBERG, *Face Recognition in Hyperspectral Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 25, Issue. 12, pp. 1552–1560, Dec 2003.
- [Pan04] Z. PAN, G. HEALEY, M. PRASAD and B.TROMBERG, *Hyperspectral face Recognition Under Variable Outdoor Illumination*. Proceedings of SPIE, Vol. 5425 pp 520-529. 2004.
- [Pan05] Z. PAN, G. HEALEY, M. PRASAD, and B. TROMBERG, *Multiband and Spectral Eigenfaces for Face Recognition in Hyperspectral Images*. Proceedings of the SPIE, Vol. 5779, pp. 144–151. 2005.
- [Pas08] R. PASCHOTTA, *Encyclopedia of Laser Physics and Technology*. Ed. Wiley-VCH. Berlin, October 2008. ISBN 978-3-527-40828-3.
- [Pav00] I. PAVLIDIS and P. SYMOSEK, *The Imaging Issue in an Automatic Face/Disguise Detection System*. Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications. Pp 15-24. June 2000.
- [Pen96] P. PENEV and J. ATICK, *Local Feature Analysis: a General Statistical Theory for Object Representation*. Neural Systems. Vol. 7, Issue 3, pp 477-500. 1996.
- [Pet06] P. S. ALEKSIC and A. K. KATSAGGELOS, *Audio-Visual Biometrics*. Proceedings of the IEEE. Vol. 94, Issue 11, pp 2025–2044. November 2006.
- [Phi99] P. J. PHILLIPS. *Support Vector Machines applied to Face Recognition*. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors. Advances in Neural Information Processing Systems. MIT press. pp 803-809. 1999.
- [Phi03] P. J. PHILLIPS, P. GROTH, R. J. MICHEALS, D. M. BLACKBURN, E. TABASSI and M. BONE, *FRVT 2002: Evaluation Report*. NIST. Tech. Rep. NISTIR 696, 2003.
- [Pla00] J. C. PLATT, N. CRISTIANINI and J. SHAWE-TAYLOR, *Large margin DAGs for Multiclass Classification*. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, nips, Vol. 12, pp 547–553. MIT Press, 2000.
- [Pod96] C. PODILCHUK and X. ZHANG, *Face Recognition Using DCT Based Feature Vectors*. IEEE, pp 2144-2147. 1996.
- [Pop10] F. M. POP, M. GORDAN, C. FLOREA and A. VLAICU, *Fusion Based Approach for Thermal and Visible Face Recognition under Pose and Expressivity Variation*. 9th RoEduNet IEEE International Conference. Pp 61-66. 2010.
- [Pro92] F. J. PROKOSKI, *Method for Identifying Individuals from Analysis of Elemental Shapes Derived from Biosensor Data*. U.S. Patent number: 5163094.

-
- [Pro99] F. J. PROKOSKI and R. RIEDEL, *Biometrics: Personal Identification in Networked Society. Chapter 9: Infrared Identification of Faces and Body Parts*. Kluwer Academic Publishers. 1999.
 - [Pro00] F. J. PROKOSKI, *History, Current Status and Future of Infrared Identification*. Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications. pp 5-14. 2000.
 - [Psi09] L. PSIHOYOS and J. CLARK, *The Cove documentary*. Oceanic Preservation Society. 2009.
 - [Qui02] J. QUIANG, *3D Face Pose Estimation and Tracking from a Monocular Camera*. *Image and Vision Computing*. Vol. 20, pp 499-511. 2002.
 - [Rag11] R. RAGHAVENDRA, B. DORIZZI, A. RAO and G. H. KUMAR, *Particle Swarm Optimization Based Fusion of Near Infrared and Visible Images for Improved Face Verification*. *Pattern Recognition*. Vol. 44 pp 401-411. 2011.
 - [Rin08] E. F. J. RING, A. JUNG, J. ZUBER, P. RUTOWSLI, B. KALICKI and U. BAJWA, *Detecting Fever in Polish Children by Infrared Thermography*. 9th International Conference on Quantitative InfraRed Thermography. Krakow, Poland. July 2-5, 2008.
 - [Rod10a] E. RODRIGUEZ, K. NIKOLAIDIS, T. MU, J. F. RALPH, J. Y. GOULERMAS, *Collaborative Projection Pursuit for Face Recognition*. Bio-Inspired Computing: Theories and Applications (BIC-TA) Fifth International Conference. pp 1346-1350. September 2010.
 - [Rod10b] I. RODRÍGUEZ-ESCANCIANO and M. HERNÁNDEZ, *Lenguaje No Verbal*. Ed. Netbiblo. 2010.
 - [Rom06] S. ROMDHANI, J. HO, *Face Recognition Using 3-D Models: Pose and Illumination*. Proceedings of the IEEE. Vol. 94, Issue 11, pp 1977-1999. November 2006.
 - [Rub97] Y. D. RUBINSTEIN and T. HASTIE, *Discriminative vs Informative Learning*. Knowledge and Data Discovery (KDD), pp 49-53. 1997.
 - [Rut01] J. P. RUTLEDGE, *They All Look Alike: The Inaccuracy of Cross-Racial Identifications*. 28 American Journal of Criminal Law. Pp 207-228. Spring 2001.
 - [Sac87] O. SACKS, *The Man Who Mistook His Wife to a Hat and Other Clinical Tales*. 5th Edition. ISBN 0-06-097079-0. 1987.
 - [Sac96] O. SACKS, *An Anthropologist on Mars*. First Vintage books edition, USA. February 1996.
 - [Sch03] B. SCHNEIER, *Beyond Fear*. Ed Springer. 2003.
 - [Sch99] B. SCHÖLKORF, A. J. SMOLA and K. R. MÜLLER, *Kernel Principal Component Analysis*. Neural Computation. Issue 10 pp 1299-1319. 1999.
 - [Sel01] A. SELINGER and D. A. SOCOLINSKY, *Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study*. Technical Report. Equinox Corporation. 2001.
 - [Sha48] C.E. SHANNON, *A Mathematical Theory of Communication*. Bell System Technical Journal, Vol. 27, pp 379-423, 623-656, July, October, 1948.
 - [She92] L.R. SHERMAN, *The Right Look can Open Doors*. Security Management. October, 1992.
 - [Sin97] D. SINLEY, *LASER and LED Eye Hazards: Safety Standards*, Optics and Photonics News, pp 32-37. September 1997.
 - [Sin06] P. SINHA, B. BALAS, Y. OSTROVSKY and R. RUSSELL, *Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know about*. Proceedings of the IEEE. Vol. 94, Issue 11, pp 1948-1962. November 2006.

-
- [Sin08] R. SINGH, M. VATSA and A. NOORE, *Integrated Multilevel Image Fusion and Match Score Fusion of Visible and Infrared Face Images for Robust Face Recognition*. Pattern Recognition. Issue 41, pp 880–893. 2008.
- [Sir87] L. SIROVICH and M. KIRBY, *Low-dimensional Procedure for the Characterization of Human Faces*. Journal of the Optical Society of America, Vol. 4, Issue 3, pp 519-524. March 1987.
- [Sla80] C.C. SLAMA, C. THEURER and S.W. HENRIKSEN, *Manual of Photogrammetry*. American Society of Photogrammetry, Falls Church, VA fourth Edition. 1980.
- [Soc01] D. A. SOCOLINSKY, L. B. WOLFF, J. D. NEUHEISEL and C. K. EVELAND, *Illumination Invariant Face Recognition Using Thermal Infrared Imagery*. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01). Kauai, December 2001.
- [Soc02] D. A. SOCOLINSKY and A. SELINGER, *A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery*. Proceedings of 16th International Conference on Pattern Recognition Vol. 4, pp217–222. 2002.
- [Soc03] D. A. SOCOLINSKY, A. SELINGER and J. NEUHEISEL, *Face Recognition with Visible and Thermal Infrared Imagery*. Computer Vision and Image Understanding. Issue 91, pp 72-114. 2003.
- [Soc04a] D. A. SOCOLINSKY and A. SELINGER, *Thermal Face Recognition in an Operational Scenario*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04) Vol. 2, pp 1012-1019. 2004.
- [Soc04b] D. A. SOCOLINSKY and A. SELINGER, *Thermal Face Recognition over Time*. IEEE Proceedings of the 17th International conference on Pattern Recognition ICPR'04. IV, pp 187-190. 2004.
- [Son99] M. SONKA, V. HLAVAC and R. BOYLE, *Image Processing, Analysis and Machine Vision. 2nd Edition*. PWS Publishing Company. 1999.
- [Str99] G. STRANG, *The Discrete Cosine Transform*. SIAM-Rev, Vol. 41, Issue 1, pp 135–147. March 1999.
- [Sub98] M. SUBBARAO, J.K. TYAN, *Selecting the Optimal Focus Measure for Autofocusing and Depth-From-Focus*. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 20, Issue 8, pp 864–870. August 1998.
- [Suo09] J. SUO, S. C. ZHU, S. SHAN X. CHEN, *A Compositional and Dynamic Model for Face Aging*. IEEE Transactions on Pattern analysis in Machine Intelligence. Vol. 2, Issue 3, pp 385-401. 2009.
- [Swe96] D. SWETS and J. WENG, *Using Discriminant Eigenfeatures for Image Retrieval*. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 8, Issue 18, pp 831-836. 1996.
- [Tem99] P. TEMDEE, D. KHAWPARISUTH and K. CHAMNONGTHAI, *Face Recognition by Using Fractal Encoding and Backpropagation Neural Network*. Fifth International Symposium on Signal Processing and its Applications (ISSPA'99). Pp 159-161. Brisbane, Australia, August 1999.
- [The06] S. THEODORISIS and K. KOUTROUMBA, *Pattern Recognition*. Ed. Elsevier. 3th edition, 2006. ISBN: 0-12-369531-7.

-
- [Tos85] TOSHIBA, *CCD Image Sensor. Data book*. 2nd Edition. Toshiba Corporation. 1985.
 - [Tra04] C. TRAVIESO, J. B. ALONSO and M. A. FERRER, *Facial Identification using Transformed Domain by SVM*. 38st IEEE-ICCST International Carnahan Conference on Security Technology. pp 321-324. Albuquerque. USA. October 2004. ISBN 0-7803-8506-3.
 - [Tu03] J. TU, T. HUANG, R. BEVERIDGE and M. KIRBY, *Orthogonal Projection Pursuit Using Genetic Optimization*. IEEE Workshop. pp 266–269. 2003. ISBN 0-7803-7997-7/03.
 - [Tur91] M. TURK and A. PENTLAND. *Eigenfaces for Recognition*. Journal Cognitive Neuroscience, Vol. 3, Issue 1, pp 71-86, Massachusetts Institute of Thecnology, 1991.
 - [Uem88] S. UETMATSU, W. R. JANKEL, D. H. EDWIN, W. KIM, J. KOZIKOWSKI, A. ROSENBAUM, and D. M. LONG, *Quantification of Thermal Asymmetry. Part 2: Application in low-back pain and sciatica*. Journal of Neurosurgery. Vol. 69, Issue 4, pp 556-561. 1988.
 - [Vet97] T. VETTER, *Recognizing Faces from a New Viewpoint*. ICASSP 97, International Conference on Acoustics, Speech and Signal Processing. Vol. 1, pp 143-146. Munich, Germany. 1997.
 - [Vio01] P. VIOLA and M. JONES, *Robust Real-time Object Detection*. Technical Report CRL 2001/01, Cambridge Research Laboratory. 2001.
 - [Vol10] M. VOLLMER and K.P. MÖLLMANN, *Infrared Thermal Imaging. Fundamentals, Research and Application*. Ed. Wiley. 2010.
 - [Wan06] J. WANG, Y. SHANG, G. SU and X. LIN, *Age Simulation for Face Recognition*. Proceedings of the 18th International Conference Pattern Recognition, ICPR Vol. 3, pp 913-916. September, 2006.
 - [Wan10] R. WANG, S. LIAO, Z. LEI and S. Z. LI, *Multimodal Biometrics Based on Near-Infrared Face Recognition*, in *Biometrics: Theory, Methods, and Applications*. Edited by Boulgouris, Plataniotis and Micheli-Tzanakou. IEEE, Inc. 2010.
 - [Wee07] C. Y. WEE and R. PARASMESRAN, *Measure of Image Sharpness Using Eigenvalues*. Inf. Sci. 177, pp 2533–2552. 2007.
 - [Wen99] J. J. WENNG and D.L. SWETS, *Face Recognition*, Biometrics Personal Identification in Networked Society, Ed. Kluwer, 1999.
 - [Wei12] T. WEINER, *Enemies: A History of the F.B.I*. Publisher Allen Lane. 2012. ISBN: 9781846143267.
 - [Wil96] J. WILDER, P. J. PHILLIPS, C. JIANG and S. WIENER, *Comparison of Visible and Infrared Imagery for Face Recognition*. Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (FG'96). Pp 182-187. Killington. 1996.
 - [Wpe51] W.P.E, *Biometrika, 1901-1951*. Vol. 38, Issue 3-4, pp 267-268. December, 1951.
 - [Yan00] M. H. YANG, N. AHUJA and D. KRIEGMAN, *Face Recognition using Ada-Boosted Gabor Features*. Proceedings of International Conference on Automatic Face and Gesture Recognition. Vancouver. 2004.
 - [Yan01] M. YANG and N. AHURA, *Face Detection and Gesture Recognition for Human-Computer Interaction*. Ed. Kluwer Academic Publishers. 2001.
 - [Yu01] H. YU and J. YANG, *A Direct LDA Algorithm for High-Dimensional data with Application to Face Recognition*. Pattern Recognition. Vol. 34, pp 2067-2070. 2001.

-
- [Yan07] A. Y. YANG, J. WRIGHT, Y. MA and S. SHANKAR SASTRY, *Feature Selection in Face Recognition: A Sparse Representation Perspective*. Electrical Engineering and Computer Sciences University of California at Berkeley Technical Report No. UCB/EECS-2007-99. August 14, 2007.
- [Zak93] R. ZAKIA and L. STROEBEL, *The Focal Encyclopedia of Photography*. Third Edition. Ed. Focal Press. 1993.
- [Zap08] N. ZAPROUDINA, V. VARMAVUO1 O. AIRAKSINEN and M. NÄRHI, *Reproducibility of Infrared Thermography Measurements in Healthy Individuals. Physiological Measurement*. Vol. 29, Issue 4, pp 515-524. 2008. DOI: 10.1088/0967-3334/29/4/007.
- [Zha00] D. ZHANG, *Automated Biometrics. Technologies and Systems*. Ed. Kluwer Academic Publishers. 2000.
- [Zha02] D. ZHANG, *Biometric Solutions. For the Authentication in an E-World*. Ed. Kluwer Academic Publishers. 2002.
- [Zha03] W. ZHAO, R. CHELLAPA, J. PHILLIPS and A. ROSENFELD, *Face Recognition: A Literature Survey*. ACM Computing Surveys. Vol. 35, Issue 4, pp 399-458. 2003.
- [Zha05] S. ZHAO and R. GRIGAT, *An Automatic Face Recognition System in the Near Infrared Spectrum*. Proceedings of the International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM'05). Pp 437-444, Leipzig, Germany. July, 2005.
- [Zha06] W. ZHAO and R. CHELLAPA, *Face Processing. Advanced Modeling and Methods*. Ed. Elsevier. 2006.
- [Zho04] S.K. ZHOU, *Face Recognition using more than One Still Image: What is More?* Lecture Notes In Computer Science LNCS 3338, A. Z. Li et al. Ed., Sinobiometrics. Springer Verlag. Pp 212-223. 2004.
- [Zou06] X. ZOU, J. KITTLER and K. MESSER, *Ambient Illumination Variation Removal by Active Near-IR Imaging*. International Conference on Biometrics (ICB'06), Hong Kong, China, January, 2006. LNCS 3832, Ed. Springer-Verlag Berlin Heidelber. pp 19-25. 2006.
- [Zor09] C. ZOR and T. WINDEATT, *Upper Facial Action Unit Recognition*. Proceedings of the Third International Conference on Advances in Biometrics (ICB'09). Ed. Springer-Verlag Berlin, Heidelberg. Pp 239-248. 2009. ISBN: 978-3-642-01792-6.